Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of *Urtica dioica* agglutinin isolectin I: dimer formation mediated by two zinc ions bound at the sugar-binding site

Kazuaki Harata,* Wolf-Dieter Schubert† and Michiro Muraki

Biological Information Research Center, AIST, Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

† Present address: Strukturforschung, Gesellschaft für Biotechnologische Forschung (GBF), Mascheroder Weg 1, Braunschweig D-38124, Germany.

Correspondence e-mail: k-harata@aist.go.jp

Received 15 February 2001
Accepted 19 July 2001

PDB Reference: agglutinin isolectin I, 1iqb.

Ultica dioica agglutinin, a plant lectin from the stinging nettle, consists of a total of seven individual isolectins. One of these structures, isolectin I, was determined at 1.9 Å resolution by the X-ray method. The crystals belong to the space group $P2_1$ and the asymmetric unit contains two molecules related by local twofold symmetry. The molecule consists of two hevein-like chitin-binding domains lacking distinct secondary structure, but four disulfide bonds in each domain maintain the tertiary structure. The backbone structure of the two independent molecules is essentially identical and this is similarly true of the sugar-binding sites. In the crystal, the C-terminal domains bind Zn^{2+} ions at the sugar-binding site. Owing to their location near a pseudo-twofold axis, the two zinc ions link the two independent molecules in a tail-to-tail arrangement: thus, His47 of molecule 1 and His67 of molecule 2 coordinate the first zinc ion, while the second zinc ion links Asp75 of molecule 1 and His47 of molecule 2.

1. Introduction

U. dioica agglutinin (UDA) from the stinging nettle is a plant lectin specific to *N*-acetylglucosamine (NAG) oligomers and chitin (Peumans *et al.*, 1984; Shibuya *et al.*, 1986). This lectin has been suggested to participate in plant defence and is known to possess both antifungal and insecticidal activities (Broekaert *et al.*, 1989; Heusing *et al.*, 1991). It binds various glycoproteins present on the membrane of T and B cells and also acts as a superantigen in T-cell activation (Galelli & Truffa-Bachi, 1993). UDA is a mixture of seven individual isolectins. Of these, isolectins II and III consist of 90 and 80 amino-acid residues, respectively, while each of the remaining five has 89. Isolectin III has been identified at the DNA level, although the mature protein has not yet been detected. Previously, we determined the crystal structures of isolectin VI and its complex with tri-*N*-acetylchitotriose (Harata & Muraki, 2000). The protein consists of two hevein-like domains, each possessing one sugar-binding site. The sugar-binding site is constituted of three aromatic amino acids which bind sugar rings through stacking contacts. The N-terminal domain incorporates a third subsite where a sugar residue is bound by hydrogen bonds.

Isolectin I differs in five of 89 amino-acid residues from isolectin VI but has identical sugar-binding activity (Van Damme *et al.*, 1988). Here, we present the crystal structure of this isolectin and discuss its structure, especially in comparison with the structure of isolectin VI. Isolectin I was crystallized in the presence of Zn^{2+}; unexpectedly, two zinc ions are observed at one of the sugar-binding sites.

2. Materials and methods

2.1. Crystallization and data collection

Isolectin I was isolated from commercially available UDA, using a pepRPC HR 5/5 column on a Pharmacia FPLC system, as reported in Muraki *et al.* (1998). The first eluted peak fraction was found to contain UDA-VI, while the second peak fraction is a mixture of isolectin I and isolectin II as indicated by MALDI-TOF/MS. This mixture was lyophilized and dissolved in water to a final concentration of 2% for crystallization. Crystals were prepared by hanging-drop vapour diffusion: 5 µl of UDA was mixed with 5 µl of reservoir solution containing 18% PEG 8000, 0.2 *M* zinc acetate and 0.1 *M* sodium cacodylate buffer pH 6.5. The droplet of the solution was equilibrated over the reservoir solution. Thick crystalline needles appeared within a few days. X-ray data were measured on an Enraf–Nonius FAST diffractometer equipped with an FR571 rotating-anode generator (40 kV, 50 mA, focal spot size 0.2 mm). Data collection proceeded at 286 K and intensity data were evaluated using the *MADNES* software (Messerschmidt & Pflugrath, 1987) incorporated in the FAST system.

2.2. Structure determination and refinement

The structure was solved by molecular replacement using the program *X-PLOR* (Brünger *et al.*, 1987). A self-rotation function indicated that the two molecules are related by non-crystallographic twofold symmetry. A cross-rotation search using the coordinates of isolectin VI supplied the orientation of the first molecule, while that of the second was derived by applying the result of the self-rotation function. The positions of both were determined by a translational search. The resulting structural model was refined using *X-PLOR* for 10 402 reflections with $F_o > 2\sigma(F)$ in the resolution range 8.0–1.9 Å. Solvent-molecule detection proceeded automatically by accepting electron-density peaks higher than 2σ, within hydrogen-bonding distance (2.5–3.3 Å^2) of protein O and N atoms, and B values less than 60 Å^2 in $F_o - F_c$ maps. The final structure was checked using the program *PROCHECK* (Laskowski *et al.*, 1993) and the atomic coordinates have been deposited with the PDB.

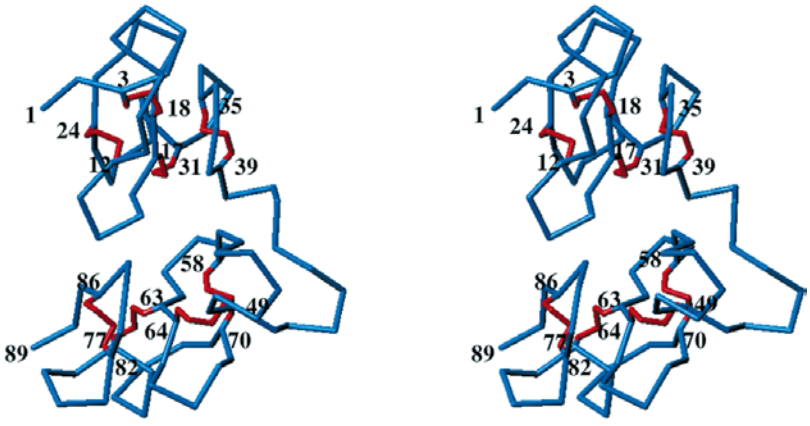

Figure 1
The backbone structure of UDA-I in stereo. Eight disulfide bridges are indicated in red.

Table 1
Statistics of data collection and structure refinement.

Data collection	
Space group	$P2_1$
Unit-cell parameters	
a (Å)	30.71 (1)
b (Å)	42.11 (1)
c (Å)	62.59 (2)
β (°)	101.55 (1)
Resolution range (Å)	25.1–1.73
Number of measured reflections	47712
Number of unique reflections	14774
$R_{\rm merge}$	0.069
Completeness (%)	
All data	89.0
$I > 3\sigma(I)$	53.8
Structure refinement	
Resolution range (Å)	8.0–1.90
No. of reflections $[F > 2\sigma(F)]$	10402
Completeness (%)	84.3
R value	0.205
Free R value	0.260
R.m.s. deviations from ideality	
Bond length (Å)	0.016
Bond angles (°)	3.32

3. Results and discussion

3.1. Data collection and structure determination

Crystal data and the statistics of data collection are given in Table 1. The UDA-I crystals belong to the space group $P2_1$ and the asymmetric unit contains two independent molecules ($V_{\rm M}$ = 2.09 Å^3 Da^{-1}). Intensity data were collected to 1.73 Å resolution and 47 712 observed reflections were merged to a set of 14 774 independent reflections (89.0% completeness) with an $R_{\rm merge}$ of 6.9%. $R_{\rm merge}$ and completeness of the highest resolution shell were 29.4 and 25.0%, respectively. Intensity data to a resolution of 1.9 Å (95.0% completeness of the highest resolution shell) were used for the structure determination.

In the crystal packing, two independent molecules were related by local twofold symmetry. The final structure model consisted of two UDA-I molecules, two zinc ions and 87 water molecules. The R value and $R_{\rm free}$ value for 10% data were 20.5 and 26.0%, respectively, for the data with $F_o > 2\sigma(F)$ and in the resolution range 8.0–1.9 Å. The Ramachandran plot (Ramakrishnan & Ramachandran, 1965) indicated that almost all (φ, ψ) values fall in the normal range.

3.2. Molecular structure

Although a mixture of isolectins I and II was used for the crystallization (see §2), the crystals exclusively contained isolectin I. This is inferred from the clearly visible N-terminal Gln residue, being particularly evident because the Gln1

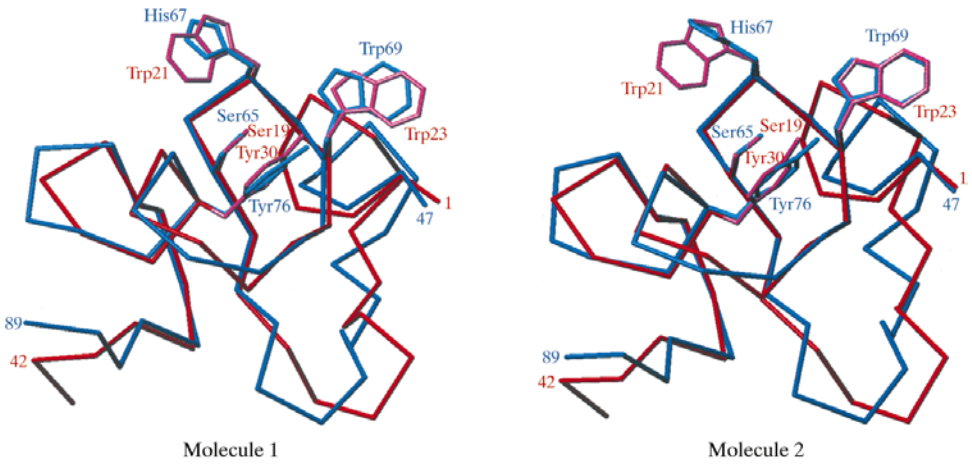

Figure 2
A superposition of the N-terminal domain (red) and the C-terminal domain (red), respectively, for the two independent molecules. Side-chain groups of amino-acid residues involving in the sugar binding are rendered in magenta (domain 1) and cyan (domain 2).

residue is converted to pyroglutamic acid. Fig. 1 shows the backbone structure of UDA-I. The protein consists of two homologous hevein-like domains. The first domain, residues 1–42, is linked to the C-terminal domain (47–89) by the sequence Glu43-Arg44-Ser45-Asp46. The molecule possesses little secondary structure and the backbone structure of each domain is maintained by four disulfide bridges (Fig. 1). Two short β-strands in the N-terminal domain, Cys17–Ser19 and Cys35–Asn37, form a β-sheet-like structure and a similar structure consisting of the strands, Cys63–Ser65 and Cys82–Tyr84, is observed in the C-terminal domain. The backbone structure of the domains is quite similar and is indicative of gene duplication in the course of its evolution (Fig. 2). Significant differences are observed in the

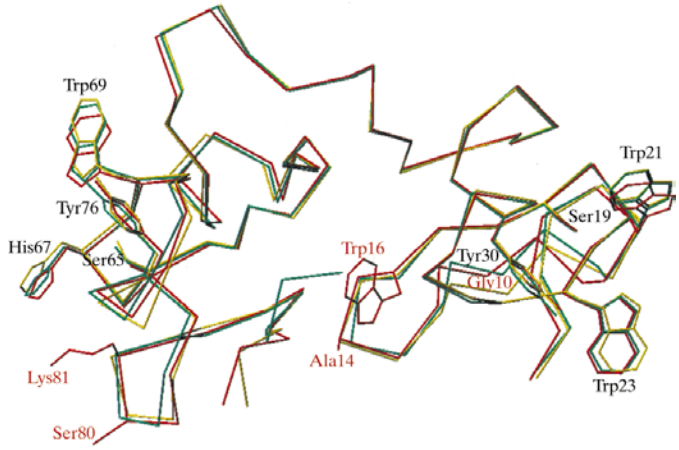

Figure 3
A superposition of the independent molecules of UDA-I (molecule 1 in red, molecule 2 in yellow) and the structure of UDA-VI (green). Amino-acid side chains involved in sugar binding are shown. Amino-acid residues differing between UDA-I and UDA-VI are depicted for molecule 1 and are labelled in red.

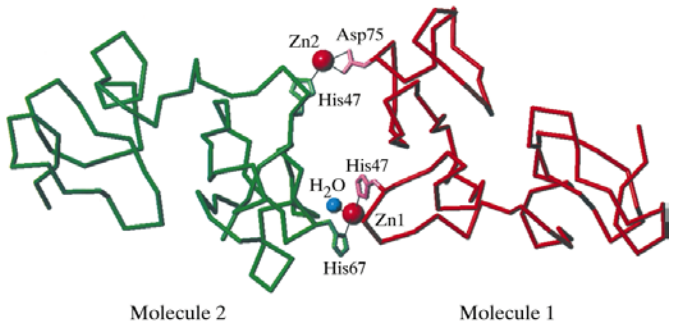

Figure 4
The pseudo-symmetric packing of the two independent UDA-I molecules: side-chain groups participating in zinc binding are drawn. The zinc ions and water molecule are shown in magenta and cyan, respectively.

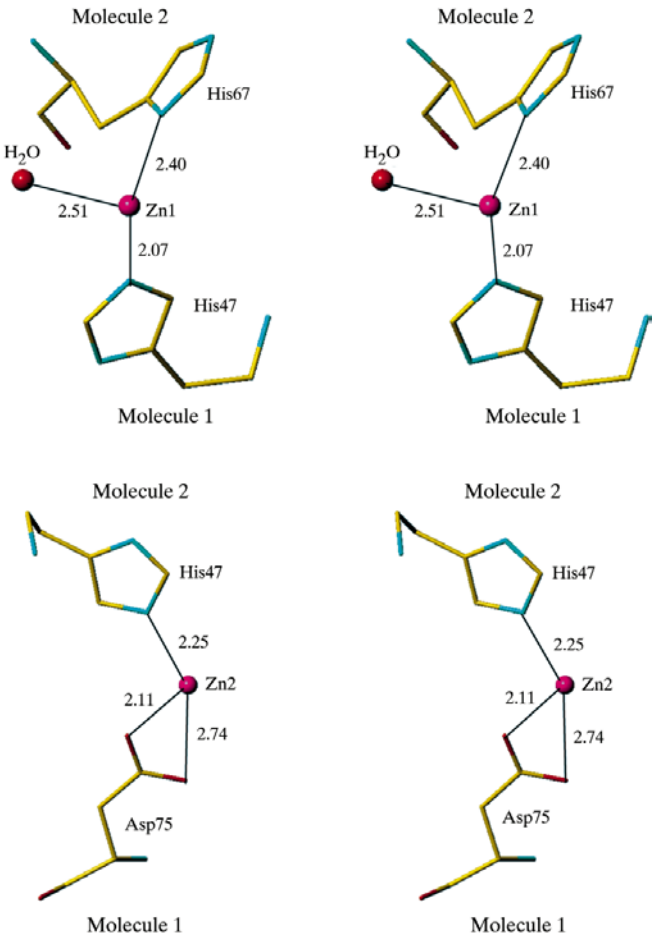

Figure 5
Stereoview of the zinc-binding sites.

region of residues 4–17 and 50–63. These regions contain non-overlapping loop structures in the superposition of the two domains.

Each domain contains a sugar-binding site specific for *N*-acetylglucosamine and its oligomers. In fact, the sugar-binding site is essentially equivalent in the two domains and the amino-acid residues involved in sugar binding superimpose particularly well. In the N-terminal domain, residues Ser19, Trp21, Trp23 and Tyr30 are involved in binding the sugar. Trp21 and Trp23 stack onto the pyranose rings of the sugar residues, while Ser19 and Tyr30 form hydrogen bonds. In the C-terminal domain, Ser65, His67, Trp69 and Tyr76 participate in the sugar binding. It should be noted that His67 corresponds to Trp21 of the N-terminal domain, *i.e.* not to Trp but to His. Therefore, the sugar-binding ability is not identical for the two domains. In total, the aromatic residues can bind two sugar units. Previously, we demonstrated that the N-terminal domain bears a third subsite where Plg1 (pyroglutamic acid) and Cys24 are hydrogen bonded to the sugar residue.

The asymmetric unit of the crystal contains two independent molecules related by local twofold symmetry. As shown in Fig. 3, the two molecules have the same backbone structure. They superimpose with an r.m.s. deviation of 0.67 Å between equivalent C^{α} atoms. No significant difference was observed in the conformation of the side-chain groups in the sugar-binding sites. The average *B* values for the main-chain and side-chain atoms in molecule 1 are 24.1 and 26.8 $Å^2$, respectively. In contrast, molecule 2 shows somewhat larger average *B* values, 28.6 $Å^2$ (main chain) and 30.7 $Å^2$ (side chain).

In Fig. 3, we compare the backbone structure of isolectin I with that of isolectin VI, highlighting the five amino-acid residues that differ in the two proteins. Residues Gly10, Ala14 and Trp16 of the N-terminal domain are replaced by Ser, Gly and Arg, respectively, in UDA-VI, while Ser80 and Lys81 of the C-terminal domain are changed to Gly and Asn, respectively. These mutations neither affect the backbone fold nor the structure of the sugar-binding sites. Saul *et al.* (2000) similarly demonstrated the identical backbone structure of isolectin I, isolectin V and isolectin VI, which they cocrystallized practically isomorphously to our isolectin VI crystals.

3.3. Zinc binding

UDA-I was crystallized in the presence of 0.1 *M* zinc acetate. Two zinc ions were identified in the structure. They are coordinated by the C-terminal domain of UDA-I and link the two independent molecules in the asymmetric unit in a tail-to-tail fashion (Fig. 4). As a consequence, the sugar-binding site of the C-terminal domain is blocked through dimer formation. As shown in Fig. 5, Zn1 is bound by the imidazolyl groups of His47 of molecule 1 and His67 of molecule 2. A water molecule is the only additional ligand visible. Zn2 is bound by a histidine (47) and an aspartate (75). The coordination geometry observed in zinc-bound proteins is tetrahedral or square planar for four-coordination, tetragonal pyramid or trigonal bipyramid for five-coordination, and octahedral for six-coordination (Harding, 2001). The two zinc ions in the present structure are not fully coordinated. Three atoms bound to Zn1 occupy the positions corresponding to the tetrahedral arrangement. On the other hand, the carboxyl group of Asp75 is chelated to Zn2; the two carboxyl O atoms, the imidazolyl N atom of His47 and Zn2 are nearly coplanar.

Numerous structures of zinc-binding proteins have been reported. Zinc-ion ligands are mostly sulfhydryl groups, imidazolyl groups, carboxyl groups and water molecules. Zinc fingers are a typical zinc-binding motif donating four cysteine and/or histidine ligand groups (Lee *et al.*, 1989). The arrangement and number of coordinating ligands vary widely in proteins such as insulin (Ciszak & Smith, 1994), thermolysin (English *et al.*, 1999), carboxypeptidase (Kim & Lipscomb, 1991), human α-lactalbumin (Ren *et al.*, 1993) *etc*. Some lectins have metal-binding sites, but metal ions are mostly involved in the recognition and binding of sugar molecules rather than the oligomer formation. Animal lectins known as C-type lectins have calcium ions that are essential for carbohydrate binding (Drickamer, 1999). Concanavalin A, a legume lectin, binds not only calcium ions but also some other divalent cations such as Zn^{2+}, Mn^{2+} and Cd^{2+}, and a manganese and a calcium ion are associated with sugar-binding activity. In the structure of zinc/calcium-substitute concanavalin A, one of the two zinc ions is located at the intermolecular interface between two tetramers (Bouckaert *et al.*, 2000). The formation of the UDA-I dimer is mediated by the two zinc ions. UDA-I consists of two homologous domains, but the N-terminal domain has no histidine residue. His47 and His67 are the residues corresponding to Plg1 and Trp21, and His67 is one of the amino-acid residues that participates in the sugar binding (Harata & Muraki, 2000). Despite dimerization, the UDA-I molecules retain agglutination activity as the dimer still possesses two unhindered sugar-binding sites at both ends (Fig. 4). Previous evidence for the zinc binding at the sugar-binding site of UDA has not been reported. As the crystal was grown in the presence of ~0.1 *M* zinc acetate, it is unlikely that zinc would bind to the sugar-binding sites under physiological conditions. Nevertheless, the present structure demonstrates the affinity of the sugar-binding site of the C-terminal domain for the zinc ion.

Many lectins have oligomer structures to achieve the agglutination activity (Rini, 1995; Bouckaert *et al.*, 1999). Legume lectins form a dimer or tetramer structure because each individual monomer has a single sugar-binding site. On the other hand, each of three homologous domains of monocot lectins has an independent mannose-binding site, although the proteins have been observed in a dimeric or tetrameric structure. Wheatgerm agglutinin (WGA) is a dimeric protein composed of four hevein-type chitin-binding domains. Each domain has one sugar-binding site that is located at the dimer interface and a sugar molecule contacts two WGA molecules. In contrast, UDA does not form an oligomer but can bind two sugar molecules. In the crystal of UDA-VI complexed with tri-*N*-acetylchitotriose, the protein molecules are arranged in a head-to-tail mode and the sugar molecule is sandwiched between the N-terminal domain and

the C-terminal domain of the next molecule. The sugar molecules bind protein molecules to form a polymeric chain extended along the crystallographic *c* axis. In the crystal of UDA-I, the zinc ions are only bound to the C-terminal domain. This resulted the formation of the tail-to-tail dimer related with the local twofold symmetry. In the trial of the crystallization using Crystal Screen and Crystal Screen 2 kits (Hampton Research), UDA-I only crystallized in the presence of Zn^{2+}. On the other hand, UDA-VI crystallized in a neutral pH region without zinc ions, and several alcohols and PEGs were effective reagents for crystallization. The dimer structure of UDA-I mediated by zinc ions is rather artificial and may be biologically less significant because no dimer formation has been detected in solution. However, the present crystal structure revealed the the role of the zinc ion in the crystallization of this isolectin of UDA.

References

Bouckaert, J., Hamelryck, T., Wyns, L. & Loris, R. (1999). *Curr. Opin. Struct. Biol.* **9**, 572–577.

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

Broekaert, W. F., Van Parijs, J., Leyns, F., Joos, J. & Peumans, W. J. (1989). *Science*, **245**, 1100–1102.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Ciszak, E. & Smith, G. D. (1994). *Biochemistry*, **33**, 1512–1517.

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

English, A. C., Done, S. H., Caves, L. S. D., Groom, C. R. & Hubbard, R. E. (1999). *Proteins Struct. Funct. Genet.* **37**, 628–640.

Galelli, A. & Truffa-Bachi, P. (1993). *J. Immunol.* **151**, 1821–1831.

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

Harding, M. M. (2001). *Acta Cryst.* D**57**, 401–411.

Heusing, J. E., Murdock, L. L. & Shade, R. E. (1991). *Phytochemistry*, **30**, 3565–3568.

Laskowski, A. R., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989). *Science*, **245**, 635–637.

Kim, H. & Lipscomb, W. N. (1991). *Biochemistry*, **30**, 8171–8180.

Messerschmidt, A. & Pflugrath, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.

Muraki, M., Morii, H. & Harata, K. (1998). *Protein Pept. Lett.* **5**, 193–198.

Peumans, W. J., De Ley, M. & Broekaert, W. F. (1984). *FEBS Lett.* **177**, 99–103.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Ren, J., Stuart, D. I. & Acharya, K. R. (1993). *J. Biol. Chem.* **268**, 19292–19298.

Rini, J. M. (1995). *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577.

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

Shibuya, N., Goldstein, I. J., Shafer, J. A., Peumans, W. J. & Broekaert, W. F. (1986). *Arch. Biochem. Biophys.* **249**, 215–224.

Van Damme, E. J. M., Broekaert, W. F. & Peumans, W. J. (1988). *Plant Physiol.* **86**, 598–601.