Volume 299, number 2, 131–134 FEBS 10798 March 1992

The primary structure of stinging nettle (*Urtica dioica*) agglutinin

A two-domain member of the hevein family

Jaap J. Beintema[a] and Willy J. Peumans[b]

[a]*Biochemisch Laboratorium, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands* and [b]*Laboratorium voor Fytopathologie en Plantenbescherming, Katholieke Universiteit Leuven, W. de Croylaan 42, B-3001 Heverlee, Belgium*

Received 10 December 1991; revised version received 22 January 1992

The primary structure of stinging nettle (*Urtica dioica*) agglutinin has been determined by sequence analysis of peptides obtained from three overlapping proteolytic digests. The sequence of 89 residues consists of two hevein-like domains with the same spacing of half-cystine residues and several other conserved residues as observed earlier in other proteins with hevein-like domains. The hinge region between the two domains is four residues longer than those between the four domains in cereal lectins like wheat germ agglutinin.

Lectin; Agglutinin; Sequence homology; Hevein; *Urtica dioica*

1. INTRODUCTION

A monomeric lectin from rhizomes of stinging nettle, *Urtica dioica* agglutinin (UDA), isolated by affinity chromatography on chitin [1], has a molecular mass of 8.5 kDa, exhibits specificity toward *N*-acetyl-D-glucosamine oligomers and inhibits growth of several phytopathogenic and saprophytic chitin-containing fungi in vitro [2].

Determination of the amino acid sequences of the N-terminus and several tryptic peptides [3] showed homology with several other chitin-binding lectins like hevein [4] and wheat germ agglutinin [5], with identical spacing of half-cystine residues. Many sequences of other members of protein families with hevein-like domains have been determined [6,7]. Here we present the complete primary structure of UDA, the first member of a protein family with two hevein-like domains per subunit.

2. MATERIALS AND METHODS

Total UDA was isolated as described [1,2]. A single isolectin (UDA_7) was obtained from the total preparation by ion-exchange chromatography as described by Van Damme and Peumans [8]. Disulfide bonds were reduced with tributylphosphine and *S*-pyridylethylated with 4-vinylpyridine [9]. The modified protein was digested in several batches with trypsin (treated with tosylphenylalanylchloromethane), chymotrypsin or endoproteinase Glu-C from *Staphylococcus aureus* V8 in 0.2 M ammonium bicarbonate at 37°C for 2–6 h, at a substrate/enzyme ratio of 100:1 (by mass). The C-terminal peptide obtained after cleavage with endoproteinase Glu-C was subdigested overnight with thermolysin. Peptides were isolated by gel filtration on a column of Sephadex G-25 (superfine) in 0.2 M acetic acid followed by reversed-phase HPLC with a linear gradient of 0–60% acetonitril in 0.1% trifluoroacetic acid [9].

Protein and peptides were hydrolyzed in 6 M HCl in vacuo at 110°C for 18 h. The hydrolyzates were analyzed with an LKB Alpha Plus amino acid analyzer. Amino acid sequences of peptides were determined by manual sequencing (on 5–10 nmol peptide) with the 4-(*N*,*N*-dimethylamino)azobenzene-phenylisothiocyanate (DABITC/PhSCN) double-coupling procedure of Chang [10] and automated Edman degradation (on 0.2–0.5 nmol peptide) on an Applied Biosystems 477A pulse-liquid sequencer with an on-line 120A phenylthiohydantoin analyzer of the sequence facility of the Institute Bioson and Eurosequence BV (Groningen). The C-terminal residue of the C-terminal tryptic peptide was identified after three Edman degradation cycles by dansylation without acid hydrolysis.

Correspondence address: J.J. Beintema, Biochemisch Laboratorium, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands.

3. RESULTS AND DISCUSSION

The amino acid sequence of UDA is presented in Fig. 1. Three sets of overlapping peptides (10–40 nmol) were obtained after gel filtration and reversed-phase HPLC of digests with three different enzymes of 100–500 nmol quantities of modified protein. The evidence for the proposed sequence has been collected from the sequences (Fig. 1) and amino acid analyses (Table I) of these peptides. The amino acid composition calculated from the sequence is in agreement with that of UDA isolectin IV (the major component [11]; identical to UDA_7 [8]), except for the presence of one more serine and two less arginines. The protein contains no methionine or phenylalanine. Most of the sequence data of UDA published earlier [3] were confirmed (T4, N-terminus of T7, T8). Other sequence data presented by Chapot et al. [3] may indicate the presence of a mixture

of UDA isolectins, as sequence analysis of the N-terminus of the protein and of peptides corresponding with peptide T2 in our study, shows the presence of arginine and serine at position 16 (tryptophan in our study), isoleucine and phenylalanine at position 20, and aspartic acid and alanine at position 26. Other heterogeneities presented by Chapot et al. [3] can be explained by sequence analysis of a peptide mixture with the N-terminal sequences of T2 and T6. The identification of the N-terminal pyroglutamic acid by digestion with pyroglutamate aminopeptidase [3] has not been repeated.

Two hevein-like domains can be recognized in the primary structure of UDA, with the same spacing of half-cystine residues as observed in other sequences with hevein-like domains (Fig. 2). Hevein-like domains have been identified as a single domain in hevein itself, wound-induced gene products (Win 1 and Win 2) of potato [14], many basic chitinases [7], and as four tandemly organized domains per subunit in several cereal lectins [5,17]. The eight half-cystine residues form four disulfide bridges. The positions of these bridges have been determined by X-ray diffraction for wheat germ agglutinin [12] and by X-ray diffraction [13] and 2D NMR spectroscopy for hevein (K. Dijkstra, R.M. Scheek and J.J. Beintema, unpublished), and are presented in Fig. 2. There is no variation in the positions of the first six half-cystine residues, which form three disulfide bridges. However, there is some variation in the spacing of the two C-terminal ones, which form a separate disulfide bridge (Fig. 2). In addition to the eight conserved half-cystine residues, there are three conserved glycine residues and one conserved serine residue in all hevein-like domains. At position 38 in the alignment of Fig. 2, UDA has glutamate instead of glutamine. Additional disulfide bridges may occur, like in rice lectin [17], which has additional half-cystine residues at position 43 in domains 1 and 3 (Fig. 2). The 3D structure of wheat germ agglutinin [12] suggests that there may be disulfide bonds between the opposing domains 1 and 3 of different subunits in the dimeric molecule from rice. This may explain the observation that unlike wheat germ agglutinin, the dimeric rice lectin does not dissociate at low pH [18].

Two *N*-acetylglucosamine-binding sites have been located in the X-ray structure of wheat germ agglutinin. Each of these sites is located between two hevein-like domains, with the tyrosine residues at position 30 in domains 2 and 4 as critical residues [12]. However, the presence of two hevein-like domains is not essential for sugar binding as monomeric hevein has affinity for oligomers of *N*-acetylglucosamine [13] and inhibits fungal growth [19], and other proteins like the Win gene products of potato [14] and the basic chitinases [7] also possess single hevein-like domains. UDA is the first sequenced lectin with two hevein-like domains per subunit in the mature protein.

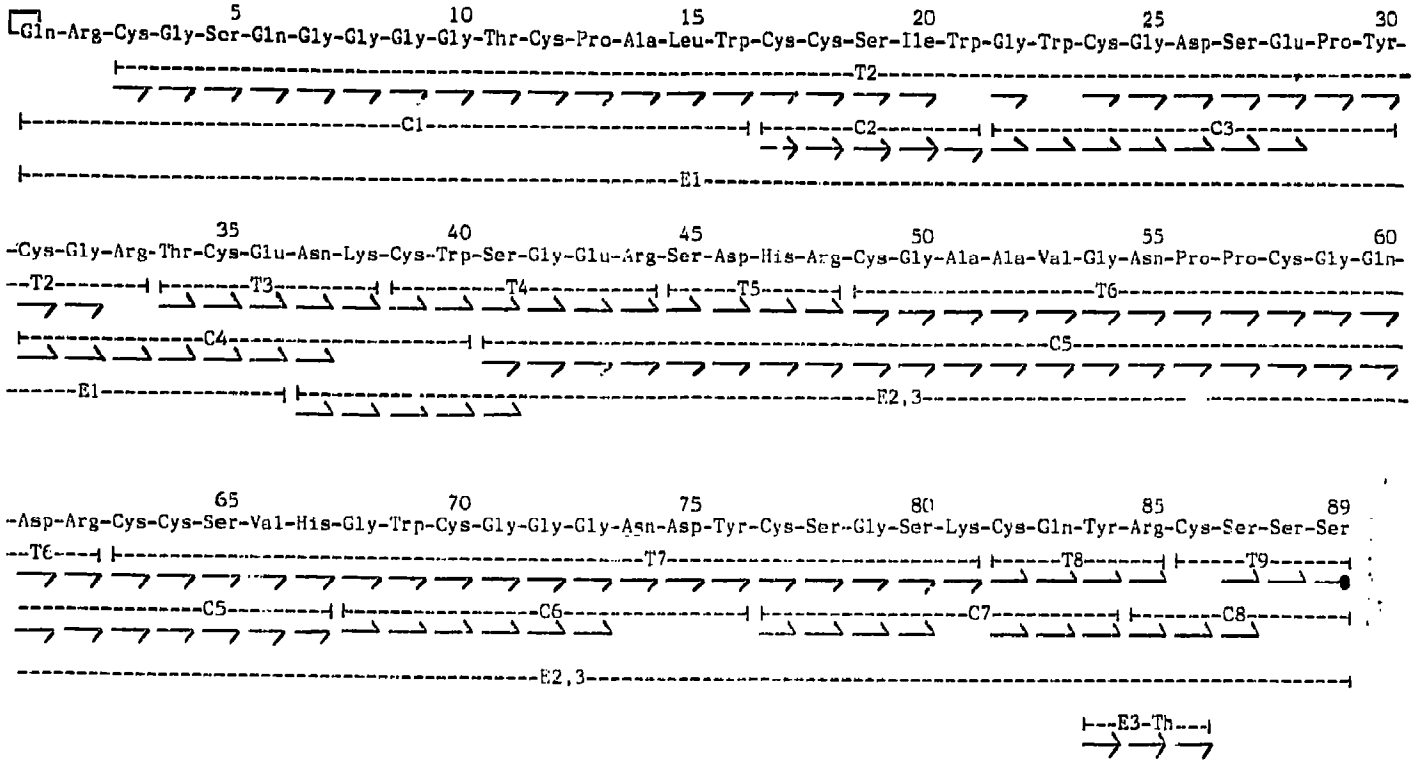

Fig. 1. Amino acid sequence of stinging nettle (*Urtica dioica*) agglutinin. |---|, analyzed peptides (Table I). Prefixes T, C, E and Th indicate tryptic, chymotryptic, endoproteinase Glu-C and thermolytic peptides, respectively. —•, identified by manual sequencing; —, identified by automated sequence analysis. —●, identified after three Edman steps as dansyl derivative without acid hydrolysis. Peptide C5 was contaminated with peptide 48–67 (the two histidine residues at positions 47 and 67 both are cleavage sites for chymotrypsin in the pyridylethylated protein).

Table I

Amino acid compositions of peptides of UDA

	T2	T3	T4	T5	T6	T7	T8	T9	C1	C2	C3	C4	C5	C6	C7	C8	E1	E2,3
Aspartic acid	1.1(1)	0.9(1)		0.9(1)	2.0(2)	1.9(2)					1.0(1)	0.9(1)	2.9(3)	2.3(2)			1.4(1)	6.6(6)
Threonine	0.8(1)	1.0(1)							0.9(1)			1.0(1)					1.9(2)	
Serine	3.5(3)		0.7(1)	1.0(1)		3.7(3)	0.2	+(3)	1.1(1)	1.0(1)	0.9(1)		2.8(3)		2.0(2)	3.1(3)	2.7(3)	9.1(8)
Glutamic acid	2.1(2)	1.0(1)	1.0(1)		1.1(1)		1.1(1)		2.3(2)		1.1(1)	1.1(1)	2.2(2)		1.1(1)		3.7(4)	3.9(3)
Proline	2.1(2)				2.0(2)				0.9(1)		0.9(1)		2.3(2)				1.8(2)	2.3(2)
Glycine	6.9(8)	0.2	1.1(1)		2.7(3)	4.5(5)	0.2		4.3(5)		2.0(2)	1.1(1)	3.7(4)	3.6(4)	1.1(1)		6.2(8)	9.8(9)
Alanine	1.1(1)				2.1(2)				1.0(1)				2.0(2)				1.1(1)	2.0(2)
Half-cystine	4.2(6)	1.1(1)	1.1(1)		1.0(2)	2.2(4)	0.7(1)	+(1)	0.7(2)	0.7(2)	+(1)	2.9(3)	1.8(4)	+(1)	0.9(2)	0.8(1)	4.2(7)	6.5(9)
Valine					1.0(1)	0.8(1)							1.6(2)					2.1(2)
Isoleucine	0.8(1)									1.0(1)							0.8(1)	
Leucine	1.2(1)					0.1			1.0(1)								0.9(1)	
Tyrosine	1.3(1)					1.1(1)	1.1(1)				1.1(1)			1.1(1)	0.9(1)		1.1(1)	1.6(2)
Lysine		1.0(1)				1.2(1)						1.0(1)			0.9(1)			1.8(2)
Histidine				1.0(1)		0.8(1)							1.3(2)					1.9(2)
Arginine	0.7(1)		1.1(1)	1.1(1)	1.2(1)		1.2(1)		1.0(1)			1.0(1)	3.2(3)			1.1(1)	2.1(2)	5.7(4)
Tryptophan	nd(3)		nd(1)			nd(1)			nd(1)	nd(1)	nd(1)	nd(1)		nd(1)			nd(3)	nd(2)
Total	(31)	(5)	(6)	(4)	(14)	(19)	(4)	(4)	(16)	(5)	(9)	(10)	(27)	(9)	(8)	(5)	(36)	(53)
Pos. in sequence	3–33	34–38	39–44	45–48	49–62	63–81	82–85	86–89	1–16	17–21	22–30	31–40	41–67	68–76	77–84	85–89	1–36	37–89

+, present, but not determined quantitatively nd, not detected. The overlapping peptide E3-Th has only be sequenced (no amino acid analysis).

The sequence similarity between the two domains of UDA is not higher than those between each of the domains and the other hevein-like domains presented in Fig. 2. This supports the hypothesis [20] that UDA has two carbohydrate binding sites with intrinsically different affinities for ligand molecules. It is also noteworthy that the hinge region between the two domains is four residues longer than those between the cereal lectins (Fig. 2) and form a more flexible connection.

Hevein is formed by post-translational processing [21] from a precursor with an open reading frame of 204 amino acids and a putative signal sequence of 17 amino acid residues followed by the hevein domain of 43 residues and a carboxyl-terminal domain of 144 residues

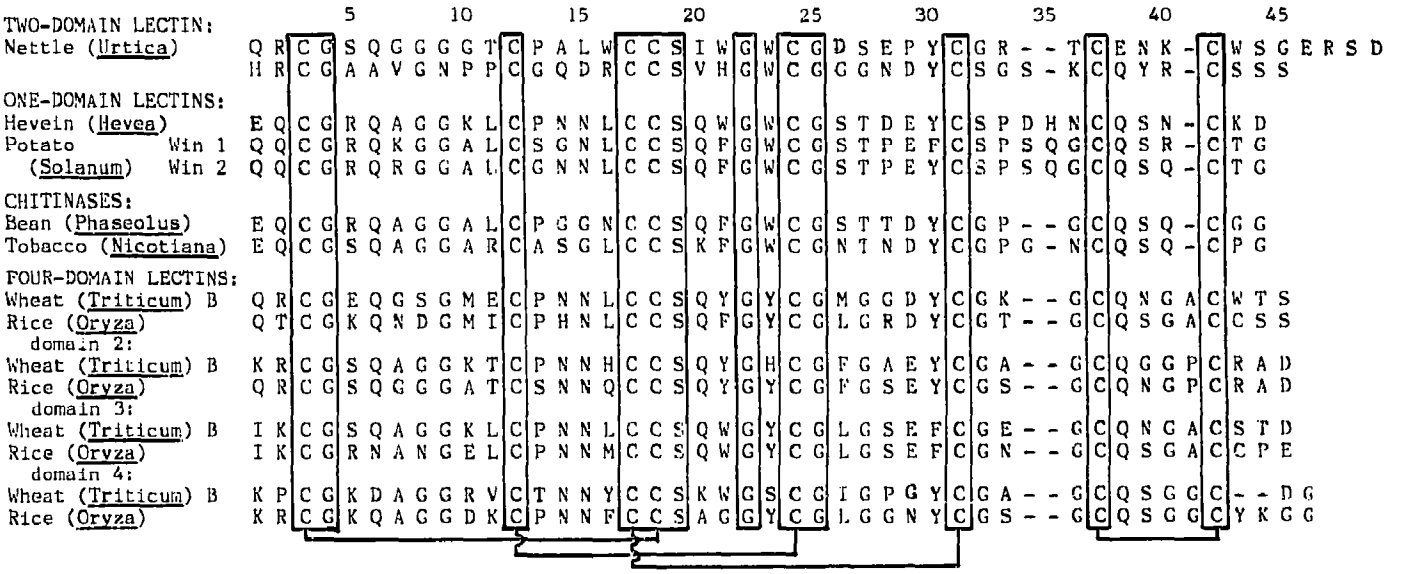

Fig. 2. Alignment of sequences of hevein-like domains of several proteins. Conserved residues are enclosed in blocks. Connections between half-cystine residues as determined by X-ray diffraction of wheat germ agglutinin [12] and hevein [13] and 2D NMR spectroscopy of hevein (K. Dijkstra, R.M. Scheek and J.J. Beintema) are indicated. Experimental evidence for the sequence of hevein has been presented by Walujono et al. [4], except for residues 34–36. These residues have been identified by performic acid oxidation of hevein — which converts the N-terminal glutamate to pyroglutamate — and then cleavage by endoproteinase Glu-C at position 29 only and automated sequence analysis of the unfractionated digest (P. Wietzes and J.J. Beintema). Other sequences: Win 1 and Win 2 [14], bean chitinase [15], tobacco chitinase [16], wheat germ agglutinin [5], rice lectin [17], stinging nettle lectin (this paper).

[22]. The Win genes of potato have open reading frames that are 75% identical with the hevein preproprotein [22] but these gene products have not yet been investigated at the protein level. Basic chitinases with N-terminal hevein-like domains do not undergo post-translational processing like the hevein precursor. The C-terminal domains have no sequence similarities to that of the hevein precursor. Basic chitinases accumulate in vacuoles. Plants that produce basic chitinases with N-terminal hevein-like domains generally also produce extra-cellular acidic chitinases (pathogenesis-related proteins), which are homologous with the basic ones but lack the hevein-like domains [7]. Cereal lectins are formed by post-translational cleavage of a short glycosylated carboxyl-terminal domain [23].

Very recently it has been communicated that the gene for UDA contains an open reading frame of 373 amino acids coding for a putative signal peptide, followed by two hevein-like domains, a spacer domain and a carboxyl domain with 45% sequence identity to the basic and acid chitinases mentioned above [24].

Acknowledgements: We thank Mr. Peter Jekel for manual sequencing of peptides, Mr. Perry Overkamp and Dr. W.W. de Jong, Laboratorium for Biochemie, Katholieke Universiteit Nijmegen, for performing amino acid analyses, and Dr. R.N. Campagne for critically reading the manuscript.

REFERENCES

[1] Peumans, W.J., De Ley, M. and Broekaert, W.F. (1983) FEBS Lett. 177, 99–103.
[2] Broekaert, W.F., Van Parijs, J., Leyns, F., Joos, H. and Peumans, W.J. (1989) Science 245, 1100–1102.
[3] Chapot, M.-P., Peumans, W.J. and Strosberg, A.D. (1986) FEBS Lett. 195, 231–234.
[4] Walujono, K., Scholma, R.A., Beintema, J.J., Mariono, A. and Hahn, A.M., in: Proceedings of the International Rubber Conference, Vol. 2, Rubber Research Institute of Malaysia, Kuala Lumpur, 1975, pp. 518–531.
[5] Wright, C.S., Gavilanes, F. and Peterson, D.L. (1984) Biochemistry 23, 280–287.
[6] Chrispeels, M.J. and Raikhel, N.V. (1991) Plant Cell 3, 1–9.
[7] Meins, F., Neuhaus, J.M., Sperisen, C. and Ryals, J. in: Genes Involved in Plant Defense (T. Boller and F. Meins Jr., Eds.) Springer Verlag, Wien, in press.
[8] Van Damme, E.J.M. and Peumans, W.J. (1987) Physiologia Plantarum 71, 328–334.
[9] Jost, W., Bak, H., Glund, K., Terpstra, P. and Beintema, J.J. (1991) Eur. J. Biochem. 198, 1–6.
[10] Chang, J.Y. (1983) Methods Enzymol. 91, 455–466.
[11] Van Damme, E.J.M., Broekaert, W.F. and Peumans, W.J. (1988) Plant Physiol. 89, 598–601.
[12] Wright, C.S. (1977) J. Mol. Biol. 111, 439–457.
[13] Rodriguez-Romero, A., Ravichandran, K.G. and Soriano-Garcia, M. (1991) FEBS Lett. 291, 307–309.
[14] Stanford, A., Bevan, M. and Northcote, D. (1989) Mol. Gen. Genet. 215, 200–208.
[15] Broglie, K.E., Gaynor, J.J. and Broglie, R.M. (1986) Proc. Natl. Acad. Sci. USA 83, 6820–6824.
[16] Shinshi, H., Mohnen, D. and Meins, F. (1987) Proc. Natl. Acad. Sci. USA 84, 89–93.
[17] Wilkins, T.A. and Raikhel, N.V. (1989) Plant Cell 1, 541–549.
[18] Peumans, W.J., Stinissen, H.M. and Carlier, A.R. (1983) Biochem. Physiol. Pflanzen 178, 423–431.
[19] Van Parijs, J., Broekaert, W.F., Goldstein, I.J. and Peumans, W.J. (1991) Planta (Berl.) 183, 258–264.
[20] Shibuya, N., Goldstein, I.J., Shafer, J.A., Peumans, W.J. and Broekaert, W.F. (1986) Arch. Biochem. Biophys. 249, 215–224.
[21] Lee, H.-I., Broekaert, W.F. and Raikhel, N.V. (1991) J. Biol. Chem. 266, 15944–15948.
[22] Broekaert, W., Lee, H.-I., Kush, A., Chua, N.-H. and Raikhel, N. (1990) Proc. Natl. Acad. Sci. USA 87, 7633–7637.
[23] Raikhel, N.V. and Wilkins, T.A. (1987) Proc. Natl. Acad. Sci. USA 84, 6745–6749.
[24] Lerner, D.R. and Raikhel, N.V. (1991) 3rd Int. Congress for Plant Mol. Biol., Tucson, Arizona, USA, 6–11 October 1991; Abstract 1356.