Extensive homologies between lectins from non-leguminous plants

M.-P. Chapot, W.J. Peumans* and A.D. Strosberg

Laboratory of Molecular Immunology, Jacques Monod Institute, CNRS and University Paris VII, 2, Place Jussieu, F75251 Paris Cedex 5, France and **Laboratorium voor Plantenbiochimie, Katholieke Universiteit Leuven, Kardinaal Mercierlaan, 92, B-3030 Leuven, Belgium*

Received 27 September 1985, revised version received 18 November 1985

Sequence studies were performed on lectins from two non-leguminous plants: rice and nettle. Extensive homologies were found between these two proteins and wheat germ agglutinin in support of the conservation of lectin sequences among non-leguminous plants. The number and positions of the cysteine residues were particularly well conserved suggesting a similar folding of the polypeptide chains.

Lectin *Sequence homology* *Non-leguminous plant* (Oryza sativa, Urtica dioica)

1. INTRODUCTION

The extensive homology between the lectins of leguminous plants is now well established [1]. This homology, first observed for the amino terminal part of these one- and two-chain proteins, has been extended over the whole length of several lectins on the basis of data obtained through protein [2–5] and DNA [6–8] sequencing.

Hitherto such homologies have only been reported for the lectins from leguminous plants. Lectins from other plants have been studied to a lesser extent, except for WGA which has been sequenced completely [9] and for which a tridimensional structure has been proposed [10]. The partial sequences of the lectins from rice and nettle, which we report here, suggest the existence of extensive homologies among lectins from cereals and other groups of non-leguminous plants.

Abbreviations: TFA, trifluoroacetic acid; TPCK, L-1-tosylamido-2-phenylethyl chloromethyl ketone; WGA, wheat germ agglutinin

2. MATERIALS AND METHODS

2.1. *Rice lectin*

Rice lectin was extracted from rice embryos *(Oryza sativa)* and purified on a column of immobilized *N*-acetylglucosamine (selectin 1 from Pierce, Rockford, IL) [11]. This lectin is composed of 2 identical subunits of M_r 18 000, which are partially cleaved into 2 smaller polypeptides of M_r 10 000 and 8000, respectively [12].

2.2. *Nettle lectin*

Nettle lectin was obtained from the rhisomes of stinging nettle *(Urtica dioica)*. It was first purified on a column of chitin, then applied on a column of sulfopropyl-Sephadex (SP Sephadex, type C50 from Pharmacia, Uppsala) [13]. This lectin appeared as a protein of M_r 9000 on SDS-PAGE.

2.3. *Tryptic hydrolysis*

The reduced and *S*-carboxymethylated (with iodo[^{14}C]acetic acid) lectin was digested at 37°C with TPCK-treated trypsin in 1% $NH_4(HCO_3)$ buffer, during 3 h. The enzyme/lectin ratio was 1:100 (w/w). The same amount of trypsin was

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

added a second time and reincubated under the same conditions.

2.4. *High-performance liquid chromatography (HPLC)*

Reverse-phase chromatography was performed with a Waters system (Waters, Milford, USA) equipped with two model 6000A pumps, a variable UV detector, using a Brownlee RP 300–10 μm C8 column. Detection was made at 206 nm. The peptides were eluted with a linear gradient made of aqueous 0.1% TFA (solvent A) and 0.1% TFA in 80% acetonitrile (solvent B). The elution started with 100% A and 0% B and ended after 3.5 h with 45% B. The flow rate was 0.8 ml/mn.

2.5. *Removal of pyroglutamic acid from the amino-terminus of proteins using pyroglutamate aminopeptidase*

A solution of nettle lectin was dialysed against the following buffer: 0.1 M Na_2HPO_4, 5 mM DTT, 5% glycerol (v/v), pH 8.0. The enzyme (from calf liver, Boehringer Mannheim)/lectin ratio was 7:100 (w/w). The reaction was performed in a screw-top vial flushed with nitrogen, capped and mixed, at 4°C for 9 h with occasional mixing. The mixture was then brought to room temperature, the same quantity of enzyme added, the vial flushed with nitrogen again and mixed. This solution was incubated at room temperature for 14 h with occasional stirring, dialysed against 0.05 M acetic acid and lyophilised.

2.6. *Amino acid sequence determination*

Sequence analyses were performed by automated Edman degradation on 0.5–1 nmol protein and peptides with a gas-phase Applied Biosystems sequencer. Phenylthiohydantoins were identified by HPLC using a Waters system and a Zorbax ODS column (5 μm, Dupont), using a linear gradient of 5 mM $NH_4(H_2PO_4)$ (solvent A) and 5 mM $NH_4(H_2PO_4)$ in 60% acetonitrile (solvent B). The separation was achieved at 55°C starting with 80% solvent A-20% solvent B up to 45% solvent A-55% solvent B.

3. RESULTS AND DISCUSSION

3.1. *Amino-terminal sequence of rice lectin*

The direct automated Edman degradation of rice lectin yielded an amino-terminal sequence which was found to be almost identical to a portion of WGA starting at residue 95 (fig.1). The comparison of the next 33 residues revealed only 8 differences, most of which corresponding to substitions between homologous residues. This unexpected similarity was easily explained by assuming that rice lectin was originally synthesized as a

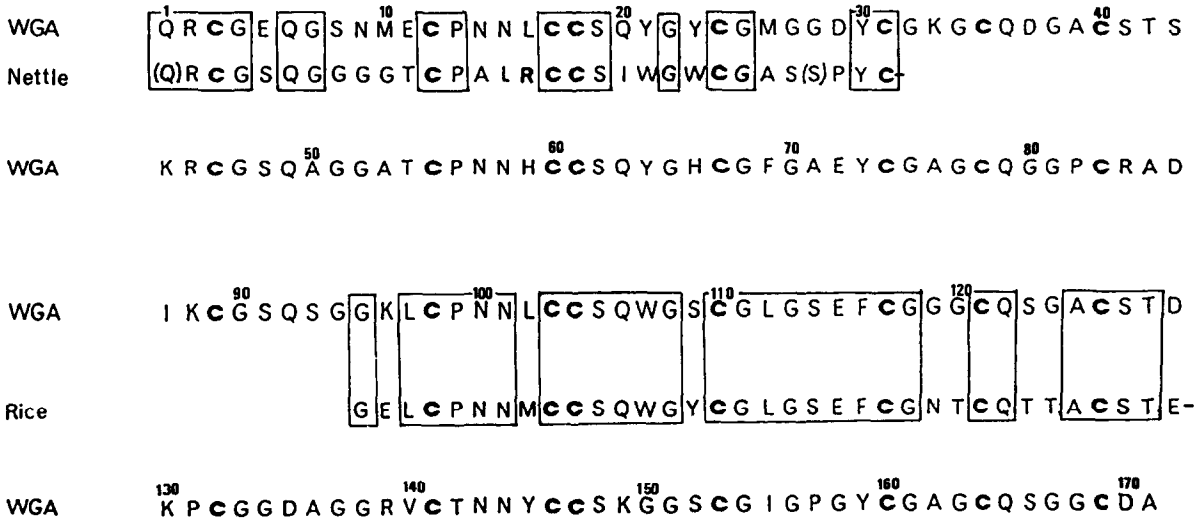

Fig.1. Comparison of the amino-terminal sequences of the rice and nettle lectins with the complete sequence of a monomer of WGA.

polypeptide chain of the same length as WGA. This single chain would have an N-terminal glutamine, as in WGA and this residue would, through cyclisation, form pyrrolidone carboxylic acid (PCA), a residue which is not accessible to Edman degradation. Post-synthetic degradation of the rice lectin chain would yield two fragments of almost equal size, one of which comprising residues 1–94, the other 95–171. This second fragment was sequenced directly. The hypothetical proteolytic cleavage would occur right between 2 of the 4 domains homologous to those existing in WGA (fig. 2) and characterized by numerous well conserved disulfide bonds.

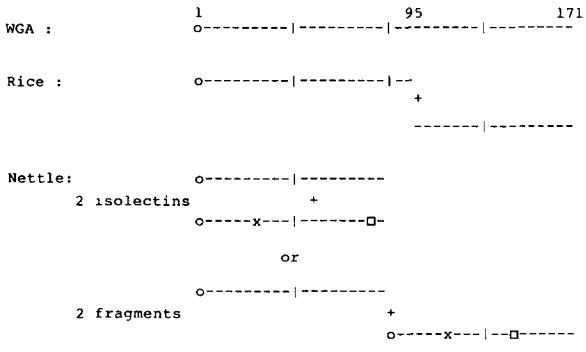

Fig.2. Schematic representation of the monomer of WGA organized in 4 domains; comparison with rice and nettle lectin. The symbol o signifies a PCA-blocked residue.

3.2. *Sequence determination of nettle lectin*

3.2.1. Amino-terminal sequence

As in WGA and rice lectin, nettle lectin appeared to be inaccessible to direct Edman degradation. However after treatment with the enzyme pyroglutamate aminopeptidase, an amino-terminal sequence could be determined, which was again homologous to that of WGA (fig.1), with 16 residues identical over 31 positions, among which were 6 cysteine residues most likely involved in disulfide bonds.

3.2.2. Sequences of tryptic peptides of nettle lectin

Five tryptic peptides well separated by HPLC (fig.3) were sequenced and positioned along the WGA sequence by using the cysteines as references (fig.4).

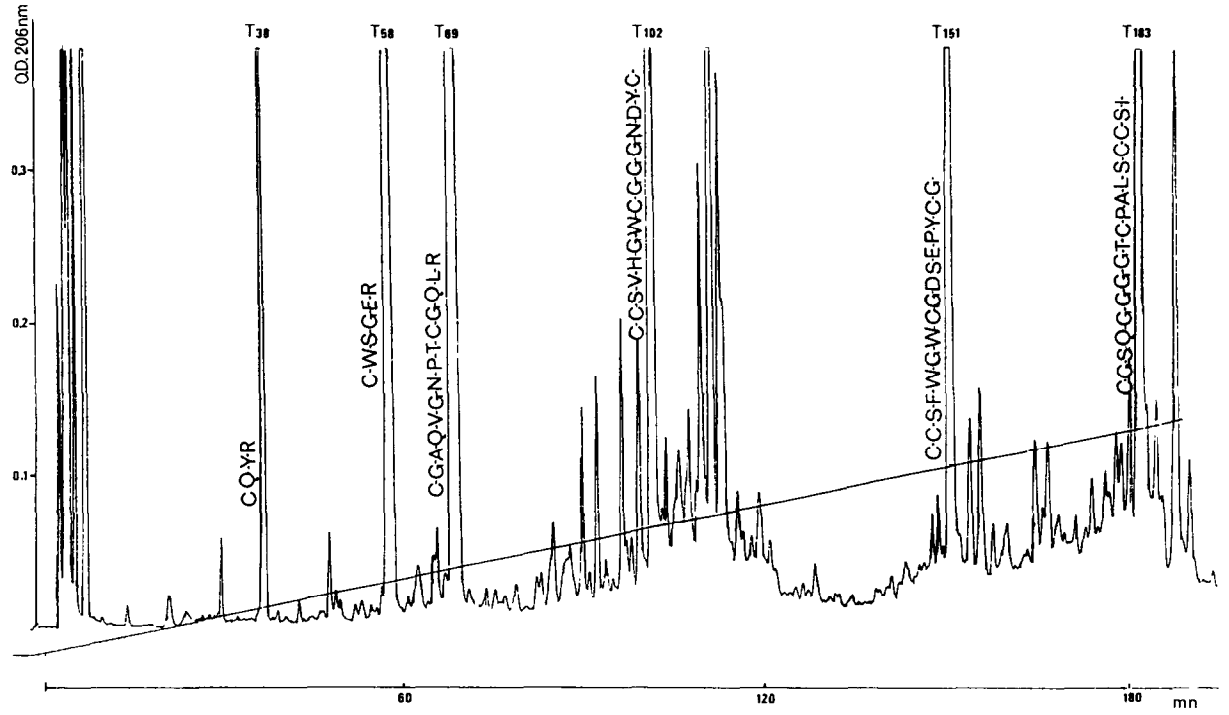

Fig.3. HPLC separation of tryptic peptides of the nettle lectin. Sequences obtained for the different peptides are indicated on the chromatogram, adjoining the corresponding major peaks.

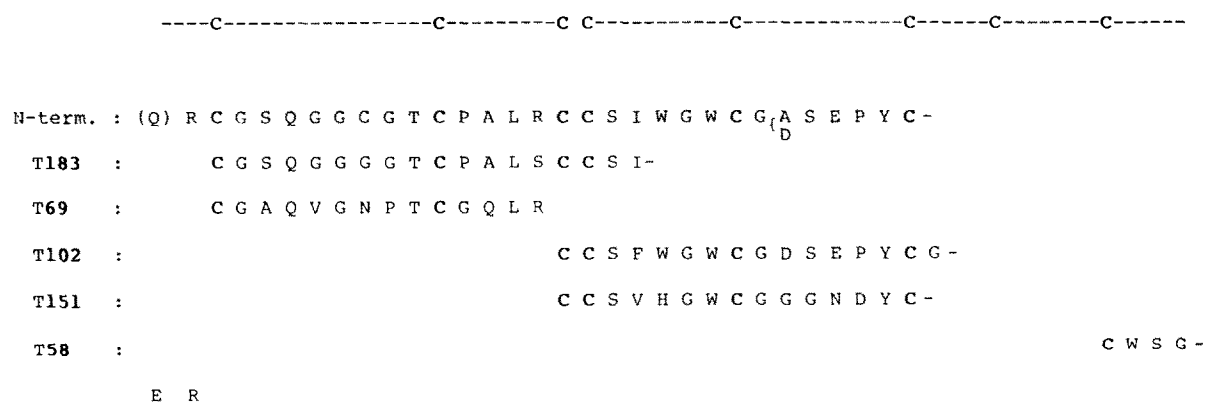

Fig.4. Sequences of the amino-terminal region and of tryptic peptides from the nettle lectin. The peptides were positioned along a single domain of WGA using the cysteine residues as reference positions.

The sequences of the peptides T183 and T151 corresponded, respectively, to positions 3–20 and 17–31 of the amino-terminal sequence of nettle lectin, except for one position for each.

Two explanations, non-mutually exclusive, may be proposed for the occurrence of these variations (fig.2): these would correspond either to substitutions between isolectins of 9 kDa or to sequence differences between 2 fragments of 9 kDa of an original lectin of 18 kDa. In contrast with the rice lectin, both fragments would have blocked amino-terminal PCA residues.

ACKNOWLEDGEMENTS

This work was supported by grants from CNRS, INSERM and the Nitrogen Fixation contract. W.J.P. is Senior Research Assistant of the National Fund for Scientific Research (Belgium).

REFERENCES

[1] Strosberg, A.D., Buffard, D., Lauwereys, M. and Foriers, A. (1986) in: The Chemistry and Biology of Lectins (Liener, I.E. et al. eds) Academic Press, New York, in press.
[2] Edelman, G.M., Cunningham, B.A., Reeke, G.N., Becker, J.W. and Waxdal, J.L. (1972) Proc. Natl. Acad. Sci. USA 69, 2580.
[3] Hemperly, J.J., Hopp, T.P., Becker, J.W. and Cunningham, B.A. (1979) J. Biol. Chem. 254, 5550.
[4] Foriers, A., Lebrun, E., Van Rappenbusch, R., De Neve, R. and Stosberg, A.D. (1981) J. Biol. Chem. 256, 5550.
[5] Young, N.M., Williams, R.E., Roy, C. and Yaguchi, M. (1982) Can. J. Biochem. 743, 212.
[6] Vodkin, L.O., Rhodes, P.R. and Goldberg, R. (1983) Cell 34, 1023.
[7] Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C. and Spencer, D. (1983) J. Biol. Chem. 258, 9544.
[8] Hoffman, L.M., Ma, Y. and Barker, R.F. (1982) Nucleic Acids Res. 10, 7819.
[9] Wright, C.S., Gavilanes, F. and Peterson, D.L. (1984) Biochemistry 23, 280–287.
[10] Wright, C.S. (1977) J. Mol. Biol. 111, 439–457.
[11] Peumans, W.J., Stinissen, H.M. and Carlier, A.R. (1983) Biochem. Physiol. Pflanzen 178, 423–431.
[12] Peumans, W.J. and Stinissen, H.M. (1983) in: Chemical Taxonomy, Molecular Biology and Function of Plants Lectins (Goldstein, I.J. and Etzler, M. eds) vol. 138, p.99, Alan Liss, New York.
[13] Peumans, W.J., De Ley, M. and Broekaert, W.F. (1984) FEBS Lett. 177, 99–103.
[14] Podell, D.N. and Abraham, G.N. (1978) Biochem. Biophys. Res. Commun. 81, 176–185.