EXTENSIVE SEQUENCE HOMOLOGIES AMONG LECTINS FROM LEGUMINOUS PLANTS

André Foriers*, Christian Wuilmart°, Nathan Sharon§ and A. Donny Strosberg*

Laboratories of Medical and Special Biochemistry, Protein Chemistry and Biochemical Pathology*, Free University of Brussels (V.U.B.), Belgium, and Departments of Chemical Immunology° and Biophysics§, The Weizmann Institute of Science, Rehovoth, Israel

Received February 7,1977

SUMMARY

The first twenty five residues of the amino terminal sequence of the β chains from lentil and pea lectins, of soybean and peanut agglutinins, and of the R and L subunits of phytohemagglutinin (PHA) were compared. Extensive homologies were found, ranging from near identity in the case of the β chains of lentil and pea lectins, to 24% identity between soybean agglutinin and L-PHA (assuming two deletions in the latter). Despite different sugar binding specificities, a common ancestry for the genes coding for leguminous lectins appears to be very likely.

INTRODUCTION

Lectins constitute a group of cell-agglutinating and sugar binding proteins, generally extracted from plants but also from lower invertebrates and even from mammals (1,2). They are being used to an increasing extent as probes for the investigation of the structure and function of cell surface saccharides, both on normal and malignant cells. Although lectins have been extensively studied during the last decade, little is known about their structure. Only for one lectin, concanavalin A, were both the sequence and the three-dimensional structure determined (3). We have undertaken to compare the primary structure of several other lectins from leguminous plants, to see whether in this large and seemingly diverse group of proteins, any structural similarities could be detected.

We report here extensive homologies between the segments of the first 25 residues of the amino terminal sequences of the β polypeptide chains from lentil and pea lectins, of soybean and peanut agglutinins, and the R and L chains of phytohemagglutinin from *Phaseolus vulgaris* (PHA) (4). These

ISSN 0006-291X

homologies suggest a common ancestry for the genes coding for these lectins and indicate that lectins may have an important physiological role in plants.

MATERIALS AND METHODS

The lectins from pea (5) and lentil (6) (both specific for D-mannose and D-glucose), soybean agglutinin (7) (specific for *N*-acetyl-D-galactosamine), and peanut agglutinin (8) (specific for D-galactose) were prepared by affinity methods as described in the literature. The α and β polypeptide chains of pea and lentil lectin were separated on Sephadex G-75 in 6 M guanidine-HCl as described (9,10).

Amino acid sequences were determined on a Beckman 890 C Sequencer using a modified 0.1 M Quadrol protein microsequence program (11). The phenylthiohydantoin derivatives were identified by gas-liquid chromatography (12), thin-layer chromatography on polyamide sheets (13) and, after back hydrolysis of the phenylthiohydantoin derivatives using HI at 150°C in vacuo for 24 hours, by amino acid analysis on a Durrum-500 Analyzer (14).

Cyanogen bromide cleavage of soybean agglutinin was carried out according to Gross and Witkop (15). The two fragments obtained were separated by gel filtration on Sephadex G-100 in 1 M acetic acid. The smaller fragment consisted of 18 amino acids, confirming the presence of a single methionine per subunit in the protein (7) and showing that this residue occupies position 18 in the chain.

RESULTS AND DISCUSSION

Pea and lentil lectins are composed of two chains, α and β, which were separated by gel filtration under denaturing conditions. The amino terminal sequences of the first 25 amino acids of the two subunits were reported previously (9,16). The amino terminal sequences of the β chains of the lentil and pea lectins, and of soybean and peanut agglutinins, are given in Table I. The soybean agglutinin sequence was determined on the intact protein and confirmed by analysis of the two fragments obtained by cyanogen bromide cleavage. It is striking that there are only two differences between the β chains of the pea and lentil lectins. The *N*-terminal sequence of soybean agglutinin is identical at 11 positions with that of the β chain of the lentil lectin; among the 14 non-identical residues, 9 could have resulted each from a single nucleotide substitution. In the case of peanut and soybean agglutinins, 11 out of the first 25 amino acids are identical, and 8 additional ones (Asn-7, Ala-17, Ile-18, Asn-19, Phe-20, Gln-21, Val-24 and Thr 25) in the peanut agglutinin could again have arisen each by a single nucleotide substitution. Considerable homologies

Table I

Amino terminal sequences of the β chains from lentil and pea lectins, the agglutinins of soybean and peanut, of the R and L subunits of PHA (4) and of concanavalin A (3). Homologies are indicated by a solid line. Deletions [ ] were introduced to maximize homology.

	1							10												20					25
Lentil β	Thr	Glu	Thr	Thr	Ser	Phe	Ser	Ile	Thr	Lys	Phe	Ser	Pro	Asp	Gln	Gln	Asn	Leu	Ile	Phe	Gln	Gly	Asp	Gly	Tyr
Pea β	———	———	———	———	———	———	———	Leu	———	———	———	———	———	———	———	———	———	———	———	———	———	———	———	Asn	———
Soybean	Ala	———	Val	———	———	———	———	Trp	Asn	———	Val	——	Lys	Glu	Pro	Asp	Met	——	Leu	Glu	———	———	———	Ala	Ile
Peanut	Ala	———	Val	———	Asn	Phe	Asn	Ser	———	Glu	Gly	Asn	Pro	Ala	Ile	Asn	———	———	———	———	———	———	———	Val	Thr
R-PHA	Ala	Ser	Glu	———	———	———	———	Phe	Glu	Arg	——	Asn	Glu	Thr	[	]	———	———	Leu	——	Arg	——	Ala	Ser	
L-PHA	Ser	Asn	Asp	Ile	Tyr	——	Asn	Phe	Glu	Arg	——	Asn	Glu	Thr	[	]	———	———	Leu	——	Arg	——	Ala	Ser	
Con A	Ala	Asp	——	Ile	Val	Ala	Val	Glu	Leu	Asp	Thr	Tyr	——	Asn	Thr	Asp	Ile	Gly	Asp	Pro	Ser	Tyr	Pro	His	Ile

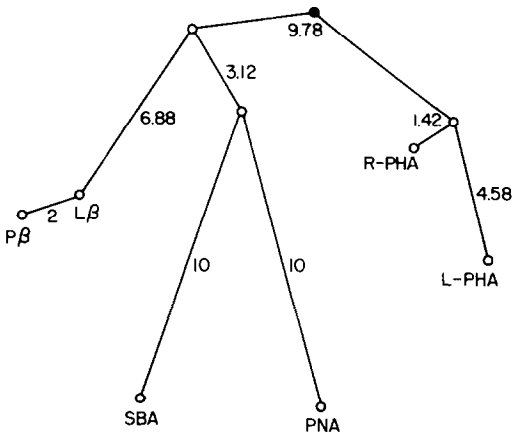

Figure 1 A phylogenetic tree as reconstructed from the values given in Table II of the different lectin sequences; the length of each leg is given. Abbreviations: Lβ and Pβ, β chains of lentil and pea lectins, respectively; SBA, soybean agglutinin; PNA, peanut agglutinin.

are also evident between these sequences and those of the R and L chains of PHA reported by Miller *et al.* (4). These sequences are, however, markedly different from the *N*-terminal sequence of concanavalin A (3) or of ricin (17).

The minimal number of nucleotide substitutions (MR) needed to interconvert the first six lectins presented in Table I are summarized in Table II. The very extensive homologies between the different lectins analyzed by us, as well as those reported by Miller *et al.* (4), strongly suggest a common genetic origin for these six polypeptide chains. A tentative phylogenetic tree was derived from Table II, based on the assumption that the different molecules evolved at a similar rate. The results are presented in Fig. 1. It is obvious that there is a close agreement between the MR values derived from the phylogenetic tree except when R-PHA is taken into account; such a deviation may only be attributed to the fact that the R-PHA and L-PHA genes are paralogous and therefore may have corresponding regions characterized by different rates of evolution.

Previous examples in which extended *N*-terminal portions appeared to be

Table II

Minimal number of nucleotide substitutions needed to interconvert the different lectin sequences (below the diagonal). The deletion introduced in the PHA sequences was taken into account (18,19) in order to construct the phylogenetic tree shown in Fig. 1. The values recalculated from the phylogenetic tree are presented above the diagonal. All values have been rounded off.

	Lentil β	Pea β	Soybean	Peanut	R-PHA	L-PHA
Lentil β		2	20	20	19	22
Pea β	2		22	22	21	24
Soybean	20	22		20	24	27
Peanut	20	22	20		23	26
R-PHA	19	20	22	22		6
L-PHA	22	24	24	22	6	

very similar (20) have shown that conclusions about homologies could later be extended to the whole chain. In this regard, it should be noted that the lentil and pea lectins, soybean and peanut agglutinins, as well as R- and L-PHA have a similar amino acid composition, being rich in hydroxyamino and dicarboxylic amino acids, and devoid of (or very poor in) cysteine. The molecular weight of four of them (soybean agglutinin, peanut agglutinin, R-PHA and L-PHA) is in the same range (110-120,000) and they are all tetramers composed of identical or very similar subunits (2). The pea and lentil lectins have a lower molecular weight (50-60,000) and a different subunit structure. These two lectins are each comprised of two subunits of molecular weight 7-8,000 (the α chains) and two of molecular weight 18,000 (the β chains) (2). It is possible that these α and β chains have been derived by proteolytic cleavage from a single polypeptide, molecular weight 25,000, similar to that found, for example, in soybean agglutinin or peanut agglutinin. This assumption is in line with findings of native cleavage fragments in preparations of concanavalin A (21) and soybean agglutinin (22).

It is also in agreement with our findings that the *N*-terminal sequences of the α chains of the pea and lentil lectins are nearly identical (16) but differ from the *N*-terminal sequences of the corresponding β chains. Concanavalin A prepared from jack bean (also classified as a legume (23)) resembles the above lectins in molecular weight, subunit structure and amino acid composition. Examination of its amino terminal sequence reveals many differences but regions of extensive homology with either the α or the β subunits of pea and lentil lectins have been described (9,16), suggesting that concanavalin A could have a common ancestor with the other leguminous lectins listed in Table II. Lectins derived from plants of other families would most likely possess different primary structures as has already been found for the *N*-terminal sequence of ricin (17). This will probably be the case also for wheat germ agglutinin and the potato lectin, which differ very markedly in their amino acid composition from the lectins obtained from leguminous plants (2).

The extensive homologies between different lectins obtained from a single plant family establish without doubt a common genetic origin for these proteins as depicted in Fig.1. It would appear therefore that lectins can be grouped in families which have conserved their primary structure even though their carbohydrate specificity and some of their biological properties may be different.

ACKNOWLEDGMENTS

This work was supported by grants from the Belgian Government, from the "ASLK Cancer Fund", the "Fonds voor Kollektief Fundamenteel Onderzoek", and by a grant to N.S. from the United States-Israel Binational Science Foundation (Jerusalem, Israel). C.W., on leave from the Laboratoire de Physiologie Animale, Université Libre de Bruxelles, is supported by an EMBO long-term fellowship.

REFERENCES

1. Sharon, N., and Lis, H. (1972) Science 177, 949-959.
2. Lis, H., and Sharon, N. (1977) in "The Antigens" Vol. 4 (ed. M. Sela), Academic Press, New York, pp.429-529.
3. Cunningham, B. A., Wang, J. L., Waxdal, M. J., and Edelman, G. M. (1975) J. Biol. Chem. 250, 1503-1512.

4. Miller, J. B., Hsu, R., Heinrikson, R., and Yachnin, S. (1975) Proc. Nat. Acad. Sci. USA 72, 1388-1391.
5. Trowbridge, I. S. (1974) J. Biol. Chem. 249, 6004-6012.
6. Howard, I. K., Sage, H. J., Stein, M. D., Young, N. M., Leon, M. A., and Dyckes, D. F. (1971) J. Biol. Chem. 246, 1590-1595.
7. Gordon, J. A., Blumberg, S., Lis, H., and Sharon, N. (1972) FEBS Lett. 24, 193-196.
8. Lotan, R., Skutelsky, E., Danon, D., and Sharon, N. (1975) J. Biol. Chem. 250, 8518-8523.
9. Van Driessche, E., Foriers, A., Strosberg, A. D., and Kanarek, L. (1976) FEBS Lett. 71, 220-222.
10. Strosberg, A. D., Foriers, A., Van Driessche, E., Mole, L., and Kanarek, L. (1976) Arch. Int. Physiol. Bioch. 84, 660-661.
11. Brauer, A. W., Margolies, M. N., and Haber, E. (1975) Biochemistry 14, 3029-3035.
12. Pisano, J. J., and Bronzert, T. J. (1969) J. Biol. Chem. 244, 5597-5607.
13. Summers, M. R., Smythers, G. W., and Oroszlan, S. (1973) Anal. Biochem. 53, 624-628.
14. Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971) Biochemistry 10, 4912-4921.
15. Gross, E., and Witkop, B. (1962) J. Biol. Chem. 237, 1856-1860.
16. Foriers, A., Wuilmart, C., Van Driessche, E., De Neve, R., Kanarek, L., and Strosberg, A. D. (1977) FEBS Lett. (in press).
17. Li, S.S.L., Wei, C. H., Lin, J.Y., and Tung, T.C. (1975) Biochem. Biophys. Res. Commun. 65, 1191-1195.
18. Fitch, W. M. (1966) J. Mol. Biol. 16, 9-16.
19. Fitch, W. M., and Margoliash, E. (1967) Science 155, 279-284.
20. Smith, G. P., Hood, L., and Fitch, W. M. (1971) Ann. Rev. Biochem. 40, 969-1012.
21. Wang, J. L., Cunningham, B. A., and Edelman, G. M. (1971) Proc. Nat. Acad. Sci. USA 68, 1130-1134.
22. Lotan, R., Lis, H., and Sharon, N. (1975) Biochem. Biophys. Res. Commun. 62, 144-150.
23. Bailey, L. H. (1942) The Standard Cyclopedia of Horticulture, MacMillan, New York, p.653.