Phytochemistry, Vol 25, No 1, pp. 115-118, 1986.
Printed in Great Britain

0031-9422/86 $3.00 + 0.00

THE AMINO ACID SEQUENCES OF THE α SUBUNITS OF THE LECTINS FROM *LATHYRUS CICERA*, *L. APHACA* AND *L. ARTICULATUS*

B. SOUSA-CAVADA, M. RICHARDSON*‡, A. YARWOOD*, D. PÉRET† and P. ROUGÉ†

Department of Biochemistry, Federal University of Pelotas, Pelotas (RS), Brazil; *Department of Botany, University of Durham, Durham DH1 3LE, U.K.; †Laboratoire de Botanique et Biologie Cellulaire, Faculté des Sciences Pharmaceutiques, Université Paul Sabatier, 34 Chemin des Maraîchers, 31062 Toulouse, France

(*Received* 29 *January* 1985)

Key Word Index—*Lathyrus cicera*; *L. aphaca*; *L. articulatus*; Leguminosae; Vicieae; lectins; amino acid sequences.

Abstract—The amino acid sequences of the α_1 and α_2 subunits of the isolectins (LcL1 and LcL2) from seeds of *Lathyrus cicera* and the α subunits of the lectins from *L. aphaca* and *L. articulatus* were determined by analysis of peptides derived from the proteins by separate digestions with chymotrypsin and the protease from *S. aureus* V8. The α_1 subunit of the *L. cicera* lectin differed from the α_2 form only in having an extra lysine inserted near the C-terminus. The *L. cicera* α chains differed from *L. aphaca* α protein in eight or nine positions and from the *L. articulatus* α subunit in seven or eight positions, whilst the *L. aphaca* and *L. articulatus* proteins also exhibited differences in seven positions. The close homology of these proteins with the α subunits of the lectins from *L. ochrus*, *L. sativus* and *L. odoratus* and various other members of the tribe Vicieae is discussed in relation to the phylogeny of the Leguminosae.

INTRODUCTION

The lectins isolated from seeds of many plants belonging to the Leguminosae exhibit a wide range of different molecular properties including size, charge and the structure of the sugar moiety that they recognize [1]. Despite this fact, it has become increasingly obvious from information on the high degree of homology in their amino acid sequences [2] that they are closely related and probably share common three-dimensional structures [3] and biological functions.

It has been suggested that the amino acid sequences of the lectins are potentially valuable tools for chemotaxonomic classification providing, for example, a useful means of delimiting the various tribes of the Leguminosae [2, 4, 5] and of studying speciation [6]. At present the application of this method is restricted by the small number of complete sequences which are available. As part of a programme to overcome this shortage we recently reported the sequences of the two light (α_1 and α_2) chains of the isolectins LoL1 and LoL2 from *Lathyrus ochrus* (L.) DC. [7] and we now wish to present the complete structures of the α subunits from the lectins of *L. aphaca* L. and *L. articulatus* L., and from the isolectins LcL1 and LcL2 of *L. cicera* L.

RESULTS AND DISCUSSION

Separate digestions of the α subunits of *L. cicera*, *L. aphaca* and *L. articulatus* with chymotrypsin and the protease from *Staphylococcus aureus* V8 yielded suitably overlapping peptides, all of which were readily separated by reverse-phase HPLC, and from which the complete sequences of the α proteins could be deduced by microsequence analysis using the DABITC/PITC double coupling method (Fig. 1).

The observed specificity of the chymotrypsin used in this study was as expected, except that it gave rise to anomalous hydrolyses of the Arg^{21}-Ile^{22} peptide bonds in the *L. cicera* α_1 and α_2 subunits and in the *L. articulatus* protein. The only examples of anomalous cleavages by the *S. aureus* V8 protease were those which occurred at low levels (not shown in Fig. 1) at Ser^3-Tyr^4 and Ser^{39}-Trp^{40} in all four proteins. Such anomalous hydrolyses of Ser-X peptide bonds have been noted previously [8, 9].

The amino acid sequences are shown in Fig. 1 together with the details of the peptides from which they were deduced. The *L. cicera* α_2 subunit and the α proteins from *L. aphaca* and *L. articulatus* all consisted of 53 amino acids, but the *L. cicera* α_1 subunit contained an additional amino acid (Lys) inserted near the C-terminus. The MW calculated from the amino acid sequences were in good agreement with the estimates of the sizes of the subunits made by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) and the sequences were entirely compatible with the amino acid compositions of the subunits. Only one minor example of microheterogeneity was observed and this occurred at position 53 in the *L. cicera* α_1 protein. The major peptides covering this region contained Lys in position 53, but small amounts (< 15%) of other peptides were also isolated which contained Val in this position. The microheterogeneity in this position was also confirmed by the results of digestions of the *L. cicera* α_1 protein with carboxypeptidase A.

The amino acid sequence of the *L. aphaca* α lectin subunit differed from the *L. cicera* subunits in nine (α_1) or eight (α_2) positions, and from the *L. articulatus* α protein in seven positions. Similarly, the *L. articulatus* α lectin also exhibited differences from *L. cicera* subunits in eight (α_1) or seven (α_2) positions.

Figure 2 shows a comparison of the amino acid se-

‡ To whom all correspondence should be addressed.

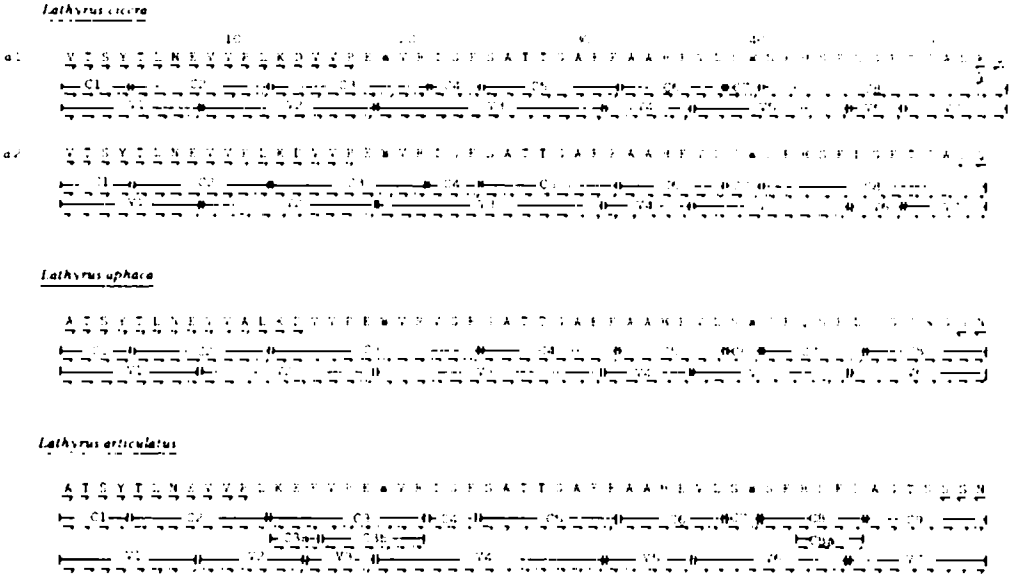

Fig. 1. The amino acid sequences of the α subunits of the lectins from *Lathyrus cicera*, *L. aphaca*, and *L. articulatus*. C, chymotryptic peptides; V, peptides from digestion with protease from *S. aureus* V8. (→) amino acids identified by the DABITC microsequencing method. (- - →) unsatisfactory identification. (←) amino acids determined by digestion with carboxypeptidase A in combination with the dansyl technique.

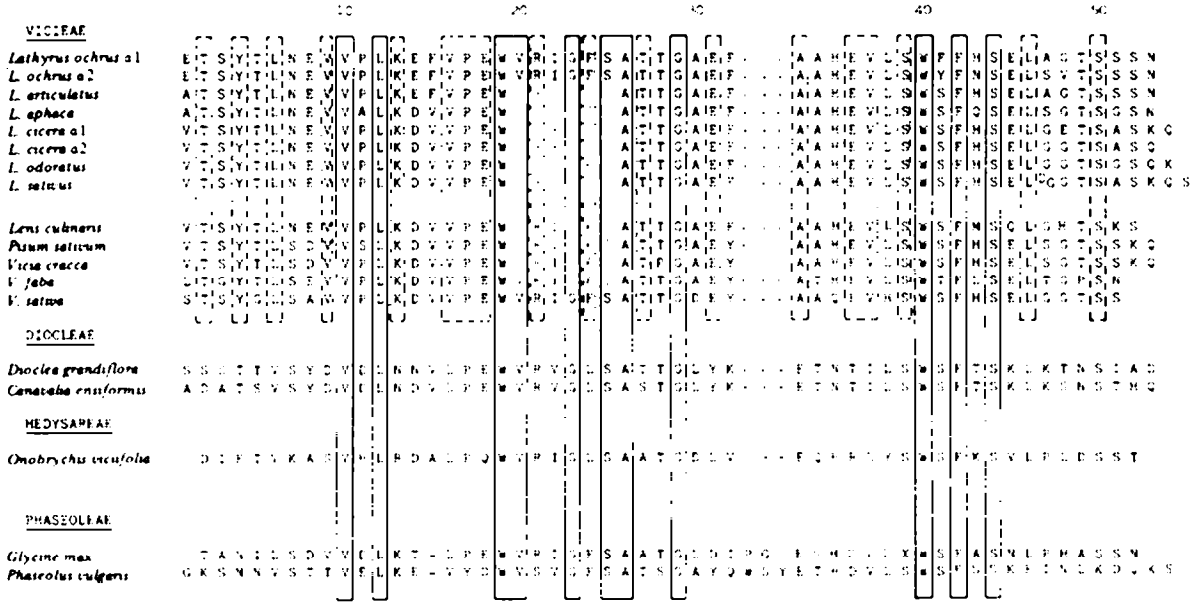

Fig. 2. Homology of the α lectins from the tribe Vicieae and the corresponding regions of the single chain lectins from other tribes of the Leguminosae. The sequences of the *Lathyrus ochrus* α chains are as in ref. [7], *L. odoratus* [10], *L. sativus* [11], *Lens culinaris* [17], *Pisum sativum* [15], *Vicia cracca* [13], *V. faba* [16], *V. sativa* [12], *Dioclea grandiflora* [5], *Canavalia ensiformis* (CON A) [18] *Onobrychis viciifolia* [19], *Glycine max* [20] and *Phaseolus vulgaris* [21]. X, unidentified amino acid; —, gaps inserted in sequence to facilitate alignment. Solid boxes enclose amino acids which are invariant in all sequences. Dashed boxes enclose amino acids which are invariant in all 13 proteins from the tribe Vicieae, but different in at least one other tribe.

quences of the *L. cicera*, *L. aphaca* and *L. articulatus* α lectins with those of the other *Lathyrus* species which have been determined [7, 10, 11]. On the basis of the eight α proteins of the genus *Lathyrus* which have been sequenced so far, it can be seen that 42 out of the 53–55 amino acids are invariant, confirming the previous suggestion from immunological studies [6] that the *Lathyrus* lectins are highly conserved proteins. The figure of 42 would be

higher except for the *L. aphaca* α sequence which is notable in that it contains amino acids in two positions which are different from the otherwise invariant ones in all of the other *Lathyrus* species. These are Ala^{11} (usually Pro) and Val^{22} (usually Ile).

Other workers have noted that the amino acid replacements which occur in the α lectins from the Vicieae tend to cluster in the terminal regions [12, 13]. This tendency is also evident in the α lectins of the genus *Lathyrus* (Fig. 2), but is less obvious because of the alterations in positions 11, 14, 15 and 22 in certain species.

The close homology of the *Lathyrus* α lectins is illustrated in Table 1 which shows the minimal number of nucleotide substitutions required to interconvert these proteins and also gives a quantitative estimate of their relatedness based on the relative substitution frequency factors according to McLachlan [14]. Inspection of these matrices suggests that the species studied may be arranged in three groups on the basis of their sequence homologies. *Lathyrus cicera* and *L. odoratus* L. appear closely related and are both similar to *L. sativus* L. *Lathyrus ochrus* and *L. articulatus* are similarly closely related to one another, but are distinct from the first group. *Lathyrus aphaca* appears to stand alone in showing relatively little affinity with either of the other groups.

An affinity tree is not presented as its significance cannot be assessed on the basis of so few *Lathyrus* α lectin sequences and is likely to be unstable on the addition of new sequence information. When the comparison of the α lectins is extended beyond the genus *Lathyrus* to include the sequences of the α chains of the other five members of the tribe Vicieae (*Pisum sativum* L. [15], *Vicia faba* L. [16], *V. sativa* L. [12], *V. cracca* L. [13] and *Lens culinaris* Medik. [17]), which are currently available, there are still 29 invariant residues (Fig. 2). Of these invariant residues (positions 2, 4, 6, 9, 13, 16, 24, 27, 31, 33, 36 and 37) 12 appear to be characteristic of the tribe Vicieae, in that they are found to be different in the homologous regions of the sequences of the single chain lectins from other tribes, such as the Diocleae (*Canavalia ensiformis* L. [18] and *Dioclea grandiflora* (Mart.) [5]) Hedysareae (*Onobrychis viciifolia* Scop. [19]) and Phaseoleae [*Glycine max* (L.) Merr [20] and *Phaseolus vulgaris* L. [21]].

Further examination of the 18 sequences shown in Fig. 2 reveals that, in 11 positions (10, 12, 19, 20, 23, 25, 26, 29, 40, 42 and 44), the amino acids found are invariant throughout the Leguminosae and that another five positions (17, 21, 39, 46 and 50) are invariant except for alterations in either the soybean (*Glycine*) or *Phaseolus* lectins. Clearly, the full significance of these findings and the value of the apparently invariant residues in delimiting the boundaries of genera and tribes within the family will only be revealed as more complete sequences become available for a wider range of species from different tribes.

EXPERIMENTAL

Purification of Lathyrus α *lectins.* The methods employed in the purification of the proteins were essentially as described in refs [7, 22]. The lectins of *L. cicera*, *L. aphaca* and *L. articulatus* were extracted from crude seed meals with 50 mM Tris, 150 mM NaCl (pH 7.6) buffer and subjected to a fractional pptn with $(NH_4)_2SO_4$. The 30–60% $(NH_4)_2SO_4$ ppts dissolved in the extraction buffer were dialysed against the same buffer and applied to Sephadex G-100 columns equilibrated with the same buffer. The lectins retained by the column were eluted by adding 0.1 M Glc to the eluting buffer, ppted with 90% (w/v) $(NH_4)_2SO_4$ and extensively dialysed against the Tris-buffered saline.

The isolectins of *L. cicera* were separated by chromatofocusing on a column (1 × 30 cm) of PBE 94 (Pharmacia AB) equilibrated with 25 mM Tris–HOAc (pH 8.4). Elution was performed with a 10-fold diluted mixture of Polybuffer 96 (30%) and Polybuffer 74 (70%, Pharmacia AB) adjusted to pH 5.0 with 1 M HOAc. Two main peaks were obtained corresponding, respectively, to LcL1 (*pI* 7.4) and LcL2 (*pI* 6.5). The isolectins were ppted with 90%

Table 1. Comparison of the sequences of the α subunits of the lectins from the genus *Lathyrus*. (a) Minimal number of nucleotide substitutions required to interconvert the α proteins (above the diagonal). (b) Comparison based on McLachlan's [14] relative substitution frequency factors (below the diagonal). Gaps at the C-termini were scored as 0

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
	0	4	4	4	10	11	11	12
(a) *L. cicera* α_1	1							
		0	5	8	7	8	8	9
(b) *L. cicera* α_2	0.973	1						
			0	6	9	9	10	12
(c) *L. odoratus*	0.959	0.959	1					
				0	12	13	13	15
(d) *L. sativus*	0.942	0.944	0.942	1				
					0	7	2	5
(e) *L. articulatus*	0.911	0.928	0.920	0.906	1			
						0	9	9
(f) *L. aphaca*	0.902	0.919	0.925	0.897	0.935	1		
							0	4
(g) *L. ochrus* α_1	0.895	0.912	0.904	0.890	0.977	0.912	1	
								0
(h) *L. ochrus* α_2	0.886	0.902	0.884	0.870	0.944	0.909	0.961	1

(w/v) $(NH_4)_2SO_4$ and extensively dialysed against H_2O.

The α and β subunits of the lectins from *L. aphaca*, *L. articulatus* and of the isolectins LcL1 and LcL2 were separated by gel filtrations on columns of Biogel P-60 equilibrated and eluted with 6 M guanidine–HCl. Two peaks corresponding, respectively, to the heavy β (first peak) and light α (second peak) subunits were obtained for all four lectins. The guanidine–HCl was eliminated by extensive dialysis against H_2O and the ppted subunits were lyophilized.

The purities of the various α subunits were confirmed by PAGE at basic [23] and acidic pH [24], SDS–PAGE [25], analytical iso-electric focusing [26] and by N-terminal analysis by the dansyl chloride and by the 4-*N*,*N*-dimethylaminoazobenzene-4'-isothiocyanate–phenylisothiocyanate (DABITC/PITC) double coupling methods [27].

Enzyme digestions and separation of peptides. Samples (3 mg) of the α subunits were digested separately with chymotrypsin and the protease from *S. aureus* V8 as described in ref. [5]. The mixtures of peptides produced by these methods were resolved by reverse-phase HPLC on a Micropac MCH-10 column (0.4 × 30 cm, Varian) in a Varian model 5000 HPLC apparatus using a linear gradient of 0–70% acetonitrile (HPLC grade S, Rathburn, Scotland, U.K.) in 0.1% trifluoroacetic acid. Peptides were detected by measuring the *A* at 214 nm.

Sequence determination. The intact α subunits and the peptides derived from them by enzymic digestion were subjected to microsequence analysis using the DABITC/PITC double coupling method [27]. The presence of tryptophan in certain peptides was confirmed by staining on paper with *p*-dimethylaminobenzaldehyde. The C-terminal residues of certain peptides and the intact α subunits were determined by digestion with carboxypeptidase A as in ref. [7] in combination with the dansyl technique as in ref. [28]. The amino acid compositions of the α subunits were determined on a Beckman model 119 BL amino acid analyser. Certain peptides were also analysed using a Varian 5000 HPLC fitted with a post-column *o*-phthalaldehyde reactor system as described in ref. [29].

Acknowledgements—We are grateful to the Science and Engineering Research Council, the Smith Kline Foundation and the CNPq (Brazilian National Council for Scientific and Technological Development) for financial support. We also thank J. Gilroy, P. Miller and D. Bown for valuable technical assistance and Professor D. Boulter for the provision of certain facilities.

REFERENCES

1. Lis, H. and Sharon, N. (1977) in *The Antigens* (Sela, M., ed.) Vol. 4, p. 429. Academic Press, New York.
2. Strosberg, A. D., Lauwereys, M. and Foriers, A. (1983) in *Chemical Taxonomy, Molecular Biology and Function of Plant Lectins* (Etzler, M. E., ed.) p. 7. A. R. Liss, New York.
3. Olsen, K. W. (1983) *Biochim. Biophys. Acta* **743**, 212.
4. Kolberg, J., Michaelson, T. E. and Sletten, K. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 655.
5. Richardson, M., Campos, F. D. A. P., Moreira, R. A., Ainouz, I. L., Begbie, R., Watt, W. B. and Pusztai, A. (1984) *Eur. J. Biochem.* **144**, 101.
6. Rougé, P. (1984) *Biochem. Syst. Ecol.* **12**, 47.
7. Richardson, M., Rougé, P., Sousa-Cavada, B. and Yarwood, A. (1984) *FEBS Letters* **175**, 76.
8. Driessen, H. P. C., Herbrink, P., Bloemendal, H. and Jong, W. W. (1983) *Eur. J. Biochem.* **121**, 83.
9. Campos, F. D. A. P. and Richardson, M. (1984) *FEBS Letters* **167**, 221.
10. Sletten, K., Kolberg, J. and Michaelsen, T. E. (1983) *FEBS Letters* **156**, 253.
11. Sletten, K. and Kolberg, J. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 1047.
12. Gebauer, G., Schiltz, E. and Rüdiger, H. (1981) *Eur. J. Biochem.* **16**, 319.
13. Baumann, C. M., Strosberg, A. D. and Rüdiger, H. (1982) *Eur. J. Biochem.* **122**, 105.
14. McLachlan, A. D. (1971) *J. Mol. Biol.* **61**, 409.
15. Richardson, C., Behnke, W. D., Freisheim, J. H. and Blumenthal, K. M. (1978) *Biochim. Biophys. Acta* **537**, 310.
16. Hopp, T. P., Hemperly, J. J. and Cunningham, B. A. (1982) *J. Biol. Chem.* **257**, 4473.
17. Foriers, A., de Neve, R., Kanarek, L. and Strosberg, A. D. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1136.
18. Cunningham, B. A., Wang, J. L., Waxdal, M. J. and Edelman, G. M. (1975) *J. Biol. Chem.* **250**, 1503.
19. Kouchalakos, R. N., Bates, O. J., Bradshaw, R. A. and Hapner, K. D. (1984) *Biochemistry* **23**, 1824.
20. Hemperly, J. J. and Cunningham, B. A. (1983) *TIBS* **8**, 100.
21. Hoffman, L. M., Ma, Y. and Barker, R. F. (1982) *Nucl. Acids Res.* **10**, 7819.
22. Rougé, P. and Sousa-Cavada, B. (1984) *Plant Sci. Letters* **37**, 21.
23. Davis, B. J. (1964) *Ann. N.Y. Acad. Sci.* **121**, 404.
24. Reisfeld, R. A., Lewis, U. J. and Williams, D. E. (1962) *Nature* **195**, 281.
25. Swank, R. T. and Munkres, K. D. (1971) *Analyt. Biochem.* **39**, 462.
26. Wrigley, C. (1968) *Sci. Tools* **15**, 17.
27. Chang, T. Y., Brauer, D. and Wittman-Liebold, B. (1978) *FEBS Letters* **93**, 205.
28. Richardson, M. (1974) *Biochem. J.* **137**, 101.
29. Campos, F. A. P. and Richardson, M. (1983) *FEBS Letters* **152**, 300.