THE SUBUNIT STRUCTURE AND N-TERMINAL SEQUENCES OF THE α- AND β-SUBUNITS OF THE LENTIL LECTIN (*LENS CULINARIS*)

A. FORIERS, E. VAN DRIESSCHE, R. DE NEVE, L. KANAREK and A. D. STROSBERG

Laboratoria Scheikunde der Proteïnen, Medische en Speciale Biochemie, Pathologische Biochemie, Vrije Universiteit Brussel, Paardenstraat, 65, B-1640 Sint-Genesius-Rode, Belgium

and

C. WUILMART

Department of Chemical Immunology, Weizmann Institute of Science, Rehovot, Israel

Received 6 January 1977

1. Introduction

Lectins are proteins or glycoproteins present in many plants and are usually recognized by their ability to agglutinate erythrocytes. Agglutination of cells is in many cases inhibited by specific sugars, suggesting that the binding is to sugar residues on the cell surface. Very little is known of the mechanism by which the lectin leads to agglutination or cell-surface alteration. In order to understand the relationship between the activity and the structure of lectins, we have undertaken the determination of the amino acid sequence of several of these proteins. The lectin studied in this work is present in seeds of the lentil, *Lens culinaris* [1,2]. The protein has a molecular weight of approximately 49 000 and is composed of two types of subunits with molecular weights of 18 000 and 8000 [3]. The isolation and characterization of this hemagglutinin was described in our previous communication [4]. We report here the N-terminal sequences of the α- and β-subunits of the lentil lectin. This protein has the same sugar-binding specificity as two other mitogenic lectins, Concanavalin A [5] and pea lectin [6], for which sequence studies were reported previously [7,8]. The comparison of these three proteins reveals surprising homologies.

2. Materials and methods

Lens culinaris phytohemagglutinin (LcH) was isolated from a commercial sample of lentil seeds by the method of Hayman and Crumpton [9]. Gel filtration, to separate the α- and β-subunits was performed at room temperature on Sephadex G-75 in the presence of 6 M guanidine–HCl. Protein in the eluted fractions was estimated spectrophotometrically at 280 nm. The fractions, containing protein were desalted by gel filtration on a Biogel P-2 column.

Polyacrylamide gel electrophoresis in the presence of SDS was carried out following the method of Weber et al. [10]. The gels were stained for protein with Coomassie Brilliant Blue R [11].

Sequence analysis was performed using automated Edman degradations on 0.05–0.1 μmol of the purified subunits with a Beckman 890C sequencer using 0.1 M Quadrol and single acid cleavage [12,13]. Phenylthiohydantoin amino acids were identified by gas–liquid chromatography [14], thin-layer chromatography [15] and amino acid analysis after back hydrolysis with 56% HI [16] on a Durrum-500 analyzer.

The sequences reported here, have been compared on the basis of the minimum number of mutations (*MR*) required to interconvert them. For each align-

ment, *MR* is computed and its probability for occurring by chance is derived from the amino acid composition [17]. If p_0, p_1, p_2, p_3 are respectively the probabilities for obtaining 0, 1, 2 and 3 mutations when 2 randomly selected amino acids are matched, one can define two equations:

$$n = a + b + c + d \qquad \text{(i)}$$

$$MR = b + 2c + 3d \qquad \text{(ii)}$$

(n is the number of amino acid matches of the alignments and a, b, c, d are respectively the number of times 0, 1, 2, 3 mutations occur when comparing the n aligned positions.)

Therefore the probabilities of each combination of a, b, c, d satisfying the equations (i) and (ii) can be derived by the multinomial equation (iii):

$$P_{a,b,c,d} = n! \frac{p_0^a \times p_1^b \times p_2^c \times p_3^d}{a!b!c!d!} \qquad \text{(iii)}$$

and the probability for obtaining by chance the given value of MR ($0 \leqslant MR \leqslant 3n$) is derived by summing the probabilities of the different combinations of a, b, c, d corresponding to MR.

3. Results and discussion

3.1. *Characterization of the lentil lectin and its subunits*

The isolated lectin displayed only one component on polyacrylamide gel electrophoresis, at pH 2.6. When subjected to polyacrylamide gel electrophoresis in the presence of SDS, three bands were visible with

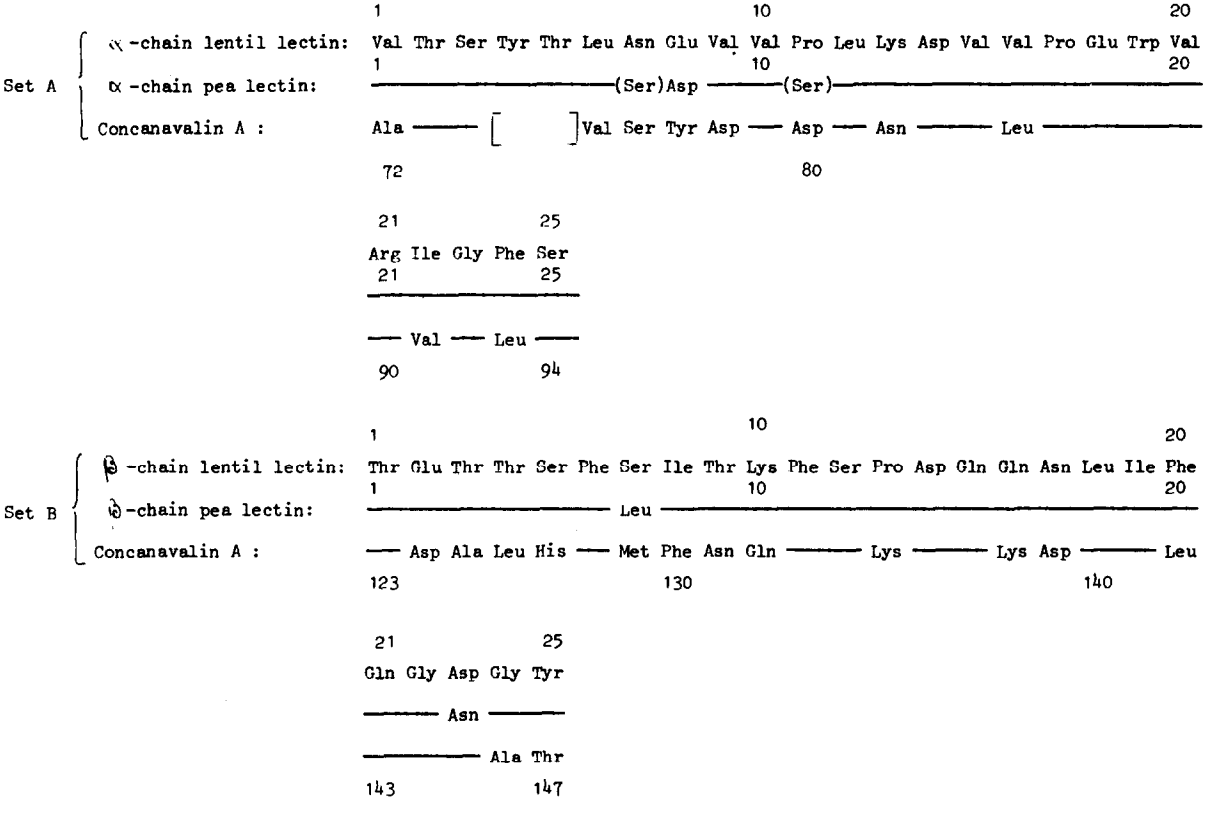

Fig.1. Amino-terminal sequences of the α- and β-chains from lentil and pea lectins as compared to Con A. Homologies are indicated by a solid line. Parentheses ( ) correspond to residues identified by only one method. Deletions [ ] were introduced to maximize homology.

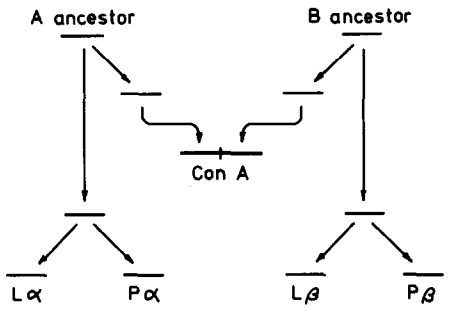

Fig.2. Proposed evolutionary tree for the genes coding for α, β and Con A.

molecular weights of 24 000 (a minor band), 17 000 and 7000. Separation of the subunits by gel-filtration on Sephadex G-75 in 6 M guanidine–HCl revealed two fragments: one heavy chain (β) and one light chain (α) with molecular weights of respectively ± 18 000 and ± 8000.

3.2. *Sequence data*

Automated Edman degradation yielded the N-terminal 25 residue sequences for the α- and β-subunits without any ambiguity in identification as shown in fig.2. No obvious homology was found between the two chains.

However when each was compared to its counterpart in pea lectin [8], it appeared that the N-terminal α-chains of the lentil and pea lectins differed only at 3 positions and the β-chains at 2 positions of the 25 residues analyzed. Homologies of the α-chains (1–25) with the primary structure of Con A is apparent for residues 72–94 (in Con A) from which the pea and lentil α-chains differ at 10 positions (fig.1). The homology of the β-chains with yet another portion of Con A, between the residues 123 and 147 becomes apparent by comparison according to Fitch [17]. The probability of chance homology between the N-terminal portions of the α-chains of the pea and lentil lectins and residues 72–94 of Con A (set A) was found by calculation to be about 1×10^{-6}, and that between the β-chains and residues 123–147 of Con A (set B) about 9×10^{-11} (table 1). The two sets A and B do not show any extensive homology.

The best result is obtained by comparing lentil α- to β-chains but the homology has a probability of 7.8×10^{-2} to occur by chance. The fact that Con A contains two segments which show a high extent of homology, respectively with sets A and B is very interesting. If substantiated by further sequencing, this result means that the actual Con A gene derives from a gene fusion event between the ancestors of sets A and B genes (fig.2). Although being quite rare in evolution, such events have been shown to occur in the case of the tryptophan synthetase gene [18].

Acknowledgements

We thank Ms Marleen Van der Linden, Mr Willy Verheulpen, Mr Ignace Caplier and Mr Urbain Lion for their excellent assistance. This work was supported by grants of the Belgian Government, of the 'ASLK Cancer Fund', the 'Fonds voor Kollektief Fundamenteel Onderzoek' and the 'Fonds voor Onderling Overlegde Aktie'. C. Wuilmart, on leave from the Université Libre de Bruxelles, is supported by an EMBO long-term fellowship.

Table 1
Probability of chance homology between the N-terminal amino acid sequences

	Lentil α	Pea α	Con A_{72-94}	Lentil β	Pea β
Pea α	3.9×10^{-23}				
Con A_{72-94}	1.2×10^{-6}	1.0×10^{-6}			
Lentil β	7.9×10^{-2}	2.4×10^{-1}	2.4×10^{-1}		
Pea β	2.0×10^{-1}	4.5×10^{-1}	2.1×10^{-1}	7.0×10^{-25}	
Con $A_{123-147}$	5.3×10^{-1}	7.7×10^{-1}	6.3×10^{-1}	8.9×10^{-11}	8.9×10^{-11}

References

[1] Entlicher, G., Ticha, M., Kostir, J. V. and Kocourek, J. (1969) Experientia 25, 17.
[2] 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.
[3] Fliegerova, O., Salvetova, A., Ticha, M. and Kocourek, J. (1974) 416–426.
[4] Strosberg, A. D., Foriers, A., Van Driessche, E., Mole, L. E. and Kanarek, L. (1976) Arch. Int. Physiol. Bioch. 84, 660–661.
[5] Powell, A. E. and Leon, A. A. (1970) Exp. Cell. Res. 62, 315–325.
[6] Trowbridge, I. S. (1974) J. Biol. Chem. 249, 6004–6012.
[7] Cunningham, B. A., Wang, J. L., Waxdal, M. J. and Edelman, G. M. (1975) J. Biol. Chem. 250, 1503–1512.
[8] Van Driessche, E., Foriers, A., Strosberg, A. D. and Kanarek, L. (1976) FEBS Lett. 71, 220–222.
[9] Hayman, M. J. and Crumpton, M. J. (1972) Biochem. Biophys. Res. Commun. 47, 923–930.
[10] Weber, K., Pringle, J. R. and Osborn, M. (1972) Method. Enzymol. 26, 3–27.
[11] Weber, K. and Osborn, M. (1969) J. Biol. Chem. 244, 4406–4412.
[12] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[13] Li, S. L., Hanlon, J. and Yanofsky, C. (1974) Biochemistry 13, 1736–1744.
[14] Pisano, J., Bronzert, T. J. and Brewer, H. B. (1972) Anal. Biochem. 45, 43–59.
[15] Summers, M. R., Smythers, G. W. and Oroszlan, S. (1973) Anal. Biochem. 53, 624–628.
[16] Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. and Ballantyne, D. L. (1971) Biochemistry 10, 4912–4921.
[17] Fitch, W. M. (1966) J. Mol. Biol. 16, 1–16.
[18] Bonner, D. M., De Moss, J. A., Mills, S. E. (1965) in: Evolving Genes and Proteins (Bryson V. and Vogel H. eds) pp. 305–318, Academic Press, NY.