Volume 295, number 1,2,3, 207–210 FEBS 10551 December 1991

A new distinct group of 2 S albumins from rapeseed

Amino acid sequence of two low molecular weight napins

Rafael I. Monsalve[1], Carlos Lopez-Otin[2], Mayte Villalba[1] and Rosalía Rodríguez[1]

[1]*Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad Complutense, Madrid, Spain* and [2]*Departamento de Biología Funcional, Facultad de Medicina, Universidad de Oviedo, Oviedo, Spain*

Received 30 July 1991; revised version received 15 October 1991

Two napins (nIa and nIb), isolated from *Brassica napus* (rapeseed) seeds, have been sequenced. The two proteins show the common structural pattern of the 2 S albumins, since they are composed of two disulfide-linked chains of different size, yet they exhibit an atypical low molecular weight (12.5 kDa vs. 14.5 kDa of the major napins). High sequence similarity has been found between these 2 proteins, but only 54% similarity can be estimated from their comparison with the 14.5 kDa major napins. Thus, nIa and nIb are considered representatives of a new distinct group of rapeseed napins since all the previously known napins exhibit 95% sequence similarity. Unexpectedly, the similarity increases when compared with the 2 S proteins from other species.

2 S albumin; Rapeseed; Primary structure

1. INTRODUCTION

2 S albumins constitute one of the most abundant components of the total protein isolated from seeds. Clinical relevance has been shown for these storage proteins, since they are associated with the induction of celiac disease, baker's asthma and other allergic responses [1–4]. Moreover, the interest in the 2 S seed proteins and their genes is increasing since they represent a good model for studying both the expression of one family of genes, and protein maturation processes in plant cells [5–8]. More than 16 genes have been estimated to exist for the 2 S albumins of *Brassica napus*. The amino acid sequences deduced from the 5 genes known to date exhibit more than 90% similarity. However, only one mature 2 S albumin from rapeseed has been sequenced [9]. This protein is a 14.5 kDa albumin whose primary structure shows strong similarity with that of the major allergen from mustard (*Sinapsis alba* L.) seeds [4]. In a previous paper [10] we reported the isolation of five 2 S albumins from *B. napus*, called napins. The amino acid compositions and mol. wts. (14 500) of the 3 major components (napins nII, nIII and nIV) closely fit with those deduced from the nucleotide sequences of the known genes. However, 2 other napins (nIa and nIb) of 12.5 kDa showed significant differences in amino acid composition, as well as in their chromatographic and electrophoretic behaviour, when compared with the 14.5 kDa napins. The present paper describes the amino acid sequence determination of these low molecular weight (LMW) napins. According to these results, the existence of a new distinct group of napins is proposed.

Abbreviations: RP, reverse phase; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid; 2-ME, 2-mercaptoethanol; PTH, phenylthiohydantoin; LMW, low molecular weight.

Correspondence address: R. Rodríguez, Departamento de Bioquímica y Biología Molecular, Facultad de Química, Universidad Complutense, 28040 Madrid, Spain.

2. MATERIALS AND METHODS

The napins, nIa and nIb, were isolated from rapeseed by the method previously described [10]. Separation of the small and large chains were carried out after reduction of the disulfide bonds and radioalkylation with iodo-[^{14}C]acetamide [4]. The carboxyamidomethylated chains were chromatographed on a nucleosil C-18 RP-column, using a Beckman HPLC-system with a 25–45% acetonitrile linear gradient in 0.1% (v/v) TFA. The eluent was continuously monitored at both 214 and 280 nm wavelengths.

Amino acid analyses were performed on a Durrum D-500 amino acid analyzer after acid hydrolysis under the conditions previously described [4]. Tryptophan was determined spectrophotometrically [11]. Protein concentrations were measured by amino acid analysis. Polyacrylamide gel electrophoresis under acidic conditions [12] was carried out in 20% (w/v) polyacrylamide gels in 0.9 M acetic acid containing 2.5 M urea. The samples were alternatively reduced with 5% (v/v) 2-ME at 80°C, for 20 min.

Small chains (20–50 nmol) were digested with trypsin and chymotrypsin. Large chains (20–80 nmol) were digested with trypsin, thermolysin, chymotrypsin and SaV8 (endoproteinase Glu-C from *Staphylococcus aureus* V8), and treated with carboxypeptidase Y, pyroglutamate aminopeptidase and CNBr. Tryptic, thermolytic and chymotryptic digestions and treatments with carboxypeptidase Y and CNBr were done as in [4]. Hydrolysis with SaV8 of the large chain was carried out at a protein concentration of 5–10 mg/ml in 0.1 M ammonium bicarbonate, pH 8.0, containing 2 mM EDTA. An enzyme/substrate ratio of 1:50 was used, and the mixture was incubated at

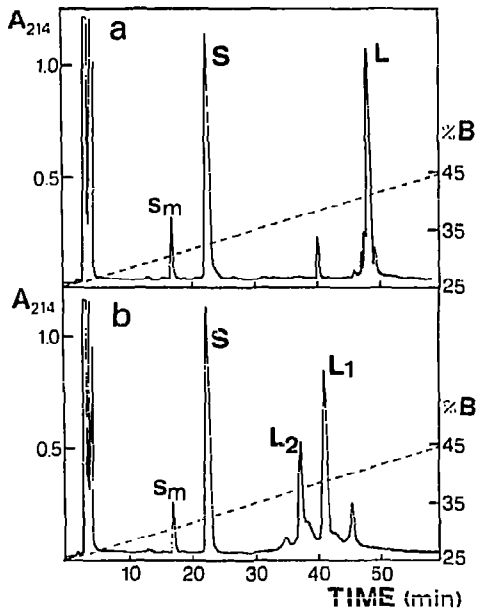

Fig. 1. HPLC elution profiles of the reduced and carboxyamidomethylated LMW-napins. (a) Napin nIa; (b) napin nIb. The small (S, S_m) and large chains (L, L_1, L_2) were eluted from the C-18 column with a gradient of buffer B (0.1% TFA in acetonitrile). The flow rate was 1.0 ml/min.

37°C for 5 h. Alternatively, large chains from both nIa and nIb (10–12 nmol) were dissolved in 0.1 ml of 0.1 M sodium phosphate buffer, pH 8.0, containing 10 mM EDTA, 5% v/v glycerol and 5 mM dithiothreitol, and pyroglutamate aminopeptidase (Boehringer, Germany) (0.02 ml at 10 mg/ml) was added. The reaction mixture was maintained for

Table I

Amino acid composition of nIa and nIb and their polypeptide chains

	Small chains		Large chains			nIa	nIb	pN2
	nIa	nIb	nIa	L1-nIb	L2-nIb			
ASX	0	0	5	5	5	5	5	2
THR	0	0	3	3	3	3	3	4
SER	1	1	3	3	3	4	4	5
GLX	12	12	15	15	15	27	27	30
PRO	2	2	8	6	6	10	8	13
GLY	1	1	5	5	5	6	6	9
ALA	2	2	4	4	4	6	6	7
CYS*	2	2	6	6	6	8	8	8
VAL	0	0	5	5	5	5	5	6
MET	0	0	1	1	1	1	1	3
ILE	1	1	5	5	4	6	6	4
LEU	2	2	4	4	4	6	6	8
TYR	0	0	2	2	2	2	2	1
PHE	2	2	4	2	2	6	4	3
HIS	1	1	1	1	1	2	2	4
LYS	1	1	4	4	4	5	5	9
ARG	3	3	3	3	3	6	6	5
TRP**	1	1	1	1	1	2	2	1
Total	31	31	79	75	74	110	106	122

The deduced amino acid sequence from the pN2 clone of *B. napus* is also included for comparison (according to [5]). The amino acid composition shown for nIb corresponds to that with the L1 large chain.

*Determined as carboxymethylcysteine.
**Determined spectrophotometrically.

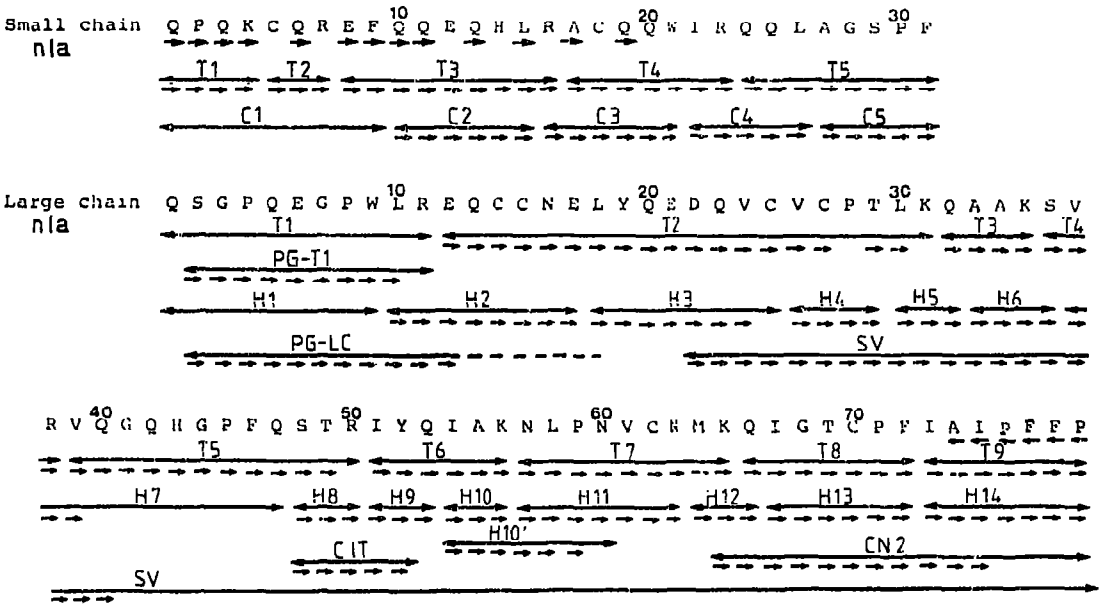

Fig. 2. Primary structure of the small and large chains of the LMW-napin nIa. Peptides obtained by using trypsin (T), chymotrypsin (C and CIT), thermolysin (H), SaV8 endoproteinase (SV), cyanogen bromide (CN), and L-pyroglutamate aminopeptidase (PG) were employed for overlapping. Right-handed arrows indicate residues identified by Edman degradation. Left-handed arrows correspond to amino acids resolved by carboxypeptidase Y-digestion. The amino acid sequence of nIb small chain was identical to that of nIa. The sequences of nIb large chains (L_1 and L_2) are the same as nIa large chain, with the exception of the differences shown in Fig. 3.

2 h at 37°C in order to cleave off the amino terminal pyroglutamic acid. Each digested sample was lyophilized and dissolved in 0.1% (v/v) TFA. The separation was performed by RP-HPLC on a μBondapak C-18 column. Different acetonitrile gradients in 0.1% (v/v) TFA were employed for the elution of peptides. Radioactive peptides were detected by counting aliquots from each collected fraction. An acetonitrile gradient in 10 mM ammonium acetate buffer, pH 6.5, was alternatively used in the HPLC system to avoid cyclization of the N-terminal glutamines.

N-terminal Edman degradations of the polypeptides were performed on an Applied Biosystems model 477A sequencer. The resulting PTH-amino acid derivatives were identified by using a model 120 A on-line PTH-analyzer and the standard Applied Biosystems program.

3. RESULTS AND DISCUSSION

Two LMW-napins (12.5 kDa napins), nIa and nIb, have been isolated from the seeds of *B. napus* (rapeseed) [10]. These napins are minor components since the 14.5 kDa napins constitute 82% of the dry weight of the total 2 S fraction of these seeds. These atypically LMW-albumins have not been found in other Brassicaceae seeds such as *B. rapa*, *B. oleracea*, *B. juncea*, *Raphanus sativus* or *Sinapis alba* [10,13,14]. Both proteins are composed of 2 different chains which can be separated after reduction with 2-ME and carboxyamidomethylation (Fig. 1). Napin nIb shows heterogeneity at the large chain, yielding 2 forms, L_1 and L_2. On the other hand, the chromatographic behaviour of the small chains of both napins is identical, as well as their amino acid compositions (Table I). But a minor component, 'S_m', (less than 7%) of the small chain, lacking 1 Phe residue, appears in nIa as well as in nIb. Multiple forms of both large and small chains have also been described for the 2-ME-treated 14.5 kDa napins [8] and for the performic acid-oxidized 2 S sulfur-rich protein from Brazil nut [15].

Napin nIa is composed of 2 chains of 31 and 79 residues (Fig. 2) whereas napin nIb contains 1 small chain of 31 amino acids and a heterogeneous large chain of 74–75 residues. Tryptic and chymotryptic digestions of the alkylated small chains render 2 sets of peptides (T and C peptides), which allow the overlapping of the sequences, in spite of the low yield of the N-terminal Edman degradation of the complete chains. The sequences of the small chains from nIa and nIb were identical. The S_m component lacks the Phe residue at the C-terminus. Thermolytic and tryptic treatments were performed to cleave the large chains, rendering the T and H peptide sets. However, the N-terminal amino acid sequence of the alkylated large chains could not be obtained by Edman degradation. Thus, pyroglutamate aminopeptidase treatments for the peptide T_1 and the complete large chains were performed. Three final cleavages were carried out to complete the alignment of the large chains; chymotryptic digestion, which renders the CIT peptide (Ile and Tyr chymotryptic-containing peptide), to overlap positions Arg^{50} and Ile^{51}; CNBr hydrolysis, rendering CN2, for positions Phe^{72} and Ile^{73}; and SaV8 digestion, rendering an SV peptide, for positions 30, 31, 32, 36 and 39. Two differences have been

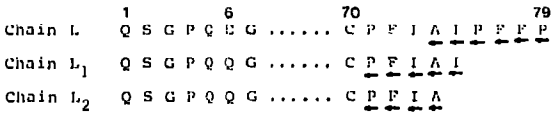

Fig. 3. N- and C-terminal amino acid sequences of the 3 large chains isolated. L corresponds to napin nIa, and both L_1 and L_2 to napin nIb. Carboxypeptidase Y-treatment data were obtained by quantitative analyses of the residues removed by the enzyme at different lengths of time, and are shown by arrows. Numbers indicate the position on the primary structure.

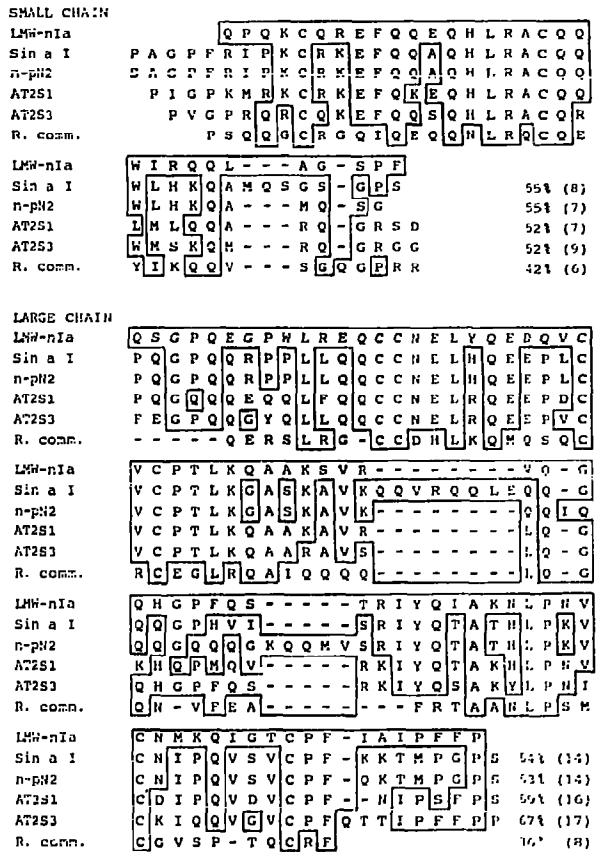

Fig. 4. Amino acid sequence similarity between nIa and several 2 S albumins from different species. Small and large chains of 2 S albumins from seeds of *B. napus* (LMW-nIa, and n-pN2, a 14.5 kDa napin), *Sinapsis alba* (Sin a 1), *Arabidopsis thaliana* (AT2S1 and AT2S3) and *Ricinus communis* (R. comm.) are considered for comparison. Gene sequences were taken into account for n-pN2 and AT2S3, and the primary structure of the mature proteins for R. comm, Sin a I and AT2S1. Processing sites assumed for n-pN2 [5] and AT2S3 [7] were considered. Alignments were obtained by using a computer program based on the algorithm of Needleman and Wunsch [18]. The percent identity value referred to LMW-nIa is given after each of the sequences aligned and the standard deviations are included in parentheses (usually a standard deviation of 3 or more is taken as demonstrating an authentic relationship; see [18]).

observed in the sequences of the 3 large chains isolated (Fig. 3): (i) the amino acid at the 6th position (Glu in L from nIa, and Gln in L_1 and L_2 from nIb); and (ii) the length of the C-termini. The length heterogeneity of these chains could be interpreted in terms of the existence of different precursors. A shift in the position of the cleavage site during the maturation process could be also considered, as suggested for the 14.5 kDa napins [8]. But, it may also be attributed to the existence of carboxypeptidases in the protein bodies of the seeds as it was discussed to explain the heterogeneity observed in the α-chains of pea seed isolectins [16].

When surveyed for the structural homology between the LMW-napins and the mature 14.5 kDa napins (Fig. 4), 55% similarity was found among the small chains, and 53% among the large chains. However, the similarity increases when napin nIa and the deduced amino acid sequence of the gene AT2S3 from *Arabidopsis thaliana* are compared (67%). The similarity is enhanced if we consider that most of the observed changes are of conservative character. Moreover, the glutamine-rich region of the 14.5 kDa napins between positions 39–60 (Fig. 4) does not appear in the LMW-napins nor in the protein from the *A. thaliana* AT2S3 gene. This fact, which reveals the structural relationship between these proteins from different species instead of the expected similarity among napins, demonstrates the existence of a new distinct group of 2 S albumins in *B. napus*, and opens new insights into the evolutionary studies of these plant storage proteins.

Finally, we must consider that in spite of the differences found, the number of conservative changes in the sequences, the conserved positions of cysteines and the identical dichroism spectra of LMW-napins and 14.5 kDa napins (data not shown) would support the idea of a required similar tri-dimensional structure for the napin-like proteins [17].

Acknowledgements: This research was supported by Grant PB89-87 from the Dirección General de Investigación Científica y Técnica, Spain. We are grateful to Dr. J.G. Gavilanes for his helpful suggestions and critical reading of the manuscript.

REFERENCES

[1] Cole, S.G. and Kagnoff, M.F. (1985) Annu. Rev. Nutr. 5, 241–266.
[2] Barber, D., Sánchez-Monge, R., Gómez, L., Carpizo, J., Armentia, A., López-Otín, C., Juan, F. and Salcedo, G. (1989) FEBS Lett. 248, 119–122.
[3] Sharief, F.S. and Li, S.S.-L. (1982) J. Biol. Chem. 257, 14753–14759.
[4] Menéndez-Arias, L., Moneo, I., Domínguez, J. and Rodríguez, R. (1988) Eur. J. Biochem. 177, 159–166.
[5] Crouch, M.L., Tenbarge, K.M., Simon, A.E. and Ferl, R. (1983) J. Mol. Appl. Genet. 2, 273–283.
[6] Allen, R.D., Cohen, E.A., Von der Haar, R.A., Adams, C.A., Ma, D.P., Nessler, C.L. and Thomas, T.L. (1987) Mol. Gen. Genet. 210, 211–218.
[7] De Clercq, A., Vandewiele, M., De Rycke, R., Van Damme, J., van Montagu, M., Krebbers, E. and Vandekerckhove, J. (1990) Plant Physiol. 92, 899–907.
[8] Monsalve, R.I., Menéndez-Arias, L., López-Otín, C. and Rodríguez, R. (1990) FEBS Lett. 263, 209–212.
[9] Ericson, M.L., Rödin, J., Lenman, M., Glimelius, K., Josefsson, L.-G. and Rask, L. (1986) J. Biol. Chem. 261, 14576–14581.
[10] Monsalve, R.I. and Rodríguez, R. (1990) J. Exp. Bot. 41, 89–94.
[11] Bencze, W.L. and Schmid, K. (1957) Analyt. Chem. 29, 299–308.
[12] Panyim, S. and Chalkley, R. (1969) Arch. Biochem. Biophys. 130, 337–346.
[13] Menéndez-Arias, L., Monsalve, R.I., Gavilanes, J.G. and Rodríguez, R. (1987) Int. J. Biochem. 19, 899–907.
[14] Laroche-Raynal, M. and Delseny, M., (1986) Eur. J. Biochem. 157, 321–327.
[15] Ampe, C., van Damme, J., de Castro, L.A.B., Sampaio, M.J.A.M., van Montagu, M. and Vandekerckhove, J. (1986) Eur. J. Biochem. 159, 597–604.
[16] Rini, J.M., Hofmann, T. and Carver, J.P. (1987) Biochem. Cell Biol. 65, 338–344.
[17] Krebbers, E., Herdies, L., de Clercq, A., Seurinck, J., Leemans, J., van Damme, J., Segura, M., Gheysen, G., van Montagu, M. and Vandekerckhove, J. (1988) Plant Physiol. 87, 859–866.
[18] Doolittle, R.F. (1981) Science 214, 149–159.