Volume 263, number 2, 209–212 FEBS 08328 April 1990

β-Turns as structural motifs for the proteolytic processing of seed proteins

Rafael I. Monsalve, Luis Menéndez-Arias, Carlos López-Otín° and Rosalía Rodríguez

Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad Complutense, 28040 Madrid and *°Departamento de Biología Funcional, Facultad de Medicina, Universidad de Oviedo, 33008 Oviedo, Spain*

Received 30 January 1990

Fifteen NH_2- and COOH-terminal ends from both small and large chains of the most abundant 2 S albumins from *Brassica napus* seeds have been sequenced. This allows the determination of the exact proteolytic maturation sites of these proteins. Each one of these proteins arises from a polypeptide precursor which is cleaved during the post-translational processing at four sites, giving two different chains linked by disulphide bridges on the mature 2 S albumin. The hydrolyzed bonds involved in the processing are located in proline and glycine-rich regions, forming tetrapeptides with a very high β-turn probability. Similar results have been found through the analysis of the 2 S albumin precursors from other seeds. These facts are interpreted in terms of the existence of a β-turn specific endoprotease activity involved in the maturation process of 2 S albumins.

Seed protein; β-Turn; Proteolytic cleavage; Rapeseed; Primary structure

1. INTRODUCTION

Several types of structural motifs have been described which are recognized by proteases during the maturation process of proteins to generate functional polypeptides. A sequence of two basic amino acids, Lys and Arg, in all the possible combinations [1–3], appears in most of the precursors for regulatory peptides (e.g. hormones, growth factors, neuropeptides, etc.), and they represent the best substrates for trypsin-like proteases. A proline-directed arginyl cleavage site has also been described in some peptides (e.g. atrial natriuretic factor, substance P, adrenorphine, etc.) [4]. Sequences such as Gly-Gly-X, Gln-Gly or X-Y-Pro-Z (where X, Y and Z are generally hydrophobic amino acids) are present in the maturation sites of protein precursors from animal and viral origin [5,6]. Although these motifs refer to the primary structure of the precursors at the processing site, the importance of its conformation has also been suggested [4–8]. Knowledge of these proteolytic events is still scarce for maturation proteases from plants.

Seed storage proteins constitute more than 10% of the total protein isolated from seeds, and they have an important role in providing nitrogen to the germinating seedling. Among these proteins, the 2 S albumins are the most significant, given their nutritional and clinical interest. The pN1 and pN2 cDNA, and the genes napA and gNa, encoding for several 2 S albumins from rapeseed (*Brassica napus*), have been recently sequenced. They appear as a multigene family that contains at least 10 and perhaps more than 16 genes [9–11]. The expression of these genes produces polypeptide precursors which must be processed by maturation proteases to render the two chains, small and large, of the native proteins (*napins*). The deduced precursor sequences encoded by these genes are homologous, with more than 90% of identity between those of pN1, pN2 and napA. Most of the differences between these proteins are observed around the proposed cleavage sites and could originate mature polypeptides of different lenghs. Therefore, we have determined the NH_2- and COOH-terminal sequences of many 2 S albumins from *B. napus* in order to find the exact processing sites of these seed storage proteins, as well as any common structural feature which could account for the protease specificity involved in the process.

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

Abbreviations: TFA, trifluoroacetic acid; HPLC, high-pressure liquid chromatography; PITC, phenylisothiocyanate

2. MATERIALS AND METHODS

The 2 S albumins from rapeseed were isolated as previously reported [12]. The native proteins were reduced in 0.1% (v/v) TFA, pH 2.1, containing 5% (v/v) β-mercaptoethanol, for 20 min, at 85°C. The samples were immediately centrifuged on a Beckman microfuge and chromatographed on a μBondapak C-18 reverse-phase column, using a Beckman HPLC-system with a gradient of acetonitrile in 0.1% (v/v) TFA. The eluent was continuously monitored at both 214 and 280 nm wavelengths.

Each native protein (50–70 nmol) was cleaved with BrCN in 70% (v/v) formic acid (5 mg/ml protein concentration and a 250–1000 BrCN molar excess) for 20 h at room temperature, in the dark and N_2 atmosphere. Samples were finally diluted with glass-distilled water

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

and lyophilized. Chymotryptic digestion of the reduced nII-small chain was performed as previously described [13], for 2 h at a 1:50 enzyme/substrate weight ratio. The peptides obtained from either BrCN or chymotrypsin treatments were separated by reverse-phase chromatography, adequating the acetonitrile gradient for each sample.

Amino acid analyses were performed on a Durrum D-500 amino acid analyzer after acid hydrolysis in the conditions previously described [13]. NH_2-terminal sequencing of peptides and reduced chains was carried out by automated Edman degradation using an Applied Biosystems model 477A sequencer. The resulting phenylthiohydantoin amino acid derivatives were identified by using a model 120A on-line PTH-analyzer and the standard Applied Biosystems program.

β-Turn prediction was done according to the rules proposed by Chou and Fasman [14]. Statistical significance of the association between a β-turn probability higher than the cut-off value ($P_T > P_\alpha$, $P_T > P_\beta$, and $P_T > 1.00$, and $p_T > 7.5 \times 10^{-5}$) and the location of an endoproteolytic processing site was estimated by using a χ^2-test.

3. RESULTS AND DISCUSSION

The three 2 S albumins mainly expressed in *Brassica napus* seeds have been isolated according to Monsalve and Rodríguez [12]. They were called nII, nIII, and nIV, and exhibit a molecular mass of 14500, in agreement with Lönnerdal and Janson [15]. All these molecules are constituted by two polypeptide chains of 4500 and 9800 Da linked by disulphide bonds. The small and large subunits were separated by a simple method which involves reduction of the native protein and HPLC of the products. By using this chromatographic procedure we have detected heterogeneity in some of the chains that constitute the mature 2 S albumins. Thus, one small chain and two large chains are obtained from nII, two small chains and one large chain from nIV, and only one chain of each molecular weight from nIII (Fig. 1).

In order to determine the sequences of the mature proteins around the cleavage sites, every isolated chain was automatically sequenced at the NH_2-terminus throughout 9 cycles. Moreover, all the native proteins were BrCN-treated in order to obtain the COOH-terminal peptides of every chain. This treatment was performed because each of the two methionines in the native proteins must be located near the COOH-terminal end of both the small and the large chains, if we assume the strong homology proposed by Crouch et al. [9] for these molecules. However, it was not possible to determine the sequence of the COOH-terminal peptide from the small chain of nII, probably due to the cyclization of the glutamine involved in the peptide bond cleaved by the BrCN treatment. The sequence of this polypeptide region was alternatively obtained after chymotryptic digestion of the reduced small chain and sequential degradation of the Met-containing peptide. The sequences obtained are shown in Table I. The Ile-Pro-Lys sequence proposed by Ericson et al. [16] was not observed at the NH_2-terminal end of the small chains. On the other hand, the NH_2-terminal sequence of the nIV-large chain could not be obtained. This can be explained if the isolated nIV is the product of the gene gNa expression [11] (the amino acid composition of the small and large nIV-chains were identical to those deduced from the corresponding gene; data not shown). Then, glutamine should appear at the NH_2-terminal position of the mature large chain, and be involved in a cyclizing reaction which would not allow the Edman degradation.

Table I shows the invariable position of the cleavage site at the amino termini of both the large and small chains, although substitutions of proline and glycine by serine were observed in S-nIII and l_1-nII. However, the maturation site at the COOH-terminal end moves on the different napins. The post-translational cleavage sites are located, in both mature chains, at proline-, glycine- and/or serine-rich regions. These amino acids are disrupters of α-helix or β-strand secondary structures [14], and might help that region to become accessible to endoprotease activity. The Chou and Fasman β-turn probability profiles along the deduced sequences from the precursors of the 2 S albumins from *B. napus*, *Arabidopsis thaliana*, and *Bertholletia excelsa* (Brazil nut) were obtained in order to evaluate the relevance of the conformation around the cleavage sites (Fig. 2). As can be observed, the processing sites are located inside or close to a region with a very high β-turn tendency. This high β-turn probability at both NH_2- and COOH-terminal ends would be determined by the presence of proline and glycine, strong inducers of that structure.

A χ^2-test was done in order to estimate the statistical significance of the association of predicted β-turn and the location of the proteolytic cleavage sites involved in the maturation of the 2 S albumins. In the rapeseed prepro-napin pN2, 27 tetrapeptides, among the total 175 overlapping tetrapeptides, are predicted as β-turns. If it is assumed that 4 bonds are cleaved to render the mature small and large chains of napin, 17 of these tetrapeptides would either contain or be near to a hydrolyzed bond. Therefore, if no association exists between β-turn and proteolytic cleavage site, only 2 or 3 predicted β-turns would be statistically located at cleavage sites. However, this number goes up to 10 assuming any of the proposals suggested for the processing of pronapins [9,16]. The corresponding value of χ^2 in this case is 24.5, revealing that the association appears to be very significant ($P < 0.005$). This level of confidence is also reached in the 2 S albumins from *Arabidopsis thaliana*, and Brazil nut, though, in the latter, the value of χ^2 is lower due to the lack of predicted β-turn at the NH_2-terminal end of the small polypeptide chain (Fig. 2).

An additional remarkable feature is deduced from the sequences shown in Table I. If we accept a strong homology between the genes encoding for the napins, the shifts found in the processing sites to render the COOH-termini of the small subunits could be explain-

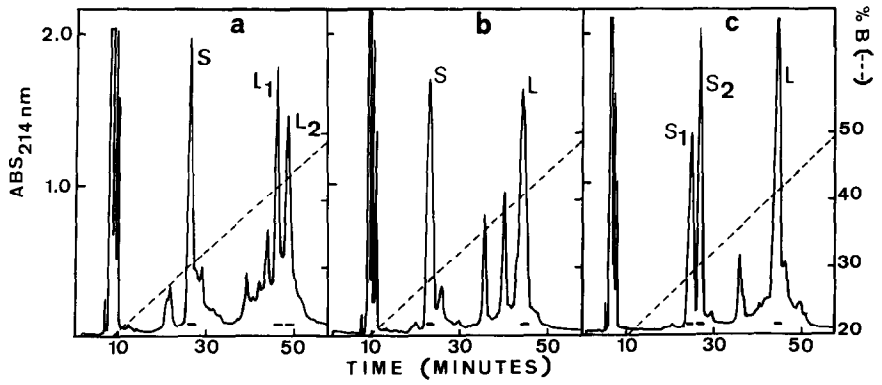

Fig. 1. HPLC elution profiles of the reduced 2 S proteins from *B. napus*. (a) nII; (b) nIII; (c) nIV. The small (S) and large (L) chains were eluted with a gradient of buffer B (0.1% TFA in acetonitrile). The flow rate was 1.5 ml/min. The fractions considered for sequence determination are represented by horizontal bars; β-mercaptoethanol elutes at the first fractions of the chromatography.

ed in terms of the amino acid changes and the resulting β-turn probabilities. Thus, substitution of serine at positions n – 3 and n – 1 of the nII- or nIII-small chains by proline and glycine, respectively, would allow an enlargement of the sequence size of the nIV-small subunits. Something similar appears at the COOH-terminal end of the large chain of nIV. The proposed sequence at this region for the protein expressed by gNa is ..PGPG, instead of ..PGPS as appears for pN1, pN2, and napA encoded proteins. That change could be responsible for the one position shift on the cleavage site of nIV that leads to the sequence ...PGP. This has also been observed for pheasant lysozyme, where the substitution of leucine by proline, two positions ahead of the NH_2-terminal end of the mature protein, causes a shift in the processing site [19].

According to the above, plant proteases involved in the post-translational processing of seed proteins cleave the polypeptide precursor at positions close to amino acids such as proline or glycine. These residues would be involved in β-turn, which would be the required conformation for the activity of the maturation proteases. This motif is similar to those found in proteins of animal origin [1–8,20], and suggests an evolutionary link between plant and animal proteases involved in the maturation of proteins.

Table I

Amino acid sequences of the NH_2– and COOH-terminal ends of the small (S) and large (L) polypeptide chains of nII, nIII and nIV, the 2 S albumins from *Brassica napus* seeds. The primary structures around the observed maturation sites, encoded by clones pN2 [9], napA [10] and gNa [11], are also included for comparison

Processed polypeptide	1 5	n-5 n
S-nII	P*A G P F R I P K	..... L H K Q A M Q S G S G
S-nIII	S A G P F R I P K	..... Q S G S G
S_1-nIV	P A G P F R I P K	..... Q P G G G S G P S
S_2-nIV	P A G P F I R P K	..... Q O G G G S G P S
Precursor		
pN2	D A T N S A G P F R I P K	..... Q A M Q S G G G - - P N W T L D
napA	D A T D S A G P F R I P K	..... Q A M Q S G G G - - P S W T L D
gNa	D A T N P A G P F R I P K	..... Q A M Q P G G G S G P S W T L D
Processed polypeptide	1 5	n
L_1-nII	P Q G*P Q Q R P P	..... P G P S
L_2-nII	P Q G P Q Q R P P	..... P G P S
L-nIII	P Q G P Q Q R P P	..... P G P S
L-nIV	----------	..... P G P
Precursors		
pN2 and napA	M E N - P Q G P Q Q R P P	..... M P G P S Y
gNa	V E N Q Q Q G P Q Q R P P	..... M P G P G F Y

*The first cycle of sequence of S-nII renders 85% proline plus 15% serine. The third NH_2-terminal position of L_1-nII contains 50% of each glycine and serine.

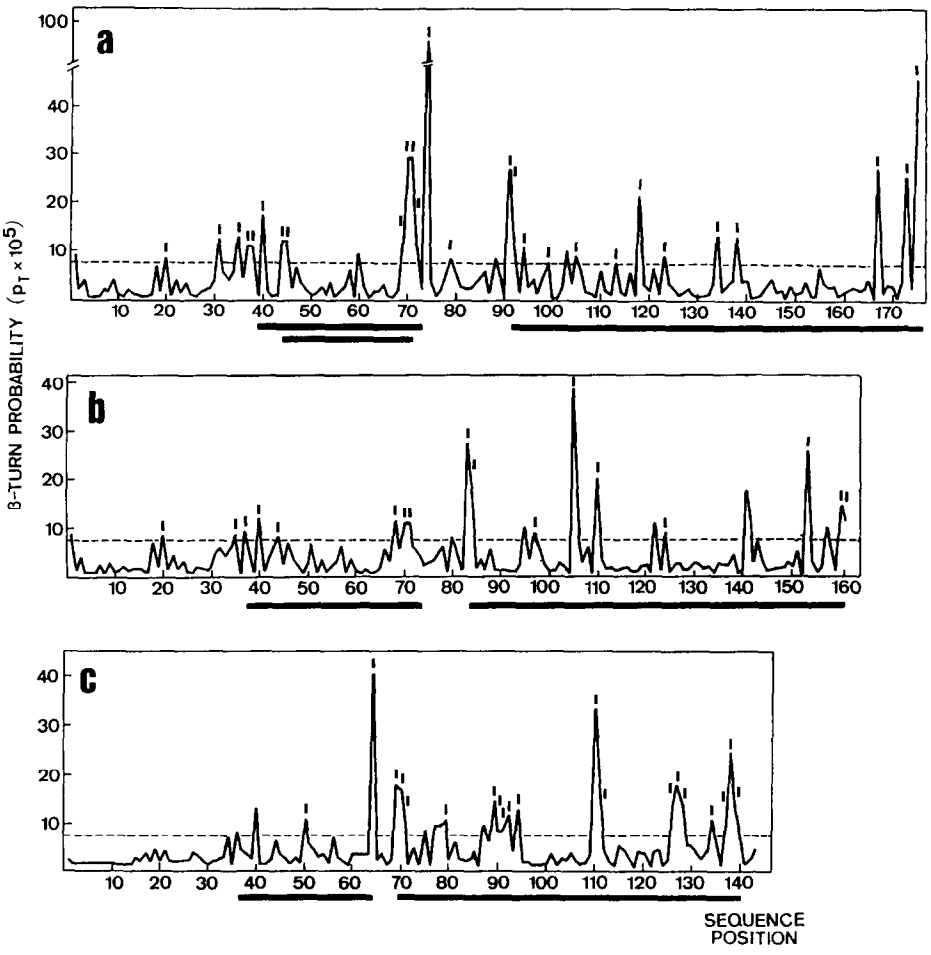

Fig. 2. β-turn probability (p_T) for the 3-residue overlapping tetrapeptides of the primary structures deduced from the nucleotide sequence of the cDNA encoding for the 2 S albumins. (a) Rapeseed (pN2 clone [8]); (b) *Arabidopsis thaliana* seeds [17] and (c) Brazil nut [18]. Vertical bars at the first residue of the tetrapeptide indicate that this is predicted as a β-turn. The mature small and large polypeptide chains are shown by horizontal bars under the sequence numbering, showing the two possibilities proposed for the small chains of pN2 [9,16].

Acknowledgement: This work has been supported by Grant PB86-0530 from the Dirección General de Investigación Científica y Técnica.

REFERENCES

[1] Tager, H.S., Patzelt, C., Assoian, R.K., Chan, S.J., Duguid, J.R. and Steiner, D.F. (1980) Ann. NY Acad. Sci. 343, 133–147.
[2] Chance, R.E., Ellis, R.M. and Bomer, W.W. (1968) Science 161, 165–167.
[3] Docherty, K. and Steiner, D.F. (1982) Annu. Rev. Physiol. 44, 625–638.
[4] Schwartz, T.W. (1986) FEBS Lett. 200, 1–10.
[5] López-Otin, C., Simón, C., Martínez, L. and Viñuela, E. (1989) J. Biol. Chem. 264, 9107–9110.
[6] Kräusslich, H.-G. and Wimmer, E. (1988) Annu. Rev. Biochem. 57, 701–754.
[7] Rholam, M., Nicolas, P. and Cohen, P. (1986) FEBS Lett. 207, 1–6.
[8] Perlman, D. and Halvorson, H.D. (1983) J. Mol. Biol. 167, 391–409.
[9] Crouch, M.L., Tenbarge, K.M., Simon, A.E. and Ferl, R. (1983) J. Mol. Appl. Genet. 2, 273–283.
[10] Josefsson, L.-G., Lenman, M., Ericson, M.L. and Rask, L. (1987) J. Biol. Chem. 262, 12196–12201.
[11] Scofield, S.R. and Crouch, M.L. (1987) J. Biol. Chem. 262, 12202–12208.
[12] Monsalve, R.I. and Rodríguez, R. (1990) J. Exp. Bot., in press.
[13] Menéndez-Arias, L., Moneo, I., Domínguez, J. and Rodríguez, R. (1988) Eur. J. Biochem. 177, 159–166.
[14] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[15] Lönnerdahl, B. and Janson, J.-C. (1972) Biochim. Biophys. Acta 278, 175–183.
[16] Ericson, M.L., Rödin, J., Lenman, M., Glimelius, K., Josefsson, L.-G. and Rask, L. (1986) J. Biol. Chem. 261, 14576–14581.
[17] Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Ghysen, G., Van Montagu, M. and Vandekerckhove, J. (1988) Plant Physiol. 87, 859–866.
[18] Altenbach, S.B., Pearson, K.W., Leung, F.W. and Sun, S.S.M. (1987) Plant Mol. Biol. 8, 239–250.
[19] Weisman, L.S., Krummel, B.M. and Wilson, A.C. (1986) J. Biol. Chem. 261, 2309–2313.
[20] Von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.