STRUCTURE, EXPRESSION AND EVOLUTION OF STORAGE PROTEIN GENES IN RADISH AND RELATED CRUCIFERS

D. Depigny, M. Raynal, F. Grellet, M. Delseny.

Université de Perpignan - 52, Avenue de Villeneuve
66860 PERPIGNAN - FRANCE

INTRODUCTION

Storage proteins are major components in evaluating seed quality in crucifers. Two types of major storage proteins have been identified in this family on the basis of their solubility and sedimentation. Napins are water-soluble proteins sedimenting at 1.7-2S of the albumin type, whereas cruciferins are salt-soluble aggregates sedimenting at 12S. Cruciferins are globulins and are similar in structure to the legumins from pea. Because there are multiple polypeptides in these proteins, their biochemical and genetical characterization have been difficult. With the use of molecular biology techniques, this situation has improved rapidly but our knowledge on these proteins is far from complete. So far, molecular studies on napins and cruciferins have been restricted to three species: rapeseed, radish, and Arabidopsis . Characterisation of the proteins and their respective genes is important, not only for rapeseed breeding, but also in related species because it can provide the genetical engineer with useful information on gene variability and with heterologous genes , with slightly different promotors. With these ideas in mind we have initiated a study to characterize radish storage proteins and their genes. Initially we purifed radish cruciferins and napins (Laroche et al. 1984), isolated specific antibodies and cDNA clones (Laroche-Raynal and Delseny M. 1986) and analyzed their in vivo protein synthesis (Raynal et al. 1987). More recently we completed this study with characterization of several cDNA coding for napins (Raynal et al. 1991) and of a genomic and several cDNA clones for cruciferins (Depigny et al. 1991).

This report will review our major findings on the organization of these two gene families.

MATERIALS AND METHODS

Plant material, extraction and purification of nucleic acids and proteins have been described in previous articles (Laroche-Raynal and Delseny 1986, Raynal et al. 1991).

A small cDNA library was prepared in pBR 322 using mRNA from immature seeds and initially screened with heterologous probes for rapeseed napin and cruciferin given to us by M. Crouch (Bloomington USA) (Crouch et al. 1983). A further screening with first strand cDNA from immature seed mRNA allowed us to pick up additional clones representative of abundant mRNA. Several individual clones were then used as probes to identify cross-hybridizing groups of clones within the library. The longest clones in each group were then sequenced and identified as cruciferin groups by sequence homology and in some case by hybrid selection and translation of the corresponding mRNA. A genomic clone for cruciferin was isolated from a radish library made by A. Boronat (Barcelona, Spain) in the Charon 35 vector. Appropriate fragments from this genomic clone were subcloned and sequenced using the Exo III deletion method (Sambrook et al. 1989). Expression was analysed using Northern blots and RNA extracted at various stages of seed formation. The 5' end of cruciferin mRNA was mapped by hybridizing mRNA to a labelled antisense

transcript derived from the 5' region. From the size of the protected region following RNase treatment, the position of the transcription initiation site was deduced.

RESULTS

Analysis of the Napin cDNA Clones

Following screening of the library with the heterologous probe pN2, 25 positive clones were recovered corresponding to 8.7% of the screened colonies. After sizing the inserts, three clones with slightly different restriction maps were further characterized and sequenced.

Comparison of the two dimensional electrophoretic patterns of total mRNA in vitro translation products immunoprecipitated with napin specific antibodies and of mRNA hybrid-selected with a radish napin cDNA clone indicated that the isolated clones were representative of all napin mRNAs. In these two-D patterns, 8 distinct napin precursors were observed, corresponding to two slightly different sizes. Sequencing of the three napin cDNA clones indicated that they were all different and belonged to two types of genes, each class corresponding to one size class of precursor. Therefore napin proteins are encoded by a small multigenic subfamily.

Comparison of the amino sequences deduced from the two types of clones show that they are 92% homologous. When radish sequences are compared with the known rapeseed sequences the average homology is 90% with one type of clone and 88% with the other. These sequence comparisons also revealed a region in the carboxy terminal part of the protein which is more variable and consists mostly of repeats of a QQX motif.

Analysis of the Cruciferin cDNA Clones

Initially our cDNA library was screened with pC1 (Crouch et al. 1983, Simon et al. 1985, Laroche-Raynal et al. 1986), the rapeseed cruciferin probe. Unexpectedly, less than 1% of the colonies hybridized with this probe. Furthermore, when the radish cDNA clones were used in hybrid-selection experiments only one class of pre-procruciferin was synthesized in vitro from selected mRNA instead of four which are detected following immunoprecipitation of total mRNA translation products with an anticruciferin serum. This puzzling result suggested to us that other cDNAs should exist in the library but that their sequence had sufficiently diverged from that of the rapeseed probe. The library was screened again with first strand homologous cDNA to reveal those clones corresponding to abundant mRNA in mature seeds. Individual clones were then hybridized back to the library so that cross-hybridizing groups of clones could be defined. Sixteen additional clones were found to belong to three major groups. All these groups have at least one clone in common, suggesting that they are related. Four representative cDNA clones were sequenced as well as two from the first screening and confirmed to be cruciferins. As a result, identified cruciferin clones account for 7% of the library. None of the clones represent a full-length one and all are different but nevertheless they could be aligned and they clearly show extensive homology both at the nucleic acid and amino acid levels. The location of these clones along the cruciferin gene sequence is shown in Fig. 1.

Table 1 shows the percentage of homology of these clones when compared together as well as the percentage of homology with clones from related species such as rapeseed and Arabidopsis (Ryan et al. 1989, Rodin et al. 1990, Pang et al. 1988). From this table, it is obvious that there are at least two major cruciferin gene families in radish. One of them represented by pAE10 (class I) is more than 90% homologous to the unique sequence (CruA) which had been reported until recently in rapeseed (Ryan et al. 1989). Homology between class I and class II sequences,

represented by pAG4 and CruRad is 60% in average. On the basis of their 85% homology at the nucleic acid level, and diverging 3' untranslated region, class II clones can be divided into two related subfamilies. As shown by a 93% homology between CruRad, pAG4 and Cru1 (Rodin et al. 1990), class II genes are also present in rapeseed.

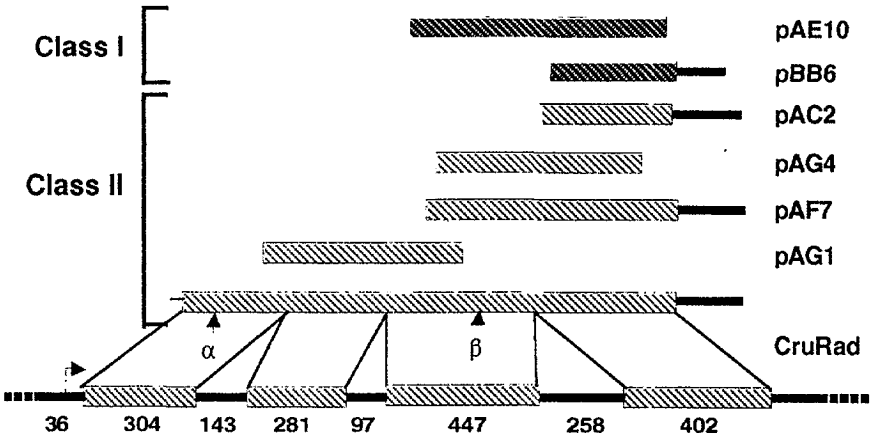

Fig. 1. Alignment of radish cDNA clones with the genomic clone (CruRad). Position and size of the exons and introns are indicated. The position of the α and β subunits is indicated on the deduced mRNA drawing.

Table 1. Homology percentages, at the aminoacids level, between cDNA (pAE10, pAG4) and genomic (CruRad) clones for radish cruciferin and available sequences for rapeseed (Cru1, CruA) and Arabidopsis (CRA1, CRB) cruciferins.

	Radish			Rapeseed		Arabidopsis	
	pAE10	pAG4	CruRad	Cru1	CruA	CRA1	CRB
pAE10							
pAG4	58						
CruRad	62	93					
Cru1	62	93	96				
CruA	98	57	57	60			
CRA1	84	58	63	62	86		
CRB	78	55	57	58	74	73	

Characterization of a Radish Cruciferin Genomic Clone

Since a single genomic clone for cruciferin (CruA), corresponding to probe pCI, had been isolated in rapeseed (Ryan et al. 1989), we isolated and characterized a radish genomic clone corresponding to the class II group of cruciferins. This clone was isolated as an Eco RI fragment, 4.9 kbp long, which hybridized to both the pAGI and pAF7 cDNA clones (Fig. 1) and was completely sequenced. It is 98% homologous in the overlapping region with pAF7 and 90% with pAGI. The overall homology with the class I gene CruA from rapeseed is 57%. A major difference between the two genes is the presence of 3 introns in the radish gene, instead of 2 in that of rapeseed. The three introns are at the same positions as in the two sequenced Arabidopsis genes (Pang et al. 1988) and as in a number of legumin genes from pea. Another difference is the presence of a small insertion coding for a 12 amino acid sequence in the class II genes. This insertion might be the footprint of a transposable element. Comparison of all the available cruciferin sequences revealed that the acidic subunit is more variable than the basic one.

Genome Organisation and Evolution of Napin and Cruciferin Genes

Genome organisation was analyzed using Southern blots and the different probes we characterized. Napins are encoded by a multigene family which can be estimated to 6-8 genes which are members of two closely related subfamilies. Analysis of existing sequences revealed that this division in two subfamilies is rather ancient since they can be recognized in rapeseed and Arabidopsis , due to diagnostic nucleotide stretches in the 3' untranslated region. On the other hand, Arabidopsis napin sequences (Krebbers et al. 1988) have diverged much more from each other than the radish and rapeseed genes suggesting extensive concerted evolution in these two species. Careful analysis of the gene sequence also leads to the conclusion that the napin gene resulted from two successive dupplication events of a motif coding for 55 amino acids (Fig. 2). One of the motifs has undergone a deletion (Raynal et al. 1991).

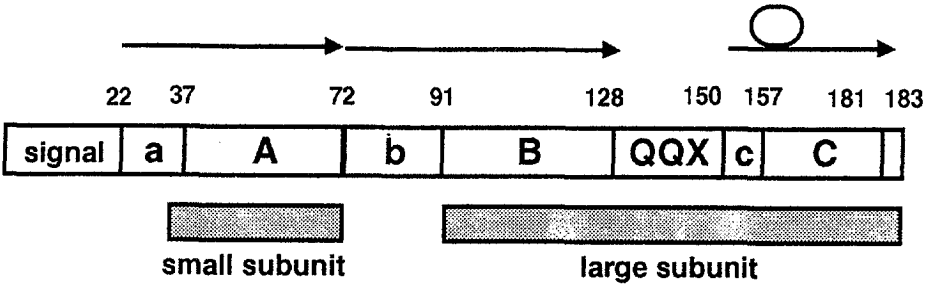

Fig. 2. Organisation of the napin gene as deduced from the analysis of internal aminoacids homology. Arrows represent the repeats. Hatched boxes represent mature subunits.

Cruciferin genes were investigated in a similar manner. Probes for class I genes do not hybridized with class II genes. Both types of probes detected multiple fragments in radish as well as in rapeseed. The gene number was estimated to 8-10 in radish and is greater for class II. Homology was lower with Arabidopsis. These data indicate that the same gene families are present in radish and rapeseed and the evidence for class II genes in rapeseed was recently substantiated with the isolation of a cDNA clone (Cru1) which is 96% homologous to the coding sequence of CruRad. (Rodin et al. 1990). The two characterized cruciferin genes from Arabidopsis are closer to the

class I genes, but it is most likely that a third Arabidopsis gene corresponds to class II. Preliminary analysis of an F2 population derived from a cross between two radish cultivars differing for the rDNA, napin and cruciferin restriction patterns indicates that there are at least two segregating loci coding for cruciferins.

Expression Pattern of Napin and Cruciferin Genes

Using specific probes for each subfamily of clones it was possible to demonstrate that both types of napin genes are expressed to a similar extent and that the maximum steady state level of the mRNA is between 25 and 35 days after flowering (Raynal et al. 1991). Almost no message is detected at 20 days. Similarly, class I and class II cruciferin genes are apparently coordinately expressed at the same period. After 35 days the steady state mRNA level decays and there is no storage protein mRNA in dry seeds (Fig. 3).

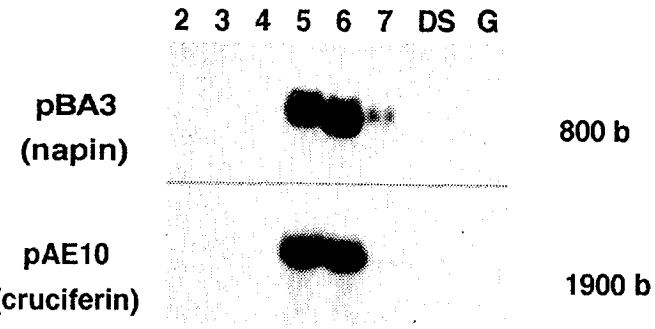

Fig. 3. Expression of napin and cruciferin genes during seed formation (Northern blots). 2,3,...7 correspond to different developmental stages from 5 to 55 DAF. DS; dry seed. G; germinated seedlings.

CONCLUSION

This work reports an extensive characterization of two important gene families in radish. It demonstrates that napins are made of two subfamilies and that there are two major groups of genes for cruciferins. Class II genes can be divided into two related subfamilies. We presented evidence that similar subfamilies also exist in other crucifers such as rapeseed and Arabidopsis . The radish gene we characterized is the first one for class II cruciferins. Careful analysis of all the coding sequences indicates that some regions are more variable than others and this is an important starting point for future engineering of storage protein genes. Analysis of the radish cruciferin gene CruRad showed that this gene codes for a protein which is richer in cysteine and methionine than the one encoded by rapeseed class I genes. Isolation of a radish genomic clone also offers an additional promoter to direct the accumulation of a foreign gene product in the seed.

REFERENCES

CROUCH, M., TENBARGE, K.M., SIMON, A.E. and FERL, R. cDNA clones for Brassica napus seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. J. Mol. Appl. Genet. 2: 273-283.

DEPIGNY, D., RAYNAL, M., GRELLET, F. and DELSENY, M. 1991. Characterization of the cruciferin gene family in radish (in preparation).

KREBBERS, E., HERDIES, L., DECLERCQ, A., SEURINCK, J., LEEMANS, J., VAN DAMME, J., SEGURA, M., GHEYSEN, G., VAN MONTAGU, M. and VAN DE KERCKHOVE, J. 1988. Determination of the processing sites of an Arabidopsis thaliana 2S albumin and characterization of the complete gene family. Plant Physiol. 87: 859-860.

LAROCHE, M., ASPART, L., DELSENY, M. and PENON, P. 1984. Characterization of radish (Raphanus sativus) storage proteins. Plant Physiol. 74: 487-493.

LAROCHE-RAYNAL, M., DELSENY, M. 1986. Identification and characterization of the mRNA for major storage proteins from radish. Eur. J. Biochem. 157: 321-327.

PANG, P.P., PRUITT R.E. and MEYEROWITZ, E.M., 1988. Molecular cloning genomic organization and evolution of 12S seed storage protein genes of Arabidopsis thaliana . Plant Mol. Biol. 11: 805-820.

RAYNAL, M., ASPART, L., THIS, P. and DELSENY, M. 1987. Biosynthesis of cruciferin polypeptides in immature radish seeds. Plant Physiol. Biochem. 25: 439-444.

RAYNAL, M., DEPIGNY, D., GRELLET, F. and DELSENY, M. 1991. Characterization and evolution of napin-encoding genes in radish and related crucifers. Gene (in press).

RODIN, J., ERICSON, M.L., JOSEFSSON, L.G. and RASK, L. 1990. Characterization of a cDNA clone encoding a Brassica napus 12S protein (cruciferin) subunit. J. Biol. Chem. 265: 2720-2723.

RYAN, A.J., ROYAL, C.L., HUTCHINSON, J. and SHAW, C.H., 1989. Genomic sequence of a 12S seed storage protein from oilseed rape (Brassica napus) CV Jet Neuf); Nucleic Acids Res. 17: 3584.

SAMBROOK, J., FRITSCH, E.F. and MANIATIS, T. 1989. Molecular cloning; A laboratory manual. Cold Spring Harbor Laboratory Press.

SIMON, A.E., TENBARGE, K.M., SCOFIELD, S.R., FINKELSTEIN, R.R., CROUCH, M.L., 1985. Nucleotide sequences of a cDNA clone of Brassica napus 12S storage protein shows homology with legumin from Pisum sativum . Plant Mol. Biol. 5: 191-201.