THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 265, No. 5, Issue of February 15, pp. 2720-2723, 1990
Printed in U.S.A.

Characterization of a cDNA Clone Encoding a *Brassica napus* 12 S Protein (Cruciferin) Subunit

RELATIONSHIP BETWEEN PRECURSORS AND MATURE CHAINS*

(Received for publication, August 29, 1989)

Joakim Rödin‡, Mats L. Ericson§, Lars-Göran Josefsson, and Lars Rask

From the Department of Cell Research, Swedish University of Agricultural Sciences, Uppsala Biomedical Center, Box 596, S-751 24 Uppsala, Sweden

Cruciferin (12 S globulin) is a large, neutral, oligomeric protein synthesized in rapeseed (*Brassica napus*) during seed development. It is the major seed protein and is composed of six subunit pairs. Each of these pairs is synthesized as a precursor containing one heavy α-chain and one light β-chain. Electrophoretic analysis of cruciferin showed that four different α- and four different β-chains exist. A cruciferin clone was selected from an embryo cDNA library. This clone, pCRU1, contains a 1518-base pair open reading frame corresponding to a truncated NH_2-terminal signal sequence followed by an α-chain of 296 and a β-chain of 190 amino acid residues. Individual cruciferin chains as well as peptides thereof were subjected to NH_2-terminal amino acid sequence analysis. The sequences obtained from a specific α- and β-chain pair (α1 and β1) showed total identity with the deduced amino acid sequence from pCRU1. Further comparisons revealed that a previously characterized cruciferin cDNA clone encodes one of the precursors for the closely related α2/α3-β2/β3 subunits. The deduced amino acid sequences of the two cDNA clones display 64% similarity.

Rapeseed (*Brassica napus*) is today one of the major oil seed crops in many parts of the world. The seeds are rich in oil, but in addition contain a large amount of protein, which constitutes some 20–25% of the dry seed weight (1). The predominant protein species are napin (1.7 S) and cruciferin (12 S). Napin is a low molecular weight, basic protein, composed of two disulfide-linked polypeptides. It constitutes some 20% of the total protein content in the mature seed. Napin is a well characterized protein (2), for which both cDNA (3, 4) and genomic clones have been isolated and analyzed (5, 6). The 12 S storage globulin cruciferin is the major protein in the mature seed, constituting 60% of the total seed protein at maturity (7). Like other 11–12 S globulins (8), cruciferin is an oligomeric protein, composed of six subunit pairs (9). Each pair consists of a heavy α- and a light β-chain with a total M_r of some 48,000–54,000. The 11–12 S globulins are synthesized as α-β precursors on the rough endoplasmic reticulum. After their translocation across the membrane, they assemble into a preform composed of three subunit pairs. These transport intermediates eventually become deposited in vacuole-derived organelles, protein bodies, where the final processing and assembly into the hexameric form take place (10). Little is known about the details of the synthesis, assembly, transport, and processing of these proteins in general and cruciferin in particular. In order to investigate these processes in detail, it is essential to understand the relationship between different cruciferin precursors and the mature chains. We have approached this question by comparing the amino acid sequence corresponding to a cruciferin precursor cDNA clone to partial sequences of cruciferin chains.

MATERIAL AND METHODS[1]

RESULTS

Characterization of Cruciferin—Cruciferin was isolated from the rapeseed lines Hanna and S. Karat and subjected to electrophoresis in SDS[2]-polyacrylamide gels to resolve subunit polypeptides. Nonreducing conditions revealed one major size class of polypeptides with apparent molecular weights of approximately 48,000–54,000 (α-β) and two minor groups with apparent molecular weights of 29,000–34,000 (α) and 20,000–23,000 (β). After reduction the polypeptides of the α-β group disappeared and the staining intensity of the other groups increased (data not shown). In total four different α- and four different β-chains were present in cruciferin from both rapeseed lines, although the staining intensity of individual polypeptides varied between the two lines. α2 and α3 migrated at almost identical positions which precluded their separation by preparative electrophoresis (Fig. 1).

In order to obtain protein sequence information on cruciferin, individual cruciferin chains were electrotransferred to polyvinylidene difluoride membranes. Peptides generated by *Staphylococcus aureus* V8 protease digestion or cyanogen bromide cleavage of cruciferin chains were also used for amino acid sequence determinations (Tables I and II). The intact α2/3 and α4 chains gave no NH_2-terminal amino acid indicating that they are blocked at the NH_2 terminus. The α1 chain is probably also blocked in cruciferin from S. Karat, but the homologous chain from the rapeseed line Hanna did

* This work was supported by grants from the Swedish Natural Science Research Council and from the Swedish Research Council for Forestry and Agriculture. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J05233.

‡ To whom correspondence should be addressed.

§ Present address: Dept. of Physiological Chemistry, Karolinska Inst., S-104 01 Stockholm, Sweden.

[1] Portions of this paper (including "Material and Methods," Tables I and II, and Figs. 2 and 3) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

[2] The abbreviation used is: SDS, sodium dodecyl sulfate.

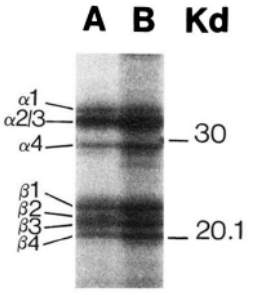

FIG. 1. **Separation of cruciferin polypeptides by SDS-polyacrylamide gel electrophoresis.** Purified cruciferin was reduced, alkylated and applied to an SDS-polyacrylamide gel (10–15% acrylamide). After electrophoresis protein was detected by staining the gel with Coomassie Brilliant Blue. Individual α- and β-polypeptides are numbered. $\alpha 2/3$ consists of two chains of similar size. *Lane A*, cruciferin from the rapeseed line Hanna; *lane B*, cruciferin from the rapeseed line S. Karat.

show the presence of an NH_2-terminal amino acid (Table I). All β-chains ($\beta 1$–$\beta 4$) yielded an NH_2-terminal amino acid. $\beta 2$ and $\beta 3$ were identical in all positions determined. Since the $\beta 2$-chain was present in higher amounts in Hanna than in S. Karat the former line was used when the $\beta 2$-chain was isolated and sequenced.

Isolation and Sequencing of a cDNA Clone Encoding a Cruciferin Precursor—Rapeseed embryonic mRNA was subjected to translation *in vitro* in the presence of [^{35}S]methionine and subsequently immunoprecipitated with the anti-cruciferin antiserum. Analysis of the precipitates by SDS-polyacrylamide gel electrophoresis showed that three [^{35}S]methionine-labeled cruciferin precursors (P1–P3) are encoded in the mRNA.[3]

A cDNA library consisting of 2000 clones was constructed from embryonic rapeseed mRNA. Fifty randomly chosen clones were screened by hybrid selection of mRNA followed by protein synthesis *in vitro* in the presence of [^{35}S]methionine and immunoprecipitation with an anti-cruciferin antiserum. Of several clones that hybridized to cruciferin mRNA, one was chosen for further characterization. This clone, pCRU1, hybridized specifically to mRNA that encodes the two largest precursors of cruciferin (P1 and P2). The other ^{35}S-labeled precursor (P3) was not detected under these conditions (data not shown). The pCRU1 insert hybridized to an embryo mRNA of 2100 nucleotides on a Northern blot (Fig. 2) and was subsequently sequenced. The endonuclease restriction map along with the sequencing strategy are shown in Fig. 3, and the DNA sequence is presented in Fig. 4. pCRU1 has an insert of 1704 base pairs in which a single open reading frame of 1518 base pairs was found. There is no initiation codon in the cDNA sequence, indicating that the clone is truncated at its 5′ end. Two consensus sequences for polyadenylation were found in the 3′-untranslated region. The one closest to the translation stop codon is apparently not used in the mRNA species corresponding to pCRU1. Due to truncation 3′ of the second one we cannot conclude that this one is used either, although it appears likely. Comparison of the amino acid sequence deduced from the pCRU1 nucleotide sequence with the amino acid sequences obtained from cruciferin chains and peptides clearly establishes that pCRU1 encodes the $\alpha 1$-$\beta 1$ precursor. It also shows that the NH_2 terminus of the α-chain corresponds to codon 21 of the pCRU1 insert and that of the β-chain to codon 317. This finding suggests that if no further processing occurs apart from the removal of the signal sequence and the proteolytic cleavage

[3] J. Rödin, unpublished results.

```
1    CTGCAGGGGGGGGGGGCGCCTTGCGAGACAGCCGTCGCAACGTTCGGGGTTCTCCTCGTC   60
                          P  C  E  T  A  V  A  T  F  G  V  L  L  V    14
                       -1 +1
61   CTCAACGGCTGTCTCGCAAGGCAGTCGCTAGGGGTTCCTCCTCAGCTAGGGAACGCGTGT   120
15    L  N  G  C  L  A  R  Q  S  L  G  V  P  P  Q  L  G  N  A  C    34

121  AACCTCGATAACTTAGACGTTCTCCAGCCTACCGAAACTATCAAGAGCGAGGCTGGTCGG   180
35    N  L  D  N  L  D  V  L  Q  P  T  E  T  I  K  S  E  A  G  R    54

181  GTCGAGTACTGGGATCACAACAATCCTCAGATCCGATGTGCTGGTGTCTCTGTCTCTCGT   240
55    V  E  Y  W  D  H  N  N  P  Q  I  R  C  A  G  V  S  V  S  R    74

241  GTTATAATCGAACAAGGCGGTCTCTACCTTCCTACCTTCTTCAGCTCCCCCAAAATTTCA   300
75    V  I  I  E  Q  G  G  L  Y  L  P  T  F  F  S  S  P  K  I  S    94

301  ATCGTTGTTCAAGGAATGGGTATTAGCGGAAGAGTGGTCCCTGGATGCGCGGAAACCTTC   360
95    I  V  V  Q  G  M  G  I  S  G  R  V  V  P  G  C  A  E  T  F   114

361  ATGGACTCGCAGCCTATGCAAGGACAACAACAAGGTCAACCATGGCAGGGACAACAAGGA   420
115   M  D  S  Q  P  M  Q  G  Q  Q  Q  G  Q  P  W  Q  G  Q  Q  G   134

421  CAACAGGGTCAGCAGGGACAACAAGGTCAACAGGGTCAGCAGGGACAACAAGGTCAACAG   480
135   Q  Q  G  Q  Q  G  Q  Q  G  Q  Q  G  Q  Q  G  Q  Q  G  Q  Q   154

481  GGTCAGCAGGGTCAACAGGGACAGCAGGGTCAGCAGCAGCAAGGGTTCCGTGACATGCAC   540
155   G  Q  Q  G  Q  Q  G  Q  Q  G  Q  Q  G  Q  Q  G  F  R  D  M  H   174

541  CAGAAGGTCGAACATGTTCGACATGGAGACATCATTGCCATTACTGCAGGCTCTTCCCAT   600
175   Q  K  V  E  H  V  R  H  G  D  I  I  A  I  T  A  G  S  S  H   194

601  TGGATCTACAACACCGGTGACCAGCCACTTGTCATTATCTGCCTTCTCGACATTGCCAAC   660
195   W  I  Y  N  T  G  D  Q  P  L  V  I  I  C  L  L  D  I  A  N   214

661  TACCAAAACCAACTCGACCGCAACCCAAGAACGTTCCGTCTGGCCGGAAACAACCCACAG   720
215   Y  Q  N  Q  L  D  R  N  P  R  T  F  R  L  A  G  N  N  P  Q   234

721  GGCGGTTCCCAGCAGCAGCAACAACAACAGAACATGTTGAGCGGGTTCGACCCTCAG     780
235   G  G  S  Q  Q  Q  Q  Q  Q  Q  Q  N  M  L  S  G  F  D  P  Q   254

781  GTCCTAGCCCAGGCATTGAAAATCGACGTTAGGTTGGCTCAGGAGCTTCAGAACCAACAA   840
255   V  L  A  Q  A  L  K  I  D  V  R  L  A  Q  E  L  Q  N  Q  Q   274

841  GACAGCAGAGGAAACATCGTTCGTGTTAAGGGACCTTTCCAGGTTGTGAGGCCGCCTCTT   900
275   D  S  R  G  N  I  V  R  V  K  G  P  F  Q  V  V  R  P  P  L   294

901  AGACAGCCATACGAGAGTGAGCAGTGGAGACACCCCCGTGGCCCACCACAAAGCCCACAA   960
295   R  Q  P  Y  E  S  E  Q  W  R  H  P  R  G  P  P  Q  S  P  Q   314

961  GACAACGGCTTGGAGGAGACTATCTGCAGCATGAGGACCCACGAGAAACATTGATGACCCA  1020
315   D  N  G  L  E  E  T  I  C  S  M  R  T  H  E  N  I  D  D  P   334

1021 GCCCGTGCTGACGTGTATAAGCCCAACCTCGGCCGTGTGACTAGCGCTAACAGCTACACT  1080
335   A  R  A  D  V  Y  K  P  N  L  G  R  V  T  S  A  N  S  Y  T   354

1081 TTACCCATCTTGCAGTATATCAGACTCAGCGCCACCCGTGGCATTCTCCAGGGTAATGCG  1140
355   L  P  I  L  Q  Y  I  R  L  S  A  T  R  G  I  L  Q  G  N  A   374

1141 ATGGTGCTTCCGAAATACAACATGAACGCGAACGAGATCTTGTACTGCACTCAAGGACAA  1200
375   M  V  L  P  K  Y  N  M  N  A  N  E  I  L  Y  C  T  Q  G  Q   394

1201 GCAAGGATTCAAGTGGTGAACGACAACGGACAGAACGTGCTGGACCAGCAGGTGCAGAAG  1260
395   A  R  I  Q  V  V  N  D  N  G  Q  N  V  L  D  Q  Q  V  Q  K   414

1261 GGACAGCTCGTGGTCATCCCACAAGGATTCGCCTATGTTGTCCAGTCCCACCAAAACAAC  1320
415   G  Q  L  V  V  I  P  Q  G  F  A  Y  V  V  Q  S  H  Q  N  N   434

1321 TTCGAATGGATTTCTTTCAAGACAAACGCTAACGCGATGGTCAGCACTTTGGCCGGTAGA  1380
435   F  E  W  I  S  F  K  T  N  A  N  A  M  V  S  T  L  A  G  R   454

1381 ACCTCGGCCTTGAGGGCATTGCCACTAGAGGTCATAACCAACGCTTTCCAAATTTCTCTC  1440
455   T  S  A  L  R  A  L  P  L  E  V  I  T  N  A  F  Q  I  S  L   474

1441 GAGGAAGCTAGAAGGATCAAGTTCAACACGCTTGAGACCACTTTGACTCGTGCGCGCGGT  1500
475   E  E  A  R  R  I  K  F  N  T  L  E  T  T  L  T  R  A  R  G   494

1501 GGACAACCCCAGTTGATCGAGGAGATAGTCGAGGCTTAAGTTAAAACGTTTACTTTTACT  1560
495   G  Q  P  Q  L  I  E  E  I  V  E  A  ***
                                          and
1561 AATAAAGTAGTGACATGGTTACTATTGTAATGGTCAGTTTGTAATCATGTCCACTCTAAG  1620

1621 TTTTTAACGTATGTGTAAAATATGTGTCTAAAGAACACGCCGGCACGTCTCTTGTATGTA  1680

1681 ACCCTTCTAATAAATACCCCCCCCC
```

FIG. 4. **Nucleotide sequence and deduced amino acid sequence of pCRU1.** Nucleotides and deduced amino acid residues are *numbered* at the *right margin*. The *arrow* at −1/+1 denotes the signal sequence cleavage site. The glycine-glutamine repeats (starting at amino acid residue 121), octaglutamine sequence (starting at amino acid residue 238), and the two direct repeats (starting at amino acid residues 379 and 439) are *underlined*. An *arrow* between amino acid residues 316 and 317 indicates the proteolytic cleavage site between the α- and β-chains. The termination codon is indicated by *three asterisks*. Two consensus polyadenylation signals are *boxed*.

of a single peptide bond between the α- and β-chain, this cDNA clone corresponds to a precursor giving rise to an α-chain of 296 amino acid residues (M_r = 32,900) and a β-chain of 190 residues (M_r = 21,200).

DISCUSSION

We have compared partial amino acid sequences of isolated cruciferin chains with the deduced amino acid sequences from a cDNA clone, pCRU1. This comparison made it possible to establish that this clone encodes a precursor that contains the $\alpha 1$-$\beta 1$ subunit. A previously characterized cDNA clone, pC1 (21), was also included in the comparison. pC1 encodes a

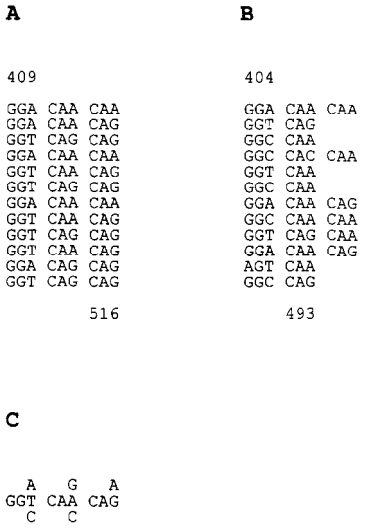

FIG. 5. **Glycine-glutamine repetitive regions in cruciferin sequences.** The nucleotide sequences corresponding to the glycine-glutamine repeats of pCRU1 (*A*) and pC1 (*B*) are shown. The consensus sequence is shown in *C*. The numbering refers to nucleotide positions in the respective cDNA clone.

distinct cruciferin precursor which displays complete identity with the available sequence of cruciferin chains $\alpha 2/3$, $\beta 2$, and $\beta 3$. Both the $\alpha 2$ and $\alpha 3$ as well as the $\beta 2$ and $\beta 3$ are highly similar. The $\alpha 2/3$ chains were analyzed as a mixture but showed virtually no sequence heterogeneity (Table I). $\beta 2$ and $\beta 3$ on the other hand which were analyzed separately were found to be identical in all determined amino acids positions. The $\alpha 4$ and $\beta 4$ chains were easily distinguished from the other chains. No cDNA clone encoding $\alpha 4$ or $\beta 4$ has so far been isolated and characterized, but peptide mapping has suggested that these two chains are synthesized together in a precursor, P3. These data also suggest that precursor P1 codes for $\alpha 1$ and $\beta 1$, whereas P2 generates the mature chains $\alpha 2/3$ and $\beta 2/3$.[3]

Three cruciferin-like genes, CRA1, CRBB, and CRC, have been isolated from another Crucifer, *Arabidopsis thaliana* (22). Two of these genes, CRA1 and CRBB, have been sequenced. Comparison of these sequences with the available sequence information on cruciferin precursors indicates that CRA1 is the homologue of the pC1 gene and that CRBB encodes a precursor homologous to cruciferin precursor P3. No sequence information is available on CRC. Accordingly, it is not yet possible to definitely say whether CRC and the pCRU1 gene are homologous. It is reasonable to assume that the different cruciferin precursor genes have arisen by multiple gene duplication events. The data above accordingly suggest that these duplications preceded the divergence of *A. thaliana* from its Brassica relatives.

The cruciferin precursor undergoes a proteolytic event as it is cleaved into the mature α- and β-chains. The site for this cleavage (Asn-Gly) is well conserved with those in other 11–12 S storage globulin precursors (23). It is not presently known whether additional proteolytic trimming of the COOH terminal of the α- and β-chains may occur.

The two cruciferin cDNA sequences are similar, but far from identical. Alignment of the nucleotide sequences as well as amino acid sequences reveals 64% identity. There is a repetitive sequence in the middle of the pCRU1 α-chain. It is 49 amino acids long and 47 of these residues are either glycine or glutamine. A similar repeat is present in almost the same position of the pC1 α-chain, although it is a few residues shorter. The nucleotide sequences encoding these two repetitive segments consist of repeated blocks of nine base pairs (Fig. 5). Many other plant storage proteins are known to contain segments composed of simple repetitive elements (24–26). It is assumed that this type of repetitive sequences originate from slippage during replication and/or by unequal crossing over. In the cruciferin family the original sequence probably was GG(A/T/C)CA(G/A/C)CA(A/G) which during evolution has been tandemly repeated more than 10 times. Some of these duplications may have involved multiple units of the basic 9-base pair repeat. Several of the repeats in pC1 are shorter by one triplet and may have duplicated separately or as a part of a longer repeat unit. It is reasonable to assume that the reiteration of this sequence was at least started before the genes corresponding to pCRU1 and pC1 diverged. During the time elapsed since then, the sequences have deviated somewhat.

It is not known whether cruciferin subunits like those of other 11–12 S proteins occur as a trimer in the endoplasmic reticulum and as a hexamer in the protein bodies. The elucidation of the relationship between the cruciferin precursors and the different mature chains as well as the availability of cruciferin cDNA clones will greatly facilitate studies on the assembly of the cruciferin molecule.

Acknowledgment—Kjell Magnusson is gratefully acknowledged for technical assistance.

REFERENCES

1. Finlayson, A. J. (1976) in *The Biology and Chemistry of the Cruciferae* (Vaughan, J. G., McLeod, A. J., and Jones, B. M. G., eds) pp. 279–306, Academic Press, New York
2. Lönnerdal, B., and Janson, J.-C. (1972) *Biochim. Biophys. Acta* **278,** 175–183
3. Crouch, M. L., Tenbarge, K. M., Simon, A. E., and Ferl, R. (1983) *J. Mol. Appl. Genet.* **2,** 273–283
4. Ericson, M. L., Rödin, J., Lenman, M., Glimelius, K., Josefsson, L.-G., and Rask, L. (1986) *J. Biol. Chem.* **261,** 14576–14581
5. Josefsson, L.-G., Lenman, M., Ericson, M. L., and Rask, L. (1987) *J. Biol. Chem.* **262,** 12196–12201
6. Scofield, S. R., and Crouch, M. L. (1987) *J. Biol. Chem.* **262,** 12202–12208
7. Crouch, M. L., and Sussex, I. M. (1981) *Planta* **153,** 64–74
8. Derbyshire, E., Wright, D. J., and Boulter, D. (1976) *Phytochemistry* **15,** 3–24
9. Schwenke, K. D., Raab, B., Plietz, P., and Damashun, G. (1983) *Die Nahrung* **27,** 165–175
10. Chrispeels, M. J., Higgins, T. J. V., Craig, S., and Spencer, D. (1982) *J. Cell Biol.* **93,** 5–14
11. Dobberstein, B., Garoff, H., Warren, G., and Robinsson, P. (1979) *Cell* **17,** 759–767
12. Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 4350–4359
13. Matsudaira, P. (1987) *J. Biol. Chem.* **262,** 10035–10038
14. Cleveland, D. W., Fischer, S. G., Kirschner, M. W. and Laemmli, U. K. (1977) *J. Biol. Chem.* **252,** 1102–1106
15. Surekha, M. Z., Shirsat, N. V., and Gothoskar, B. P. (1986) *Anal. Biochem.* **155,** 10–13
16. Auffray, C., and Rougeon, F. (1980) *Eur. J. Biochem.* **107,** 303–314
17. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
18. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
19. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene (Amst.)* **33,** 103–119
20. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
21. Simon, A. E., Tenbarge, K. M., Scofield, S. R., Finkelstein, R. R., and Crouch, M. L. (1985) *Plant Mol. Biol.* **5,** 191–201
22. Pang, P. P., Pruitt, R. E., and Meyerowitz, E. M. (1988) *Plant Mol. Biol.* **11,** 805–820
23. Borroto, K., and Dure III, L. (1987) *Plant Mol. Biol.* **8,** 113–131
24. Spena, A., Viotti, A., and Pirrotta, V. (1983) *J. Mol. Biol.* **169,** 799–811
25. Rafalski, J. A., Scheets, K., Metzler, M., Peterson, D. M., Hedgcoth, C., and Soll, D. G. (1984) *EMBO J.* **3,** 1409–1415
26. Forde, J., Malpica, J.-M., Halford, N. G., Shewry, P. R., Anderson, O. D., Greene, F. C., and Miflin, B. J. (1985) *Nucleic Acid Res.* **13,** 6817–6832

Supplementary material to:

Characterization of a cDNA clone encoding a Brassica napus 12S protein (cruciferin) subunit: Relationship between precusors and mature chains.

by

Joakim Rödin, Mats L. Ericson, Lars-Göran Josefsson and Lars Rask

MATERIAL AND METHODS

Plants

Seeds of a dihaploid line of the B. napus variety Svalöfs Karat were generously provided by Dr. Lena Bengtsson, (Svalöf AB, Svalöv, Sweden). This rapeseed line was used throughout these studies, except for some of the amino acid sequence determinations where the rapeseed line Hanna (Weibulls AB, Landskrona, Sweden) was used. Seeds of this line were a kind gift of Dr. Hans Svensk.

Extraction and purification of cruciferin.

Extraction of seeds was done as described by Crouch and Sussex (7). Cruciferin was subsequently purified from the crude extract by gel chromatography on an S-300 Sephacryl column (100x1cm) (Pharmacia, Uppsala, Sweden) equilibrated with extraction buffer (10mM sodium-phosphate, pH 7.5; 0.5M NaCl; 1mM phenylmethylsulfonylfluoride). Fractions containing cruciferin were pooled and dialysed extensively against 20mM Tris-HCl, pH 8.0. The pool was then applied to a Mono-Q ion exchange chromatography column (5x0.5 cm) (Pharmacia), equilibrated with the dialysis buffer. A linear gradient of 0-0.5 M NaCl was used to elute the bound proteins. The purity of the eluted cruciferin was investigated as described below.

SDS-polyacrylamide gel electrophoresis and electrotransfer.

Crude rapeseed extract was subjected to SDS-polyacrylamide gel electrophoresis according to Dobberstein et al. (11). The electrophoretically separated proteins were transferred to Polyvinylidene difluoride (PVDF) membranes (Millipore, Bedford, MA, USA) as described by Towbin et al. (12) with some minor modifications. After electrophoresis, the gels were soaked in transfer buffer (50mM sodium-phosphate buffer, pH 6.5) for 15 min. Transfer was accomplished by electrophoresis at 400 mA for 4h at +4°C. The PVDF membranes were thereafter treated as described by Matsudaira (13). Bands of interest were localized by Coomassie Brilliant Blue staining prior to their excision and use for amino acid sequence determination.

Proteolytic digestion of cruciferin chains using S.aureus V8 protease.

Peptides were generated by limited proteolysis with S. aureus V8 protease as described (14). Crude rapeseed extract was applied on an SDS-polyacrylamide gel. After electrophoresis the gel was stained with Coomassie Brilliant Blue and bands of interest were excised. The gel slices were then incubated in soaking buffer (0.125M Tris-HCl, pH 6.8, 0.1% SDS and 1mM EDTA) for 60 min at room temperature. They were thereafter inserted in the sample slots of a second SDS-polyacrylamide gel (15-20% polyacrylamide) and overlaid with S. aureus V8 protease (5µg/lane). Electrophoresis was interrupted for 60 min when the tracking dye was approximately 1 cm from the separating gel, to allow the proteolysis to occur in the stacking gel. Electrophoresis was then continued and the resulting gel treated as described in the above section.

Cyanogen bromide cleavage of cruciferin chains.

Cruciferin chains were isolated by SDS-polyacrylamide gel electrophoresis and then processed for cyanogen bromide cleavage as described (15). Resulting peptides were separated by SDS-polyacrylamide gel electrophoresis, electrotransferred to PVDF membranes and used for amino acid sequence determination as described above.

Preparation of antiserum against cruciferin.

Purified cruciferin (100µg) was injected into the popliteal lymph-nodes of a rabbit. After 4 additional subcutaneous booster injections each with 100µg of cruciferin the rabbit was bled. The resulting antiserum (K452) gave in double immunodiffusion analysis against total rapeseed protein extract rise to a single precipitation line. Western blotting analysis of total rapeseed protein extract showed that the antiserum reacted exclusively with the four cruciferin α-chains.

Amino acid sequence determination.

Peptides immobilized on PVDF membranes by electrotransfer were sequenced with an Applied Biosystem 470 C sequencer using the PROFTA program (Applied Biosystems, Foster City, CA, USA). Phenylthiohydantoin-amino acids (PTH-amino acids) were analyzed by on-line microbore HPLC identification using an Applied Biosystem 120 A analyzer according to Applied Biosystem's Users manual with some minor modifications. 50-100 pmole of protein were used for the sequence determinations.

Northern blotting.

Total RNA was isolated from developing B.napus seeds 35 days after pollination (7,16). Polyadenylated mRNA was enriched by two passages through an oligo-dT cellulose column. Denaturing agarose gels were prepared and run as described (17). 2 µg of denatured mRNA were loaded per lane on a 1% agarose/formaldehyde gel and electrophoresed at 1.5 V/cm for 16h. Transfer of mRNA to nitrocellulose filters and the subsequent treatment of the filters were also according to standard procedures (17).

Random priming and hybridization to Northern blots.

0.1 µg of pCRU1 cDNA were random primed to obtain radioactively labelled probe. Prehybridizations and hybridizations were done with formamide-containing solutions according to standard protocools (17). Washing of filters was done at high stringency, ie. 3mM sodium citrate-HCl, pH 7.1; 30mM NaCl; 0.5% SDS at 58°C two times 1h. Filters were exposed on X-ray film with intesifying screens at -70°C.

Construction of a cDNA library and screening of clones.

The cDNA library construction and screening of clones were as described (4) with the exception that an anti-cruciferin rabbit antiserum (K452) was used for the immunoprecipitations.

Nucleotide sequence determination.

Nucleotide sequencing was performed according to Sanger et al. (18) with [^{35}S]α-thio-dATP as the labelled nucleotide. The M13 vector used was mp19 (19). The clones used for sequencing were obtained by cloning fragments derived by restriction enzyme digestions into the polylinker of M13 mp19. Parts of the cDNA were also sequenced according to Maxam and Gilbert (20). Synthetic primers used were synthesized with an Applied Biosystem 380 nucleotide synthesizer.

TABLE I: Amino acid sequences of cruciferin heavy chains and peptides (generated by Staph. aureus V8 protease and cyanogen bromide cleavage)

	α1-chain		α1;V8-2		α2/3;V8-1		α2/3;V8-2		α4;V8-4		α4;CNBr-1		α4;V8-1	
Degradation cycle	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)
1	R	5	A	8	A	26	T	12	Q(L)	6(10)	H	3	T	35
	Q	11	G	14	G	23	F	6	G	6	Q	11	F	17
	S	12	R	9	R	7	Q	8	Q	4	K	3	M	8
	L	13	V	8	I	20	D	10	E	3	V	10	D	14
5	G	14	E	9	E	20	S	6	G	5	E	12	S	11
	V	12	Y	7	V	16	S	6	Q	2	X	-	P	14
	P	6	X[1]	-	W	3	P(V)	2(2)	G	4	L	9	V	4
	P	10	D	1	D	16	F	2	Q	3	R	<1	F	10
	Q	7	H	<1	R	<1	Q	4	G	4	H	<1	G	13
10	L	12	N	1	K	<1	F	4	Q	4	G	7	G	12
	G	13	N	7	A	14	X	-	G	4	D	5	G	11
	N	3	P	3	P	19	G	4	F	2	T	7	Q	8
			Q	2	Q	10	G	4			I	4		
			I	4	L	13	S	<1			A	8		
15			X	-	R	<1	P	3			T	5		
			C	NQ[2]	X	-	F	<1						
			A	2	S	3	G	4						
			G	1	G	9	E	2						
			V	1	V	6	G	3						
20			S	<1	S	<1								
					F	4								
					V	3								
					X	-								
25					T	3								
					I	1								
					I	2								
					E	1								

1) X denotes unidentified amino acid
2) NQ denotes unquantitated amino acid

TABLE II: Amino acid sequences of cruciferin light chains and peptides (generated by Staph. aureus V8 protease and cyanogen bromide cleavage)

	β1-chain		β1;V8-1		β2-chain		β3-chain		β3;CNBr-1		β4-chain		β4;V8-1		β4;CNBr-1	
Degradation cycle	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)	Amino acid	Yield (pmole)
1	G	38	I	23	G	31	G	33	V	29	G	25	I	10	V	8
	L	44	L	27	L	39	L	30	L	33	L	28	S	12	L	8
	E	23	Y	19	E	17	K	15	P	53	E	17	Q	6	P	16
	E	25	C	NQ	E	28	E	31	Q	40	E	27	G	13	Q	10
5	T	71	T	37	T	46	T	54	W	19	T	30	Q	11	W	3
	I	33	Q	14	I	36	I	26	N	29	L	21	L	14	N	8
	C	NQ[1]	G	18	C	NQ	C	NQ	A	38	C	NQ	X	-	V	4
	S	43	Q	11	S	33	S	21	N	27	T	35	V	6	N	7
	M	24	A	20	A	18	A	15	A	34	M	19	V	7	A	7
10	R	4	R	4	R	7	R	3	A	25	R	4	P	14	N	5
	T	31	I	18	C	NQ	C	NQ	A	41	C	NQ	Q	10	A	5
	X[2]	-	Q	12	T	18	T	23	V	17	T	26	G	10	A	11
	E	6	V	11	D	12	D	7	L	30	E	12	F	7	L	6
15	N	10	V	20	N	10	N	9	Y	22	N	11	A	9	Y	5
	I	22	N	7	L	29	L	14	Y	10	L	12	V	3	V	5
	D	6	D	10	D	5	D	3	T	22	D	6	V	9	T	10
	D	17	N	6	D	21	D	11	D	13	D	15	K	<1	X	-
	P	16	G	10	P	12	P	12	G	15	P	7	R	<1	G	6
20	A	16	Q	7	S	23	S	19	K	13	S	8	A	8	X	-
	R	7	N	5	N	5	N	5	A	16	S	11	T	8	A	5
	A	15	V	9	A	11	A	10	H	NQ	A	8	S	2	X	-
	D	9	L	12	D	6	D	4	V	11	D	6	Q	2	I	2
	V	15	D	6	V	7	V	8	Q	18	V	7	Q	7	Q	5
	Y	12	Q	4	Y	8	Y	4	V	11	Y	6	F	1	M	<1
25	K	9	Q	7	K	5	K	7	V	13	K	3			V	2
	P	10	V	8	P	6	P	8			P	5				
	N	5	Q	4	Q	5	Q	8			X	-				
	L	18	K	7	L	11	L	10			L	7				
	G	7	G	7	G	7	G	8			G	5				
30			Q	4	Y	6	Y	3			Y	3				
			L	8			I	6			I	3				
			V	4			S	5								
			V	9			A	6								
35			I	7			L	8								
			P	5			N	2								
			Q	4			X	-								
			G	5			Y	2								
			F	5			D	2								
40			A	5												
			Y	3												
			V	6												
			V	8												
			Q	4												

1) NQ denotes unquantitated amino acid
2) X denotes unidentified amino acid

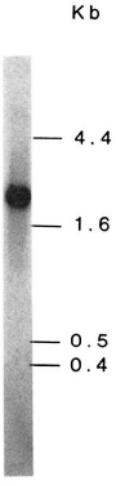

Fig. 2: Autoradiogram showing hybridization of labelled cruciferin cDNA clone pCRU1 with B. napus embryo mRNA.

Polyadenylated mRNA was purified and separated in a denaturing agarose gel. After transfer to nitrocellulose filter the immobilized mRNA was hybridized to a randomly primed [^{32}P]-labelled cDNA probe for cruciferin. The molecular weight marker used was a denatured HinfI digest of pBR322.

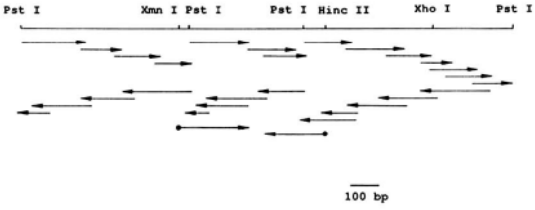

Fig. 3: Restriction map and strategy for the sequence analysis of the cDNA insert of pCRU1.

Restriction sites relevant for the sequence analysis are indicated. Horizontal arrows denote direction and extent of sequence determination. Sequence information was obtained on 100% and 87% of the coding and non-coding strands, respectively. Arrows starting with a dot denots sequence reactions carried out accordingly to Maxam and Gilbert (20).

Characterization of a cDNA clone encoding a Brassica napus 12 S protein (cruciferin) subunit. Relationship between precursors and mature chains.

J Rödin, M L Ericson, L G Josefsson and L Rask

J. Biol. Chem. 1990, 265:2720-2723.