Structure of the Rapeseed 1.7 S Storage Protein, Napin, and Its Precursor

Mats L. Ericson, Joakim Rödin, Marit Lenman, Kristina Glimelius, Lars-Göran Josefsson, and Lars Rask

From the Department of Cell Research, The Wallenberg Laboratory, and the Department of Plant Breeding, The Swedish University of Agricultural Sciences, S-751 22 Uppsala, Sweden

Napin (1.7 S protein) is a basic, low molecular weight storage protein synthesized in rapeseed (Brassica napus) embryos during seed development. Napin is composed of two polypeptide chains with molecular weights of 9000 and 4000 that are held together by disulfide bonds.

Comparison of the deduced amino acid sequence of a napin cDNA clone with that of napin peptide fragments established that napin is initially synthesized as a precursor of 178 residues. This polypeptide is subsequently processed through several proteolytic events, which ultimately generate two mature napin chains, of 86 and 29 residues, respectively. Protein biosynthesis in vitro showed that the initial translation product (Mr 20,000) contains a signal sequence which is removed during transfer of the protein into the endoplasmic reticulum. Two additional peptides, of 22 and 19 residues, as well as the COOH-terminal residue, are also removed during maturation of napin, as deduced from the sequence comparison. Comparisons of the napin sequence with other known protein sequences established that there is a significant homology between napin and two other small seed proteins, the castor bean storage protein and a trypsin inhibitor from barley.

Rapeseed (Brassica napus) has been used as a cultivated plant for 4000 years and is today a major oil-seed crop in many parts of the world. In addition to a high lipid level, the seeds also have a significant protein content, which constitutes some 20-25% of the dry seed weight (1). The predominant protein species are two seed storage proteins: the 12 S and 1.7 S proteins. The 12 S protein, or cruciferin, is a high molecular weight, neutral complex, composed of several polypeptide chains (2). In contrast, the 1.7 S protein, or napin, is a low molecular weight, basic protein, composed of two disulfide-linked polypeptide chains (3, 4).

The expression of both cruciferin and napin appears to be strictly regulated, as their synthesis is confined to embryonal and axis cells and occurs only during a limited time of seed development (5). Storage proteins are synthesized on and translocated across the rough endoplasmic reticulum membrane. Eventually they become deposited in distinct cellular vesicles, protein bodies, that most probably are derived from the endoplasmic reticulum (6). The possibility of engineering these proteins to improve their amino acid composition from a nutritional point of view has been given much attention. To this end, more has to be learned about the specific mechanisms underlying storage protein transport to, and deposition in, the protein bodies. No study has so far been able to reveal any general signal common to storage proteins that could serve for these purposes. As a prerequisite for studies on the biosynthesis and intracellular transport of storage proteins we have determined the primary structure of napin. We report here the sequence analyses of a cDNA clone and of napin peptide fragments. Sequence comparisons have enabled us to deduce the primary structure of the precursor and of the two mature napin chains.

MATERIALS AND METHODS AND RESULTS

Sequence analyses of napin polypeptide chains and of a cDNA clone encoding napin have enabled the elucidation of the primary structure of both the precursor and the mature napin chains. Napin is synthesized as a precursor consisting of 178 residues on membrane-bound ribosomes. During the transfer of the polypeptide into the microsomal lumen, a signal peptide is removed. Our data do not unambiguously establish where the cleavage of the signal sequence occurs. Nevertheless, we have tentatively assigned the alanine (−I in Fig. 5) as the processing site based on (i) general features of signal sequences, (ii) the "−1, −3 processing rule" (21, 22), and (iii) the shift in molecular weight of the products in vitro made in the absence or presence of microsomal vesicles (Fig. 3). The Ala (−I) residue constitutes the best compromise to meet these three criteria.

In addition to the above-mentioned processing, sequence comparisons reveal that amino acids +1−22, 52−70, as well as the carboxylterminal residue are removed during maturation of napin. Fig. 6 gives a schematic representation of the proc-
essing steps that the initial translation product undergoes to yield the mature napin. The order, intracellular location, and mechanism of the events involved in these latter processing steps of the napin precursor are not known. Enzymatic cleavage of other types of storage protein precursors takes place in the protein bodies (23). However, although the embryonal cells of B. napus do contain protein bodies (24), it has not actually been shown that napin is located in these vesicles. It is nevertheless quite likely that the napin precursor, from which the signal sequence has been removed, is transported to and deposited in protein bodies prior to the final processing. One experimental finding in support of this postulate is that the maturation of pulse-labeled napin is slow in vivo, whereas the removal of the signal peptide is fast (25).

Other storage proteins, e.g., those of legumes, have been found to possess conserved sequence motifs around the processing sites (26) which could be an indication of a conserved specificity of the processing enzymes. We were not able to find any analogous sequence motifs in the napin precursor and, consequently, we do not know whether the final processing steps of the napin precursor occur by endoproteolytic events alone or by exoproteolysis following an endoproteolytic cleavage at a single site within the removed peptide. The involvement of exoproteolytic activity in the protein maturation process gains some support from the observation that a fraction of the heavy chains are lacking two NH₂-terminal amino acids and that all of the heavy chains seem to have lost their COOH-terminal tyrosine residue. We cannot presently strictly rule out the possibility that these modifications occur during the purification of napin. An interesting feature of the napin precursor is that the charged amino acids are unevenly distributed and that 12 out of 16 of the negatively charged residues are removed during maturation. This gives the mature protein (pI = 11, Ref. 3) a substantially higher pI than that of the precursor (pI = 7, Ref. 4). The significance, if any, of this fact for the processing and deposition of the napin precursor remains to be established.

Analyses of the protein fragments, the cDNA clone in this work, and two other cDNA clones characterized by Crouch et al. (4) have confirmed that napin is heterogeneous in sequence. Although this heterogeneity appears to be quite limited, it points to the existence of a small gene family encoding napin variants. Napin has previously been classified among a vast, diverse members as secalin from rye, hordein from barley, the so-called CM-proteins, α-amylase inhibitors, and trypsin inhibitors from different species (27). Considering the molecular weight and overall primary structure of napin, the most obvious kinship seems to be that of napin and the castor bean storage protein or that of napin and some seed protease inhibitor proteins.

Napin and the castor bean storage protein are both small, basic, glutamine-rich proteins made up of two subunits linked together by disulfide bonds (28). They show no homology to the 7 S, 11 S, and 12 S storage globulins so far sequenced from Brassica and other dicots. Furthermore, by analogy with napin, the mature storage protein of the castor bean, M, 10,900, is derived from a M, 34,000 precursor form (29). A trypsin inhibitor from barley, M, 13,000 (30), together with the castor bean storage protein, gave high alignment scores when compared to napin in a computer analysis. The results of these comparisons are shown in a schematic drawing (Fig. 7). An Arg-Leu sequence in position 33 and 34 is the assigned reactive site of the trypsin inhibitor, the position of which is indicated in Fig. 7. The castor bean storage protein has Arg-Arg in the same positions, 33 and 34, and it has been specu-
lated whether the protein might in fact retain some protease inhibitory function (31). Napin, on the other hand, seems to lack such a site in the aligned sequence portion.

In two of the compared proteins, napin and the castor bean storage protein, we have found an imperfect repeat. There is one copy in the small subunits and the other in the NH2-terminal part of the large subunits (Fig. 8a). We have also noted that the repeat core in napin consists of amino acids with strong α-helix-forming tendency (data not shown). The amino acid substitutions found within the repeats when comparing different napin sequences are all conservative with regard to α-helix-forming tendency. A short, direct repeat with glutamine residues was also found in the napin large subunit, possibly reflecting a selective pressure, because nitrogen-containing amino acids as repetitive, glutamine-rich, sequence motifs are frequent among storage proteins (Fig. 8b). A close comparison between napin and the castor bean storage protein might perhaps give hints as to what sequence portions are relevant to questions of transport to, and storing in, the protein bodies of this class of seed proteins. However, we doubt whether a satisfactory answer can be drawn from protein primary structure alone.

Acknowledgments—Elisabeth Westergren and Eva Murén are gratefully acknowledged for technical assistance.

REFERENCES

1. Finlayson, A. J. (1976) in The Biology and Chemistry of the Cruciferæ (Vaughan, J. G., McLeod, A. J., and Jones, B. M. G., eds) pp. 279–306 Academic Press, New York
2. Schwenke, K. D., Raab, B., Linow, K.-J., Patz, W., and Uhlin, J. (1981) Die Nahrung 25, 271–280
3. Lönnrdal, B., and Janson, J.-C. (1972) Biochim. Biophys. Acta 278, 175–183
4. Crouch, M. L., Tenbarge, K. M., Simon, A. E., and Ferl, R. (1983) J. Mol. Appl. Genet. 2, 273–283
5. Crouch, M. L., and Sussex, I. M. (1981) Planta (Berl.) 153, 64–74
6. Larkins, B. A., and Hurkman, W. J. (1978) Plant Physiol. 62, 256–263
7. Gross, E., and Witkop, K. (1962) J. Biol. Chem. 237, 1856–1862
8. Sundelin, J., Laurent, B. C., Anundi, H., Traghrbl, L., Larhammar, D., Björck, L., Eriksson, U., Åkerström, B., Jones, A., Newcomer, M., Peterson, P. A., and Rask, L. (1985) J. Biol. Chem. 260, 6472–6480
9. Fohlman, J., Rask, L., and Peterson, P. A. (1980) Anal. Biochem. 106, 22–26
10. Auffray, C., and Rougeon, F. (1980) Eur. J. Biochem. 107, 303–314
11. Pelhara, H. R. B., and Jackson, R. J. (1976) Eur. J. Biochem. 67, 247–256
12. Kvist, S., Wiman, K., Claesson, L., Peterson, P. A., and Dodberstein, B. (1982) Cell 29, 61–69
13. Dodberstein, B., Garoff, H., Warren, G., and Robinson, P. (1979) Cell 17, 759–769
14. Gubler, U., and Hoffman, B. J. (1983) Gene (Amst.) 25, 263–269
15. Vieira, J., and Messing, J. (1982) Gene (Amst.) 19, 269–268
16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Iah-Horowicz, D., and Burke, J. F. (1981) Nucleic Acids Res. 9, 2989–2990
18. Ricciardi, R. P., Millers, J. S., and Roberts, B. E. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 4927–4931
19. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
20. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
21. von Heijne, G. (1984) EMBO J. 3, 2315–2318
22. von Heijne, G. (1984) J. Mol. Biol. 173, 243–251
23. Chrispeels, M. J., Higgins, T. J. V., Craig, S., and Spencer, O. (1982) J. Cell Biol. 93, 5–14
24. von Hofsten, A. (1974) Sven. Bot. Tidskr. 68, 153–163
25. Simon, A. E., Tenbarge, K. M., Scofield, S. R., Finkelstein, R. R., and Crouch, M. L. (1985) Plant Mol. Biol. 5, 191–201
26. Gatehouse, J. A., Croy, R. R. D., and Boulter, D. (1984) CRC Crit. Rev. Plant Sci. 1, 287–314
27. Kreis, M., Forde, B. G., Rahman, S., Miflin, B., and Shewry, P. R. (1985) J. Mol. Biol. 185, 498–502
28. Sharief, F. S., and Li, S. S.-L. (1982) J. Biol. Chem. 257, 14753–14757
29. Butterworth, A. G., and Lord, J. M. (1983) Eur. J. Biochem. 137, 57–65
30. Odani, S., Koide, T., and Ono, T. (1983) J. Biol. Chem. 258, 7998–8000
31. Odani, S., Koide, T., Ono T., and Ohnishi, K. (1983) Biochem. J. 213, 543–545
32. Dayhoff, M. O., Schwartz, R. M., and Greutt, B. C. (1979) in Atlas of Protein Sequence and Structure (Dayhoff, M. O., ed) Vol. 5, Suppl. 3, pp. 345–362 National Biomedical Research Foundation, Washington, D. C.

APPENDIX

Fig. 5. Nucleotide sequence and deduced amino acid sequence of the coding region of the pNP1 cDNA. The differences in two previously sequenced napin cDNA clones (4) and corresponding amino acid residues are also indicated. The numbering of the amino acid residues refers to a tentatively assigned signal peptide processing (see "Discussion"). Open arrows show a direct imperfect repeat and solid arrows a direct perfect repeat. Dashes indicate regions not sequenced. A deletion in PNP1 is denoted by asterisks. Vertical arrows denote processing sites.
Chromatography in SDS-polyacrylamide gel electrophoresis of the napin preparation revealed denaturing conditions (Fig. 11). The molecular weights of the heavy and light napin chains were determined and the apparent molecular weight of the heavy chain was 13,000. Reduction and alkylation of the sample revealed two bands with apparent molecular weights of 9,500 and 4,500, respectively.

Preparation of the immunoprecipitated polypeptide was carried out by the calcium phosphate precipitation method of Maxam and Gilbert (1977). The tailed cDNA was synthesized by oligo-dC primer and terminal deoxynucleotidyl transferase reaction. It was allowed to proceed for 16 h at 37°C. The tailed cDNA was then purified by niacin chromatography. The nucleotide sequence was determined and the total RNA isolated from developing rapeseed and cell-free translation was performed in accordance with the calcium phosphate precipitation method of Maxam and Gilbert (1977).

The specificities of the antiserum for napin preparation and autoradiography (131. The specificity of the antiserum for napin was verified by immunodiffusion and Western blotting analyses against purified napin as well as against a crude rapeseed protein preparation. The antiserum was found to react exclusively with the two polypeptides containing both chains of the mature napin.

Total RNA was isolated from developing rapeseed seeds and analyzed for its ability to direct protein synthesis in vitro in a gelatinose system. The hybridization profile as determined by hybridization with polyacrylamide gel electrophoreses, more than a dozen components were evident. One of these polyacides could be immunoprecipitated using an antiserum against napin (Fig. 3, lane A). This antibody appeared to precipitate two polypeptides with apparent molecular weights of 11,500 and 4,500, respectively (Fig. 3, lane A). This suggested that the in vitro product contained an amino-terminal sequence corresponding to amino acids 1 to 315 of the mature sequence.

The specificity of the antiserum for napin reacted exclusively with the heavy and light chains of napin (data not shown). It thus appeared that the single polypeptide made in vitro is a precursor, consisting between chains of the mature napin. Several independent lines of evidence indicate that napin possesses diasomal properties in sequence. First, although most amino acids showed close agreement in the heavy chain, it was not possible to rule out the alternative explanation that products related to only one of the mature polyacides were made in vivo.

Isolation and sequencing of the cDNA clone encoding a napin housekeeping gene from rapeseed mRNA and cell-free translation was performed in accordance with the calcium phosphate precipitation method of Maxam and Gilbert (1977). The tailed cDNA was synthesized by oligo-dC primer and terminal deoxynucleotidyl transferase reaction. It was allowed to proceed for 16 h at 37°C. The tailed cDNA was then purified by niacin chromatography. The nucleotide sequence was determined and the total RNA isolated from developing rapeseed and cell-free translation was performed in accordance with the calcium phosphate precipitation method of Maxam and Gilbert (1977).

The specificities of the antiserum for napin preparation and autoradiography (131. The specificity of the antiserum for napin was verified by immunodiffusion and Western blotting analyses against purified napin as well as against a crude rapeseed protein preparation. The antiserum was found to react exclusively with the two polypeptides containing both chains of the mature napin.

The nucleotide sequence was determined and the total RNA isolated from developing rapeseed seeds and analyzed for its ability to direct protein synthesis in vitro in a gelatinose system. The hybridization profile as determined by hybridization with polyacrylamide gel electrophoreses, more than a dozen components were evident. One of these polyacides could be immunoprecipitated using an antiserum against napin (Fig. 3, lane A). This antibody appeared to precipitate two polypeptides with apparent molecular weights of 11,500 and 4,500, respectively (Fig. 3, lane A). This suggested that the in vitro product contained an amino-terminal sequence corresponding to amino acids 1 to 315 of the mature sequence.

The specificity of the antiserum for napin reacted exclusively with the heavy and light chains of napin (data not shown). It thus appeared that the single polypeptide made in vitro is a precursor, consisting between chains of the mature napin. Several independent lines of evidence indicate that napin possesses diasomal properties in sequence. First, although most amino acids showed close agreement in the heavy chain, it was not possible to rule out the alternative explanation that products related to only one of the mature polyacides were made in vivo.

The results are summarized in Table I. The data indicate that the heavy and light chains of napin have close amino acid sequences, and the two chains have close amino acid sequences. The heavy chain contains a single open reading frame consisting of 534 base pairs. The light chain contains a single open reading frame consisting of 534 base pairs. The data indicate that the heavy and light chains of napin have close amino acid sequences, and the two chains have close amino acid sequences. The heavy chain contains a single open reading frame consisting of 534 base pairs. The light chain contains a single open reading frame consisting of 534 base pairs.
### Table III

| Lys | Pro | Ser | Gly | Val | Gly | Tyr | His | Leu | Gln | Asp | Glu | Lys | Gln | Phe | Leu |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0   | 1   | 0   | 1   | 1   | 0   | 0   | 0   | 0   | 1   | 0   | 1   | 1   | 0   | 0   | 0   |

### Figure 1

**Molecular weight determination of napin subunits.** A reduced and alkylated sample of napin was analyzed by use of an LKB Ultropac TSK-G2000SW column (3.2 x 60 cm), which was equilibrated with a buffer containing 0.1 M sodium acetate (pH 5.5) and 6 M urea. The column was calibrated with reduced and alkylated IgG, IgM, and bovine serum albumin (100,000) and calf thymus histones (500,000). The molecular weights of the empty gel columns were determined by use of blue dextran and standard proteins. The positions of the void volume, the molecular weight markers in the above order, and the total volume are indicated by the positions of the arrows from left to right.

### Figure 2

**Absorbance at 280 nm.**

| Effluent volume (ml) | Absorbance at 280 nm |
|----------------------|----------------------|
| 5                    | 0.15                 |
| 10                   | 0.30                 |
| 15                   | 0.45                 |
| 20                   | 0.60                 |
| 25                   | 0.75                 |

### Figure 3

**Translation in vitro of expressed napin cDNA.** The 1.2-kb cDNA encoding the napin precursor was excised from the cosmids clone pNAP1 and inserted into the Bam HI site of the baculovirus transfer vector pAc5. Cotransfection of the recombinant baculovirus with an SV40 early promoter construct expressing luciferase in sf9 insect cells gave rise to transgenic SF9 lines. The translated product was fractionated by gel electrophoresis. Molecular weights were determined by use of blue dextran and standard proteins. The positions of the void volume, the molecular weight markers in the above order, and the total volume are indicated by the positions of the arrows from left to right.

### Figure 4

**Restriction map and cloning strategy of pNAP1.** The 718 base pair long napin cDNA, pNAP1, was cloned in pUC19.