Carlsberg Res. Commun. Vol. 48, p. 187–199, 1983

NUCLEOTIDE SEQUENCES OF cDNA CLONES FOR B1 HORDEIN POLYPEPTIDES

by

SØREN K. RASMUSSEN, H. ESTEBAN HOPP[1)]

and

ANDERS BRANDT

Department of Physiology, Carlsberg Laboratory,
Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby

[1)]Fellow of the Consejo Nacional de Investigaciones Cientificas y Técnicas, on leave from the Departamento de Genética, INTA, CC 25, Castelar, República Argentina

Keywords: Barley, storage protein, prolamin, amino acid sequence, polyadenylation signal

Nucleotide sequences of four cDNA clones coding for the carboxy-terminal portion of at least two different B1 hordein polypeptides are presented. The open reading frame in the nucleotide sequence of the largest clone (pc hor2-4, 720 nucleotides) translates into the 181 carboxy-terminal amino acids of a polypeptide chain showing close homology to the previously determined primary structure of B1 hordein peptides. Of the 74 amino acid residues which can be compared 61 proved to be identical. The second cDNA clone (pc hor2-3, 257 nucleotides) encodes the 54 carboxy-terminal amino acids of a different B1 hordein polypeptide, which is revealed by 21 nucleotide substitutions resulting in 9 amino acid changes.

The two other analysed cDNA clones contained sequences of 54 and 41 nucleotides respectively for the carboxy-terminal end of the same B1 hordein polypeptide as that coded for by pc hor2-3. The latter two clones (pc hor2-1, 254 nucleotides and pc hor2-2, 153 nucleotides) comprised the entire 3' noncoding region of the B1 hordein messenger RNA including a poly(A) tail. Clone pc hor2-2 measures 99 nucleotides between the stop codon and the poly(A) tail. A putative polyadenylation signal AATAAA is located 15 residues upstream from the poly(A) site. These 99 nucleotides of the 3' noncoding region are extended by a sequence of 63 nucleotides in clone pc hor2-1. This additional sequence contains two AATAAA sequences located 13 and 24 nucleotides respectively from its poly(A) tail.

Abbreviations: BCIG = 5-bromo-4-chloro-3-indolyl-β-galactoside; c in plasmid designation = cDNA; cDNA = DNA complementary to mRNA; IPTG = isopropyl-β-thiogalactoside; p = plasmid; poly(A) = polyadenylic acid; SDS = sodium dodecylsulphate.

0105/1938/83/0048/0187/$02.60

1. INTRODUCTION

Hordein, the alcohol soluble storage protein of the barley endosperm, can be separated into a number of polypeptide bands by SDS-gel electrophoresis (14, 15). Peptide mapping and amino acid sequence analysis have revealed structural homology as well as distinct differences among the separated polypeptides (23, 41, 42, 46). Genetic analysis of the polymorphisms in polypeptide band patterns has shown that one locus on barley chromosome 5, designated Hor-2, determines the structure of the B hordein polypeptides and another linked locus designated Hor-1, encodes the larger C hordein polypeptides (14, 16, 34, 45, 47). The hordein polypeptides are synthesized by the ribosomes of the endoplasmic reticulum in the endosperm cell as larger sized precursors (8, 29). Subsequently they are processed and transported into the vacuoles of the cell where their final deposition takes place (7, 9). Several barley mutants defective in hordein biosynthesis are available. Mutant Risø 1508 (lys3a) reduces the overall synthesis of hordein polypeptides, whereas mutant Risø 56 (hor2ca) preferentially inhibits the synthesis of B hordein polypeptides (14, 27). The abundant formation of messenger RNA encoding hordein polypeptides during the development of the endosperm cell makes this messenger RNA readily available for isolation and in vitro translation (5, 29). By reverse transcription and insertion of the complementary double stranded DNA (cDNA) into bacterial plasmids, clones encoding hordein polypeptides have been obtained (4, 19, 24).

In order to assess if the extensive heterogeneity suggested by peptide mapping and amino acid sequence analysis of B and C hordein polypeptides is reflected in the messenger RNA population, we have determined the partial nucleotide sequence of four cDNA clones encoding portions of B1 hordein polypeptides.

2. MATERIALS AND METHODS

2.1. Chemicals

Restriction endonucleases, Alu I, Eco RI and Pst I were obtained from Boehringer Mannheim, Sau 3A from Bethesda Research Laboratories (USA), Acc I and Hinc II from New England Biolabs (USA) or Amersham International (UK). The 2'-deoxyribonucleotide triphosphates, DNA polymerase (Klenow fragment), calf intestinal alkaline phosphatase, T4-polynucleotide kinase and T4-DNA ligase were obtained from Boehringer Mannheim and 2',3'-dideoxynucleotide triphosphates and 2'-deoxyinosine triphosphate were obtained from P-L Biochemicals (USA). Radionucleotides α-^{32}P-dATP (>800 Ci/mmol) were obtained from New England Nuclear and γ-^{32}P-ATP (>5000 Ci/mmol) were from Amersham International. Synthetic M13 specific primer d(T-C-A-C-G-A-C-G-T-T-G-T) was obtained from Collaborative Research (USA) and the oligo-dT primer d$(T)_{12}$-G from New England Biolabs. The DNA Sequencing System (Maxam-Gilbert procedure) was supplied from New England Nuclear. 5-Bromo-4-chloro-3-indolyl-β-galactoside (BCIG) and isopropyl-β-thiogalactoside (IPTG) were obtained from Sigma Chemicals (USA). Low melting temperature agarose was obtained from Bethesda Research Labs. X-ray film (X-omat RP) and X-omatic regular intensifying screens were from Kodak (USA). Amberlite MB-1 and polyethyleneglycol 6000 were obtained from BDH Chemicals (UK). Capillar tubes 1.5-1.8 × 150 mm were supplied by Bie & Berntsen (Denmark). The 0.2 mm thick PVC foil for spacers and slot formers in sequencing gels was supplied by RIAS (Denmark).

2.2. Strains

E. coli HB101 cells were used for transformation with recombinant pBR322 plasmids. E. coli JM101 cells served for propagation of phage M13 mp701. This phage was constructed and kindly supplied by DAVID BENTLEY (Department of Pathology, University of Oxford, UK).

2.3. Media and stock solutions

Growth media for bacteria and phages consisted of 2x YT (Bacto tryptone 1.6%, Bacto yeast extract 1.0%, NaCl 0.5%) and LA-plates (Bacto tryptone 1.0%, Bacto yeast extract 0.5%, Bacto agar 1.0%, NaCl 1.0%, glucose 0.1%). Soft agar was made of 0.7% agar in water. Buffer: 10x TE contained 100mM-Tris-HCl, 10mM-EDTA, pH 7.5. Formamide was deionized with Amberlite MB-1 and stored at -20 °C until use. Phenol was redistilled, equilibrated with 10x TE buffer, and

kept at -20 °C until use. All chemicals used in the media were analytical grade.

2.4. Isolation of DNA for nucleotide sequencing

The recombinant plasmids for nucleotide sequencing were purified from 1 liter cultures by $CsCl_2$ density gradient centrifugation according to the procedure of CLEVELL and HELINSKY (11) or from 10 ml cultures according to BIRNBOIM and DOLY (3). Restriction endonuclease fragments were separated by electrophoresis in low melting temperature agarose, and the appropriate fragments isolated by melting the excised gel slices in an equal volume of 10x TE buffer at 70 °C for 10 minutes. Following phenol extraction, the DNA was ethanol precipitated. The DNA was resuspended in TE buffer and the fragments chromatographed through a Sephadex G-50 column and ethanol precipitated.

2.5. Nucleotide Sequencing by the dideoxy method

The isolated cDNA fragment was either ligated directly into the appropriate site of the replicative form of M13 mp701 or restricted with a second restriction endonuclease and thereafter inserted into the vector using T4-DNA ligase. E.coli JM101 was transformed after the cells had been made competent for transformation by $CaCl_2$ treatment (28) and plated in the presence of 0.5 mM-IPTG and 0.03 % BCIG. Harvesting of phages, purification of single stranded phage DNA, and the chain termination reactions were essentially as described in (32, 38, 39).

2.6. Nucleotide sequencing by the base specific chemical degradation method

When premature termination of the polymerase reaction occurred probably due to the formation of secondary structures in the cDNA, the cDNA was sequenced by the chemical degradation method (30). The 5' phosphate group of the DNA fragment was removed by treatment with calf intestinal alkaline phosphatase (10) followed by labelling with γ-^{32}P-ATP using polynucleotide kinase (31). The 5' end labelled DNA was then restricted with a second endonuclease to cleave it into two fragments of unequal size, which can be separated and isolated on 6% polyacrylamide gel (31).

2.7. Nucleotide sequencing gels

The labelled DNA fragments were separated electrophoretically with a constant voltage of 2000 V, at 60 °C on 0.2 mm thick polyacrylamide gels (200x500mm) containing 7.3 M-urea (37). The gel was cast on to silicone coated glass plates according to the procedure of ANSORGE, DE MAYER and GAROFF (1, 20). After electrophoresis, the gel was soaked in 10% acetic acid for 10 minutes in order to remove urea from the gel, and then dried at 70 °C for 1 hour. Autoradiography was performed at -70 °C with an intensifying screen.

2.8. Containment

The experiments were registered with the Committee on Genetic Engineering of the Danish National Research Councils, and carried out under P1 laboratory conditions (22).

3. RESULTS

3.1. Nucleotide sequence determination of the pc hor2-4 cDNA insert

The cDNA containing plasmid pc hor2-4 was chosen from a cDNA library for nucleotide sequence analysis because of an insert size of more than 500 base pairs. It cross hybridized with plasmid pc hor2-3 which by sequence analysis had revealed the presence of an open reading frame for the carboxy-terminal amino acid sequence (42) of the B1 hordein polypeptide.

As the cDNA had been inserted into the Pst I restriction site of the bacterial plasmid pBR322 by GC-tailing and did not contain an internal Pst I restriction site, the insert could be recovered by digestion with the restriction endonuclease Pst I as a single fragment. This fragment was inserted into the Pst I restriction site of the replicative form of the phage M13 mp701. These subcloning experiments are expected to give rise to single stranded M13 mp701 recombinants containing one or the other DNA strand depending on the orientation with which the cDNA is inserted in the double stranded vector.

The strategy for sequencing the 720 nucleotide long cDNA insert is outlined in Figure 1. Approximately 300 nucleotides could be sequenced starting from both Pst I restriction sites. This revealed the location of the Sau 3A, Alu I

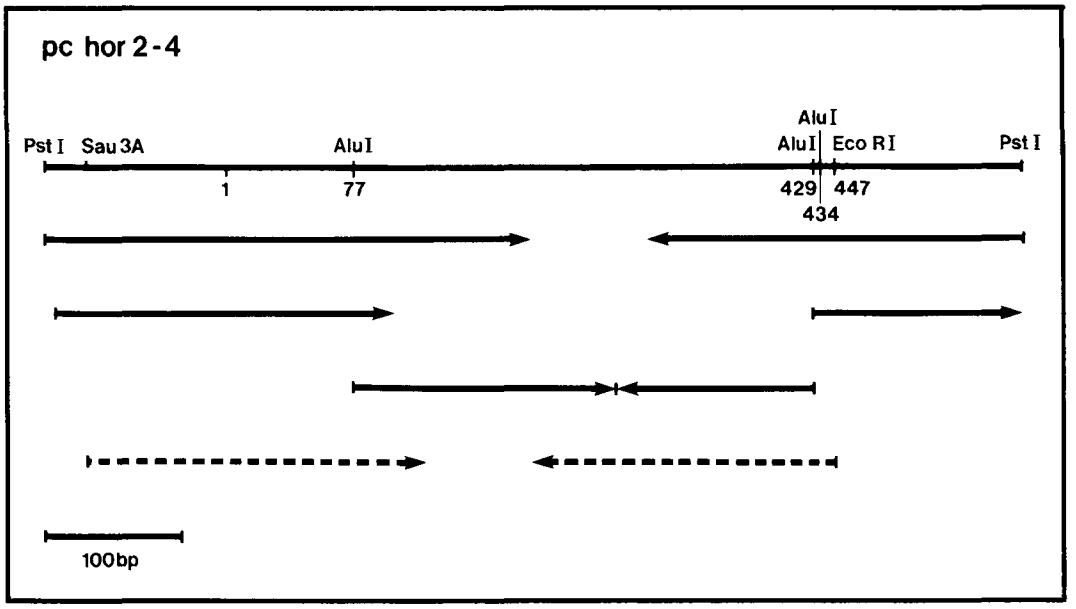

Figure 1. Sequencing strategy for the 720 nucleotide long cDNA clone pc hor2-4, encoding a B1 hordein polypeptide.

The indicated restriction endonuclease sites served as starting point for sequencing. Horizontal arrrows indicate the direction of sequencing and the lenght of the sequences obtained. Solid lines indicate the sequences determined by the dideoxy chain termination method and broken lines those obtained by the chemical degradation method. The nucleotide numbering used in Figure 2 is given under the restriction site map.

and Eco RI restriction sites shown in Figure 1. In the region between the left Pst I and the Sau 3A site a stretch of 16 C's and 11 A's gave several times rise to double band patterns indicating a mixture of two fragments one being a base shorter. In order to avoid ambiguity in reading the sequence, this stretch of nucleotides was used as annealing site for an alternative primer, oligo $d(T)_{12}G$ (left arrow in third line of Figure 1).

For further sequence information, the Pst I restriction fragment was digested with restriction endonuclease Alu I, and the resulting fragments ligated into the Hinc II restriction site of M13 mp701 DNA. One recombinant contained in tandem the right Alu I - Pst I and the left Pst I - Alu I restriction fragments. This is due to the generation of a central Pst I restriction site by ligation of the two fragment ends. The fragment permitted the determination of the nucleotide sequence from the two Alu I sites through the Eco RI site to the right Pst I restriction site (right arrow in third line Figure 1). Recombinant phages containing the large internal Alu I restriction fragment allowed the determination of the sequence in both orientations (fourth line in Figure 1). Overlapping sequences in the central part of the fragment were not readable due to consistent compression of the sequencing ladders in this region, a situation possibly due to secondary structures in the DNA. Readability was not im-

Figure 2. The nucleotide sequence of the coding region of pc hor2-4 region coding for the carboxy-terminal end of a B1 hordein polypeptide.

The amino acid sequence predicted by the genetic code is shown below the nucleotide sequence and compared to six B1 hordein peptide sequences previously established by automated EDMAN degradation (42), a) CNBr III; b) T III; c) SV8 II; d) SV8 IV; e) CNBr V, are fragments obtained by CNBr-cleavage and digestion with trypsin or Staphylococcus aureus V8 protease; f) carboxy-terminus. The solid lines indicate identity between the determined and the predicted amino acid sequence. At positions where differences were encountered the amino acid residues are also written out in the peptide sequence.

```
            1                                                  30
pc hor2-4   CTA GCA TGT GAG GTG TTG AAA GAA CGT TCT GTA TGT ATA CTT
            leu ala cys glu val leu lys glu arg ser val cys ile leu
B1 peptides
                        60                          Alu I          90
CAA CTT CAC TAT GTT CAA CCA TCT ATT TTG CAA CAG CTA AAC CCA TGC AAG
gln leu his tyr val gln pro ser ile leu gln gln leu asn pro cys lys

                                    120
GTA TTC CTC CAG CAG CAG TGC AGC CCC GTG CGA ATG CCA CAA CTT ATT GCT
val phe leu gln gln gln cys ser pro val arg met pro gln leu ile ala

    150                                             180
AGG TCG CAA ATG TTG CAG CAG AGC AGT TGC CAT GTG TTG CAG CAA CAA TGT
arg ser gln met leu gln gln ser ser cys his val leu gln gln gln cys
              a|--------------------X-------------------------lys
                  210                                   240
TGC CAG CAA CTG CCG CAA ATC CCC GAA CAA TTC CGC CAT GAG GCA ATC CGT
cys gln gln leu pro gln ile pro glu gln phe arg his glu ala ile arg
-X-----glu----------|

                                270
GCA ATC GTC TAC TCT ATC TTT CTG CAA GAA CAA CCC CAA CAG TCG GTC CAA
ala ile val tyr ser ile phe leu gln glu gln pro gln gln ser val gln
              b|-------------------------------------------leu-------

300                                         330
GGT GCC TCC CAA CCC CAA CAA CAG TTG CAG GAG GAG CAA GTC GGA CAA TGT
gly ala ser gln pro gln gln gln leu gln glu glu gln val gly gln cys
----val-----------------glu---------(trp)|
c|---val-----------------glu---------trp-pro-gln-----------gly---X-

            360                                         390
TAT TTC CAA CAA CCT CAA CCA CAA CAA CTT GGT CAA CCA CAA CAG GTA CCA
tyr phe gln gln pro gln pro gln gln leu gly gln pro gln gln val pro
leu|

                        420          Alu I  Alu I           Eco RI
CAG AGT GTT TTC TTG CAG CCA CAC CAG ATA GCT CAG CTT GAG GCC ACG AAT
gln ser val phe leu gln pro his gln ile ala gln leu glu ala thr asn
                                                  d|---------X-

                                    480
TCC ATT GCG CTG CGT ACC CTA CCA ACG ATG TGC AAT GTT AAT GTG CCA TTG
ser ile ala leu arg thr leu pro thr met cys asn val asn val pro leu
-------------------------------val|   e|-X---ser-------------------

        510                                         540           552
TAT GAC ATC ATG CCA TTC GGC GTT GGC ACT AGA GTT GGT GTC TAA TGA TAA
tyr asp ile met pro phe gly val gly thr arg val gly val
----arg----(leu)|                                  f|-------|
```

proved by either raising the temperature from 24 °C to 37 °C during the polymerase reaction or by substitution of deoxyguanosine triphosphate with deoxyinosine triphosphate in the incubation mixture. This part of the nucleotide sequence was therefore analysed by chemical degradation of the DNA according to MAXAM and GILBERT (30, 31). The 5'-OH ends of the Eco RI restriction site in the recombinant plasmid were labelled with ^{32}P employing T4 polynucleotide kinase. Following Pst I digestion, the labelled Eco RI - Pst I restriction fragment was purified and subjected to chemical degradation. This procedure allowed the determination of 200 nucleotides from the Eco RI restriction site towards the center of the nucleotide sequence and a 250 nucleotide long sequence determination from the Sau 3A restriction site. As will be evident from the location of the restriction sites in the sequence depicted in Figure 2 the carboxy-terminus is located to the right of the Eco RI site. An open reading frame was not present beyond the position marked as 1, which thus was used as the start point for numbering the nucleotides of the clone.

3.2. Alignment of the amino acid sequence of six B1 hordein peptides in the open reading frame of the pc hor2-4 insert

The nucleotide sequence of pc hor2-4 presented in Figure 2 predicts a 181 amino acid long polypeptide, into which can be fitted all previously reported (42) amino acid sequences of B1 hordein peptides. The N-terminal eighteen amino acids of the CNBr III fragment (a) are colinear with the amino acid sequence deduced from the codons starting with nucleotide 157 and ending with nucleotide 213. One glutamic acid in the peptide sequence is replaced by a glutamine, a substitution possible by a single nucleotide change from GAA to CAA. The single lysine residue in the peptide sequence is replaced by a cysteine, an alteration requiring three nucleotide changes in the codon. Both of the undetermined amino acid residues in the CNBr III fragment (X) were found to be cysteine. The CNBr V peptide (e) consists of 10 amino acid residues, which line up with the amino acids coded by nucleotides 481 to 510. One serine is replaced by an asparagine, requiring a codon change from AGT to AAT. The change from arginine to an aspartic acid requires two nucleotide changes in the codon. One undetermined residue in the CNBr V peptide again turned out to be cysteine.

The two peptides SV8 II (c) and SV8 IV (d) obtained by cleavage with Staphylococcus protease SV8 can be aligned with the deduced amino acid sequence. The codons for the SV8 II peptide span the nucleotides 298 to 351. SV8 protease cleaves after a glutamic or aspartic acid. However, it is seen that the peptide starts after a codon for glutamine indicating that the cleavage at this position is due either to partial deamidation of the glutamine during the purification of the B1 hordein polypeptide or due to a polymorphism. Of the 18 amino acid residues in the peptide, 7 are different from those predicted by the nucleotide sequence. Three of the changes can be obtained by single nucleotide substitutions : valine to alanine, glutamic acid to glutamine, and glutamine to glutamic acid. The remaining replacements are: tryptophan to glutamine, proline to glutamic acid, glycine to glutamine and leucine to tyrosine requiring two nucleotide changes. One undetermined residue reads as cysteine. The SV8 IV peptide sequence comprises 12 amino acid residues of which 11 are identical to those predicted by the nucleotide sequence. Two nucleotide changes are required to account for the conversion of valine to threonine. One undetermined residue in the peptide sequence reads as an asparagine.

The tryptic peptide sequence T III (b) can also be aligned with the amino acid sequence predicted by nucleotides 259 to 327. Of the 21 amino acid residues in this peptide sequence, four are different from the nucleotide sequence prediction. The SV8 II peptide overlaps the last 10 amino acid residues of the T III peptide, a region already discussed. Of the remaining 11 residues in the T III peptide, one leucine is replaced by serine, a change requiring only a single nucleotide substitution. The predicted amino acid sequence contains a tyrosine residue at the position after which trypsin cleaved the B1 polypeptide chain. Cleavage at this residue may be explained by yet another polymorphic site or by chymotryptic activity in the trypsin. Finally the predicted carboxy-terminal glycine-valine di-

peptide is identical to that determined by carboxypeptidase Y digestion of the B1 hordein polypeptide (40).

In conclusion, the open reading frame of pc hor2-4 codes for a B1 hordein polypeptide but differs from that sequenced by EDMAN degradation in 13 amino acid positions out of 74 which can be compared. The nucleotide sequence comprising 148 nucleotides upstream of nucleotide 1 contains a stop codon. The sequence could be converted into an open reading frame by a deletion or addition of a single base. Since no peptide sequences are known in this region and a mistake in reverse transcription cannot be excluded, we have not given these sequence data. The carboxy-terminal valine is followed by three stop codons (ochre, opal, ochre).

3.3. Nucleotide sequence heterogeneity of cDNA clones

Clone pc hor2-3 hybridized at medium stringency with the above described pc hor2-4. The insert of 257 nucleotides comprised the coding information for the carboxy-terminal 54 amino acid residues of a B1 hordein polypeptide and the three stop codons followed by 85 nucleotides of the 3' noncoding region. Comparing the nucleotide sequences for the carboxy-terminal end of the polypeptide, (Figure 3), pc hor2-3 was distinguished from pc hor2-4 by twenty one nucleotide alterations. Twelve of these involved A - G transitions, two C - T transitions, four A - C transversions and three C-G transversions. Eleven of the nucleotide changes occurred in the third position of the codons and did not alter the amino acid in question. The remaining nucleotide substitutions resulted in 9 amino acid changes. We therefore conclude that pc hor2-4 and pc hor2-3 encode two different B1 hordein polypeptides.

The latter clone hybridised at highly stringent conditions to the cDNA clones pc hor2-1 and pc hor2-2, comprising 254 and 153 nucleotides, respectively, and containing besides a stretch of

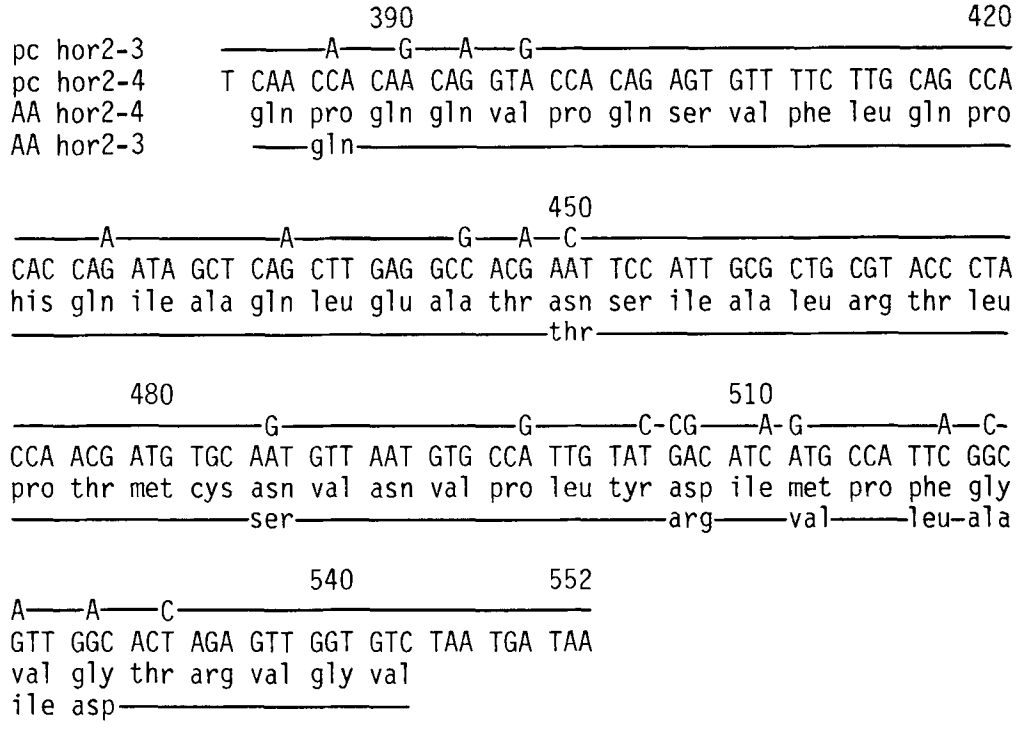

Figure 3. The nucleotide sequences for the carboxy terminal portions of B1 hordein polypeptides encoded by pc hor2-3 and pc hor2-4.

The pc hor2-3 insert contains the coding sequence for the last 54 amino acids of a B1 hordein polypeptide. The solid lines indicate identity in the two sequences. Nucleotide substitutions in pc hor2-3 are indicated. Ten of the 21 substitutions lead to a changed amino acid residue.

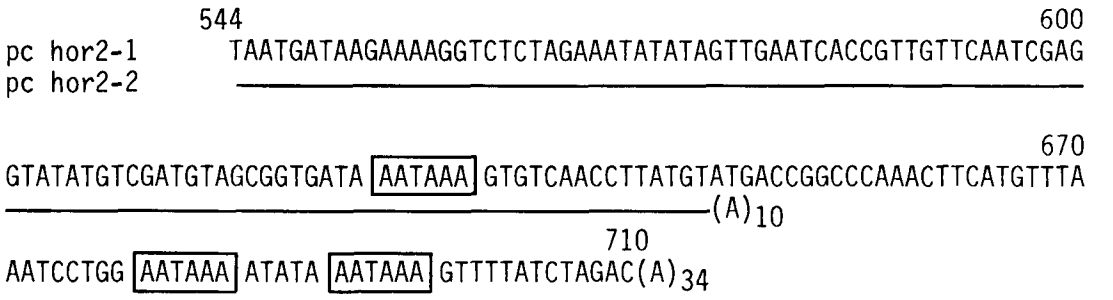

Figure 4. The nucleotide sequences of the 3'-noncoding regions of pc hor2-1 and pc hor2-2.

The solid line denotes identity between the two sequences. Putative polyadenylation signals are marked by boxes and $(A)_n$ identifies the position and number of deoxyadenine nucleotides in the poly(A) tail.

codons for carboxy-terminal amino acid residues, the three stop codons and a complete 3' noncoding region with a poly(A) tail. The 18 respectively 13 carboxy-terminal codons of the 2 clones were identical to those of pc hor2-3. Clone pc hor2-1 contained 162 nucleotides in the 3' noncoding region followed by a 34 nucleotide long poly(A) tail (Figure 4). Three putative polyadenylation signals (36), AATAAA, were found 13, 24 and 78 nucleotides upstream from the poly(A) tail. In contrast, pc hor2-2, contained only the first 99 nucleotides of the 3' noncoding region, followed by a 10 nucleotide long poly(A) tail (Figure 4). In this sequence only the first of the putative polyadenylation signals was found 15 residues upstream from the poly(A) tail. It can be seen from Figure 4, that the first and the third polyadenylation signal is flanked at left by an ATA and at right by a GT base sequence. The second polyadenylation signal is flanked by different bases. Since the nucleotide sequences of these two clones are identical in the region where they overlap, they could originate from two transcripts of the same B1 hordein gene. Either the first or the third polyadenylation signal was used in processing the two transcripts.

3.4. Frequency of codons used and features of the primary structure of B1 hordein

The amino acid codons found in the open reading frame of pc hor2-4 are tabulated in Figure 5. The most abundant amino acids in the polypeptide are glutamine (28%), leucine (10%), proline (9%), valine, serine and isoleucine, comprising 64% of all amino acids. The glutamine codon CAA was more frequently used than the codon CAG. All codons for proline are used, but the codon CCA is preferred as it was used in 10 out of 16 cases. All triplets encoding leucine except TTA are used. The coding capacity of this cDNA stretch is 181 amino acids. Assuming a molecular weight of 27,000 dalton for a B1 hordein polypeptide, pc hor2-4 may contain 2/3 or more of the required coding capacity. The deduced polypeptide sequence is made up of equal amounts of polar and nonpolar amino acids, in agreement with the amino acid composition of the B1 hordein polypeptide group (42). The presence of two lysine, nine cysteine and four methionine residues is revealed by the nucleotide sequence.

Homologous heptapeptides, containing the amino acid sequence -X-Leu-Gln-Gln-Gln-X-X- were found at four different places in the center of the amino acid sequence (Figure 6A). Towards the carboxy-terminal another four heptapeptides, containing the sequence, -Gln-Pro-Gln-Gln-X-X-Gln- were present (Figure 6B). In the nucleotide sequences encoding the last four heptapeptides, close conservation of the nucleotide sequence is observed.

4. DISCUSSION

From nucleotide sequence analysis of cDNA clones for hordein B1 polypeptides three polypeptide chains with unique carboxy-terminal amino acid sequences have been identified. These are the 181 amino acid long carboxy-terminal peptide encoded in pc hor2-4, the 53 residue long peptide encoded in pc hor2-3 and the

37 residue long peptide (19) encoded in pc16. Comparison of the codons for the 53 carboxy-terminal amino acids of pc hor2-3 with those of pc hor2-4 reveals 21 nucleotide changes and 9 amino acid residue substitutions. The codons for the 37 carboxy-terminal amino acids of pc16 (19) differ by a single nucleotide transition (C instead of T) from those of pc hor2-3 not giving rise to an amino acid alteration.

By automated EDMAN degradation the amino acid sequence of 6 unique peptides of B1 hordein has been established (42). The purification procedures used to obtain the peptides and the fact that single amino acid sequences were obtained attest to the high purity of these peptides. On the other hand it is not known whether they belong to a single B1 polypeptide or a mixture of different B1 polypeptides. All six could be located in the carboxy-terminal portion of the reading frame of pc hor2-4. However, all 6 peptides differ by 1 to 7 amino acid residues from the corresponding peptides deduced from the nucleotide sequence. Two of the peptides can be compared with the 37 carboxy-terminal amino acids encoded in pc hor2-3 and the same 37 carboxy-terminal residues represented in pc16 (19). A peptide of 11 amino acid residues deduceable with the aid of CNBr fragment V is found unaltered in the sequence deduced from these two clones. The other peptide (SV8 II) differs by a single amino acid from the corresponding peptide encoded by the two clones. These comparisons allow the conclusion that a minimum of 3 unique B1 polypeptide chains occur in addition to the three identified by the cDNA clones. Two of these are represented by the polymorphic trypsin peptides T III and T IV (42) assuming that the other 4 unique peptides belong to either one of the two chains. The third is represented by a polypeptide chain combining CNBr V and 4 of the other unique peptides. Excepting CNBr V which is represented in the cDNA clones one

	T		C		A		G		
T	Phe	1	Ser	3	Tyr	3	Cys	4	T
	Phe	5	Ser	2	Tyr	1	Cys	5	C
	Leu	0	Ser	0	ochre	1	opal	0	A
	Leu	7	Ser	2	amber	0	Trp	0	G
C	Leu	5	Pro	1	His	2	Arg	3	T
	Leu	1	Pro	4	His	2	Arg	1	C
	Leu	3	Pro	10	Gln	26	Arg	1	A
	Leu	3	Pro	1	Gln	16	Arg	0	G
A	Ile	3	Thr	1	Asn	3	Ser	2	T
	Ile	5	Thr	1	Asn	1	Ser	2	C
	Ile	2	Thr	0	Lys	1	Arg	1	A
	Met	4	Thr	2	Lys	1	Arg	1	G
G	Val	5	Ala	2	Asp	0	Gly	3	T
	Val	4	Ala	2	Asp	1	Gly	2	C
	Val	3	Ala	3	Glu	3	Gly	1	A
	Val	4	Ala	1	Glu	5	Gly	0	G

Figure 5. Codons used in pc hor2-4.

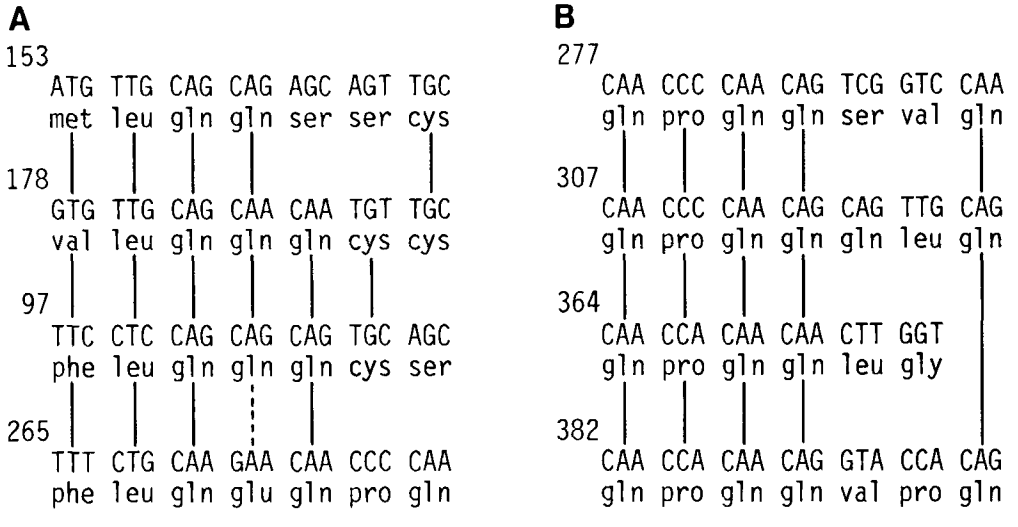

Figure 6. Internally repeated amino acid sequences found in the B1 hordein clone pc hor2-4.

could alternatively assume each of the other 5 unique peptides to represent unique B1 hordein polypeptide chains, which will bring the count to 5 additional chains.

The polymorphism found in the 3' noncoding region of the messenger RNA transcribed into respectively pc hor2-1 and pc hor2-2 consisted of an extra 63 nucleotide stretch in front of the poly(A) tail in one of them. This could be due to alternative processing of the messenger RNA from the same gene in connection with its polyadenylation.

The present analysis would allow to estimate that the B1 hordein consists of 6 to 8 unique polypeptide chains coded for by a corresponding number of genes in the Hor-2 locus. In addition to the cDNA clones presented in this paper we have nucleotide sequenced two clones containing 247 and 332 nucleotides, respectively. These clones contained 3' noncoding regions homologous to pc hor2-1 but in the vicinity of the stop codons nucleotide sequences were obtained which represented inversions of the carboxy-terminal portion of the B1 hordein polypeptide code. Such inversions have been described as accidents occurring during reverse transcription and second strand synthesis (18, 49, 50). The interpretable sequences of these two clones were identical to those of pc hor2-1 and pc hor2-3. This may indicate that in the messenger RNA sample used for reverse transcription the messenger RNA coding for the B1 polypeptide represented by pc hor2-3 was more frequent than others.

The hordein B1 polypeptide encoded by pc hor2-4 contains 8 heptapeptides (Figure 6) which could have evolved from 2 sets of internal repeats or be considered as domains of functional significance. They are characterised by clusters of glutamines with proline, leucine and cysteine as flanking amino acids. In zein, the prolamine of maize similar clusters of glutamines are observed but often in groups of four (26, 35). Internal repeated sequences involving these blocks have been used to construct a possible model for the folding of the zein polypeptide and its packaging in the protein bodies (2).

Nine cysteine residues are found in the B1 hordein polypeptide represented by clone pc hor2-4. Unreduced B hordein polypeptides are known to aggregate and form high molecular weight multimers (15). This property could result from the ability of the cysteine residues to form intermolecular disulphide bonds. These may also play an important role during the condensation of hordein polypeptides in the vacuole of the endosperm cell. In an ultrastructural study of the developing barley endosperm, CAMERON-MILLS and VON WETTSTEIN observed alterations of storage protein condensation in the vacuoles of the B hordein deficient mutant Risø 56 (9). As shown in the accompanying paper this mutant does not contain messenger RNA encoding B1 hordein polypeptides (24). It is therefore likely that the cysteines in the B1 hordein polypeptides may play a role in the final packaging of the storage proteins in the vacuole by forming high molecular weight multimeric networks.

The nucleotide sequence analysis of the B1 hordein cDNA revealed that the 3' noncoding region is highly conserved. Two cDNA sequences covered the 3' noncoding regions including the poly(A) tail. pc hor2-1, contains 3 putative polyadenylation signals, AATAAA, one of which is located 13 nucleotides from the polyadenylation site. Among known plant genes, the one coding for the sweet protein thaumatin from the fruit of Thaumatococcus danielli has been reported to contain three such polyadenylation signals (17). Multiple polyadenylation sites are also characteristic for plant nuclear transcripts of Agrobacterium genes (13). The zein genes of maize and the leghaemoglobin genes of soybean (21, 25, 26, 35, 51) contain variants of the polyadenylation signal 20-30 nucleotides upstream from the poly(A) tail. In the 3' noncoding region of the genes for storage proteins of legumes (12,43,44) the sequence AATAAATAAA was found 20-30 nucleotides from the polyadenylation site and is thought to be a double polyadenylation signal. In the B1 hordein clone pc hor2-1 the sequence AATAAAATATAAATAAA could be interpreted as a double signal (Figure 5). It has been noted that the putative polyadenylation signal in animal cells is frequently followed by a guanine nucleotide (36). In the B1 hordein cDNA pc hor2-1 two of the three polyadenylation signals are followed by a guanine nucleotide (Figure 5). Interestingly the homology extends further on both sides of the polyadenylation sig-

nals. They are both flanked by ATA upstream and GT downstream of the relevant hexanucleotide (Figure 4).

Sequence analyses of the nuclear genes encoding the B1 hordein polypeptides are needed to determine their exact number and organisation. This will also decide whether introns occur in these genes like in the gene for phaseolin, the storage globulin of the French bean (48) and the genes for the 7S storage protein of soybean (44) or whether introns are absent like in the prolamin of maize (26, 35).

ACKNOWLEDGEMENTS

We are very grateful to Professor D. VON WETTSTEIN for his continuous encouragement and critical comments and help in preparation of the manuscript. We thank J. INGVERSEN for helpful discussions, T. BEKTVED and B. HANSEN for excellent technical assistance, and N. RASMUSSEN for drawing the figures. We would like to thank J.J. HYLDIG-NIELSEN and Professor K. MARCKER for introduction to the chain termination sequencing method. This work was supported by the Biomolecular Engineering Programme of the Commission of the European Communities with contract No. GB 1-4-024-DK(G) to Professor D. v. WETTSTEIN.

REFERENCES

1. ANSORGE, W. & L. DE MAYER: Thermally stabilised very thin (0.02-0.3 mm) polyacrylamide gels and their application for electrophoresis. J. Chrom. 202, 45-53 (1980)
2. ARGOS, P., K. PEDERSEN. M. D. MARKS & B. A. LARKINS: A structural model for maize zein proteins. J. Biol. Chem. 257, 9984-9990 (1982)
3. BIRNBOIM, H. C. & J. DOLY: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7, 1513-1523 (1979)
4. BRANDT, A.: Cloning of double stranded DNA coding for hordein polypeptides. Carlsberg Res. Commun. 44, 255-267 (1979)
5. BRANDT, A. & J. INGVERSEN: Isolation and translation of hordein messenger RNA from wild type and mutant endosperms in barley. Carlsberg Res. Commun. 43, 451-469 (1978)
6. BURR, B. & F. A. BURR: Zein storage protein gene family of maize. An assessment of heterogeneity with cloned messenger RNA sequences. J. Mol. Biol. 159, 33-49 (1982)
7. CAMERON-MILLS, V.: The structure and composition of protein bodies purified from barley endosperm by silica sol density gradients. Carlsberg Res. Commun. 45, 557-576 (1980)
8. CAMERON-MILLS, V. & J. INGVERSEN: In vitro synthesis and transport of barley endosperm proteins: Reconstitution of functional rough microsomes from polyribosomes and stripped microsomes. Carlsberg Res. Commun. 43, 471-489 (1978)
9. CAMERON-MILLS, V. & D. VON WETTSTEIN: Protein body formation in the developing barley endosperm. Carlsberg Res. Commun. 45, 577-594 (1980)
10. CHACONAS, G. & J. H. VAN DE SANDE: 5'-^{32}P labeling of RNA and DNA restriction fragments. Methods in Enzymology 65, 75-85 (1980)
11. CLEWELL, D. B. & D. K. HELINSKI: Supercoiled circular DNA protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci U.S.A. 62, 1159-1166 (1966)
12. CROY, R. R. D., G. W. LYCETT, J. A. GATEHOUSE, J. N. YARWOOD & D. BOULTER: Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature 295, 76-79 (1982)
13. DHAESE, P., H. DE GREVE, J. GIELEN, J. SEURINCK, M. VAN MONTAGU & J. SCHELL: Identification of sequences involved in the polyadenylation of higher plant nuclear transcripts using Agrobacterium T-DNA genes as models. The EMBO Journal 2, 419-426 (1983)
14. DOLL, H.: A nearly non-functional mutant allele of the storage protein locus *Hor 2* in barley. Hereditas 93, 217-222 (1980)
15. DOLL, H. & B. ANDERSEN: Preparation of barley storage protein, hordein for analytical sodium dodecyl sulfate polyacrylamide gel electrophoresis. Anal. Biochem. 115, 61-66 (1981)
16. DOLL, H. & A. D. H. BROWN: Hordein variation in wild (*Hordeum spontaneum*) and cultivated (*Hordeum vulgare*) barley. Can. J. Genet. Cytol. 21, 391-404 (1979)
17. EDENS, L., L. HESLINGA, R. KLOK, A. M. LEDEBOER, J. MAAT, M. Y. TOONEN, C. VISSER & C. VEPRIPS: Cloning cDNA encoding the sweet-tasting plant protein thaumatin and its expression in *Escherichia coli*. Gene 18, 1-12 (1982)
18. FAGAN, J. B., I. PASTAN & B. DE CROMBRUGGHE: Sequence rearrangements and duplication of double stranded fibrionectin cDNA occurring during probably cDNA synthesis by AMV re-

verse transcriptase and *Escherichia coli* DNA polymerase I. Nucl. Acids Res. 8, 3055-3064 (1980)
19. FORDE, B. G., M. KREIS, M. B. BAHRAMIAN, J. A. MATTHEWS & B. J. MIFLIN, R.D. THOMPSON, D. BARTELS & R.B. FLAVELL: Molecular cloning and analysis of cDNA sequences derived from poly A+ RNA from barley endosperm: identification of B hordein related clones. Nucl. Acids Res. 9, 6689-6707 (1981)
20. GAROFF, H. & W. ANSORGE: Improvements of DNA sequencing gels Anal. Biochem. 115, 450-457 (1981)
21. GERAGHTY, D. E., J. MESSING & I. RUBINSTEIN: Sequence analysis and comparison of cDNAs of the zein multigene family. The EMBO Journal 1, 1329-1335 (1982)
22. Guidelines for research involving DNA molecules. Federal Register U.S. 47(77) 17180-17198 (1982)
23. HOLDER, A. A. & J. INGVERSEN: Peptide mapping of the major components of in vitro synthesized barley hordeins. Evidence of structural homology. Carlsberg Res. Commun. 43, 177-184 (1978)
24. HOPP, H. E., S. K. RASMUSSEN & A. BRANDT: Organization and transcription of B1 hordein genes in high lysine mutants of barley. Carlsberg Res. Commun. 48, 201-216 (1983)
25. HYLDIG-NIELSEN, J.J., E.Ø. JENSEN, K. PALUDAN, O. WIBORG, R. GARETT, P. JØRGENSEN & K.A. MARCKER: The primary structures of two leghemoglobin genes from soybean. Nucl. Acids Res. 10, 689-701 (1982)
26. HU, N.-T., M.A. PFEIFER, G. HEIDECKER, J. MESSING & I. RUBINSTEIN: Primary structure of a genomic zein sequence of maize. The EMBO Journal 1, 1337-1342 (1982)
27. KØIE, B. & H. DOLL: Protein and carbohydrate components in the Risø high lysine barley mutants. In: Seed Protein Improvement in Cereals and Grain legumes, Vol I IAEA., Vienna pp 205-215 (1979)
28. LEDERBERG, E.M. & S.N. COHEN: Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. J. Bacteriol. 119, 1072-1074 (1974)
29. MATTHEWS, J.A. & B.J. MIFLIN: In vitro synthesis of barley storage proteins. Planta 149, 262-268 (1980)
30. MAXAM, M.A. & W. GILBERT: A new method for sequencing DNA. Proc. Natl. Acad. Sci U.S.A. 74, 560-564 (1977)
31. MAXAM, M.A. & W. GILBERT: Sequencing end-labeled DNA with base-specific chemical-cleavages. Methods in Enzymology 65, 497-560 (1980)
32. MESSING, J., B. GRONENBORN, B. MÜLLER-HILL & P.H. HOFSCHNEIDER: Filamentous coliphages M13 as a cloning vehicle: insertion of a Hind III fragment of the lac regulatory region in M13 replicative form in vitro. Proc. Natl. Acad. Sci. U.S.A. 74, 3642-3646 (1977)
33. MILLER, J.: in "Experiments in Molecular Genetics" Cold Spring Harbor, New York. (1972)
34. ORAM, R.N., H. DOLL & B. KØIE: Genetics of two storage protein variants in barley. Hereditas 80, 53-58 (1975)
35. PEDERSEN, K., J. DEVEREUX, D.R. WILSON, E. SHELDON & B.A. LARKINS: Cloning and sequence analysis reveals structural variation among related zein genes in maize. Cell 29, 1015-1026 (1982)
36. PROUDFOOT, N.J. & G.G. BROWNLEE: 3' Noncoding region sequence in eukaryotic messenger RNA. Nature 263, 211-214 (1976)
37. SANGER, F. & A.R. COULSON: The use of thin acrylamide gels for DNA sequencing. FEBS Lett. 87, 107-110 (1978)
38. SANGER, F., A.R. COULSON, B.G. BARRELL, A.J. SMITH & B.A. ROE: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143, 161-178 (1980)
39. SANGER, F., S. NICKLEN & A.R. COULSON: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467 (1977)
40. SCHMITT, J. M.: Purification of hordein polypeptides by column chromatography using volatile solvents. Carlsberg Res. Commun. 44, 431-438 (1979)
41. SCHMITT, J.M. & I. SVENDSEN: Amino acid sequences of hordein polypeptides. Carlsberg Res. Commun. 45, 143-148 (1980)
42. SCHMITT, J.M. & I. SVENDSEN: Partial amino acid sequence from hordein polypeptide B1. Carlsberg Res. Commun. 45, 549-555 (1980).
43. SCHULER, M.A., B.F. LADIN, J.C. POLLACO, G. FREYER & R.N. BEACHY: Structural sequences are conserved in the genes coding for the α, α' and β-subunit of the soybean 7S seed storage protein. Nucl. Acids Res. 10, 8245-8261 (1982)
44. SCHULER, M.A., E.S. SCHMITT & R.N. BEACHY: Closely related families of genes code for the α and α' subunits of the soybean 7S storage protein complex. Nucl. Acids Res. 10, 8225-8244 (1982)
45. SHEWRY, P.R., A.J. FAULKS, R.A. PICKERING, I.T. JONES, R.A. FINCH & B.J. MIFLIN: The ge-

netic analysis of barley storage proteins. Heredity 44, 383-389 (1980)
46. SHEWRY, P.R., J.R. MARCH & B.J. MIFLIN: N-terminal sequence of C hordein. Phytochem. 19, 2113-2115 (1980)
47. SHEWRY, P.R., H.M. PRATT, R.A. FINCH & J. MIFLIN: Genetic analysis of hordein polypeptides from single seeds of barley. Heredity 40, 463-466 (1978)
48. SLIGHTOM, J.L., S.M. SUN & T.C. HALL: Complete nucleotide sequence of a French bean storage protein gene: Phaseolin. Proc. Natl. Acad. Sci. U.S.A. 80, 1897-1901 (1983)
49. VOLCKAERT, G., J. TAVERMIER, R. DERYNCK, R. DEROS & W. FIERS: Molecular mechanism of nucleotide-sequence rearrangements in cDNA clones of human fibroblast interferon mRNA. Gene 15, 215-223 (1981)
50. WEAVER, C.A., D.F. GORDON & B. KEMPER: Introduction by molecular cloning of artifactual inverted sequences at the 5' terminus of the sense strand of bovine parathyroid hormone cDNA. Proc. Natl. Acad. Sci. U.S.A. 78, 4073-4077 (1981)
51. WIBORG, O., J.J. HYLDIG-NIELSEN, E.Ø. JENSEN, K. PALUDAN & K.A. MARCKER: The nucleotide sequence of two leghemoglobin genes from soybean. Nucl. Acids Res. 10, 3487-3494 (1982)