Cloning and nucleotide sequence of a cDNA encoding the precursor of the barley toxin α-hordothionin

Fernando PONZ, Javier PAZ-ARES, Carlos HERNÁNDEZ-LUCAS, Francisco GARCÍA-OLMEDO and Pilar CARBONERO

Departamento de Bioquimica, Escuela Técnica Superior de Ingenieros Agrónomos, Universidad Politecnica Madrid

(Received November 6, 1985) – EJB 85 1215

A cDNA library, prepared from developing barley endosperm, was screened for thionin recombinants. Clone pTH1 was that with the largest insert out of three identified. The longest reading frame in the 610-base-pair insert codes for a protein of 127 amino acids that includes an internal sequence of 45 amino acids, which is identical to that obtained for the α-hordothionin by direct protein sequencing. The deduced thionin sequence is preceded by a leader sequence of 18 residues and followed by a sequence that corresponds to an acidic protein of 64 amino acids. This structure supports previous evidence indicating that thionin is synthesized as a much larger precursor, which undergoes two processing steps: the cotranslational cleavage of a leader sequence and the post-translational one of a larger peptide. The size of the mRNA was estimated to be about 950 bases by Northern analysis. Thionin concentration in mature endosperm of barley cv. Bomi was about twice that of its high-lysine mutant Risø 1508. The same difference was observed in thionin mRNA in the corresponding developing endosperms, indicating that gene expression is partially blocked in the mutant at a pretranslational level.

Thionins are cereal proteins of molecular mass ≈ 5000 Da with high lysine and cystine contents, which belong to a homology group that also includes the viscotoxins from the mistletoes (Loranthaceae) and crambin from the abyssinian cabbage (Cruciferae). Extensive structural studies have been carried out with this group of proteins: many members of the group have been sequenced [1–8], detailed ^{1}H NMR spectra at 600 MHz have been obtained for crambin and barley thionins [9], and the crystal structure of crambin has been determined by X-ray diffraction to a resolution of 0.088 nm, the highest to date for any protein [10]. The thionins of cereal endosperm are of interest because they are lysine-rich proteins in a lysine-deficient tissue, which is the main world-wide edible plant product, and because, although orally innocuous, they are toxic to laboratory animals (intraperitoneally), cultured mammalian cells, yeasts and phytopathogenic bacteria (for a review see [11]).

Genetic and biochemical studies have shown that in hexaploid wheat (*Triticum aestivum* L.) and in rye (*Secale cereale* L.) there is one thionin variant per diploid genome [12, 13], whereas in diploid barley (*Hordeum vulgate* L.) there is evidence for at least a major and a minor component, respectively designated α and β-hordothionin, which are not separated from each other by electrophoresis [9, 14]. Recent evidence from our laboratory shows that barley thionins are synthesized by membrane-bound polysomes as much larger precursors that are processed in two steps; the first one is presumably cotranslational and the second one is post-translational [14]. We have now obtained a cDNA clone that encodes the precursor of α-hordothionin and have investigated the effect of a high-lysine mutation (Risø 1508) on the expression of the corresponding gene.

Correspondence to P. Carbonero, Escuela Técnica Superior de Ingenieros Agrónomos, Ciudad Universitaria, E-28040 Madrid, Spain

We dedicate this work to Dr J. M. Sala-Trepat (1942–1985) *in memoriam*.

Abbreviations. SSC, standard saline citrate; SDS, sodium dodecyl sulphate.

Enzymes (IUB Recommendations, 1984). Terminal transferase (EC 2.7.7.31); S1 nuclease (EC 3.1.30.1); DNA pol I (EC 2.7.7.7); Klenow fragment (EC 2.7.7.7); DNase (EC 3.1.21.1); RNases (EC 3.1.27.5); T4 DNA ligase (EC 6.5.1.1); AMV reverse transcriptase (EC 2.7.7.49); restriction enzymes *Pst*I, *Bam*HI, *Hae*III, *Alu*I (EC 3.1.24.4); trypsin (EC 3.4.21.4).

MATERIALS AND METHODS

Biological material

Hordeum vulgare cv. Bomi and its mutant Risø 1508 were the kind gift of H. Doll (Risø National Laboratory, Denmark). Endosperms were collected at 20 days after anthesis by mechanical extrusion into liquid nitrogen, using the device designed by O'Dell and Thompson [15].

Reagents

The wheat germ cell-free translation system was prepared by us from commercial wheat germ (General Mills, USA) as previously described [14]. Oligo(dT)-cellulose was from Collaborative Research; terminal transferase and S1 nuclease were from P-L Biochemicals. Restriction endonucleases, DNA pol I, Klenow fragment, DNase, RNases, T4 DNA ligase etc. were from Boehringer, Amersham, New England Bio-Labs or Bethesda Research Laboratory, and avian myeloblastosis virus reverse transcriptase was from Life Sciences Inc. [^{35}S]Cysteine, 900 Ci/mmol, was from New England Nuclear. [α-^{32}P]dATP, 400 Ci/mmol, and [γ-^{32}P]-ATP, 3000 Ci/mmol, were from Amersham.

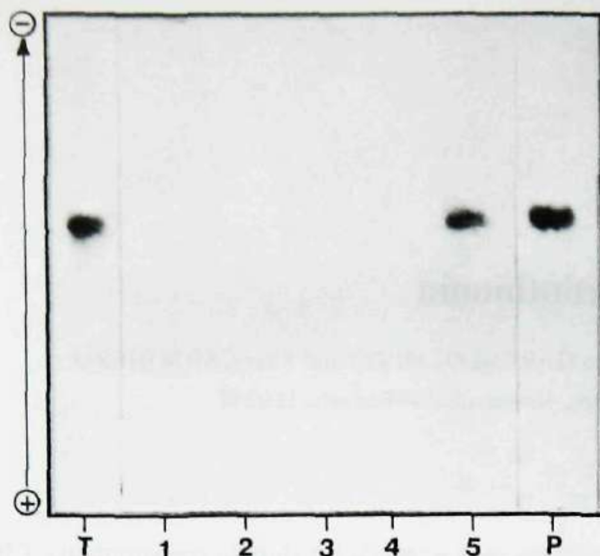

Fig. 1. *Electrophoresis (pH 3.2) and fluorography of products obtained by hybrid-selected translation.* T, thionin precursor immunoprecipitated from translation products of total poly(A)-rich RNA. P, translation product from RNA selected by paper-bound DNA from a pool of five clones. 1 – 5, second round of hybrid-selected translation with the individual clones of the previous pool. The positive is that of clone pTH1

Purification of mRNA and in vitro *translation*

Total polysomal RNA, membrane-bound polysomal RNA and poly(A)-rich RNA were prepared as described [14]. Gradient centrifugation of RNA was also carried out as before, except that CH_3HgOH was included. Gradient fractions were analysed by *in vitro* translation, using [^{35}S]cysteine, followed by immunoprecipitation of products, electrophoresis at pH 3.2, under reducing conditions and fluorography, as previously reported [14].

Construction and analysis of a cDNA bank

Total polysomal RNA from developing barley endosperm (cv. Bomi) was the source of poly(A)-rich RNA used as the template for synthesis of cDNA. After double-strand cDNA synthesis with the Klenow fragment and treatment with S1 nuclease, dscDNA was size-fractionated by preparative electrophoresis in low-melting agarose; dscDNA of more than 350 base pairs, purified by the method of Langridge et al. [16], was used for homopolymeric tailing [poly(G)/poly(C)] and cloning in the *Pst*I site of pBR322. Cells of *Escherichia coli* strain Mc1061, made competent by the $CaCl_2$ procedure, were transformed with the recombinant plasmid and selected for tetracycline resistance. About 10^4 colonies were obtained from approximately 300 ng double-stranded cDNA.

Colony hybridization was carried out according to Grunstein and Hogness [17]. Probes were prepared by synthesis of radioactive sscDNA by reverse transcription, using selected ultracentrifugation RNA fractions as templates.

Hybrid-selected translation

Plasmid DNA was covalently bound to diazobenzyloxymethyl paper [18]. Filters were incubated in the hybridization buffer (50% formamide, 800 mM NaCl, 100 mM Tris/HCl, pH 7.5, 2 mM EDTA, 0.4% sodium dodecyl sulphate (SDS) 2 mg/ml yeast tRNA) for 1 h at 37°C before adding total polysomal RNA (100 – 500 μg/ml) and further incubation for 1 h at 37°C. After repeated washings with incubation buffer, the filters are washed with the same buffer at 150 mM NaCl, and the selected mRNA was finally released with buffer without salt, heating for 5 min at 100°C and rapidly cooling at 0°C. *In vitro* translation was carried out as indicated above.

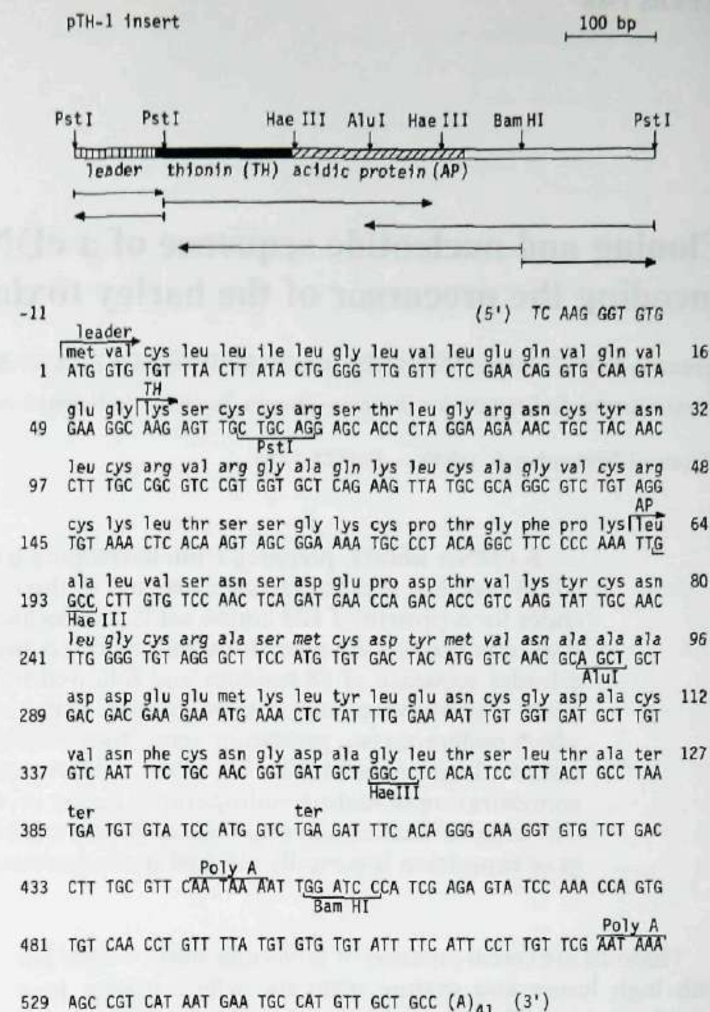

Fig. 2. *Restriction map, sequencing strategy and complete nucleotide sequence of the cDNA insert in clone pTH1, with deduced amino acid sequence of the longest open reading frame.* The d(G-C) tails are excluded. Nucleotides are numbered from the 5′ end starting with those encoding the initial methionine. Amino acids are numbered from the N terminal. Presumed polyadenylation signals are indicated as underlined 'Poly A'. The sequence of mature α-hordothionin begins at nucleotide 55 (indicated as TH) and ends at nucleotide 190. The beginning of the sequence corresponding to the putative acidic protein is at nucleotide 191 (indicated as AP)

Restriction maps and DNA sequencing

Restriction maps were carried out by standard techniques [19] and DNA sequencing was according to Maxam and Gilbert [20] with minor modifications. Each of the two *Pst*I fragments from the insert in clone pTH1 were subcloned into the *Pst*I site of plasmid pUC12 [21] and sequenced from both ends. The bigger fragment was also sequenced from the internal *Bam*HI site.

RNA electrophoresis, Northern and dot blots

Polysomal and total RNA from 20-day developing barley endosperms from cultivars Bomi and Risø 1508, obtained as

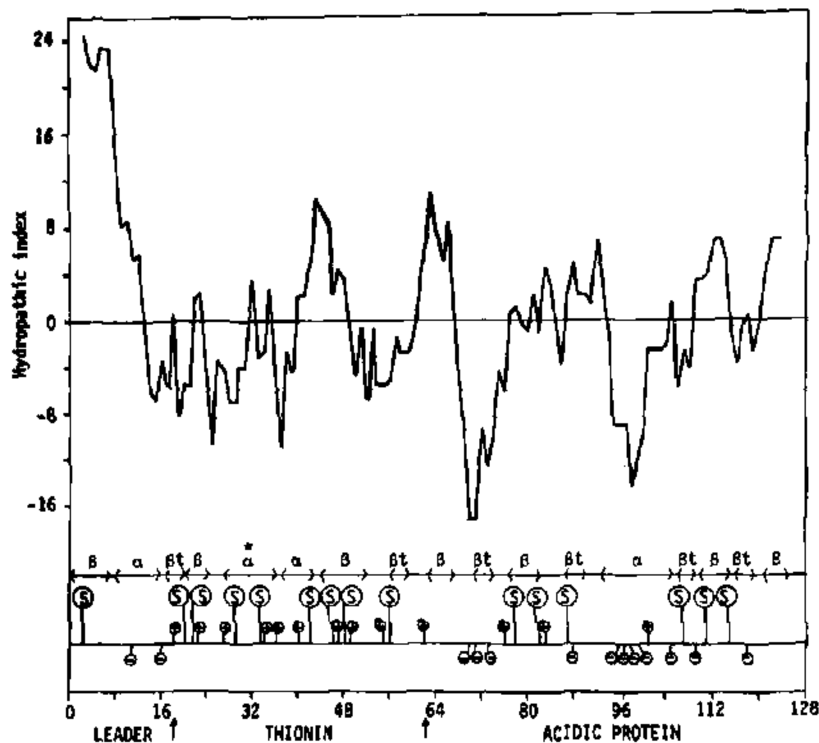

Fig. 3. *Characteristics of the α-hordothionin precursor.* Top, hydropathy profile calculated according to Kyte and Doolittle [35] with a span of seven positions. Middle, secondary structure prediction as indicated by Chou and Fasman [36]: α, α helix; β, β sheet, βt, β turn. The stretch marked with a star (*) was not predicted by the above method but assumed by comparison with the crystalline structure of crambin [10]. Bottom, charge and cysteine (circled S) distribution along the precursor molecule

described [14], were denatured by heating for 60 min at 50°C in 1 M glyoxal, 10 mM phosphate (pH 7) and quickly chilling [22]. Electrophoresis was in 1.1% agarose gels at 90 V for 5 h in 10 mM NaH_2PO_4, pH 7. After electrophoresis, the gels were blotted directly into nitrocellulose in 20 × SSC. Filters were treated in 20 mM Tris, pH 8, for 5 min at 100°C and prehybridized, hybridized etc. as described for RNA dot blots.

RNAs for dot blots were serially diluted in the same glyoxal/phosphate solution used for denaturation. Yeast RNA was spotted in the same amounts as the barley RNAs as a negative hybridization control. Spots were of 8, 4, 2 and 1 µg RNA respectively, all of them in a final volume of 8 µl. The nitrocellulose filters were previously equilibrated in 20 × SSC and air dried. After spotting, filters were baked at 80°C for 2 h. Prehybridization was at 42°C in 50% formamide, 5 × SET (0.6 M NaCl, 8 mM EDTA, 0.15 M Tris/HCl, pH 7.8), 10 × Denhardt's [23], 0.1% SDS, 0.1% pyrophosphate, 250 µg/ml sonicated and denatured salmon sperm DNA in 100 mM phosphate, pH 7. Hybridization was also at 42°C with 10^6 cpm nick-translated pTH1 probe in buffer similar to that used for prehybridization with the exception of 1 × Denhardt's.

Filters were washed in 2 × SET, 0.1 × SDS, 0.1% sodium pyrophosphate twice at room temperature and three more times at 60°C for 30 min.

Radial immunodiffusion

Quantification of hordothionin in mature kernels of cv. Bomi and mutant Risø 1508 was carried out by radial immunodiffusion [24]. Monospecific antibodies were obtained as described [14] and purified hordothionin was used as standard.

RESULTS

Screening of the cDNA bank from developing barley endosperm

A bank of about 10^4 clones was obtained as described above and subjected to a preliminary screening. The fraction from sucrose gradient ultracentrifugation of poly(A)-rich RNA (membrane-bound polysomes), that included most of the messenger activity for the thionin precursor, was used to synthesize radioactive sscDNA. The collection was screened with this probe and about 12% of the clones gave a positive signal. About 1/3 of these were discarded because they also hybridized with a second sscDNA probe prepared from high-molecular-mass RNA.

The identification of cDNA clones encoding the hordothionin precursor, among those obtained from the preliminary screening, was carried out in two rounds of hybrid-selected translation: DNAs from pools of five clones were used for the hybridization in the first round, and clones from those pools that gave positive results were individualized for the second round (Fig. 1). Three clones were thus identified out of the first 25 that were processed. These clones selected a mRNA that upon *in vitro* translation yielded a protein of the same electrophoretic mobility as the hordothionin precursor and this protein was also recognized by monospecific antibodies raised against mature hordothionin (Fig. 1).

Characterization of clone pTH1

All three clones had an insert with an internal *Pst*I site and that with the largest one, pTH1, was selected for further study. The restriction map, sequencing strategy and complete nucleotide sequence of the pTH1 insert are presented in Fig. 2. The longest open reading frame in this sequence codes for a

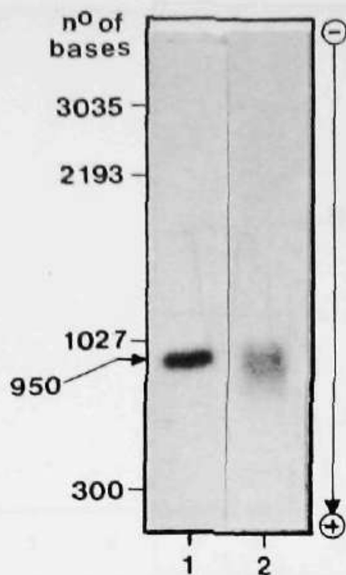

Fig. 4. *Northern blot analysis of total polysomal RNA (10 µg) from barley endosperms: (1) cv. Bomi, (2) mutant Risø 1508.* Hybridization was with a 550-base-pair *Pst*I-*Pst*I insert of clone pTH1. Molecular mass markers were RNAs from cucumber mosaic virus (gift of F. Garcia-Arenal)

protein of 127 amino acids that includes an internal sequence of 45 amino acids, which is identical with that obtained by Ozaki et al. [7] for hordothionin II (α-hordothionin) by direct protein sequencing, and differs at three positions from that proposed for the same protein by Mak [3]. The thionin sequence is preceded by that of a putative signal peptide of 18 amino acids and followed by a sequence that would correspond to an acidic protein with 64 amino acids. The hydropathy profile, secondary structure predictions and the distribution of charges and cysteines of the hordothionin precursor are schematically represented in Fig. 3. A Northern blot of total polysomal RNA from developing endosperm, using the pTH1 insert as a probe, is shown in Fig. 4. A size of about 950 bases was thus estimated for the thionin mRNA.

Hordothionin in Bomi barley and in its high-lysine mutant Risø 1508

Hordothionin was quantified by radial immunodiffusion, using monospecific antibodies raised against the purified protein, both in cv. Bomi barley and in its high-lysine mutant Risø 1508, as shown in Fig. 5A. The net amount of this protein accumulated in the mature kernel in the wild-type Bomi is about twice that in the mutant. Total RNA was extracted from developing endosperms of Bomi and Risø 1508 collected at 20 days after anthesis and the relative levels of hordothionin messengers were estimated by dot hybridization and densitometry, using a radioactive probe obtained by nick-translation of the pTH1 insert (Fig. 5B). The level of thionin mRNA was also about twice as high in Bomi.

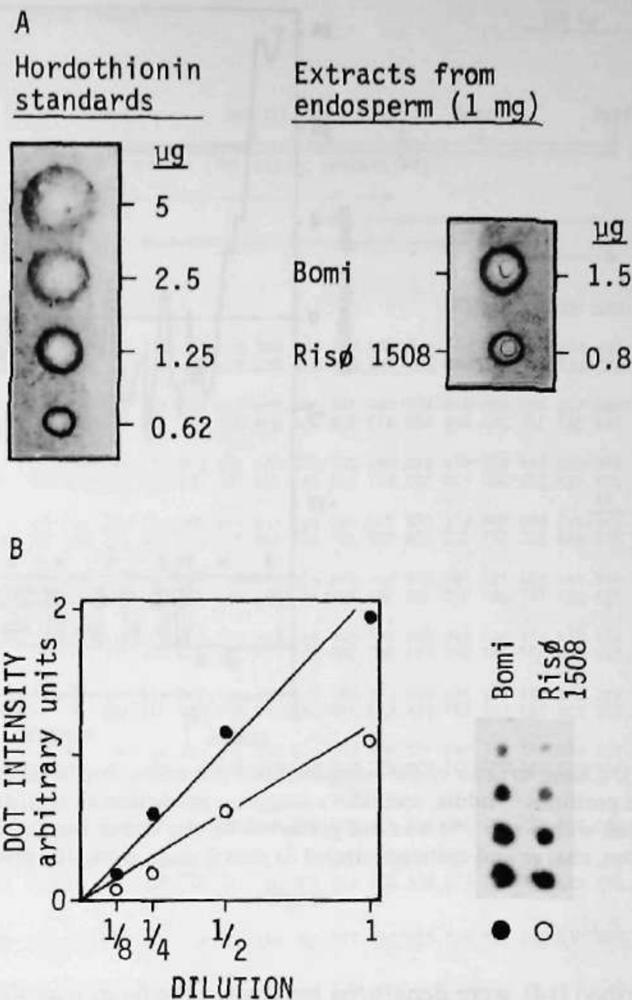

Fig. 5. *(A) Quantification of hordothionin by radial immunodiffusion in the mature endosperms of Bomi barley and its high-lysine mutant Risø 1508. Monospecific antibodies raised against mature thionin were used. Purified protein was used as standard. (B) Quantification of relative hordothionin mRNA levels in developing endosperms (20 days after anthesis) of Bomi and Risø 1508 by dot-blot hybridization, autoradiography and densitometry.* The amounts of RNA (about 8 µg for dilution 1) in the two samples correspond to the same fresh weight of endosperm

DISCUSSION

Two types of thionin precursors, THP1 and THP2, have been previously identified *in vitro* and *in vivo*, respectively, using monospecific antibodies raised against the mature protein [14]. The two precursors differed in apparent size by about 400 Da as judged by SDS electrophoresis, and in electrophoretic mobility at acid pH, but both of them were much larger than the final protein. The cotranslational conversion of THP1 into THP2 was proposed on the basis of our failure to detect THP1 *in vivo*, whereas the conversion of THP2 into mature protein was clearly shown in a pulse-chase experiment. The structure of the cloned cDNA fits well with the above finding: THP1 and THP2 would differ by the 18 amino acids of the putative signal peptide, which would be cotranslationally excised, while in the second processing step the C-terminal acidic peptide would be separated. It is true that the size of the leader sequence was underestimated, but in proteins with peculiar charge distributions, such as that shown in Fig. 3, deviant behaviour in SDS electrophoresis is to be expected. The stretch of 18 amino acids preceding the N-terminal lysine of hordothionin has the hydrophobic core, which is characteristic of leader sequences, and the cleavage site is that with the highest probability according to Heijne [25]. However, it is unusual in that there is no charged amino acid close to the initial methionine, a feature that is only shared by two other known leader sequences, both of which

correspond to plant proteins that are post-translationally processed, taumatin and vicilin [26, 27]. This lack of charge might be also related to the fact that hordothionin ends up externally associated with membrane structures and is not accumulated in protein bodies [14].

The sequence of α-hordothionin found by Ozaki et al. [7] differed from that proposed by Mak [3] at positions 36, 38 and 41, and had been thought to be in error by Lecomte et al. [9]. Our present results fully support those of Ozaki et al. [7] and imply that the assignment of ^{1}H NMR signals reported by Lecomte et al. [9] might have to be revised.

We have repeatedly attempted, without success, to identify a small acidic protein with the characteristics expected from the sequence starting at the C-terminal of hordothionin, using gel filtration, preparative electrophoresis at alkaline pH and high-performance liquid chromatography, so it is uncertain whether such a peptide does accumulate in endosperm. We have no information concerning the untranslated 5′ end of the mRNA, whereas the untranslated 3′ end includes the following characteristics: two stop signals in tandem, a feature that is shared with vicilin mRNA [27] and is also present in B hordein mRNA [28]; two canonical d(AATAAA) polyadenylation signals, at −30 and −110 from the poly(A) tract, and a length of 177 bases, with a slight excess of dA + dT with respect to the coding region (59% versus 50%). No peculiar trait in codon usage seems to be worth mentioning.

Risø 1508 is a mutant of cv. Bomi with an increased overall lysine content, due to a decrease of the lysine-poor hordein fraction and an enhancement of lysine-rich salt-soluble proteins and glutelins [29−31]. Our present results show that the synthesis of hordothionin, a lysine-rich protein of barley endosperm, is decreased in the mutant. The gene in Risø 1508 (*lys3a*), which is located in chromosome 7, seems to block not only the expression of genes encoding specific hordein components (B hordeins; $<1\%$ Lys), as reported by Shewry et al. [30], but also that of genes for proteins with intermediate or high lysine content, such as trypsin inhibitor CMe (1.7% Lys [32]), Z protein (5% Lys [33]) and β-amylase (7% Lys [34]). The lysine content of the mature thionin is very high ($>11\%$) and even if it is assumed that the acidic component of the precursor does accumulate at equimolar amounts, the overall lysine proportion would still be high (6.4%). The present evidence indicates that the expression of the hordothionin gene is also partially blocked at a pretranslational level in mutant Risø 1508.

We would like to thank Drs C. Escarmís and A. González for useful advice with the DNA sequencing, and C. Rojas and L. Lamoneda for technical assistance. We are also indebted to L. J. Elía for help with the RNA work. Financial support from the *Comisión Asesora de Investigación Científica y Técnica* (grant 1344) is also acknowledged.

REFERENCES

1. Samuelsson, G., Seger, L. & Olson, T. (1968) *Acta Chem. Scand. 22*, 2624−2642.
2. Ohtani, S., Okada, T., Kagamiyama, H. & Yoshizumi, H. (1975) *Agric. Biol. Chem. 39*, 2269−2270.
3. Mak, A. S. (1975) Ph.D. Thesis University of Manitoba, Winnipeg (Manitoba), Canada.
4. Mak, A. S. & Jones, B. L. (1976) *Can. J. Biochem. 54*, 835−842.
5. Ohtani, S., Okada, T., Yoshizumi, H. & Kagamiyama, H. (1977) *J. Biochem. (Tokyo) 82*, 753−767.
6. Jones, B. L. & Mak, A. S. (1977) *Cereal Chem. 54*, 511−523.
7. Ozaki, Y., Wada, K., Hase, T., Matsubara, H., Nakanishi, T. & Yoshizumi, H. (1980) *J. Biochem. (Tokyo) 87*, 549−555.
8. Teeter, M. M., Mazer, J. A. & L'Italien, J. J. (1981) *Biochemistry 20*, 5437−5443.
9. Lecomte, J. T. J., Jones, B. L. & Llinas, M. (1982) *Biochemistry 21*, 4843−4849.
10. Hendrickson, W. A. & Teeter, M. M. (1981) *Nature (Lond.) 290*, 107−113.
11. García-Olmedo, F., Carbonero, P. & Jones, B. L. (1982) in *Advances in cereal sciences and technology* (Pomeranz, Y., ed.) vol. 5, pp. 1−47, Am. Ass. Cereal Chem. Inc., St Paul, MN.
12. Fernández de Caleya, R., Hernández-Lucas, C., Carbonero, P. & García-Olmedo, F. (1976) *Genetics 83*, 687−699.
13. Sánchez-Monge, R., Delibes, A., Hernández-Lucas, C., Carbonero, P. & García-Olmedo, F. (1979) *Theor. Appl. Genet. 54*, 61−63.
14. Ponz, F., Paz-Ares, J., Hernández-Lucas, C., Carbonero, P. & García-Olmedo, F. (1983) *EMBO J. 2*, 1035−1040.
15. O'Dell, M. & Thompson, R. D. (1982) *J. Sci. Food Agric. 33*, 419−420.
16. Langridge, J., Langridge, P. & Bergquist, P. L. (1980) *Anal. Biochem. 103*, 264−271.
17. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl Acad. Sci. USA 72*, 3961−3965.
18. Christophe, D., Brocas, H. & Vassart, G. (1982) *Anal. Biochem. 120*, 259−261.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular cloning. A laboratory manual*, pp. 374−381. Cold Spring Harbor Laboratory Press, New York.
20. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol. 65*, 499−560.
21. Messing, J. (1983) *Methods Enzymol. 101*, 20−78.
22. Thomas, P. S. (1983) *Methods Enzymol. 100*, 255−266.
23. Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun. 23*, 641−646.
24. Garvey, J. S., Cremer, N. E. & Sussdorf, D. H. (1977) in *Methods in Immunology*, 3rd edn, pp. 313−327, Benjamin, Reading, MA.
25. Heijne, G. (1983) *Eur. J. Biochem. 133*, 17−21.
26. Edens, L., Heslinga, L., Klok, R., Ledeboer, A. M., Maat, J., Toonen, M. Y., Visser, C. & Verrips, C. Th. (1982) *Gene 18*, 1−12.
27. Lycett, G. W., Delauney, A. J., Gatehouse, J. A., Gilroy, J., Croy, R. R. D. & Boulter, D. (1983) *Nucleic Acids Res. 11*, 2367−2380.
28. Forde, B. G., Kreis, M., Bahramian, M. B., Matthews, J. A. & Miflin, B. J. (1981) *Nucleic Acids Res. 9*, 6689−6707.
29. Doll, H., Køie, B. & Eggum, B. O. (1974) *Radiat. Bot. 14*, 73−80.
30. Shewry, P. R., Faulks, A. J. & Miflin, B. J. (1980) *Biochem. Genet. 18*, 133−151.
31. Doll, H. (1984) *Philos. Trans. R. Soc. Lond. B. Biol. Sci. 304*, 373−380.
32. Lázaro, A., Barber, D., Salcedo, G., Mendez, E. & García-Olmedo, F. (1985) *Eur. J. Biochem. 149*, 617−623.
33. Hejgard, J. (1982) *Physiol. Plant. 54*, 174−182.
34. Balasaraswathi, R., Køie, B. & Doll, D. (1984) *Hereditas 100*, 225−231.
35. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105−132.
36. Chou, P. Y. & Fasman, G. D. (1977) *Trends Biol. Sci. 2*, 128−131.