Nucleotide sequence of barley chymotrypsin inhibitor-2 (CI-2) and its expression in normal and high-lysine barley

Martin S. WILLIAMSON, Janice FORDE, Bernadette BUXTON and Martin KREIS

Biochemistry Department, Rothamsted Experimental Station, Harpenden

(Received October 8/December 12, 1986) – EJB 86 1065

cDNA clones for chymotrypsin inhibitor-2 (CI-2) have been isolated from an endosperm-specific library of barley using a synthetic oligonucleotide probe. The nucleotide sequences of several of the cDNAs predict an open reading frame (beginning with an ATG codon) which encodes a protein of 84 residues (M_r 9380). In the longest clone another ATG codon is present, a further 69 nucleotides upstream. The nucleotide sequence between these two ATG codons predicts an amino acid sequence with the characteristics of a signal peptide, as found in other cloned plant protease inhibitors. However, it contains an in-frame TAA stop codon, which is also present in all of the shorter cDNAs which extend into this region. From *in vitro* translation experiments, using mRNAs synthesized from cDNAs, we conclude that, *in vitro*, translation of all or the vast majority of CI-2 mRNAs begins at the second ATG codon, 31 nucleotides downstream from the ochre stop codon.

Southern blotting of genomic DNA shows that CI-2 is encoded by a small multigene family, while sequence analysis of the cDNAs shows that at least two sub-families of mRNAs, which are more than 90% homologous, are present in the endosperm. Northern blotting analysis shows that related but different sequences are present in leaf and shoot RNA populations.

Further Northern blot hybridizations using RNA from the normal line, Sundance, and the high-lysine barley mutant, Hiproly, show that endosperms of the latter contain greatly increased levels of CI-2 mRNA. This correlates with the increased amount of CI-2 protein deposited in Hiproly, and demonstrates that the differential expression of CI-2 in the two genotypes is controlled at the level of transcription and/or stability of the mRNA. In contrast, the abundance of CI-2 mRNAs in leaves and shoots is not affected.

Barley endosperm contains two chymotrypsin inhibitors called CI-1 and CI-2 which belong to a family of proteins including the potato inhibitor I and the leech inhibitor elgin [1]. CI-2 normally accounts for about 0.25% of the total salt-soluble proteins of the barley grain (calculated from [2]). However, the spontaneous high-lysine barley mutant Hiproly [3] has a 20-fold increased content of CI-2 [4, 5]. Because of its high lysine content (11.5 g %), CI-2 accounts for 37% of the difference in lysine content between Hiproly and normal barley grains and so contributes significantly to the improved nutritional quality of Hiproly. The protein has therefore received considerable attention. It has been purified from Hiproly endosperms and characterized in detail [5–7].

The increased amount of CI-2 in Hiproly is controlled by a single recessive gene called *lys* on chromosome 7 [3, 8], whereas the structural genes for CI-2 are located at the *Ica*-2 locus on chromosome 5 [9]. It is not known how the product of the *lys* gene affects the regulation of expression of the *Ica*-2 locus, though a recent study by Rasmussen [10] suggests that the increased amount of CI-2 in Hiproly may stem from an earlier synthesis of CI-2 during endosperm development.

Correspondence to M. Kreis, Biochemistry Department, Rothamsted Experimental Station, Harpenden, England AL5 2JQ

Abbreviations. SDS, sodium dodecyl sulphate; bp, base pairs; SSC, 0.15 M NaCl, 0.015 M trisodium citrate, pH 7.0.

In order to gain a better understanding of the processes which result in the increased deposition of CI-2 in Hiproly, we have isolated several cDNA clones and used them to study the organization and expression of the genes for CI-2.

MATERIALS AND METHODS

Plant material

Barley plants (*Hordeum vulgare* L. cvs. Sundance and Hiproly) were grown at Rothamsted. Endosperms for nucleic acid extraction were isolated 18–20 days after flowering, unless otherwise stated, frozen in liquid nitrogen and stored at −80 °C.

Extraction and separation of salt-soluble proteins from mature barley grain

Grain was milled in a Glen Creston hammer mill to pass a 0.5-mm sieve. Duplicate 1-g samples were stirred for 1 h at 20 °C in 0.1 M phosphate buffer pH 8, 5 mM DL-dithiothreitol to extract the salt-soluble fraction [11].

Isolation, fractionation and in vitro *translation of polysomes and RNA*

Total, free and membrane-bound polysomes and the RNAs from these were prepared from developing barley

endosperms, shoots and green leaves as previously described [12]. Poly(A)-rich RNA was purified by affinity chromatography using oligo(dT)-cellulose [13]. Polysomes and RNAs were translated *in vitro* using wheat-germ cell-free extracts supplemented with [^{3}H]leucine and [^{3}H]proline [14]. The products were analyzed on a 16% urea/SDS/polyacrylamide gel and fluorographed. Poly(A)-rich RNA (20 μg) was size-fractionated on 5–30% (w/v) sucrose gradients [15]. Fractions enriched for CI-2 translation products were pooled and stored at −80 °C.

Construction of Hiproly endosperm cDNA libraries

Size-fractionated Hiproly poly(A)-rich RNA from membrane-bound polysomes, enriched for CI-2 mRNA, was used as template for cDNA synthesis. Two methods were used for the construction of cDNA libraries. Method A was based on the procedure of Efstratiadis et al. [16] with certain modifications [17]. 400 transformants were obtained from 40 ng recombinant plasmid. Method B followed the cloning procedure of Heidecker and Messing [18]. Competent cells of JM83 were prepared [19] and transformants were plated at high colony density (1000–2000 colonies/filter) on 82-mm Biodyne A membrane filters (Pall Process Filtration Ltd) [20]. Replica filters were prepared for *in situ* hybridization and long-term storage [20].

Screening cDNA libraries for CI-2 sequences

Screening with an oligonucleotide probe. The library generated by cloning method A was screened by *in situ* colony hybridization [21] using an $(N)_{14}$ nucleotide mixture specific for CI-2 sequences as the hybridization probe. The oligonucleotide mixture contained the 16 possible nucleotide sequences which could encode the known CI-2 protein sequence from amino acids 21–25 [7], (3′-TTT/C-TGN-CTT/C-ACC-GG-5′). The oligonucleotide was 5′-end-labelled with [γ-^{32}P]ATP [22]. Hybridization was at 35 °C for 18 h in 5 × Denhardt's solution, 6 × SSC, 50 mM Tris/HCl pH 7.5, 5 mM EDTA, 0.05% sodium pyrophosphate, 0.1% SDS, 200 μg/ml denatured herring sperm DNA, 120 μg/ml tRNA and approximately 60 ng labelled oligonucleotide (6×10^6 cpm). Denhardt's solution is 0.02% Ficoll, 0.02% polyvinylpyrollidone, 0.02% bovine serum albumin. The filters were washed extensively in 6 × SSC and 0.1% SDS at 35 °C and autoradiographed.

Screening with the Hin*dIII insert of pcIH 161.* Replica filters prepared from cDNA cloning method B were screened by colony hybridization using the ^{32}P labelled insert of pcIH 161. Hybridization was for 18 h at 34 °C in 30% formamide, 10 × Denhardt's solution, 6 × SSC, 20 mM Tris/HCl pH 7.5, 200 μg/ml herring sperm DNA and 100 ng probe (2.5×10^6 cpm). The filters were washed in 6 × SSC and 0.1% SDS at 54 °C (equivalent stringency in the absence of formamide).

Analysis of cDNA clones

Recombinant plasmid DNA was isolated by the method of Holmes and Quigley [23]. Restriction enzyme digests were performed according to the suppliers conditions (Bethesda Research Laboratories) and analyzed by polyacrylamide gel electrophoresis [24]. Selected fragments for nucleotide sequencing or for use as hybridization probes were recovered by electro-elution from gel slices [22]. Probes were labelled by nick translation using [α-^{32}P]dATP and DNA polymerase I [25]. Overlapping restriction fragments were sub-cloned in M13 vectors [26] and sequenced in both orientations by the dideoxy method of Sanger et al. [27].

Cloning of CI-2 cDNA sequences into an in vitro *transcription vector and* in vitro *synthesis of mRNA*

The *Eco*RI/*Hin*dIII cDNA inserts of clones pcIC 34 and pcIC 38 (see Fig. 2) were sub-cloned into the Riboprobe Gemini vector pGEM1 (Promega Biotech). Recombinants were identified by colony hybridization to the insert of pcIC 38. Plasmid DNA was prepared and linearized downstream from the T7 promoter using *Hin*dIII. Messenger RNA was transcribed using T7 RNA polymerase according to the method recommended by the suppliers. The mRNA was translated in the wheat-germ system containing ribonuclease inhibitor. The products were analyzed on an 18% urea/SDS/polyacrylamide gel and fluorographed.

Genomic blot analysis

Total barley DNA was isolated from etiolated shoots and purified by CsCl gradient centrifugation [28]. DNA samples (10 μg) were digested to completion with 40 U restriction enzyme and separated on 0.8% agarose gels. DNA was transferred electrophoretically to Gene Screen Plus membrane (Du Pont) according to the manufacturers recommendations. The hybridization and washing conditions are described in the legend to Fig. 4.

Northern blot analysis

Barley endosperm RNA was denatured using glyoxal and dimethylsulphoxide [29] separated on a 1.4% agarose slab gel and transferred to Biodyne A membrane (Pall Process Filtration Ltd) [30]. Hybridization and washing conditions are described in the legend to Fig. 5.

RESULTS

Analysis in vivo *and* in vitro *of CI-2 in normal and high-lysine barley endosperms*

Total salt-soluble proteins were extracted from mature grains of normal barley (cv. Sundance) and high-lysine barley (cv. Hiproly) and analyzed by SDS-PAGE (Fig. 1A). The greatly increased level of CI-2 in Hiproly grain is clearly shown, confirming previous results [4, 5, 10]. CI-2 levels comparable to those found in Sundance were observed in other normal barley varieties (results not shown). Analysis of *in vitro* translation products of membrane-bound poly(A)-rich RNA from Sundance and Hiproly (Fig. 1B) shows a clear correlation between the levels of *in vitro* products of CI-2 mRNA from developing endosperm, and the amounts of the protein in the mature seed. The results also confirm that CI-2 mRNA is present on membrane-bound polysomes [31].

Identification of CI-2 cDNA clones

An initial cDNA library of 400 transformants was constructed in pUC8 using *Hin*dIII linkers (see Materials and Methods). Size-fractionated poly(A)-rich RNA derived from the membrane-bound polysomes of Hiproly endosperm was used as template for cDNA synthesis. The library was screened for CI-2 sequences by colony hybridization using

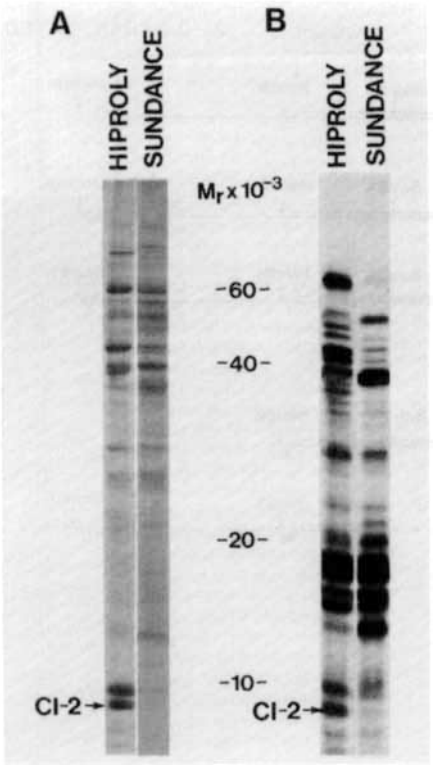


Fig. 1. *Comparison of* in vivo *extracted salt-soluble proteins and the* in vitro *translation products of mRNA from endosperms of normal and high-lysine barley.* (A) Salt-soluble proteins from mature grain separated on a 16% polyacrylamide gel and stained with Coomassie blue. (B) *In vitro* translation products of poly(A)-rich RNA purified from membrane-bound polysomes from 20-day-old endosperms. RNA was translated in the wheat germ system using [^{3}H]leucine and [^{3}H]proline and the labelled products separated on a 16% polyacrylamide gel and fluorographed. The migration of CI-2 is known from its immunoprecipitation with CI-2 monospecific antibodies (not shown)

a ^{32}P-labelled synthetic oligonucleotide probe; one colony hybridized strongly. This clone (pcIH 161) was found to contain only a partial copy of a CI-2 mRNA, the cDNA insert being truncated at its 3′ end by an internal *Hin*dIII site. A second library was therefore constructed using the 'tailing' procedure of Heidecker and Messing [18] to overcome the problem of internal restriction sites. A high-colony-density plating and screening technique [20] was employed, allowing large numbers of colonies (500 – 1000 recombinants per filter) to be screened quickly and efficiently. Replica filters were screened with the ^{32}P-labelled insert of pcIH 161 and 40 colonies were identified (designated pcIC 1 – 40). Ten clones from this second library and pcIH 161 were taken for further analysis.

Nucleotide sequences of CI-2 cDNAs

Nucleotide sequencing of these 11 cDNA inserts has revealed two types of CI-2 cDNA clones which we have called CI-2A and CI-2B. The partial restriction maps of five cDNAs are aligned in Fig. 2A and their sizes and orientation within the vector are indicated. The nucleotide and deduced amino acid sequences of CI-2A and CI-2B are presented in Fig. 2B along with the published protein sequence.

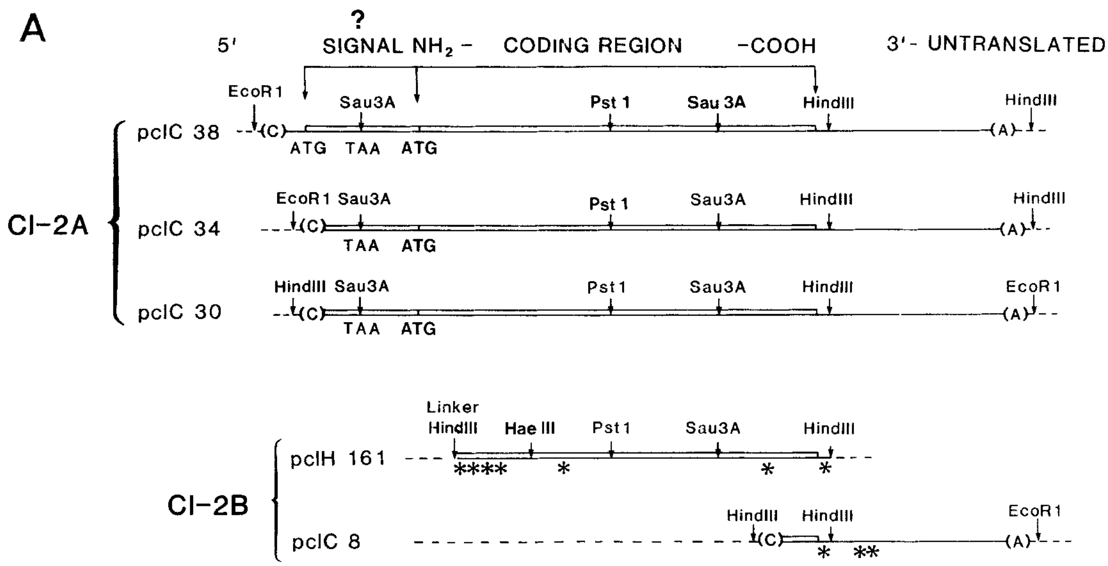
The majority of our cDNAs belong to the CI-2A subfamily of which nine inserts have been sequenced. Although these inserts differ in length, they all have identical sequences. The longest clone pcIC 38 has an insert of 520 bp and is probably a full-length copy of a CI-2 mRNA. The sequence contains a single open reading frame starting wih the ATG codon at nucleotide position 85 and terminating with a TAG codon at position 337. The deduced protein of 84 residues (M_r 9380) has a sequence identical to the amino acid sequence directly determined by Svendson et al. [7]. The N-terminus of the mature inhibitor was blocked to Edman degradation and so its exact position is not clear. The identity of four N-terminal residues was deduced from the amino acid composition of the corresponding tryptic fragment as Glx, Val, Ser, Ser [7]. The derived amino acid sequence of CI-2A agrees with this data and establishes this putative N-terminal sequence as Ser-Ser-Val-Glu.

Because CI-2 has been reported to be synthesized on membrane-bound polysomes [31] (Fig. 1B), it was assumed that the CI-2 mRNA would encode a signal peptide to direct the nascent protein to the membrane, consistent with the reported mode of synthesis of the chymotrypsin inhibitors in tomato [32, 33]. It is therefore interesting to note that the open reading frame of the cDNA pcIC 38 is preceded by a short nucleotide sequence starting with an in-frame ATG at position 16 which could encode a hydrophobic peptide with the characteristics of a signal sequence [34, 35]. However, the nucleotide sequence has an ochre codon (TAA) at position 52 in-frame with the two ATG codons at positions 16 and 85 respectively. Two other cDNA clones (pcIC 34 and pcIC 30, Fig. 2A) which result from independent cloning events, but are shorter copies of the same mRNA, also contain the TAA termination codon. This suggests that the stop codon might be present in all the CI-2A mRNAs (see below) and is not a cloning artefact.

The CI-2A sequence has a 3′ untranslated region of 111 nucleotides including a putative polyadenylation signal AATAAG (boxed in Fig. 2B), 17 bases upstream from the polyadenylation site. Although this deviates from the consensus polyadenylation signal (usually AATAAA), this variant signal has also been reported for the zein genes of maize [36] and for the tomato inhibitor I and II genes [33].

Only two CI-2B clones, pcIH 161 and pcIC 8 have been identified, neither of which is full length. pcIH 161 covers most of the coding region but terminates short of the N-terminus with a linker *Hin*dIII site. This cDNA is also truncated at its 3′ end by an internal *Hin*dIII site as a result of the original cloning procedure and so lacks virtually all the 3′ noncoding region. pcIC 8 however contains a full 3′ untranslated region and partially overlaps with pcIH 161 within the coding region. Alignment of these two sequences is based somewhat tentatively on a C to G substitution from the CI-2A sequence immediately after the TAG termination codon. Both sequences contain a number of substitutions when aligned with the CI-2A sequence (marked by an asterisk in Fig. 2A). Most of the heterogeneity is found in the extreme 5′ region of pcIH 161 (Fig. 2B). The most surprising feature of this region is that it would encode three cysteines, although the purified inhibitor has been shown to be devoid of cysteine [7]. Attempts to isolate further CI-2B clones using the 40-bp *Hin*dIII/*Hae*III fragment from pcIH 161 (Fig. 2A) as a probe were unsuccessful, suggesting that either the abundance of this subfamily is extremely low or that this 5′ region represents a cloning artefact. However, because the remaining substitutions occur randomly within otherwise identical sequences, it is clear that CI-2A and CI-2B do represent two distinct subfamilies of CI-2 mRNA.

A



B

```
                      10        20        30        40        50        60
CI-2A (NUC)  [C]ATTAAACTGATGACATGACAGTTCAAGATCTCACAGTCACATCGGCGATCTAATCAGTC
CI-2A (PRO)                 MetThrValGlnAspLeuThrValThrSerAlaIle * SerVal
                                     -20                            -10
          SIGNAL ?--┐
                70        80 ¦      90       100       110       120
CI-2A (NUC)  TCACAGGAAGCGAGCGTAACAAGGATGAGTTCAGTGGAGAAGAAGCCGGAGGGAGTGAAC
CI-2A (PRO)  SerGlnGluAlaSerValThrArgMetSerSerValGluLysLysProGluGlyValAsn
DIRECT PRO                          (SerSerValGlx)- - - - - - -
                                  -1  1                      10

               130       140       150       160       170       180
CI-2A (NUC)  ACCGGTGCTGGTGACCGTCACAACCTGAAGACAGAGTGGCCAGAGTTGGTGGGGAAATCG
CI-2B (NUC)  GATT-CCTGT-----T----A----A------------------------A-------
CI-2A (PRO)  ThrGlyAlaGlyAspArgHisAsnLeuLysThrGluTrpProGluLeuValGlyLysSer
CI-2B (PRO)  AspCysLeuCys - CysGln - Gln - - - - - - - - Glu - -
DIRECT PRO   ThrGlyAlaGly - ArgHis - Leu - - - - - - - - Gly - -
                                  20                         30

               190       200       210       220       230       240
CI-2A (NUC)  GTGGAGGAGGCCAAGAAGGTGATTCTGCAGGACAAGCCAGAGGCGCAAATCATAGTTCTG
CI-2B (NUC)  ------------------------------------------------------------
CI-2A (PRO)  ValGluGluAlaLysLysValIleLeuGlnAspLysProGluAlaGlnIleIleValLeu
CI-2B (PRO)  - - - - - - - - - - - - - - - - - - - -
DIRECT PRO   - - - - - - - - - - - - - - - - - - - -
                                  40                         50

               250       260       270       280       290       300
CI-2A (NUC)  CCGGTGGGGACAATTGTGACCATGGAATATCGGATCGACCGCGTCCGCCTCTTTGTCGAT
CI-2B (NUC)  ------------------------------------------------------------
CI-2A (PRO)  ProValGlyThrIleValThrMetGluTyrArgIleAspArgValArgLeuPheValAsp
CI-2B (PRO)  - - - - - - - - - - - - - - - - - - - -
DIRECT PRO   - - - - - - - - - - - - - - - - - - - -
                                  60                         70

               310       320       330       340       350       360
CI-2A (NUC)  AAACTCGACAACATTGCCCAGGTCCCCAGGGTCGGCTAGCAAGCTTGAGAGCTAGCCTGC
CI-2B (NUC)  -G-------------------------------G--------------------------
CI-2A (PRO)  LysLeuAspAsnIleAlaGlnValProArgValGly *
CI-2B (PRO)  Arg - - - - - - - - - - - *
DIRECT PRO   Lys - - - - - Glu - - - - -
                                  80

               370       380       390       400       410       420
CI-2A (NUC)  TGCTGGCGTGTATGTATTGCAGCTTCACCATCTCTTCTTGGCTATAGCAAGATTGAGATT
CI-2B (NUC)  ---------A----[     ]---------------------------------------

               430       440       450
CI-2A (NUC)  TATAAATCATATACAATAAGAGTTGCTGCGG[A]
CI-2B (NUC)  -------------------------------[A]
```

Fig. 2. *The two sub-families of CI-2 cDNAs: their partial restriction maps, nucleotide and deduced amino acid sequences.* (A) Partial restriction map of the major CI-2A and CI-2B cDNAs. The orientation of the cDNAs in the vector is shown. Nucleotide substitutions between CI-2A and CI-2B are indicated by asterisks. (B) Nucleotide and deduced amino acid sequences of the CI-2A and CI-2B cDNAs. The positions where CI-2B differs from CI-2A are indicated. The short deletion in the CI-2B 3′ non-coding region is flanked by square brackets. Stop codons are marked by asterisks. A possible cleavage site for the putative signal sequence is indicated by an arrow. The putative polyadenylation signal is boxed

Cloning of CI-2 cDNA into an in vitro *transcription vector and translation of mRNA transcripts*

It was important to determine whether the CI-2A cDNA produces a functional mRNA and, if so, to define the AUG codon at which protein synthesis starts *in vitro*. To do this we cloned the inserts of pcIC 38 and pcIC 34 into an *in vitro* transcription vector. pcIC 38 is the full-length cDNA and contains two ATG codons, which might function as initiation codons, separated by a TAA stop codon (Fig. 2A). pcIC 34 is a shorter copy of the same RNA and does not contain the ATG codon upstream from the TAA codon (Fig. 2A). Therefore pcIC 34 has only one ATG codon where initiation of translation is likely. Messenger RNA transcripts were made from both these cDNAs using T7 polymerase and were used to prime the wheat-germ *in vitro* translation system. Both mRNAs translated *in vitro* to give a single polypeptide which co-migrated with the CI-2 synthesized by poly(A)-rich RNA from Hiproly (Fig. 3, compare lanes e and f with lane c). This result has several implications. Firstly, it indicates that the cDNAs are probably derived from a functional mRNA. Secondly, because the translation product of the mRNA from pcIC 38 is the same size as that synthesized by mRNA from pcIC 34, we can conclude that the ATG codon (at position 85) downstream from the TAA codon, is the site at which

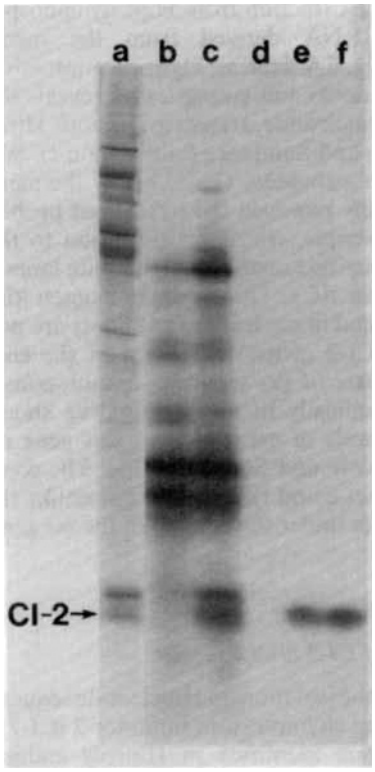


Fig. 3. *Translation products of CI-2 mRNA transcripts synthesized from cDNA.* Lane a, salt-soluble proteins from Hiproly grain; lanes b and c, *in vitro* translation products of Sundance and Hiproly poly(A)-rich RNA respectively; lane d, no RNA control; lanes e and f, *in vitro* translation products of transcripts from pcIC 38 and pcIC 34, respectively. The RNA was translated in the wheat-germ system and the ^{3}H-labelled products separated on a 18% polyacrylamide gel and fluorographed

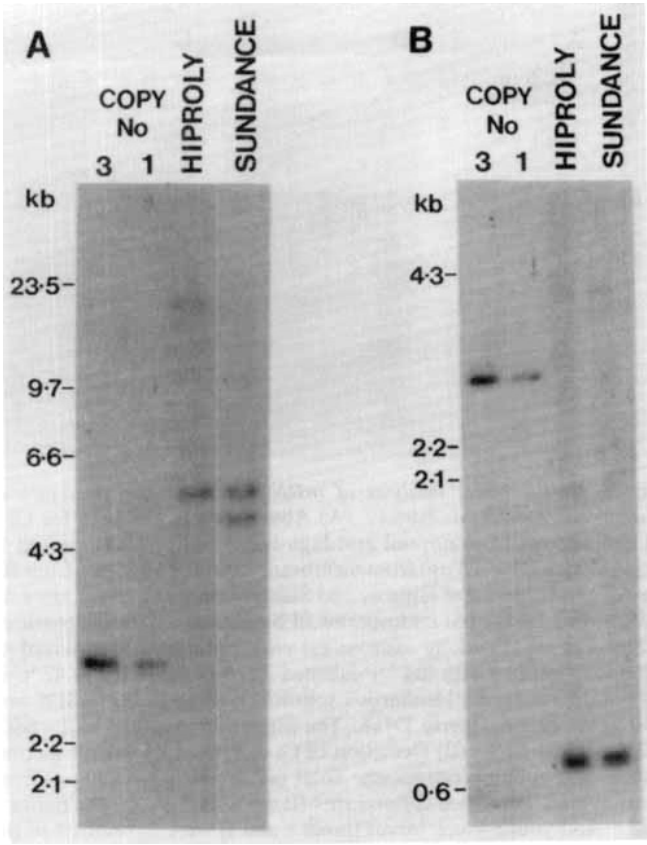


Fig. 4. *Southern-blot analysis of genomic DNA from normal and high-lysine barleys.* 10 µg of DNA from Sundance and Hiproly was digested with *Eco*RI (A) and *Hin*dIII (B), separated on an 0.8% agarose gel and transferred to nylon membranes. Hybridization to the ^{32}P-labelled *Hin*dIII insert of pcIH 161 was at 65°C for 20 h in 10 × Denhardt's solution, 50 mM Tris/HCl, pH 7.4, 2 × SSC, 200 µg/ml herring sperm DNA, 1% SDS and 10 ng/ml probe (specific activity 5×10^{8} dpm/µg). The membranes were washed in 2 × SSC, 1% SDS at 65°C. *Hin*dIII-digested lambda phage DNA (BRL) was included as size markers. *Bam*HI digested pcIH 161 was used to estimate copy number; 5.4 pg plasmid is equivalent to 1 gene copy/haploid barley genome

translation initiates *in vitro*. Thirdly, because the CI-2 synthesized by the poly(A)-rich RNA from Hiproly also has the same mobililty (compare lane c with lanes e and f) it seems likely that with most, if not all, of the CI-2 mRNAs, protein synthesis begins at the same place. This would indicate that all, or the vast majority of CI-2 mRNAs (both CI-2A and CI-2B) are translated *in vitro* with no N-terminal signal sequence extension. This was confirmed by showing that the CI-2 protein extracted from the grain has the same mobility as the *in vitro* translation products (compare lane a with lane c).

Two technical points are worth noting. The mRNA synthesized from the cDNA was not capped and it contained very little 3′ flanking sequence, only 10 nucleotides from the first nucleotide of the termination codon. Thus capping and a long 3′ untranslated region are not essential requirements for the translation of CI-2 mRNA. We cannot, however, rule out the possibility that these factors could improve the efficiency of translation.

Organization of the structural genes for CI-2 in normal and high-lysine barley

Genomic blots were used to investigate the copy number and organization of CI-2 structural genes in Sundance and Hiproly. High-molecular-mass DNA was extracted from endosperms [28], digested to completion with either *Eco*RI or *Hin*dIII, electrophoresed on a 0.8% agarose gel and electroblotted onto Gene-screen Plus membranes. The filters were hybridized to the ^{32}P-labelled insert of pcIH 161 under conditions of moderately low stringency (25°C below the melting temperature). Under these conditions the pcIH 161 insert hybridizes to one *Hin*dIII fragment and two *Eco*RI restriction fragments from both genotypes (Fig. 4). Comparison with the reconstruction experiments suggests that there are at least four copies of CI-2 related genes per haploid genome of Sundance and Hiproly. The 0.8-kb *Hin*dIII fragment and 4.6-kb *Eco*RI fragment are both conserved in the two genotypes while the second *Eco*RI fragment is polymorphic having sizes of 8 kb and 20 kb in Sundance and Hiproly respectively (Fig. 4A).

Expression of CI-2 in normal and high-lysine barley

The differential expression of CI-2 in Hiproly and Sundance endosperms was studied by determining the abundance of CI-2 mRNA on membrane-bound polysomes by Northern blot hybridization. Glyoxylated poly(A)-rich RNA derived from initially membrane-bound polysomes from 20-day-old developing endosperms of Hiproly and Sundance respectively was probed with the cDNA insert of pcIH 161

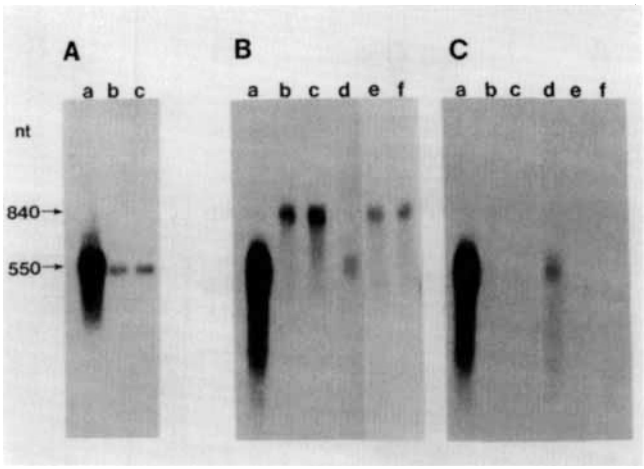


Fig. 5. *Northern-blot analysis of mRNAs from different tissues of normal and high-lysine barleys.* (A) Abundance of mRNAs for CI-2 in endosperms from normal and high-lysine barley. Lanes a and b, poly(A)-rich RNA (3 μg) from membrane-bound polysomes from 20-day-old endosperms of Hiproly and Sundance respectively. Lane c, as b but from 26-day-old endosperms of Sundance. Glyoxal-denatured RNA was separated by agarose gel electrophoresis, transferred to filters and probed with the ^{32}P-labelled insert of pcIH 161 at 42 °C in 50% formamide, 5 × Denhardt's solution, 5 × SSPE, 0.2% SDS and 250 μg/ml herring sperm DNA. The filters were washed in 2 × SSC, 0.2% SDS at 50 °C. (B) Detection of CI-2 transcripts in endosperms, shoots and young green leaves. Total polysomal RNA (40 μg) from Hiproly and Sundance endosperms (lanes a and d), shoots (lanes b and e) and young green leaves (lanes c and f) were hybridized to the ^{32}P-labelled insert of pcIC 38 at 60 °C in 5 × Denhardt's solution, 5 × SSPE, 0.1% sodium pyrophosphate, 1% SDS and 250 μg/ml herring sperm DNA. The filters were washed in 2 × SSC, 0.5% SDS at 60 °C. (C) The filter shown in C was washed at higher stringency in 0.2 × SSC, 0.1% SDS at 65 °C (1 × SSPE = 0.15-M NaCl, 10-mM NaH_2Po_4, 1-mM EDTA, pH 7.4)

(Fig. 5A, lanes a and b). The greatly increased levels of CI-2 mRNA in Hiproly is clearly demonstrated, correlating with the *in vitro* translation results of these RNA fractions and the *in vivo* accumulation of the inhibitor protein in mature barley grain (Fig. 1). RNA was also isolated from 24–26-day-old endosperms of Sundance (Fig. 5A, lane c). Although this is a much later stage of development, the level of Sundance CI-2 mRNA has only increased around twofold, and is still many times lower than that observed for Hiproly at 20 days. These results suggest not only an earlier, but also a much stronger activation of the CI-2 gene(s) in Hiproly.

Using glyoxylated *Hae*III fragments of ΦX174 as markers, we estimate the size of the endosperm CI-2 mRNA to be 550 nucleotides. The longest cDNA, pcIC 38, contains a total of 450 nucleotides. Allowing for a poly(A) tail of 50–100 bases this would indicate that pcIC 38 is virtually a full-length copy of the CI-2 mRNA.

Tissue specificity of CI-2 expression

We have also investigated the tissue-specific expression of CI-2 in the developing endosperm, etiolated shoots and young green leaves. Total polysomal RNA was extracted from these tissues and analyzed by Northern blot hybridization. A cDNA fragment covering the entire coding region of pcIC 38 was used as a probe and hybridized at low stringency to the filter (35 °C below the melting temperature). Lane a of Fig. 5B shows the same increased level of CI-2 mRNA in the total polysomal RNA fraction from Hiproly endosperms as in the poly(A)-rich RNA derived from the membrane-bound polysomes (Fig. 5A, lane a). Hybridization of CI-2 cDNA to RNAs from shoots and young leaves reveals the presence of a larger, 830-nucleotide, transcript in both Hiproly (Fig. 5B, lanes b and c) and Sundance (lanes e and f). Washing of this filter at higher stringency (15 °C below the melting temperature) completely removed the hybridized probe to the shoot and leaf transcripts, whilst hybridization to the endosperm transcripts remained unaffected (compare lanes a and lanes d of Fig. 5B and 5C). These results suggest that the larger transcripts found in the leaves and shoots are not the product of the same CI-2 gene(s) expressed in the endosperm, but are the products of (a) separate, homologous CI-2 gene(s) expressed specifically in the germinating shoot/young leaf. The similar levels of expression of this gene in the shoots/leaves of Hiproly and Sundance (Fig. 5B; compare lanes b and c with lanes e and f), indicate that unlike the endosperm gene(s), it is not under the control of the *lys* gene in Hiproly.

DISCUSSION

Sub-families of CI-2 cDNA clones

We report the isolation and nucleotide sequences of cDNA clones encoding chymotrypsin inhibitor-2 (CI-2) from barley. The clones were identified in Hiproly endosperm cDNA libraries, initially using a synthetic oligonucleotide probe. The cDNAs have been divided into two types, CI-2A and CI-2B, on the basis of their nucleotide sequences. The sequences are over 90% homologous with substitutions occurring in both the coding and 3′ non-coding regions.

Two lines of evidence suggest that CI-2A represents the major sub-family of mRNA in Hiproly endosperm. Firstly the abundance of the CI-2A cDNAs is much higher; 9 out of 11 of the clones characterized contained this sequence. Secondly the deduced amino acid sequence of CI-2A agrees exactly with the published protein sequence for CI-2 from Hiproly, the CI-2B sequence differing at several positions (Fig. 2B).

Only one CI-2 band is resolved by one-dimensional gel electrophoresis (results not shown), which suggests that either the products of the CI-2A and CI-2B mRNAs co-migrate, or that CI-2B protein is in such low abundance that it is not detected on the gel.

Does CI-2 possess a signal sequence?

Recently cDNAs for several other plant protease inhibitors have been identified. These include the tomato inhibitors I and II [33], the potato proteinase inhibitor II [37] and the soybean Kunitz inhibitor [38]. Their nucleotide sequences all predict an N-terminal signal sequence extension, consistent with the synthesis of these proteins on membrane-bound polysomes and their deposition in vacuoles. Similarly CI-2 has been shown to be synthesized on membrane-bound polysomes [31] and thus its mRNAs were expected to encode an N-terminal signal sequence extension to direct the nascent protein to the rough endoplasmic reticulum.

It is therefore surprising that the single open reading frame of pcIC 38 begins with the ATG codon at position 85 immediately preceding the putative N-terminus of the mature protein (Fig. 2B). Translation experiments using mRNA transcripts made from cDNA, show that protein synthesis, *in vitro*, starts at this ATG. Is it possible, therefore, that CI-2

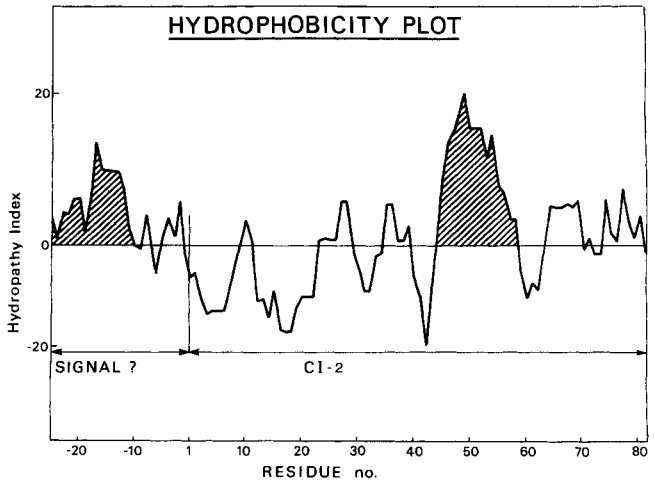


Fig. 6. *Hydrophobicity plot of the deduced protein sequence of pcIC 38.* The ordinate gives the hydrophobicity index [46] and the abcissa shows the amino acid position (labelled according to Fig. 2B). The two main hydrophobic domains are hatched

does not contain a signal sequence and that the protein is really cytosolic? This would be an unusual result in view of the location of other protease inhibitors (see above). It is, however, possible that a signal sequence for CI-2 is not located beyond the N-terminus, but internally, within the coding sequence. The hydrophobicity plot shown in Fig. 6 shows that a region of the coding sequence, extending from position 46 (Ala) to position 59 (Met), is indeed highly hydrophobic. However this region is close to the C-terminus and is unlikely to function as a signal peptide, because the ribosomes would have disengaged from the mRNA before this sequence was exposed [39]. Therefore we think it unlikely that CI-2 contains a sequence within the coding regions which acts as a signal sequence. An alternative explanation is that, *in vivo,* initiation begins at the ATG codon at position 16 (Fig. 2B). Both ATG codons are in comparable contexts suggesting that either could initiate translation [40, 41]. If initiation occurs at this first ATG codon and the TAA codon is read-through, this would produce a sequence with the properties of a functional signal peptide [34, 35]. Since the N-terminus of the mature protein is blocked we cannot deduce the precise cleavage site. We have therefore applied the rules of von Heijne [35] and suggest that a possible processing site is between the threonine and the arginine (see Fig. 2B). This would give a size of 22 amino acids, which is of the right order for a signal peptide.

The synthesis of this signal sequence would be totally dependent on the suppression of the stop codon. Nonsense suppressor tRNAs have recently been found in plants [42, 43] and they may be present in barley endosperm. We are currently investigating this possibility.

Organization and expression of CI-2 genes

Southern hybridizations to genomic DNA show that CI-2 is encoded by a small multigene family of approximately 4–6 copies per haploid genome both in normal and high-lysine barley (Fig. 4). This means that the differential accumulation of CI-2 is not due to an increased gene copy number in Hiproly as compared to normal barley but is controlled by a *trans*-acting regulatory gene. Similar high-lysine regulatory mutants are known in maize [44]. The abundance of CI-2 mRNA in endosperms from normal and high-lysine barley was determined by Northern analysis using both poly(A)-rich RNA from membrane-bound polysomes and RNA from total polysomes. A similar increase of CI-2 mRNAs was obtained in Hiproly when compared to Sundance irrespective of the source of RNA. The abundance of CI-2 mRNA correlated well with the amount of CI-2 synthesized *in vitro* and the deposition of CI-2 *in vivo*. Therefore the deposition of CI-2 is chiefly controlled by the abundance of CI-2 mRNA and may be due to more efficient transcription and/or stability of the mRNA. Comparison of the abundance of mRNAs in 20- and 26-day-old endosperms of Sundance (Fig. 5A) shows only an about twofold increase. This would again suggest a higher rate of transcription of the genes in Hiproly or greater stability of the mRNA.

We have been able to distinguish between mRNAs that are closely related to our cDNA clones and those that are more distantly related by varying the stringency of hybridization. Fig. 5 shows that the larger transcripts in shoots and leaves are the products of different but homologous CI-2 gene(s). This result agrees with the work of Kirsi [45] who demonstrated that the chymotrypsin inhibitors from shoots and endosperms have different activities, inferring they must be different proteins. At present it is unclear whether the genes expressed in the leaves and shoots hybridized in the genomic analysis. Interestingly the levels of CI-2 mRNAs from shoots and leaves are similar in normal and high-lysine barleys, indicating that, unlike the endosperm genes, the expression of the shoot and leaf genes is not under the control of the *lys* gene.

It is clear from these results that the CI-2 multigene family of barley is under complex regulatory control. In addition to the strict developmental and tissue-specific expression of individual members of this family, the expression of at least one of the endosperm-specific genes can be controlled by the action of a mutant 'regulatory' gene on a separate chromosome. The chymotrypsin inhibitor 2 genes therefore provide a useful system for studying the control of gene expression in higher plants.

We are grateful to Drs B. J. Miflin, B. G. Forde and P. R. Shewry for helpful discussions and to P. R. S. for critical reading of the manuscript. This work was supported partly by EEC Contract GBI-4-023-UK(H) of the Biomolecular Engineering Programme and EEC Grant 4712 of the Plant Productivity Programme. We thank Dr L. Munck for the generous gift of Hiproly seed and for CI-2 antibodies.

REFERENCES

1. Svendsen, I., Boisen, S. & Hejgaard, J. (1982) *Carlsberg Res. Commun. 47,* 45–53.
2. Hejgaard, J. & Boisen, S. (1980) *Hereditas 93,* 311–320.
3. Munck, L., Karlson, K. E., Hagberg, A. & Eggum, B. O. (1970) *Science (Wash. DC) 168,* 985–987.
4. Jonassen, I. (1980) *Carlsberg Res. Commun. 45,* 59–68.
5. Boisen, S., Andersen, C. T. & Hejgaard, J. (1981) *Physiol. Plant. 52,* 167–176.
6. Jonassen, I. (1980) *Carlsberg Res. Commun. 45,* 47–58.
7. Svendsen, I., Martin, H. B. & Jonassen, I. (1980) *Carlsberg Res. Commun. 45,* 79–85.
8. Karlsson, K. E. (1976) in *Barley genetics III* (Gaul, H., ed.) pp. 536–541, Verlag Karl Thiemig, München.
9. Hejgaard, J., Bjorn, S. E. & Nielsen, G. (1984) *Theor. Appl. Genet. 68,* 127–130.
10. Rasmussen, U. (1985) *Carlsberg Res. Commun. 50,* 83–93.
11. Giese, H. & Hejgaard, J. (1984) *Planta (Berl.) 161,* 172–177.

12. Rahman, S., Kreis, M., Forde, B. G., Shewry, P. R. & Miflin, B. J. (1984) *Biochem. J. 233,* 315–322.
13. Bantle, J. A., Maxwell, I. H. & Hahn, W. E. (1976) *Anal. Biochem. 72,* 413–427.
14. Kreis, M., Rahman, S., Forde, B. G., Pywell, J., Shewry, P. R. & Miflin, B. J. (1983) *Mol. Gen. Genet. 191,* 194–200.
15. Beachy, R. N., Batron, K. A., Thompson, J. F. & Madison, J. T. (1980) *Plant Physiol. 65,* 990–994.
16. Efstratiadis, A., Maniatis, T., Kafatos, F. C., Jeffrey, A. & Vournatis, J. N. (1975) *Cell 4,* 367–378.
17. Forde, B. G., Kreis, M., Williamson, M. S., Fry, R. P., Pywell, J., Shewry, P. R., Bunce, N. & Miflin, B. J. (1985) *EMBO J. 4,* 9–15.
18. Heidecker, G. & Messing, J. (1983) *Nucl. Acids Res. 11,* 4891–4906.
19. Hanahan, D. (1983) *J. Mol. Biol. 166,* 557–580.
20. Hanahan, D. & Meselson, M. (1980) *Gene 10,* 63–67.
21. Grunstein, M., Hogness, D. S. (1975) *Proc. Natl Acad. Sci. USA 73,* 3146–3150.
22. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol. 65,* 499–560.
23. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem. 114,* 193–197.
24. Peacock, A. C., Dingman, C. W. (1967) *Biochemistry 6,* 1818–1827.
25. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol. 113,* 237–251.
26. Messing, J., Crea, R. & Seeburg, D. H. (1981) *Nucleic Acids Res. 9,* 309–321.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl Acad. Sci. USA 74,* 5463–5467.
28. Kreis, M., Shewry, P. R., Forde, B. G., Rahman, S. & Miflin, B. J. (1983) *Cell 34,* 161–167.
29. McMaster, G. K. & Carmichael, G. C. (1977) *Proc. Natl Acad. Sci. USA 11,* 4835–4838.
30. Thomas, P. S. (1980) *Proc. Natl Acad. Sci. USA 77,* 5201–5205.
31. Jonassen, I., Ingversen, J. & Brandt, A. (1981) *Carlsberg Res. Commun. 46,* 175–181.
32. Walker-Simmons, M. & Ryan, C. A. (1977) *Plant Physiol. 60,* 61–63.
33. Graham, J. S., Pearce, G., Merryweather, J., Titani, K., Ericsson, L. & Ryan, C. A. (1985) *J. Biol. Chem. 260,* 6555–6560.
34. Watson, M. E. E. (1984) *Nucleic Acids Res. 12,* 5154–5164.
35. Von Heijne, G. (1983) *Eur. J. Biochem. 133,* 17–21.
36. Hu, N. T., Pfeifer, M. A., Heidecker, G., Messing, J. & Rubenstein, I. (1982) *EMBO J. 1,* 1337–1342.
37. Sanchez-Serrano, J., Schmidt, R., Schell, J. & Willmitzer, L. (1986) *Mol. Gen. Genet. 203,* 15–20.
38. Hoffman, L. M., Sengupta-Gopalan, C., Paaren, H. E. (1984) *Plant Mol. Biol. 3,* 111–117.
39. Hortsch, M. & Meyer, D. E. (1984) *Biol. Cell 52,* 1–8.
40. Kozak, M. (1983) *Microbiol. Rev. 47,* 1–45.
41. Kozak, M. (1986) *Cell 44,* 283–292.
42. Barcisewski, J., Barciszewska, M., Suter, B. & Kubli, E. (1985) *Plant Sci. 4,* 193–196.
43. Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R. & Gross, H. J. (1984) *EMBO J. 3,* 351–356.
44. Soave, C., Tardani, I., Di Fonzo, N. & Salamini, F. (1981) *Cell 27,* 403–410.
45. Kirsi, M. & Mikola, J. (1977) *Physiol. Plant. 39,* 110–114.
46. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol. 157,* 105–132.