Identification of barley and wheat cDNA clones related to the high-M_r polypeptides of wheat gluten

J. Forde, B.G. Forde, R.P. Fry, M. Kreis, P.R. Shewry and B.J. Miflin

Department of Biochemistry, Rothamsted Experimental Station, Harpenden, Herts AL5 2JQ, England

Received 24 August 1983

We have identified 3 cDNA clones related to the high-M_r group of storage proteins in barley endosperm, the D-hordeins. A cDNA library has been constructed from wheat endosperm poly(A^+)-RNA and screened using one of the D-hordein cDNA clones. Two wheat clones which cross-hybridised to the barley clone have been identified, by hybrid-release translation and nucleotide sequence analysis, as partial copies of mRNAs encoding the high-M_r gluten polypeptides of wheat.

D-Hordein *Molecular cloning* *DNA-sequencing*

1. INTRODUCTION

The high-M_r gluten polypeptides of wheat occur in large, disulphide-linked aggregates [1,2] and they are thought to be important determinants of the viscoelastic properties of the gluten (review [3]). The high-M_r polypeptides are specified by the complex *Glu1* locus on the long arm of chromosome 1 [4]. The D hordeins in barley are specified by the *Hor3* locus which maps at an approximately equivalent location to *Glu1*, on the long arm of chromosome 5 [5]. The D hordeins, like the high-M_r wheat polypeptides, are a glycine-rich group of high-M_r prolamin polypeptides. Knowledge of the amino acid sequences of these proteins may help us to understand their role in determining the properties of gluten and to explain the differences in breadmaking quality between cereal species and between wheat cultivars. The complete sequences of individual high-M_r polypeptides are likely to be most readily obtained via nucleotide sequencing of the genes which encode them. As a first step, we report here the identification of barley and wheat cDNA clones related to the D-hordeins and to the high-M_r gluten polypeptides of wheat.

2. MATERIALS AND METHODS

2.1. *RNA purification, cDNA synthesis and cloning*

Poly(A^+)-RNA was isolated from membrane-bound polysomes of developing wheat (*Triticum aestivum* cv. Sentry) and barley (*Hordeum vulgare* cv. Sundance) endosperms as in [6,7]. Construction of the barley cDNA library has been described [7]. Double-stranded cDNA was synthesized from wheat RNA as in [7], ligated to pUC8 using *Hin*dIII linkers [8] and cloned in *Escherichia coli* K12 strain JM83.

2.2. *Plasmid purification and hybridizations*

Plasmid DNA was purified either as in [9] or, for larger culture volumes, by banding DNA from cleared lysates in CsCl density gradients [10]. The cDNA insert was excised from pHvE–c135 by *Pst*1 digestion, separated by electrophoresis in a 6% (w/v) acrylamide gel [11], electroeluted [12] and ^{32}P-labelled by nick translation [13]. In situ colony hybridizations were as in [14] except that the hybridization solution contained 30% (v/v) formamide and 1 M NaCl (at 35°C) and the washing solutions contained 0.9 M NaCl and 0.09 M

Published by Elsevier Science Publishers B.V.

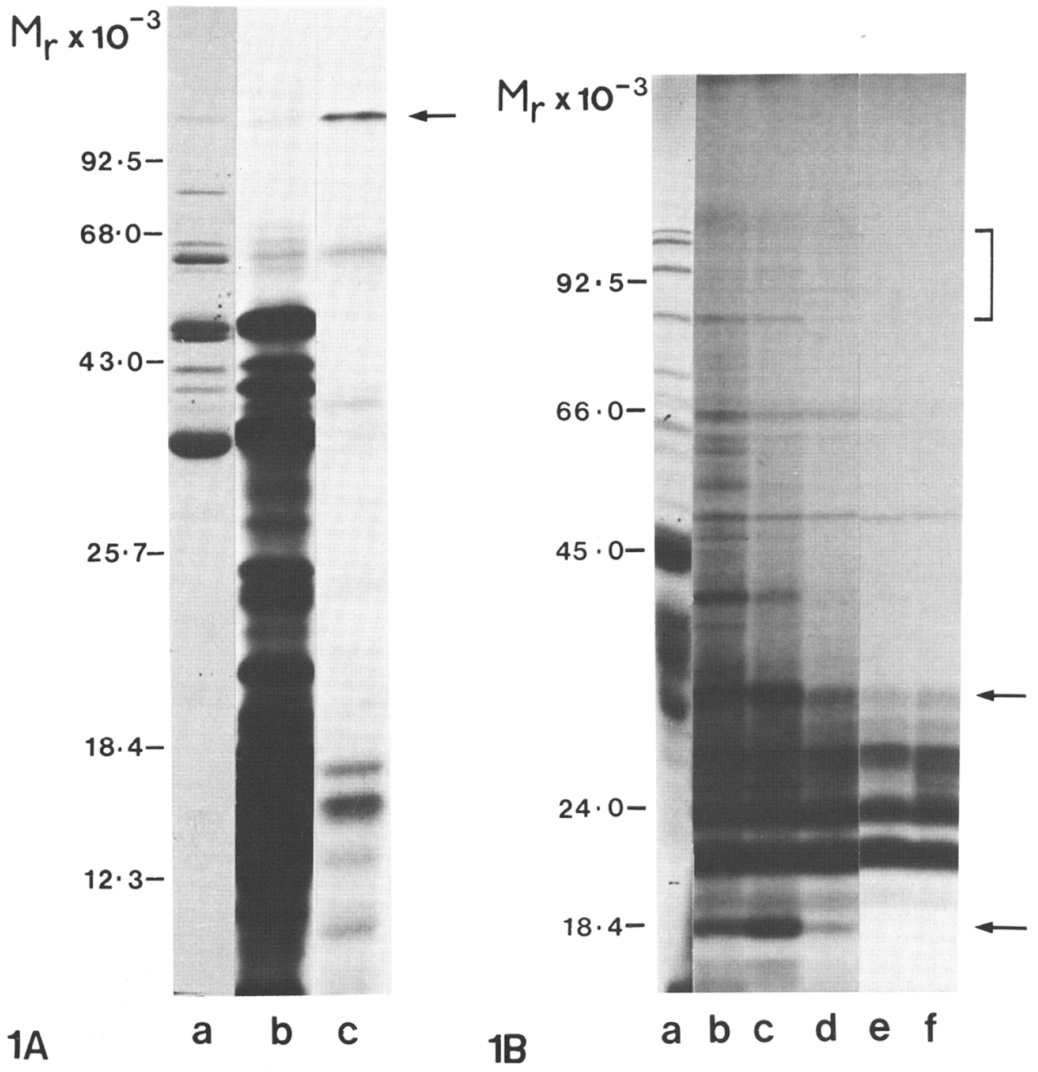

Fig.1. (A) Identification of a putative D-hordein clone by hybrid-release translation: (a) authentic hordein, stained with Coomassie blue; (b) in vitro translation products of endosperm poly(A^+)-RNA; (c) translation products of RNA selected from barley endosperm poly(A^+)-RNA by hybridization to the cDNA clone pHvE–c135. The mRNAs were translated using a wheat germ extract, and the ^{3}H-labelled products were analysed by SDS–PAGE and fluorography. The arrow indicates the position of the D-hordein polypeptide. Similar results were obtained with two other cDNA clones, pHvE–c133 and pHvE–c155.

Fig.1. (B) Characterization of putative cDNA clones for the high-M_r prolamin polypeptides of wheat using hybrid-release translation: (a) authentic wheat prolamin polypeptides; (b–d) translation products of RNA selected from wheat endosperm poly(A^+)-RNA by hybridization to pTaE–c256. Identical results were obtained with plasmids isolated from the 4 other cDNA clones identified from fig.2. The hybridized RNA was eluted successively in water at increasing temperatures; (b) 70°C; (c) 75°C; (d) 85°C; (e) translation products of RNA selected by hybridization to pUC8; (f) translation products arising from the addition of tRNA alone. The RNAs were translated in vitro using a reticulocyte lysate with [^{35}S]methionine as the labelled amino acid, and the products analyzed by SDS–PAGE and fluorography. The bracket indicates the position of the high-M_r polypeptides. The arrows refer to polypeptides discussed in the text. Additional polypeptides in (b–d) are common to the pUC8 control (e) and originate from the tRNA preparation used as carrier (f) and from endogenous mRNAs in the reticulocyte lysate.

Na_3-citrate (at 57°C) [15]. Dot hybridizations were based on the method in [16]. Poly(A^+)-RNA was bound to nitrocellulose filters using the Bethesda Research Laboratories Hybri-Dot apparatus. Nick-translated plasmid DNA was hybridized to the mRNA under identical conditions to those used for hybrid-release translations, except that the final wash with EDTA was not performed. Hybrid-release translations were done as in [7]. The selected mRNAs were translated in a reticulocyte lysate [6,17] containing [^{35}S]methionine, or a wheat-germ extract [18] containing [^{3}H]leucine and [^{3}H]proline. The translation products were analysed by SDS–PAGE [19] and the gels were fluorographed [20].

2.3. *Sub-cloning and nucleotide sequence analysis*

*Hin*dIII digests of pTaE–c237 and pTaE–c256 were sub-cloned in M13mp5, or M13mp9, using *E. coli* K12 strain JM101. Phage DNA was sequenced by the dideoxy method [21] using DNA polymerase 1, or, in some instances, avian myeloblastosis virus reverse transcriptase [22]. Procedures for transformation, propagation of M13 and isolation of phage DNA were as in [23].

2.4. *Amino acid sequencing*

The carboxy-terminal sequence of the high-M_r gluten polypeptide-termed subunit 2 of the wheat cultivar Highbury [24] was determined by following the kinetics of digestion with carboxypeptidase A [25].

3. RESULTS

3.1. *Preliminary identification of D hordein cDNA clones*

A selection of clones from a barley endosperm cDNA library have been analyzed by hybrid-release translation [26] and a group of B hordein clones has been identified [7,27]. Three plasmids from the same library (pHvE–c133, pHvE–c135 and pHvE–c155) hybridize to mRNA which encodes a polypeptide of similar electrophoretic mobility to the D hordeins (which migrate as a single band on SDS–PAGE). The results obtained with one of these plasmids, pHvE–c135, are shown in fig.1A.

3.2. *Identification of cDNA clones for high-M_r* gluten polypeptides

3.2.1. Cross-hybridisation to barley cDNA clones

The cDNA insert from pHvE–c135 was hybridized at low stringency ($T_m \sim -45$°C) to a wheat endosperm library using in situ colony hybridization [14]. Five of the 324 wheat cDNA clones hybridized to the barley cDNA probe. Two of these (pTaE–c237 and pTaE–c256) were chosen for further study.

3.2.2. Hybrid-release translation

Wheat endosperm RNA which hybridize to pTaE–c256 was eluted successively in water at increasing temperatures (70, 75 and 85°C) and translated in a reticulocyte lysate (fig.1B). The mRNAs which eluted at 75°C encoded polypeptides in the same M_r range as high-M_r polypeptides (bracketed in fig.1B), as well as two smaller polypeptides of M_r 30000 and 18000 (arrowed in fig.2). The synthesis of these two smaller polypeptides, in addition to the high-M_r polypeptides, suggested that the cloned cDNA might be derived from mRNA for a lower M_r homologue of the high-M_r polypeptides. We therefore determined the approximate size of the mRNA to which pTaE–c256 is most closely related.

3.2.3. Size-fractionation of mRNA

RNA from the developing wheat endosperm was fractionated on denaturing sucrose gradients ([6], fig.2A) and a dot hybridization assay used to estimate the abundance of sequences related to pTaE–c256. ^{32}P-Labelled plasmid DNA was hybridized at high stringency ($T_m \sim -5$°C) to aliquots of each RNA fraction. The strongest hybridizations (fig.2B) were to RNAs of highest M_r particularly those in fraction 13. RNA from this fraction directed the synthesis of a range of polypeptides (M_r 50000–130000), including the high-M_r polypeptides (fig.2A). Fig.2A also shows that two polypeptides, M_r 18000 and 30000, are present in the translation products of fractions 13 and 14 which give the highest proportion of high-M_r polypeptides. These were not synthesized by fractions 9 and 10, which contained mRNAs for other polypeptides of similar sizes. We conclude that the M_r 18000 and 30000 polypeptides are most likely to be premature termination products of the high-M_r polypeptide mRNA.

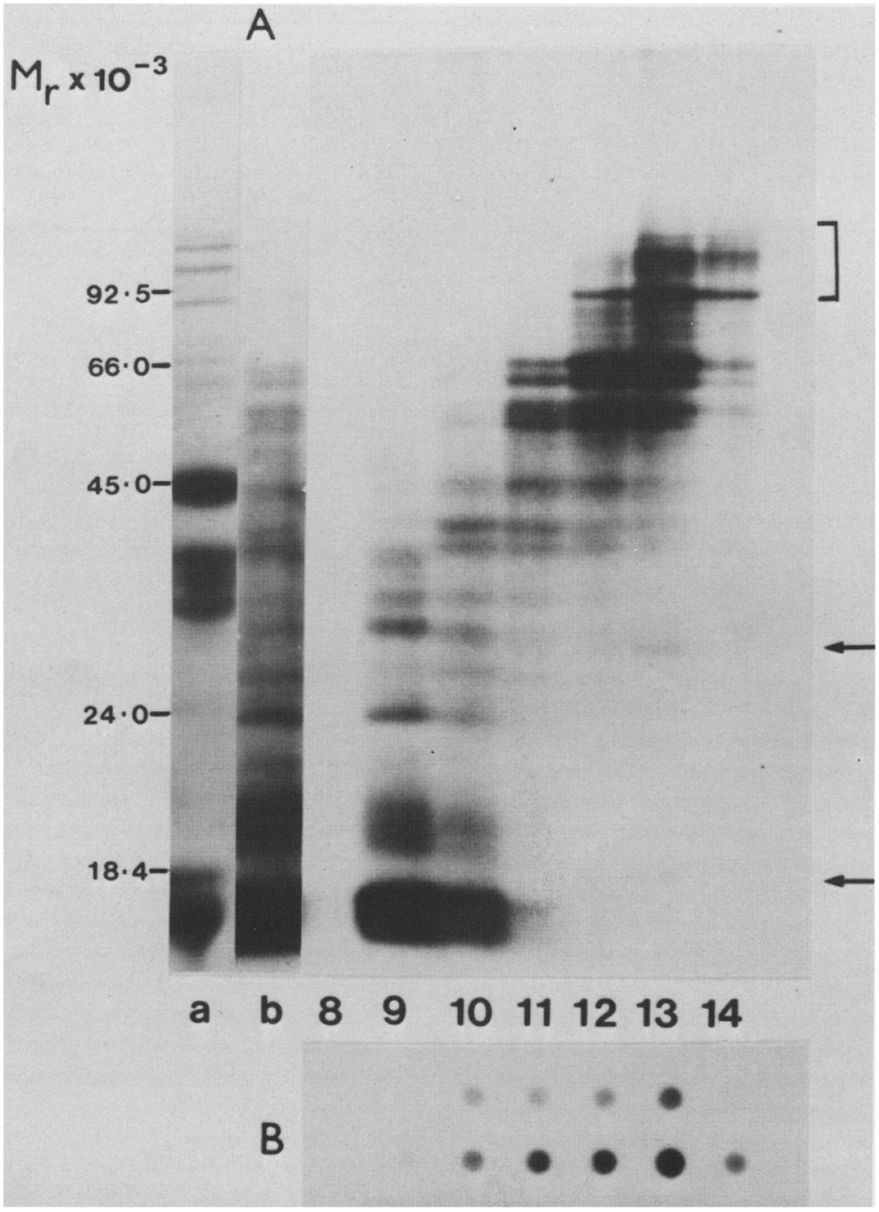

Fig.2. Abundance of (A) RNA encoding high-M_r gluten polypeptides and (B) RNA hybridizing to pTaE–c256, in fractions from a denaturing sucrose gradient. Poly(A^+)-RNA from membrane-bound polysomes of wheat endosperm was partially enriched for large mRNAs by fractionation in a 10–40% sucrose gradient in DMSO/formamide [6]. RNA sedimenting above 18 S was pooled and fractionated in a 5–20% sucrose gradient in DMSO/formamide. (A) Aliquots from each fraction were translated in a reticulocyte lysate containing [^{35}S]methionine and analyzed by SDS–PAGE and fluorography: (a) authentic ^{14}C-labelled wheat prolamins; (b) translation products of unfractionated poly(A^+)-RNA; (8–14) translation products obtained from fractions 8–14. Approximately equal amounts of radioactivity were loaded on each track for fractions 8–14. The location of the high-M_r polypeptides is indicated by a bracket and the arrows point to polypeptides discussed in the text. (B) One-twentieth (bottom row) and one-fortieth (top row) of each sucrose gradient fraction was dotted onto a nitrocellulose filter and hybridized to ^{32}P-labelled pTaE–c256. After washing, the filter was autoradiographed.

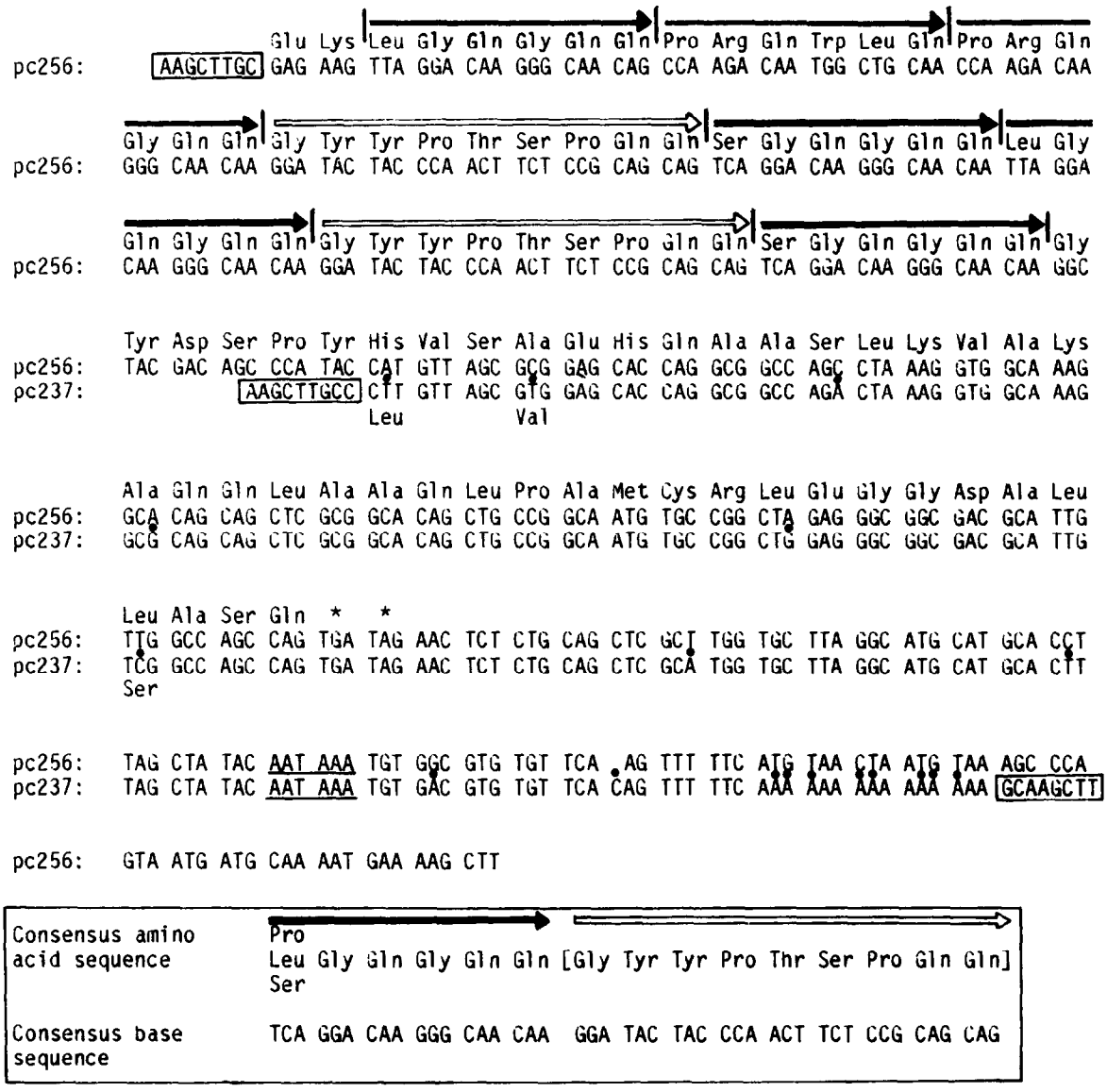

Fig.3. The nucleotide sequences of two wheat cDNA clones, pTaE–c237 and pTaE–c256. Above the pTaE–c256 sequence (p256; upper line) are shown the amino acids which it encodes, while below the pTaE–c237 sequence (pc237; lower line) are shown the amino acids encoded by it where they differ from these. The boxed sequences at either end of pTaE–c237, and at the start of pTaE–c256, are derived from *Hin*dIII linkers used to clone the cDNA. The putative polyadenylation signal sequences are underlined. Dots between the sequences indicate nucleotide differences between pTaE–c237 and pTaE–c256. The translation termination codons are indicated by asterisks. Closed arrows mark a repeated sequence of 6 amino acids. Open arrows mark another repeating unit of 9 amino acids. The consensus amino acid and nucleotide sequences for the two repeats are shown at the bottom of the figure.

3.2.4. Nucleotide sequencing

The nucleotide sequences of pTaE–c237 and pTaE–c256 are shown in fig.3. Both clones are derived from the 3′-region of the mRNA and are 95% homologous. The pTaE–c237 sequence contains 15 residues of a poly(A) tail and 24 nucleotides upstream from this lies a putative polyadenylation sequence (AATAAA). Although

pTaE–c256 extends further in the 3′ direction than pTaE–c237, it has no poly(A) tail, being truncated by a naturally occurring *Hin*dIII site. The 440 basepairs of pTaE–c256 represent about 20% of the estimated minimum size of the high-M_r polypeptide mRNA, based on an M_r of 69000 estimated for the protein by velocity sedimentation [28].

Using carboxypeptidase A we have determined the C-terminal sequence of a purified high-M_r polypeptide [28], classified as subunit 2 [24], as Leu–Ala–Ser–Gln-COOH. This is identical to the C-terminal sequence predicted from pTaE–c256 (fig.3), although pTa–c237 predicts a slightly variant form. The high proportion of glutamate + glutamine, proline and glycine in the first 61 amino acids of the sequence deduced from pTaE–c256 (36.0, 11.5 and 19.7 mol%, respectively) is also highly characteristic of the high-M_r polypeptides (32.6, 12.8 and 14.8 mol%, respectively [28]). The next 40 amino acids, up to the C-terminus, have a very different composition and appear to make up a separate C-terminal domain.

The region of the deduced amino acid sequence which is rich in glutamine, proline and glycine is made up of short repeated sequences, which are of two types (fig.3). The two longer repeats (9 amino acids) are identical at both the amino acid and nucleotide levels, while even the most divergent of the 6 shorter repeats (6 amino acids) is 72% homologous, at the nucleotide level, to the consensus sequence (fig.3).

4. DISCUSSION

We have identified a number of cDNA clones as partial copies of mRNAs encoding the high-M_r gluten polypeptides of wheat and their homologues in barley. The nucleotide sequences of the wheat cDNA clones have revealed the occurrence of short, interspersed repeated sequences in the primary structure of the high-M_r polypeptides (fig.3). The amino acid composition of the region containing the repeats is very similar to that of the complete protein, suggesting that the repeats may extend throughout the protein. Repeated amino acid sequences have been found in two other groups of cereal prolamins [29–31] but there is no apparent homology with those reported here.

The predicted amino acid sequence also supports previous evidence that the 3–4 cysteine residues in the high-M_r polypeptides are located at either end of the protein. Direct protein sequencing has located two cysteines in the first 25 residues of the N-terminus [28], and the cDNA sequences of pTaE–c237 and pTaE–c256 predict the presence of a single cysteine residue 13 amino acids from the C-terminus. Full-length cDNA or genomic sequences will be needed to confirm the hypothesis that all the cysteines are localized at the extremities of the high-M_r polypeptides. This arrangement of cysteines would be compatible with the 'linear glutenin' hypothesis [31], since it could allow the formation of long chains of disulphide-bonded high-M_r polypeptides (see [3]).

ACKNOWLEDGEMENTS

The authors thank Jackie Pywell and Saroj Parmar for experimental assistance, Jeane Hutchins and Sue Wilson for their help in preparing the manuscript and J.D. Marriage for the wheat germ. The work was partly supported by grant no.470 from the EEC.

REFERENCES

[1] Huebner, F.R. and Wall, J.S. (1976) Cereal Chem. 53, 258–269.
[2] Field, J.M., Shewry, P.R. and Miflin, B.J. (1983) J. Sci. Food Agric. 34, 370–377.
[3] Miflin, B.J., Field, J.M. and Shewry, P.R. (1983) in: Seed Proteins (Daussant, J.C. et al. eds) pp.255–319, Academic Press, London, New York.
[4] Payne, P.I., Holt, L.M., Worland, A.J. and Law, C.N. (1982) Theor. Appl. Genet. 63, 129–138.
[5] Shewry, P.R., Finch, R.A., Parmar, S., Franklin, J. and Miflin, B.J. (1983) Heredity 50, 179–189.
[6] Forde, J. and Miflin, B.J. (1983) Planta 157, 567–576.
[7] Forde, B.G., Kreis, M., Bahramian, M.B., Matthews, J.A., Miflin, B.J., Thompson, R., Bartels, D. and Flavell, R.B. (1981) Nucleic Acids Res. 9, 6689–6707.
[8] Goodman, H.M. and MacDonald, R.J. (1979) Methods Enzymol. 68, 75–90.
[9] Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193–197.
[10] Clewell, D.B. (1972) J. Bacteriol. 110, 667–676.
[11] Peacock, A.C. and Dingman, C.W. (1967) Biochemistry 6, 1818–1827.

[12] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[13] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.
[14] Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3965.
[15] Howley, P.M., Israel, M.A., Law, M.-F. and Martin, M.A. (1979) J. Biol. Chem. 254, 4876–4883.
[16] Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201–5205.
[17] Pelham, H.R.B. and Jackson, R.J. (1976) Eur. J. Biochem. 67, 247–256.
[18] Matthews, J.A. and Miflin, B.J. (1980) Planta 149, 262–268.
[19] Laemmli, V.K. (1970) Nature 227, 681–685.
[20] Chamberlain, J.P. (1979) Anal. Biochem. 98, 132–135.
[21] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[22] Smith, A.J.H. (1980) Methods Enzymol. 65, 560–580.
[23] Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309–321.
[24] Payne, P.I., Holt, L.M. and Law, C.N. (1981) Theor. Appl. Genet. 60, 229–236.
[25] Shewry, P.R., Field, J.M., Lew, Ellen, J.-L. and Kasarda, D.D. (1982) J. Exp. Bot. 33, 261–268.
[26] Ricciardi, R.P., Miller, J.S. and Roberts, B.E. (1979) Proc. Natl. Acad. Sci. USA 76, 4927–4931.
[27] Kreis, M., Rahman, S., Forde, B.G., Pywell, J., Shewry, P.R. and Miflin, B. (1983) J. Mol. Gen. Genet., in press.
[28] Field, J.M., Shewry, P.R., Miflin, B.J. and March, J.F. (1982) Theor. Appl. Genet. 62, 329–336.
[29] Geraghty, D., Peifer, M.A., Rubenstein, I. and Messing, J. (1981) Nucleic Acids Res. 9, 5163–5174.
[30] Pedersen, K., Devereux, J., Wilson, D.R., Sheldon, E. and Larkins, B.A. (1982) Cell 29, 1015–1026.
[31] Bartels, D. and Thompson, R.D. (1983) Nucleic Acids Res. 11, 2961–2976.
[32] Ewart, J.A.D. (1977) J. Sci. Food Agric. 28, 191–199.