Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 2081-2085, April 1986
Botany

Primary structure of the (1→3,1→4)-β-D-glucan 4-glucohydrolase from barley aleurone

(cell-wall degradation/cDNA/codon usage/tissue specificity)

GEOFFREY B. FINCHER*, PETER A. LOCK*, MARGARET M. MORGAN*, KLAUS LINGELBACH*, RICHARD E. H. WETTENHALL*, JULIAN F. B. MERCER†, ANDERS BRANDT‡, AND KARL KRISTIAN THOMSEN‡

*Department of Biochemistry, La Trobe University, Bundoora, Victoria, 3083, Australia; †Birth Defects Units, Royal Children's Hospital, Parkville, Victoria, 3052, Australia; and ‡Department of Physiology, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500, Copenhagen Valby, Denmark

Communicated by Diter von Wettstein, October 28, 1985

ABSTRACT **During germination of barley grains, the cell walls of the starchy endosperm are degraded by (1→3,1→4)-β-glucanases (EC 3.2.1.73) secreted from the aleurone and scutellar tissues. The complete sequence of the aleurone (1→3,1→4)-β-glucanase isoenzyme II comprises 306 amino acids and was determined by sequencing nine tryptic peptides (110 residues) and aligning them with the amino acid sequence deduced from a cDNA clone encoding the 291 NH_2-terminal residues. Although no amino acid sequence homology with a bacterial (1→3)(1→4)-β-glucanase is apparent, close to 50% homology is found with two large regions of a (1→3)-β-glucanase from tobacco pith tissue. The gene for barley (1→3,1→4)-β-glucanase isoenzyme II shares with that for the α-amylase isoenzyme 1 a strongly preferred use of codons with G and C in the wobble position (94% and 90%, respectively). Both enzymes are secreted from the aleurone cells during germination. Such one-sided codon usage is not characteristic for the gene encoding the (1→3)-β-glucanase of tobacco pith tissue or the *hor2-4* gene encoding the B_1 hordein storage protein in the endosperm.**

An important early process in the germination of cereal grains is the degradation of cell walls of the nonliving starchy endosperm. This allows the amylases and proteases secreted from the surrounding aleurone or scutellar tissues to reach their starch and storage-protein substrates inside the endosperm cells. Partly as a result of its importance in the malting and brewing industries, barley has been used extensively in studies of the physiology and enzymology of cereal grain germination. The major constituent of barley endosperm cell walls is (1→3,1→4)-β-D-glucan, which represents a family of long, linear polysaccharides. Amongst the most important enzymes in the depolymerization of these cell walls during germination are the (1→3,1→4)-β-D-glucan 4-glucanohydrolases (EC 3.2.1.73).

Two (1→3,1→4)-β-glucanase isoenzymes have been purified from extracts of germinating barley, and their substrate specificities, chemical properties, and kinetic parameters have been defined (1, 2). Based on the sequence homology of 37 of the 40 NH_2-terminal amino acids, it has been concluded that the two isoenzymes are the products of two genes, which originated by duplication of a common ancestral gene (3). A putative third isoenzyme has been detected in secretions from isolated scutella and aleurone layers (4). There is a marked tissue specificity for the preferential synthesis and secretion of the (1→3,1→4)-β-glucanase isoenzyme II from isolated aleurone layers. Moreover, in the presence of Ca^{2+}, gibberellic acid seems to enhance secretion of isoenzyme II from the aleurone (4).

Recently, mRNA encoding (1→3,1→4)-β-glucanase has been identified in both aleurone and scutellar tissue of germinating barley (5). In the present work, we have isolated a cDNA clone encoding barley (1→3,1→4)-β-glucanase isoenzyme II. Sequence analysis of the cDNA, in conjunction with amino acid sequence analysis of tryptic peptides, have enabled us to deduce the complete amino acid sequence of the enzyme.

MATERIALS AND METHODS

Amino Acid Sequence Analysis. Purified (1→3,1→4)-β-glucanase isoenzyme II (2) at 1 mg/ml in 10 mM Tris·HCl (pH 8.0) was hydrolyzed with L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin (Worthington) at 20 μg/ml for 1 hr and 24 hr at room temperature. The reaction was stopped by the addition of trifluoroacetic acid to 0.2% (vol/vol). Tryptic peptides were fractionated by reversed-phase HPLC on a Bakerbond C_8 column using a gradient of 0–50% acetonitrile in aqueous 0.1% (vol/vol) trifluoroacetic acid, increasing at 0.56% per min at a constant flow rate of 1 ml/min. Fractions were dried and the peptides were dissolved in 50% (vol/vol) trifluoroacetic acid in preparation for sequencing (6). After Edman degradation of purified tryptic fragments, the phenylthiohydantoin derivatives of the amino acids were analyzed by reversed-phase HPLC (7).

Isolation of mRNA. After incubation of half grains from *Hordeum vulgare* cv. Himalaya for 36 hr in the presence of 1 μM gibberellin and 10 mM Ca^{2+}, aleurone tissue was ground under liquid nitrogen, and the RNA isolated as described (8) except that 1% (wt/vol) insoluble polyvinylpyrrolidone (Polyclar AT) was included in the extraction buffer. Poly(A)$^+$ mRNA was isolated by affinity chromatography on poly(U)-Sepharose 4B (Pharmacia) and fractionated by ultracentrifugation in an 8–20% linear sucrose gradient. Gradient fractions enriched in mRNA coding for (1→3,1→4)-β-glucanase were identified by *in vitro* translation and $NaDodSO_4$/polyacrylamide gel electrophoresis (5).

Generation of cDNA Library. The cDNA was synthesized by the RNase H method, including DNA ligase in the reaction mixture, as described by Gubler and Hoffman (9) except that the second-strand synthesis was performed at 12°C for 2–5 hr until its amount was equal to that of the first-strand template. The oligo(dC)-tailed cDNA was annealed to *Pst* I-cut oligo(dG)-tailed plasmid pUC9 (10) (Pharmacia P-L Biochemicals) and cloned by transformation of *Escherichia coli* K-12 strain JM101 by the procedure of Hanahan (11), and recombinants were identified as described (10). The (1→3,1→4)-β-glucanase cDNA clone was selected by colony hybridization and dot blot and Southern blot analyses with a probe consisting of a mixed pentadecameric oligonucleotide.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

The probe was labeled by use of T4 polynucleotide kinase (Pharmacia P-L Biochemicals) and [γ-^{32}P]ATP (5000 Ci/mmol; 1 Ci = 37 GBq). Hybridization was for 16 hr at 25°C in 6× standard saline citrate (1× SSC is 0.15 M NaCl/15 mM sodium citrate, pH 7)/0.2% bovine serum albumin/0.2% Ficoll/0.2% polyvinylpyrrolidone 4000/0.05% sodium pyrophosphate; filters were washed extensively in 6× SSC/0.05% sodium pyrophosphate at 30°C and finally for 10 min at 37°C in the same buffer (12, 13).

Synthesis of Oligonucleotides. The mixed oligonucleotide (15-mer) used to screen the cDNA library was synthesized manually using solid-phase phosphoramidite chemistry (14). The product was purified by reversed-phase HPLC on a Bio-Sil ODS-5S column (Bio-Rad) using a 20–30% acetonitrile gradient in 0.1 M triethylammonium acetate (pH 7.0). Oligonucleotides used as sequencing primers were synthesized on an Applied Biosystems model 380A DNA synthesizer.

Nucleotide Sequence Analysis. The dideoxynucleotide chain-termination procedure of Sanger *et al.* (15) was used to sequence the cDNA clone. Plasmid DNA was prepared (16) and cDNA fragments were subcloned in M13 mp9 (10).

Southern Blot Analysis of Barley DNA. DNA was prepared from immature endosperms of the barley variety Carlsberg II according to Hopp *et al.* (17), digested to completion with the restriction endonucleases *Eco*RI, *Hin*dIII, and *Bam*HI, separated by agarose gel electrophoresis, and blotted onto nitrocellulose filters (BA-85, Schleicher & Schuell) as described by Southern (18). Hybridization was in 2× SSC/0.2% bovine serum albumin/0.2% Ficoll/0.2% polyvinylpyrrolidone 4000/0.1% $NaDodSO_4$ (19) for 18 hr at 65°C in the presence of denatured herring sperm DNA (50 μg/ml) and poly(A) (50 μg/ml). The isolated large *Pst* I fragment of the (1→3,1→4)-β-glucanase clone was nick-translated to a specific activity of 2×10^8 cpm/μg (20) and used as hybridization probe at a concentration of 2.5 ng/ml. Filters were washed extensively at 65°C in 2× SSC/0.1% $NaDodSO_4$, followed by two 30-min washes at 65°C in 0.1× SSC. Kodak XAR-5 x-ray films were exposed at −70°C for 11 days, using Kodak Regular intensifying screens.

RESULTS AND DISCUSSION

The complete amino acid sequence of barley (1→3,1→4)-β-glucanase isoenzyme II (Fig. 1) was obtained in the following way. Amino acid sequences of nine tryptic peptides were determined. A 15-base mixed oligonucleotide probe of 32-fold redundancy, encoding the sequence Phe-Tyr-Asn-Gln-His (tryptic peptide 11, Fig. 1), was used to screen a cDNA library prepared with mRNA from gibberellic acid-treated aleurone layers. A cDNA clone comprising 874 nucleotide pairs [excluding the oligo(dG·dC) tails] was selected with the probe, and the amino acid sequence was deduced from its nucleotide sequence (for cDNA sequencing strategy, see Fig. 2). The cDNA has one open reading frame encoding 291 amino acids on each strand. One of the open reading frames encoded the 40 NH_2-terminal residues of isoenzyme II (3) as well as the sequences of the eight isoenzyme II tryptic peptides (110 amino acids). Since the directly determined and the deduced amino acid sequences are identical, we conclude

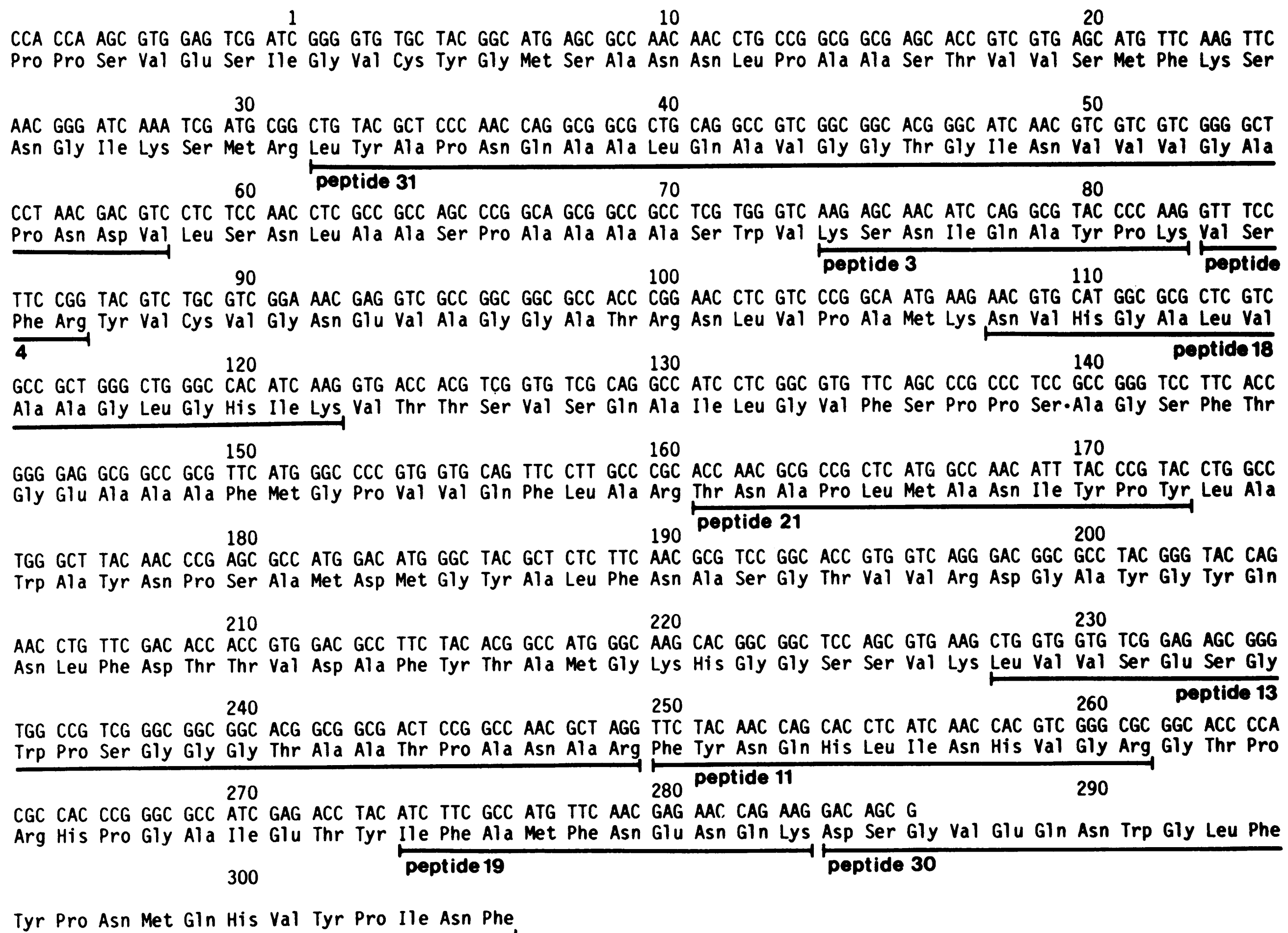

FIG. 1. Partial cDNA and complete amino acid sequences for barley aleurone (1→3,1→4)-β-glucanase isoenzyme II. Sequences corresponding to tryptic peptides are identified.

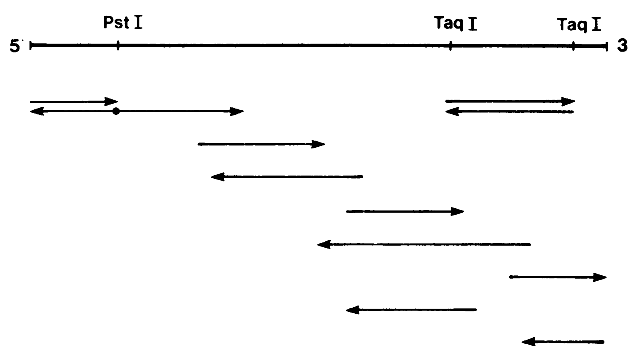

FIG. 2. Strategy for nucleotide sequencing of the cDNA clone encoding 285 of the 306 amino acids of the (1→3,1→4)-β-glucanase, using various oligonucleotide primers. Arrows below the restriction map indicate the length and direction of the DNA strands sequenced.

that the cDNA encodes the (1→3,1→4)-β-glucanase isoenzyme II. The NH_2-terminal residue of the enzyme is at the seventh codon (Fig. 1); the first six amino acids encoded by the cDNA may represent a signal peptide. The 3′ end of the cDNA clone does not extend to the stop codon of the open reading frame. However, the last two codons encode an Asp-Ser dipeptide, a sequence that is also found at the NH_2 terminus of tryptic peptide 30 (Fig. 1). The COOH terminus of isoenzyme II was determined by carboxypeptidase Y digestion to be Val-Tyr-Pro-Ile-Asn-Phe, a sequence located at the COOH terminus of tryptic fragment 30. Since it is likely that the Asp-Ser dipeptide encoded at the 3′ end of the clone corresponds to the NH_2-terminal end of tryptic fragment 30, a complete sequence of 306 amino acids can be assigned to this enzyme. This compares well with both the number and the composition of the 309 amino acids previously reported for (1→3,1→4)-β-glucanase isoenzyme II (2).

Sequence Analysis of Tryptic Peptides. The HPLC elution profile of tryptic peptides of (1→3,1→4)-β-glucanase isoenzyme II after 24-hr tryptic hydrolysis is shown in Fig. 3, and the amino acid sequences obtained from the peptides found in peaks 31, 3, 4, 18, 21, 13, 11, 19, and 30 are indicated in Fig. 1. The initial yields of phenylthiohydantoin derivative from individual tryptic peptides ranged from 50 to 200 pmol. Peptide 30 from the COOH-terminal end was a major peak in the elution profile after 1-hr tryptic hydrolysis (data not shown) rather than the minor component seen in the 24-hr digest. The amino acid sequence of peptide 30 was determined from a 1-hr digest. The decrease in relative abundance of peak 30 with prolonged trypsin digestion was due to secondary, nonideal cleavage between Phe-294 and Tyr-295, as well as between Asn-297 and Met-298. The fragment Met^{298}–Phe^{306} was identified in peak 20, whereas the cleavage products Asp^{284}–Asn^{297} and Tyr^{295}–Phe^{306} were in peak 23. Individual amino acid sequences for the two latter components could be deduced from the yields of phenylthiohydantoin derivatives. Peptide 19 also results from nonideal trypsin digestion, at Tyr-273.

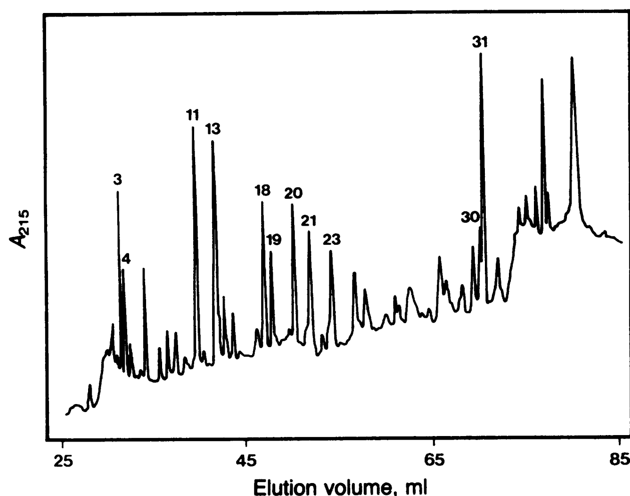

FIG. 3. Separation of tryptic peptides of (1→3,1→4)-β-glucanase isoenzyme II by reversed-phase HPLC after 24-hr digestion. Numbers indicate the fractions from which the peptides were sequenced.

Southern Blot Analysis of Barley DNA. Restriction endonuclease fragments of barley DNA hybridizing to the (1→3,1→4)-β-glucanase cDNA insert are shown in Fig. 4. The restriction endonucleases *Eco*RI, *Hin*dIII, and *Bam*HI each generated two such fragments (14 and 0.9, 4 and 3, and 11 and 7 kilobases, respectively). Since the cloned cDNA lacks restriction sites for the three enzymes used, the hybridizing fragments could correspond to the two genes encoding different (1→3,1→4)-β-glucanase isoenzymes (3). However, the absence of introns containing restriction sites for the three endonucleases remains to be demonstrated.

Sequence Homology Between the Barley (1→3,1→4)-β-Glucanase Isoenzyme II and Other Proteins. The only other (1→3,1→4)-β-glucanase for which the primary structure is known is the *Bacillus subtilis* enzyme (21). No amino acid sequence homology can be detected between the barley and the *Bacillus* enzymes. However, comparison of the amino acid sequence of the (1→3,1→4)-β-glucanase from barley aleurone with that of a (1→3)-β-glucanase (EC 3.2.1.39) from tobacco reveals extensive homologies (Fig. 5). The sequence of the tobacco (1→3)-β-glucanase, an enzyme of 322 amino acids that accumulates in pith tissue cultures in the absence of auxin or cytokinin (22–25), was determined by H. Shinshi (personal communication) from a cDNA clone (26) encoding the enzyme and from tryptic peptides. After introducing four gaps in the (1→3)- and nine gaps in the (1→3,1→4)-β-glucanase sequences, respectively, 45% of the amino acids in the two enzymes occupy identical positions. The region

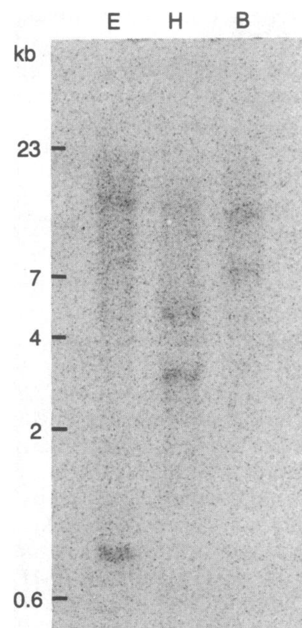

FIG. 4. Hybridization of the ^{32}P-labeled large *Pst* I fragment of the cDNA clone for (1→3,1→4)-β-glucanase to Southern blots of barley endosperm DNA digested with *Eco*RI (lane E), *Hin*dIII (lane H), or *Bam*HI (lane B) reveals the presence of the sequence in two fragments. The probe does not contain restriction sites for the three enzymes. *Hin*dIII restriction fragments of λWES/gt have been used as size markers. kb, Kilobases.

FIG. 5. Amino acid sequence homologies between the (1→3,1→4)-β-glucanase of barley aleurone and the (1→3)-β-glucanase of tobacco cells. Asterisks indicate conserved acidic amino acid residues that may play a role in the catalytic function. Standard one-letter abbreviations for amino acids are used.

Table 1. Codon usage (% for each of 16 amino acids) in the genes for (1→3,1→4)-β-glucanase of barley aleurone, α-amylase 2 of barley aleurone (27), and B_1 hordein of barley endosperm

	Glucanase	Amylase	Hordein		Glucanase	Amylase	Hordein
Leu	[16]	[34]	[23]	Ala	[43]	[47]	[13]
TTA	0	0	0	GCT	14	9	38
CTT	6	6	17	GCA	5	9	46
CTA	0	3	17	GCC	51	52	23
TTG	0	3	22	GCG	30	30	15
CTC	50	47	22	Gly	[31]	[48]	[8]
CTG	44	41	22	GGT	0	6	38
Ser	[23]	[22]	[12]	GGA	3	8	13
TCT	0	5	17	GGC	68	65	38
TCA	0	0	0	GGG	29	21	13
AGT	0	0	25	Tyr	[13]	[16]	[10]
AGC	44	32	17	TAT	0	0	60
TCC	28	45	17	TAC	100	100	40
TCG	28	18	25	His	[6]	[13]	[3]
Arg	[8]	[17]	[6]	CAT	17	8	67
CGT	0	0	33	CAC	83	92	33
CGA	0	0	17	Gln	[8]	[12]	[73]
AGA	0	0	17	CAA	0	17	74
AGG	24	24	17	CAG	100	83	26
CGC	38	52	17	Asn	[21]	[18]	[4]
CGG	38	24	0	AAT	0	11	75
Val	[28]	[28]	[19]	AAC	100	89	25
GTT	4	0	26	Lys	[9]	[23]	[2]
GTA	0	0	11	AAA	11	9	0
GTC	48	43	36	AAG	89	91	100
GTG	48	57	42	Asp	[6]	[37]	[1]
Pro	[16]	[17]	[42]	GAT	0	11	0
CCT	6	6	2	GAC	100	89	100
CCA	17	18	76	Glu	[5]	[13]	[6]
CCC	22	35	16	GAA	0	15	50
CCG	56	41	5	GAG	100	85	50
Thr	[15]	[17]	[10]	Cys	[2]	[8]	[9]
ACT	7	0	10	TGT	0	12	44
ACA	0	18	30	TGC	100	88	56
ACC	63	59	20				
ACG	30	23	40				

Number of residues of each amino acid in each protein is given in brackets.

spanning positions 53–101 shows a low degree of homology (≈20%), but the NH_2- and COOH-terminal segments reach homology values of 50%.

Preferential Codon Usage in the Gene for (1→3,1→4)-β-Glucanase Isoenzyme II. In Table 1, we have compared the codon usage for amino acids 1–285 of the (1→3,1→4)-β-glucanase isoenzyme II with that for the 438 amino acids of barley α-amylase isoenzyme 1 (27) and the 271 amino acids of the B_1 hordein encoded by the *hor2-4* gene (33). These three coding regions have G+C contents of 67%, 64%, and 47%, respectively. Rogers and Milliman (27) and Chandler *et al.* (28) have observed a strong preference for C and G in the third position of degenerate codons used for barley α-amylase isoenzymes. The marked bias toward C and G in the wobble position is also found in the (1→3,1→4)-β-glucanase isoenzyme II (Table 1). This is in sharp contrast to the generally nonpreferential codon usage in genes for proteins that are synthesized in the developing endosperm, as exemplified by the B_1 hordein storage protein of barley. In the latter gene, a preferred codon usage is seen only for the two predominant amino acids proline (CCA) and glutamine (CAA). Preference in this case is for A in the third position, which balances the overall G+C content at 47% in spite of the predominance of proline and glutamine residues in the protein. Protein Z, another protein from the barley endosperm with a "balanced" amino acid composition, has a G+C content of 49% and nonpreferential codon usage (29). The genes for sucrose synthase (30) and zein storage proteins (31, 32) expressed in developing maize endosperm show a very limited codon preference.

Thus, only the two enzymes α-amylase and (1→3,1→4)-β-glucanase, which are synthesized during germination in aleurone cells of barley, show the strongly preferred use of G and C in the wobble position (90% and 94%, respectively). It will be interesting to determine whether or not this pronounced preferential codon usage is characteristic of genes for other enzymes expressed in aleurone cells of barley and whether the genes for the isoenzymes expressed in scutellar tissue also have this property.

We are indebted to Dr. Hideaki Shinshi (Friedrich Miescher Institut, Basel) for communicating to us the amino acid sequence for the tobacco (1→3)-β-glucanase. We thank Dr. Liz Jackson, Dr. Nick Hoogenraad, Ms. Lynda Graf, Ms. Kathy Auddino, and Dr. Ib Svendsen for their invaluable assistance with various aspects of this project. The work was supported by grants from the Australian Research Grants Scheme (to G.B.F.) and by Grant 1982-133/001-82.112 from Teknologistyrelsen of the Danish Ministry of Industry to Professor Diter von Wettstein.

1. Woodward, J. R. & Fincher, G. B. (1982) *Carbohydr. Res.* **106**, 111–122.
2. Woodward, J. R. & Fincher, G. B. (1982) *Eur. J. Biochem.* **121**, 663–669.

3. Woodward, J. R., Morgan, F. J. & Fincher, G. B. (1982) *FEBS Lett.* **138**, 198–200.
4. Stuart, I. M., Loi, L. & Fincher, G. B. (1985) *Plant Physiol.* **79**, 867–871.
5. Mundy, J., Brandt, A. & Fincher, G. B. (1985) *Plant Physiol.*, in press.
6. Nick, H. P., Wettenhall, R. E. H., Hearn, M. T. W. & Morgan, F. J. (1985) *Anal. Biochem.* **147**, in press.
7. Zimmermann, C. L., Apella, E. & Pisano, J. J. (1977) *Anal. Biochem.* **77**, 569–573.
8. Brandt, A. & Ingversen, J. (1978) *Carlsberg Res. Commun.* **43**, 451–469.
9. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
10. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
11. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580.
12. Meinkoth, J. & Wahl, G. (1984) *Anal. Biochem.* **138**, 267–284.
13. Suggs, S. V., Wallace, R. B., Hirose, T., Kiwashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6613–6617.
14. Matteucci, M. D. & Caruthers, M. H. (1981) *J. Am. Chem. Soc.* **103**, 3185–3191.
15. Sanger, F., Nicklen, S. & Coulson, A. K. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
17. Hopp, H. E., Rasmussen, S. K. & Brandt, A. (1983) *Carlsberg Res. Commun.* **48**, 201–216.
18. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
19. Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641–644.
20. Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
21. Murphy, N., McConnell, D. J. & Cantwell, B. A. (1984) *Nucleic Acids Res.* **12**, 5355–5367.
22. Eichholz, R., Harper, J., Felix, G. & Meins, F., Jr. (1983) *Planta* **158**, 410–415.
23. Felix, G. & Meins, F., Jr. (1985) *Planta* **164**, 423–428.
24. Shinshi, H. & Kato, K. (1983) *Agric. Biol. Chem.* **47**, 1455–1460.
25. Shinshi, H. & Kato, K. (1983) *Agric. Biol. Chem.* **47**, 1275–1280.
26. Mohnen, D., Shinshi, H., Felix, G. & Meins, F., Jr. (1985) *EMBO J.* **4**, 1631–1635.
27. Rogers, J. C. & Milliman, C. (1983) *J. Biol. Chem.* **258**, 8169–8174.
28. Chandler, P. M., Zwar, J. A., Jacobsen, J. V., Higgins, T. J. V. & Inglis, A. S. (1984) *Plant Mol. Biol.* **3**, 407–418.
29. Hejgaard, J., Rasmussen, S. K., Brandt, A. & Svendsen, I. (1985) *FEBS Lett.* **180**, 89–94.
30. Werr, W., Frommer, W.-B., Maas, C. & Starlinger, P. (1985) *EMBO J.* **4**, 1373–1380.
31. Hu, N.-T., Pfeifer, M. A., Heidecker, G., Messing, J. & Rubinstein, I. (1982) *EMBO J.* **1**, 1337–1342.
32. Pedersen, K., Devereux, J., Wilson, D. R., Sheldon, E. & Larkins, B. A. (1982) *Cell* **29**, 1015–1026.
33. Brandt, A., Montembault, A., Cameron-Mills, V. & Rasmussen, S. K. (1985) *Carlsberg Res. Commun.* **50**, 333–345.