A single gene encodes two different transcripts for the ADP-glucose pyrophosphorylase small subunit from barley (*Hordeum vulgare*)

Tine THORBJØRNSSEN*, Per VILLAND*,‡, Leszek A. KLECZKOWSKI*† and Odd-Arne OLSEN*

*Plant Molecular Biology Laboratory, Agricultural University of Norway, P.O. Box 5051, 1432 Ås, Norway, and †Department of Plant Physiology, Umeå University, 901-87 Umeå, Sweden

ADP-glucose pyrophosphorylase (AGPase), a heterotetrameric enzyme composed of two small and two large subunits, catalyses the first committed step of starch synthesis in plant tissues. In an attempt to learn more about the organization and expression of the small-subunit gene of AGPase, we have studied the small-subunit transcripts as well as the structure of the gene encoding these transcripts in barley (*Hordeum vulgare* L. cv. Bomi). Two different transcripts (bepsF1 and blps14) were identified: bepsF1 was abundantly expressed in the starchy endosperm but not in leaves, whereas blps14 was isolated from leaves but was also found to be present at a moderate level in the starchy endosperm. The sequences for the two transcripts are identical over approx. 90% of the length, with differences being confined solely to their 5' ends. In blps14, the unique 5' end is 259 nt long and encodes

a putative plastid transit peptide sequence. For the 178-nt 5' end of bepsF1, on the other hand, no transit peptide sequence could be recognized. A lambda clone that hybridized to the AGPase transcripts was isolated from a barley genomic library and characterized. The restriction map has suggested a complex organization of the gene, with alternative exons encoding the different 5' ends of the two transcripts followed by nine exons coding for the common part of the transcripts. The sequence of a portion of the genomic clone, covering the alternative 5'-end exons as well as upstream regions, has verified that both transcripts are encoded by the gene. The results suggest that the small-subunit gene of barley AGPase transcribes two different mRNAs by a mechanism classified as alternative splicing.

INTRODUCTION

In higher plants, ADP-glucose pyrophosphorylase (AGPase; EC 2.7.7.27), a heterotetrameric enzyme composed of two small and two large subunits, catalyses the conversion of glucose 1-phosphate and ATP to ADP-glucose and PP_i. The reaction is the first committed step of starch synthesis in plants and is believed to be important in regulating starch accumulation [1,2]. AGPase from barley leaves [3], as well as from most other plant tissues examined [1,2], is regulated by 3-phosphoglyceric acid and inorganic phosphate. By contrast, the enzyme from barley seed starchy endosperm seems to be insensitive to these effectors [3,4]. The different properties of AGPase from leaves and grains of barley suggest that different genes encode the enzyme in these two organs.

Expression of AGPase genes in barley tissues was previously examined by PCR experiments, using oligonucleotide primers that recognized transcripts encoding both large and small subunits of the enzyme [5]. This analysis, supported also by genomic Southern blot data [6], indicated at least two different genes for the large subunit of AGPase, but only one for the small subunit. So far, single transcripts for the small-subunit gene have been reported for potato [7], rice [8], wheat [9], and *Arabidopsis* [10]. On the other hand, for maize [11] and *Vicia faba* [12] two distinct transcripts for this subunit have been characterized. Several cDNAs have also been described for the large subunit of AGPase from various plants and tissues (see, for example, [1,2,5,10,12,13]). The small-subunit genes are very conserved between tissues and species, but those encoding the large subunit have diverged so that they may be divided into two groups,

represented by genes expressed in either photosynthetic or non-photosynthetic tissues [13].

In the present study we report the characterization of a barley AGPase small-subunit (AGPase B) gene that is expressed in both leaves and starchy endosperm. The gene seems to be differentially expressed in the two tissues, via an alternative splicing mechanism not yet described for plants, to yield transcripts of distinct coding capacities. Possible implications for targeting of the AGPase protein to plastids, and thus the spatial organization of the starch synthetic pathway in barley tissues, are discussed.

MATERIALS AND METHODS

Recombinant DNA techniques

Standard methods for DNA analysis and manipulation were used [14]. Plasmid and M13 templates were sequenced with the Sequenase kit (Amersham), and the sequences were analysed with the Sequence Analysis Software Package of the University of Wisconsin Genetics Computer Group [15] at the EMBnet node in Oslo.

cDNA cloning by PCR

The RACE (rapid amplification of cDNA ends) technique was used for the isolation of cDNA clones from starchy endosperm. Poly(A)⁺ RNA was extracted directly from 100 mg of barley (cv. Bomi) starchy endosperm (20 days after pollination) by using magnetic oligo(dT) microspheres (Dynal) as described by Jakobsen et al. [16]. For details on the histology of the barley

Abbreviations used: AGPase, ADP-glucose pyrophosphorylase; AGPase B, ADP-glucose pyrophosphorylase small subunit.

‡ To whom correspondence should be addressed at Department of Plant Physiology, Umeå University, 901-87 Umeå, Sweden.

bepsF1

```

1  AAAGTGAACACACATCACTCAATATCTATATCCTTCCATTTATATCCCTCGGTGATGGATGTACCTTGGCA
1  M D V P L A
71 TCTAAAGTTCCCTTGCCCTCCCTTCCAAAGCATGAACAATGCAACGTTTATAGTCATAAGAGCTCATCGAAGCAT
71 S K V P L P S P S K H E Q C N V Y S H K S S S K H
141 GCAGATCTCAATCCCATGCTATTGATAGTGTTCGCT....
32 A D L N P H A I D S V L G ....

```

blps14

```

1  CTCGTATCCACCTCAATGGCGATGGCCGGCGCGCTCCCTTCCAAAGATCCTGATCCCTCCGACCGAGCCTC
1  M A M A A A A S P S K I L I P P H R A S
76 CGCGGTGACCGCTGCCGCGTCCACCTCTGCGACTCCCTCCGCTCCTTGCGCGCCACGAGGACGGCCAGGCC
26 A V T A A A S T S C D S L R L L C A P R G R P G
151 CGCGCGGTGGTCCGCGTCCGCTTCCGCGCGCGCTTCTTCTTCTCCCGACGTCGCGTGTGACAGTCCAAAG
51 R G L V A R P V P R R P F F F S P R A V S D S K S
226 CTCCAGACTTGTCTCGACCCGACGCAAGCAGCAGTGTTCGCTATCATTCTTGGAGGTGGTGCAGGGACTAG
76 S Q T C L D P D A S T S V L G I I L G G G A G T R
301 ATTGTATCCCTGACGAAGAAGCGTGCAAGCCTGCAGTGCCTGGTGGCAACTACAGGCTTATTGATATTC
101 L Y P L T K K R A K P A V P L G A N Y R L I D I P
376 TGTCAATAATTGTCTGAACAGCAACATATCAAAGATCTATGTCTTACACAGTTCAACTCAGCTTCTCTTAATCG
126 V S N C L N S N I S K I Y V L T Q F N S A S L N R
451 TCATCTCTCAGAGCCTATGGGAGCAACATTGGAGGTACAAGAATGAAGGATTGTGAAGTCTTGTGTCACA
151 H L S R A Y G S N I G G Y K N E G F V E V L A A Q
526 GCAGAGCCAGATAACCTGACTGGTCCAGGGTACTGCAGATGCTGTAAGGCAGTACTTGTGGCTATTCGAGGA
176 Q S P D N P D W F Q G T A D A V R Q Y L W L F E
601 GCATAATGTTATGGAGTATCTAATTCTTGTCTGGAGTACCTGTACCGAATGGACTATGAAAAGTTTATTCAGGC
201 H N V M E Y L I L A G D H L Y R M D Y E K F I Q A
676 ACACAGAGAAACGGATGTGATATTACTGTTGCTGCCTTGCCCATGGATGAGGAACGTGCAACTGCATTTGGCCT
226 H R E T D A D I T V A A L P M D E E R A T A F G L
751 TATGAAATCGATGAAGAAGGGAGGATAATTGAATTCGAGAGAAACAAAAGGAGAACAGTTGAAAGCTATGAT
251 M K I D E E G R I I E F A E K P K G E Q L K A M M
826 GGTGATACGACCATACTTGGCCTTGAAGATGCGAGGGCAAAGGAAATGCCTTATATTGCTAGCATGGGTATCTA
276 V D T T I L G L E D A R A K E M P Y I A S M G I Y
901 TGTATTAGCAACATGTGATGCTTCACTTCTCCGTGAGCAATTCCTGGAGCTAATGACTTCGGAAGTGAAGT
301 V I S K H V M L Q L L R E Q F P G A N D F G S E V
976 TATTCCTGGTGAACATAGCACTGGCATGAGGTACAAGCATACCTATACGACGTTACTGGGAAGATATTGGTAC
326 I P G A T S T G M R V Q A Y L Y D G Y W E D I G T
1051 AATTGAGGCATTCTATAATGCAAAATTTGGGAATTACCAAAAAACCAATACCTGATTTTCAGTTTCTATGACCGTTC
351 I E A F Y N A N L G I T K K P I P D F S F Y D R S
1126 TGCTCCCATTTACACACAACCTCGACACTTGCCCTCCTTCAAAGGTTCTTGATGCTGATGTGACAGACAGTGAAT
376 A P I Y T Q P R H L P P S K V L D A D V T D S V I
1201 TGGTGAAGGATGTGTTATTAAGAACTGCAAGATACACCATTCACTAGTTGGACTCCGTTCTGCTATCTGAAGG
401 G E G C V I K N C K I H H S V V G L R S C I S E G
1276 TCGAATAATAGAGGACACGTTGCTAATGGGTGCGGACTACTATGAGACTGAAGCTGATAAGAACTCCTTGCTGA
426 A I I E D T L L M G A D Y Y E T E A D K K L L A E
1351 AAAAGGTGGCATTCCCATTTGGTATTGGAAAGAATTCACACATCAAAGAGCAATATTGACAAGAATGCTCGTAT
451 K G G I P I G I G K N S H I K R A I I D K N A R I
1426 TGGAGATAACGTGATGATAATCAATGTTGACAATGTTCAAGAAGCGGCGAGGGAGACAGATGGATACTTCATCA
476 G D N V M I I N V D N V Q E A A R E T D G Y F I K
1501 AAGTGGCATCGTAACTGTGATCAAGGATGCTTTACTCCCTAGTGGAAACAGTCATATGAAGCAGATGTGAAATGTA
501 S G I V T V I K D A L L P S G T V I *
1576 TGCCAAAAGACAGGGCTACTTGCGTCAGTCTGGAATCAACCAACAAGGCCGCGAAGAGATCATAAAAAAAGG
1651 AGTGCCATGCGAGTCACTTCTACACCTTTTCCCGCTTGATGATTAGGAAGTGTGATGTAAGCAACTGTGA
1726 TGCCTTACGCGAAGTGCCTTGGATTGAGCTTCTCTTCTGCTTGTAACTGGTTCCAGCAGACCATGCTATTG
1801 TTGTATGGTTCGTGCAAAACCTTCCGATGCTTTATATATGCTTTATATATAAAACAAGATGAATCCCGCGCTG
1876 CTGCGGCACAAAAA

```

Figure 1 Nucleotide sequences and deduced amino acid sequences for the AGPase B cDNA clones isolated from barley leaves (blps14) and endosperm (bepsF1; only the unique 5' end is shown)

From nt 260 in blps14 and nt 178 in bepsF1, the sequences of the two cDNAs are identical. The unique regions are in bold letters.

endosperm, see [17]. One-fifth of the RNA preparation was used for cDNA synthesis by the method of Huang et al. [18], with an AGPase-specific primer as previously described [5]. A poly(dG) tail was introduced onto the 3' end of the first-strand cDNA by the method of Loh et al. [19]. An AGPase B-specific primer (GCCAGTGTAGTTGCACC), based on a partial cDNA sequence [5], and an oligo(dC) primer [CCGGAATT(C)₁₆] were

used for PCR amplification. The reaction mixture (50 µl) contained cDNA, 100 ng each of the primers, 50 mM each of dNTP, and 2 units of Taq DNA polymerase in a standard reaction buffer. Denaturation of DNA was at 96 °C for 0.5 min, annealing at 50 °C for 1 min, and extension at 72 °C for 2 min. After 40 cycles of amplification, a final extension at 72 °C was performed for 10 min. The PCR product was blunt-ended with T4 DNA

polymerase and cloned into the *EcoRV* site of pBluescript (Stratagene) for sequence analysis. To amplify the full-length transcript, a new primer was prepared for the 5' end of the cDNA (AACTCACACATCACTCAATATCT) and used together with an oligo(dT) primer [CGCGATCC(T)₁₇]. Conditions for PCR were as above, except that the extension time per cycle was changed to 5 min. The final PCR product was inserted into pBluescript and the sequences of independent clones were determined. Two clones, designated bepsF1 and bepsF2, represented insertion of the PCR product in opposite directions in the vector.

Isolation of cDNA and genomic clones from barley libraries

From a custom-made (Clontech) lambda ZAPII cDNA library constructed by using poly(A)⁺ RNA of barley (cv. Bomi) leaves, 14 clones hybridizing to bepsF1 were isolated. Analysis by restriction cleavage and partial sequencing indicated that nine of the clones contained putative full-length transcripts. The other five seemed to represent truncated or composite clones. A putative full-length clone, blps14, was chosen for complete sequencing.

Seventeen clones hybridizing to beps, a barley AGPase B cDNA fragment [5], were isolated from a lambda EMBL3 genomic library containing DNA from barley (cv. Bomi). The clones were reprobated with fragments of the cDNAs described in this paper. Three clones seemed to contain sequences hybridizing to the entire cDNAs. Of these, the clone designated PS11 was chosen for further analysis.

Isolation and analysis of RNA

Poly(A)⁺ RNA was extracted from barley (cv. Bomi) leaves 2–3 weeks old and from endosperm 15 days after pollination by the method of Jakobsen et al. [16]. Northern blot analysis was performed by separating 6 µg of poly(A)⁺ RNA in 1.0% (w/v) agarose, transferring to a nylon filter and hybridizing at high stringency by the method of Galau et al. [20]. Transcript-specific probes were made by isolating the 150-nt and 230-nt fragments upstream of a *SacI* site in bepsF1 and blps14, respectively.

RESULTS

Presence of different transcripts for AGPase B in barley endosperm and leaves

For the starchy endosperm, a probably full-length cDNA was amplified from mRNA by using a RACE (rapid amplification of cDNA ends) technique. Sequence information for the design of the specific oligonucleotide primer was obtained from a partial AGPase B cDNA clone (beps) previously described [5]. Sequence analysis of the 1820-nt bepsF1 clone (Figure 1) revealed a 1416-nt open reading frame corresponding to a protein with 472 amino acid residues. A calculated molecular mass of 52.0 kDa agrees well with the 51 kDa estimated for the AGPase B protein from barley endosperm [4]. The cDNA sequence of bepsF1 is 80% and 90% identical with the corresponding cDNAs of rice [8] and wheat [9], respectively.

Corresponding AGPase B clones were isolated from leaves by screening a cDNA library, with the bepsF1 clone as a hybridization probe, resulting in nine clones of similar length and with identical restriction patterns. One of these clones, blps14 (Figure 1), is 1903 nt in length and has an open reading frame encoding a 518-residue protein with a molecular mass of 55.9 kDa. The corresponding mature AGPase B protein extracted from barley leaves has an estimated molecular mass of 51 kDa [3], indicating a possible processing of the protein.

Figure 2 Northern blot analysis of RNA from barley leaves and endosperm for the bepsF1 and blps14 cDNAs

Poly(A)⁺ RNA (6 µg) from starchy endosperm 15 days after pollination (E) and leaves 2–3 weeks old (L) was separated by gel electrophoresis, transferred to a nylon membrane and probed with radioactively labelled DNA from the unique parts of bepsF1 (bepsF1 5') and blps14 (blps14 5') as well as from the full-length bepsF1. For the bepsF1 5', blps14 5' and bepsF1 probes, the autoradiogram was exposed for 1 h, 3 days and 1 day, respectively. Up to 10 days of exposure of the radiogram did not reveal any signal for the bepsF1 5' probe in leaf RNA.

Surprisingly, when the sequences of the bepsF1 and blps14 cDNAs were compared, they were found to be identical over 90% of their lengths, with the differences confined solely to the 5' ends of the transcripts. The identical regions cover 1296 nt of coding sequence as well as 350 nt of 3' untranslated sequence. This explains why we could detect only one AGPase B transcript when we used PCR primers hybridizing to this region [5]. In contrast, a significant level of identity could not be detected between the 178-nt 5' sequence of bepsF1 and the 259-nt 5' end of blps14. The putative start codons for both transcripts are located within this region, leading to deduced amino acid sequences with different N-termini.

The deduced N-terminal amino acid sequences were analysed for the presence of transit peptide motifs characteristic of plastid-targeted nuclear-encoded proteins [21,22]. For the amino acid sequence of blps14, all the criteria for a transit peptide are met. The sequence VTAA tentatively locates the cleavage site between residues 24 and 25. If this site were active *in vivo*, the processed polypeptide would be expected to have a molecular mass of 53.6 kDa, which is close to the molecular mass of approx. 51 kDa reported for the mature AGPase B of barley leaves [3]. For the endosperm sequence, however, no transit peptide sequence could be identified. Again, on the basis of the rules of Bairoch [22], the criteria for a transit peptide are not met owing to the shortage of G, A, S, T and C residues in the first 15 residues of the sequence, as well as the lack of a consensus cleavage site (I/V-X-A/C-↓-A).

Expression of bepsF1 and blps14

The steady-state level of the two AGPase B transcripts was analysed by Northern blot analysis of poly(A)⁺ RNA extracted from leaves and starchy endosperm. To discriminate between the bepsF1 and blps14 transcripts, the unique 5' regions of the two cDNAs were used as probes. The analysis (Figure 2) shows that the bepsF1 transcript is present at a high level in the starchy

Figure 3 Restriction map for the lambda clone PS11 (14.7 kb) containing the AGPase B gene

Exons are indicated by boxes. The location of exons 2–10 are based on restriction analysis. An open box represents the first bepsF1 exon, a shaded box the first blps14 exon and filled boxes represent common exons for the two transcripts.

```

-240  TGT TTT TGT GTG AATA AACT GTT GCG CAA TAA AGCG AAG AGC ATAT GTAG TAC GCC AAA AACT TTA CAG CTG TCA CAT GCG AACT AAT TCG TCG CAC ATG GATA TCA TGT GCT C
-120  TTT TTT GTA CGT GCAT ATA CTTC CTTC GCG C TATA AATA AAG AAG GTT CTCT ATG AC TTT CAAA AGT GAA CT CAC ACA TCA CTCA ATA TCT ATA TCTCT TCC ATT TAT AT CCCC TCG GT G
1      ATGG ATG TAC CTT TGG CAT CTA AAG TTC CTT GCG C TCC CTT CCA AGC ATG AAC AAT GCA ACG GTT ATAG TCATA AGA GCT CAT CGA AGC ATG CAG ATCT CAAT CCCC ATG CTATT GA T
      M D V P L A S K V P L P S P S K H E Q C N V Y S H K S S S K H A D L N P H A I D
121   G T A A G T G G T G C T A T C T T A A C T A T G A T T T C G T T T C T G T T C C A T C T T G A G T A T C A T A T G A G T A A T A T T A T T T T A G G A A G T C T T A G G A A A G C T T C T T G G G G A G C T C A A G C A T A A
241   T T A A A G A C A C T C C A G A G C A C A T C A C A T G C A T G C A T A T A C A C A C A C A C A C A T G A A G T G A G C G A T A T C T T T T A G T T T T C G A A C T C C A A T T T T T T C T C T C A A T A A A A A
361   T A C G A T T A A A A T C C A T T C C A T C A T A A A T C T G C T T C C A C A G A T T T T A A A C T A C A T C G C A T T T G A A G T T T C A A C G A T A T T T T T G G C C A A A A G T T A T C A T G A T G C T T A C C C T
481   C A A G T T A A C A T A G T G T T A C T A C T A A A G T T G C A T G A T A T G T T T A T A T A T A A T T T T C A A G A T T T A A A G C T A C C A T G G A A C A T G A C A A A T T T A A G T A A T C A C C A T G A C A A T T T T A A T
601   T T A T G G T T C A C G T C A A A A T A A G T C A T T G A C T A T G C A T T T T A A G T A A T T T A A C G A C A T A T T T G T T T T A G T T A T G A A C C G T G C C A A A A T A T T T C A T G G A T C A T G G C A A T T T T C A T A
721   A T T C A C C A T G A T A A G T T T A G T T C T A T T T T T A T A A C A T G A C A A A A T A C T T T A A A T G T A G A A A A A A A T T G T T A A A T A T A T C A T G A A C T T T A G T G T A A A A A C A A T T T A T G T G
841   C A C A A A C A T G A C A G C T T T A A C C C A A A A G C C A T T G A A C A T A T G A A A T C A G T T T C A G A A A T C G T T G T G A C G A A T T A A T G A T G A A A A T A T G T T T C A A T C G G A T T T T A A T T T A
961   A A A G A T A A A T C A T T T T A A A A G C T A A A T T T T A A A A A A A T A T T G C A T C A T A G T T T C A T C A T T A T G C A T A C A T G C A T G A T A T A G T G T G A T G T G C G G A T A A T C G T T C G G T C G A T
1081  G A T C T C C C C A T T G A A T G T G T G A A C G C A T C A G A G T C T T A T C T T C C A C T G T T C T A T C T A T A T A T A C G T T T T T T A A A A T A T T G T C T T A A A T T A T C T A G C T A C A A A T A T A T C T A A
1201  C C T T A A A C A C G A T T A G A T A C A T C C G T T A G A T A A A T C C A T G A T A A T T T T T A A G A C G G A G A T A T G T G C T A T G C C T G T C A A G C A C A A A A T C T G A A A A G A C A A T A A G A G A A A C A A G
1321  T T A C C A T T G A T A T T C C A A A A T C A T C G T G C T A T G T A C T C T C A A G G A G C A T T C C A A A A G T G C G C T G C T C A C G A T C T C T C A C T C C C A C G C A G C T C T C T A A A A G A A A A T G G C A
1441  A A A A C T G A A A T G G A A A G A T C T T A C G A A A G A T T A G T T A A A T T A C T C A G C C A C A T G C A C C A C T C G G T G T C A G G C G T A T C T C T C C C T T A C C C C T G T G A T C T C T C C C A C G G A G C
1561  C C C G T G A C T C G A G C T C G T C A T C C A C C T C A A T G G C A T G G C C G G C G C G C T C C C C T T C C A A G A T C C T G A T C C C T C C G C A C C G C C G C A C C G C C G C C G T G A C C G C T G C C G C G T C A C C T C C T G C G
      M A M A A A A S P S K I L I P P H R A S A V T A A A S T S C D
1681  A C T C C C T C C G C C T C T G C G C C C A C G A G A C G G C C A G C C G C G G G T T G T G C G C G T C C G G T T C C G C G C G C C C T T C T C T C T C C C A C G T G C G T C A G A C T C C A A G A G C T
      S L R L L C A P R G R P G R P L V A R P V R P R P F F F S R P R A V S D S K S S
1801  C C C A G A C T T G T C T C G A C C C G A C G C A A G C A C G T A C G C C C C T C G C T A G C C A A T G C G C G C T T C T G C C G C T A G T C T T G T C T C G C T C C C T G A T C G T T C G T C C G T A T C T T T C C G
      Q T C L D P D A S T
1921  G A T G A G A A T T G A C A C A T G C G G G A A G T A T T G C C T C G G T A A T T T A G A T G C G C A A A T G T G G T T C G C G T C T T G T G T T C A T G T G G A C A T T T C T T A G A G A T G A T A C A A A A A A A T A C T A C T C T
2041  A C T T T G C T A C T A G C G A T A G C A T G A C T T T T G A C T A A T C A T G T G G A C A T T T C A T T T G C T C C A A T T A T T A T T T G T A C T A G A T T C A G T A A A T G T G G C A A C T G T G G C G C T T T G T G T A C A
2161  G T C C T A C T T A A T C A A T T G G T G T G G G T C T G A C A C G T G A G C C C C A T A A G A A A T T A T A A A A G A A G T G G C A T G A T C C A T G A C A T G T G G C T C T T A A G A C C A C T G G G C A A T C A G A T C A A G T T
2281  C T T G G A T T T A T T T G T A A C T A T T C A G T T T T C T T T G A G T T T T G C T T C A G T A C C C C T T A T A A A A A C A T T A C G A T T T T G G A T G T G G T G G A C A T T A A A A C T A C C C T T T C A T T T T A A
2401  T T A A A G G G T A G T A G A G T C T A T T C C A C G T G T G C A A A T C A A T G G T G G T T G C C T C T T C C T C A A A C C G C T T G G C A T C A A C A C A A C T A A C A A A G T A A T A C A C A C G C G A G T T T A G
2521  G C G A G A T A A T A G A T T G G G T C T T C T G T C T G T G G T G A A G A C T C A C G T G A T G T G A A A A A G T T A T A C A C A A C A G G A A A C T T G G A T C A C A T C T C A G C T G C A A T G C T G A T T C A G A C A T C T G A C
2641  G T C C A A T T A A A C C C A T A T T C G G A A A A A A A C T A A A C A T A A G T T C C A G C T T G A T T G A T A A A A A A A G G T C A G A A C T A T T C A C C A G C C A G A T C C C A G T A T C T A C A A A G A T T G G T A G A C A
2761  T T G A G C T T C A G T T A T T G G A T A A A C T T G T G C C C A T G T C A C A T T C A T C A T G A T C T T T T G T G A T A A A C A A G C T G A A C T C A G T G C C A C C G T C T G A A C A C T T G T T C T C T G T T C
2881  T T T G T T T G A T T A C T T T G C A A T A G G C A A T T G A T G A G T T T T G C T G T T T G T G C A G T G T C C G G A T T C T T G G A G T G T G C A G G A C T A G A T T G T A T C C C C T G A C G A A G A A G C G T
      S V L G I I L G G G A G G T R L Y P L T K K R
3001  G C A A A G C C T G C A
      A K P A

```

Figure 4 Nucleotide sequence for part of the AGPase B gene

A portion of upstream sequence, the two alternative 5' exons (1a and 1b), intron sequences and a part of the first common exon (2) are shown. The deduced amino acid sequences for the exons are shown beneath the nucleotide sequence. A tentative TATA box is underlined; GT and AG residues in the boundaries of exons are in bold; the sites corresponding to the 5' ends of the transcripts are marked with an arrowhead.

endosperm but not in leaves. A prolonged exposure of the autoradiogram (10 days) did not reveal any signal for the bepsF1-specific probe in leaf RNA. In contrast, the blps14 transcript was found to be expressed in both tissues although the steady-state level was considerably lower in both tissues than for the bepsF1 transcript in the endosperm. Using a probe which covers the common part of the AGPase B transcripts as a standard (Figure 2), we estimated bepsF1 to be 20-fold more abundant than blps14 in the starchy endosperm.

Isolation and structure analysis of AGPase B gene

Using a sequence from the common region of blps14 and bepsF1 as probe, we isolated 15 positive lambda clones from a barley

genomic library. Further hybridization of these clones to probes from the unique 5' ends of blps14 and bepsF1 cDNAs identified three clones that hybridized strongly to both probes. In Figure 3 we present a restriction map prepared for the clone PS11, which was chosen for further analysis. During the initial characterization by Southern blot analysis, sequencing of selected intron/exon borders and by PCR with exon-specific primers, we tentatively identified a total of 11 exons. Of these, the sequence labelled 1a, which hybridized to the unique 5' end of the bepsF1 transcript, is located in the 5' end of the coding region. Similarly, the unique part of blps14 hybridized to the sequence labelled 1b, which is located approx. 1.5 kb further downstream. The common sequences of the AGPase B transcripts are found to be located on nine putative exons separated by introns. The coding

region of the putative gene covers approx. 10 kb of genomic DNA sequence. To verify the suggested sequence identity between the two transcripts and the isolated gene, 6 kb of the proximal part of the gene were sequenced. The results show that the proposed exons 1a and 1b match perfectly with the unique 5' ends of the bepsF1 and the blps14 transcripts, respectively. Similarly, exon 2 is identical with the sequence of the first part of the common region of the two transcripts, suggesting strongly that the two AGPase B transcripts are encoded by this gene. A portion of the gene sequence with potentially important elements identified is shown in Figure 4. These sequence elements include GT residues at the 3' boundary of the proposed exons 1a and 1b and AG residues at the 5' boundary of exon 2. A TATA-box consensus sequence is underlined upstream of exon 1a. The sites corresponding to the 5' end of the transcripts are marked with arrowheads in Figure 4.

DISCUSSION

Distinct 5' ends of the two barley transcripts correspond to two divergent groups of AGPase B amino termini

In this paper we report the isolation of two different cDNAs for the AGPase small subunit from barley endosperm and leaves (Figure 1). One of these transcripts is abundantly expressed in the starchy endosperm (bepsF1) and the other is moderately expressed in both endosperm and leaves (blps14) (Figure 2). Sequence alignments show that the transcripts consist of two parts, i.e. a 3' region that is common to bepsF1 and blps14, and a shorter 5' region that is different (Figure 1). When the sequences for the transcripts were compared with other AGPase B sequences reported, the highest level of conservation at the derived protein level (96.1–99.5% amino acid similarity; results not shown) was observed in the common region. For the 5' unique regions, on the other hand, the deduced amino acid sequence of bepsF1 has an overall similarity of 88%, 66% and 48% to cDNAs for AGPase from wheat [9], rice [8] and maize [11] respectively. For the 5' region of blps14, the highest similarity figures are: 49% to two *Vicia faba* clones [12] (one expressed in all starch-synthesizing organs and the other in cotyledons [12]), 54% to an *Arabidopsis thaliana* expressed sequence tag [23] (clone 38E3T7 in dbEST [24]), and 53% to a potato tuber sequence [25]. These clones are all of the full-length sequences reported, and AGPase clones can thus be divided into two groups based on similarities. One group contains the clones that were isolated from the starchy endosperm of cereals, and they are similar to bepsF1. The other group includes all other sequences, and these sequences are similar to blps14. The alignment of sequences across the two groups is uncertain, partly because of length differences of the derived N-termini (40–47 residues for the bepsF1 group, 76–88 residues for the blps14 group). Comparison of all sequences in a multiple alignment (results not shown) suggests that the sequences of the two groups are not homologous. The division of clones into two distinct groups is illustrated by the dendrogram presented in Figure 5.

The derived amino acid sequences of bepsF1 and blps14 differ only in their N-termini (Figure 1), i.e. where transit peptide sequences of nuclei-encoded plastid-targeted proteins should be present [21]. Unfortunately, the presence of a transit peptide can be neither verified nor refuted by analysing the amino acid sequence of a protein. The method of Bairoch [22] is expected to detect an estimated 77.4% of the sequences known to be involved in targeting peptides to plastids, but it can also predict transit peptide sequences incorrectly with a frequency of 5.8%. The identification of a transit peptide in blps14 but not bepsF1 based on the amino acid sequences is therefore tentative, and will have

Figure 5 Dendrogram for AGPase B sequences with similarities to the unique 5'-end regions of bepsF1 and blps14

The *A. thaliana* clone is an expressed sequence tag derived from a mixed library [24]. For *V. faba*, the P type is expressed in all starch-synthesizing tissues, and the C type only in cotyledons [12]. The *Solanum tuberosum* transcript is expressed in photosynthetic and non-photosynthetic tissues [7]. The three cereal clones (*Triticum aestivum*, *Oryza sativa* and *Zea mays*) were all from libraries constructed from starchy endosperm RNA (see discussion for details on the different clones). Phylogenetic distances between the N-terminal amino acid sequences were calculated by the GCG program PILEUP [15].

to be verified experimentally. Preliminary results [26], from comparing proteins expressed in *E. coli* with the proteins *in vivo*, have suggested that there is no cleavage of transit peptides *in vivo* for the proteins derived from either bepsF1 or bepl10, an AGPase large-subunit transcript that is abundantly expressed in the starchy endosperm but not leaves [27]. Similar observations were made for the corresponding transcripts of maize [28], but analogous experiments have not yet been done for a postulated plastidic AGPase. The apparent absence of deduced transit peptide sequences in the AGPase transcripts abundantly expressed in the starchy endosperm coincides with the suggestion that there might be a cytosolic AGPase in this tissue of barley [26] and maize [29]. Possible roles for a putative cytosolic isoform of AGPase have recently been discussed [26,30]. To our knowledge, there are no reports on alternative mechanisms (not involving a transit peptide) for targeting proteins to plastid stroma.

A single AGPase B gene encodes two different transcripts

The presence of sequences corresponding to the unique 5' regions of both transcripts within a restricted area of genomic DNA strongly suggests that they were derived from the same gene. The bepsF2 cDNA clone, identical in sequence with the bepsF1 one, has previously been mapped to the barley genome [6]. The restriction pattern observed for the barley genomic DNA [6] is in full agreement with the presence of the relevant restriction sites in the PS11 clone (Figure 3), supporting the presence of only one AGPase B gene in barley. Furthermore, the sequences of the two

Figure 6 A hypothetical model for the expression of the AGPase B gene of barley

The long bepsF1 primary transcript is initiated from an upstream promoter (open arrow), and probably spliced according to the 5'-most donor site. For the shorter bps14 primary transcript, initiated from a downstream promoter (solid arrow), a different donor site has to be used, and hence a different exon is retained. Note that the introns are compressed in this Figure.

transcripts and the location of exons in the putative AGPase B gene (Figures 3 and 4), suggest the following hypothetical model for the expression of the gene (Figure 6). To generate the bepsF1 transcript, transcription is initiated upstream of exon 1a, and a large intron, comprising the blps14-specific exon (1b), is spliced out from the primary transcript. To make the blps14 transcript, on the other hand, transcription is initiated downstream of exon 1a and a shorter intron is consequently spliced out. Because the coinciding right splicing site (acceptor) is used for the splicing of the two primary transcripts, the processed transcripts will be identical downstream of exon 2 (Figure 6). The mechanism presupposes two independent promoters, upstream of each of the alternative 5' exons. The specificity for generating the two different transcripts is thus at the transcription level rather than in the processing of the primary transcripts. This type of mechanism, classified as alternative splicing, has previously been described for a number of mammalian genes (e.g. mouse α -amylase gene) [31]. Other alternative splicing mechanisms described [31] are not compatible with the structure of the bepsF1 and blps14 cDNAs. These mechanisms all generate alternative transcripts with identical 5' ends, but in bepsF1 and blps14 these are clearly different. Thus, provided that the bepsF1 and blps14 cDNAs are close to full-length, the other mechanisms for alternative splicing can be discarded.

Searching the sequence upstream of exon 1a for conserved promoter motifs, a TATA-box characteristic of eukaryotic promoters is present 33 nucleotides upstream of the 5' end of exon 1a (Figure 4). Although no such motif is present in the upstream sequence of exon 1b, the presence of a promoter in this region cannot be ruled out because, for example, many constitutively expressed 'housekeeping' genes are known to lack TATA-box sequences in their promoter region. The hypothetical model (Figure 6) is also consistent with the presence of sequences conforming to the GT-AG rule for splicing in the predicted intron/exon borders for both splicing events. Because the same right splicing site (acceptor) is used for splicing of both primary transcripts, processed transcripts would be expected to be identical downstream from exon 2, i.e. the sequence corresponding to the common region of the transcripts.

As indicated by the close relationship between the deduced N-termini of the cereal AGPase B sequences (Figure 5), a similar organization of the gene in these species may be expected. Indeed, clear support for this view is found when the barley gene is compared with the corresponding rice gene, the only other cereal AGPase B gene so far published [32]. There is a high level of identity between the sequences corresponding to exons 1a and 2, reflecting the high level of identity between the transcripts isolated from endosperm tissue of the two species. Also, there is

a significant level of identity (80 %) between the barley sequence corresponding to the 3' end of exon 1b and a 70-nt element in the first intron of the rice gene. This similarity includes the proposed splicing donor site for exon 1b of the barley gene. Unfortunately, this region of the rice gene is located immediately downstream of a 700-bp region that was not sequenced, so a more definite evaluation of the collinearity between the barley and rice AGPase B genes cannot be done at present. For maize, alignments of two AGPase B cDNA sequences show differences throughout the length of the sequences, indicating that they are encoded by different genes [11]. Further studies are therefore needed to decide the universality of the barley-type organization of the AGPase B gene in cereals.

We thank Kjetill Jakobsen and Robin Stacy for assistance with the Northern blots, and Johannes Einset and Alison M. Smith for useful comments on the manuscript. This work was supported by grants from The Norwegian Research Council.

REFERENCES

- Kleczkowski, L. A., Villand, P., Lønneborg, A., Olsen, O.-A. and Lüthi, E. (1991) *Z. Naturforsch. Teil C* **46**, 605–612
- Preiss, J. (1991) in *Oxford Surveys of Cellular and Molecular Biology*, vol. 7 (Mifflin, B. J., ed), pp. 59–114, Oxford University Press, Oxford
- Kleczkowski, L. A., Villand, P., Preiss, J. and Olsen, O.-A. (1993) *J. Biol. Chem.* **268**, 6228–6233
- Kleczkowski, L. A., Villand, P., Lüthi, E., Olsen, O.-A. and Preiss, J. (1993) *Plant Physiol.* **101**, 179–186
- Villand, P., Aalen, R., Olsen, O.-A., Lüthi, E., Lønneborg, A. and Kleczkowski, L. A. (1992) *Plant Mol. Biol.* **19**, 381–389
- Kilian, A., Kleinhofs, A., Villand, P., Thorbjørnsen, T., Olsen, O.-A. and Kleczkowski, L. A. (1994) *Theor. Appl. Genet.* **87**, 869–871
- Nakata, P. A., Anderson, J. M. and Okita, T. W. (1994) *J. Biol. Chem.* **269**, 30798–30807
- Anderson, J. M., Hnilo, J., Larson, R., Okita, T. W., Morell, M. and Preiss, J. (1989) *J. Biol. Chem.* **264**, 12238–12242
- Einsworth, C., Tarvis, M. and Clark, J. (1993) *Plant Mol. Biol.* **23**, 23–33
- Villand, P., Olsen, O.-A. and Kleczkowski, L. A. (1993) *Plant Mol. Biol.* **23**, 1279–1284
- Prioul, J.-L., Jeannette, E., Reyss, A., Gregory, N., Giroux, M., Hannah, L. C. and Causse, M. (1994) *Plant Physiol.* **104**, 179–187
- Weber, H., Heim, U., Borisjuk, L. and Wobus, U. (1995) *Planta* **195**, 352–358
- Smith-White, B. J. and Preiss, J. (1992) *J. Mol. Evol.* **34**, 449–464
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Genetics Computer Group (1991) *Program Manual for the GCG Package: Version 7*, Madison, WI
- Jakobsen, K. S., Breivold, E. and Hornes, E. (1990) *Nucleic Acids Res.* **18**, 3669
- Jakobsen, K. S., Klemsdal, S. S., Aalen, R. B., Bosnes, M., Alexander, D. and Olsen, O.-A. (1989) *Plant Mol. Biol.* **12**, 285–293
- Huang, N., Koizumi, N., Reinl, S. and Rodriguez, R. L. (1990) *Nucleic Acids Res.* **18**, 7007–7014
- Loh, E. Y., Elliot, J. F., Cwirla, S., Lanier, L. L. and Davis, M. M. (1989) *Science* **243**, 217–220
- Galau, G. A., Hughes, D. W. and Dure, L., III (1986) *Plant Mol. Biol.* **7**, 155–170
- Gavel, Y. and von Heijne, G. (1990) *FEBS Lett.* **261**, 455–458
- Bairoch, A. (1992) *PC/GENE User and Reference Manual: Release 6.7*, IntelliGenetics Inc.
- Newman, T., de Bruijn, F. J., Green, P. et al. (1994) *Plant Physiol.* **106**, 1241–1255
- Boguski, M. S., Lowe, T. M. and Tolstoshev, C. M. (1993) *Nature Genet.* **4**, 332–333
- du Jardin, P. and Berhin, A. (1991) *Plant Mol. Biol.* **16**, 349–351
- Villand, P. and Kleczkowski, L. A. (1994) *Z. Naturforsch.* **49c**, 215–219
- Villand, P., Olsen, O.-A., Kilian, A. and Kleczkowski, L. A. (1992) *Plant Physiol.* **100**, 1617–1618
- Giroux, M. J. and Hannah, L. C. (1994) *Mol. Gen. Genet.* **243**, 400–408
- Hannah, L. C., Giroux, M. and Boyer, C. (1993) *Sci. Hortic.* **55**, 177–197
- Kleczkowski, L. A. (1994) *Phytochemistry* **37**, 1507–1515
- Breitbart, R. E., Andrealis, A. and Nadal-Ginard, B. (1987) *Annu. Rev. Biochem.* **56**, 467–495
- Anderson, J. M., Larsen, R., Laudencia, D., Kim, W. T., Morrow, D., Okita, T. W. and Preiss, J. (1991) *Gene* **97**, 199–205