Mol Gen Genet (1991) 228:240–248
002689259100235V

MGG

Sequence of the structural gene for granule-bound starch synthase of potato (*Solanum tuberosum* L.) and evidence for a single point deletion in the *amf* allele

Feike R. van der Leij [1], Richard G.F. Visser [2], Anne S. Ponstein [3], Evert Jacobsen [2], and Will J. Feenstra [1]

[1] Department of Genetics, University of Groningen, Kerklaan 30, NL-9751 NN Haren, The Netherlands
[2] Department of Plant Breeding, University of Wageningen, P.O. Box 386, NL-6700 AJ Wageningen, The Netherlands
[3] Biochemistry Department, University of Groningen, Nijenborgh 16, NL-9747 AG Groningen, The Netherlands

Received February 23, 1991

Summary. The genomic sequence of the potato gene for starch granule-bound starch synthase (GBSS; "waxy protein") has been determined for the wild-type allele of a monoploid genotype from which an amylose-free (amf) mutant was derived, and for the mutant part of the *amf* allele. Comparison of the wild-type sequence with a cDNA sequence from the literature and a newly isolated cDNA revealed the presence of 13 introns, the first of which is located in the untranslated leader. The promoter contains a G-box-like sequence. The deduced amino acid sequence of the precursor of GBSS shows a high degree of identity with monocot waxy protein sequences in the region corresponding to the mature form of the enzyme. The transit peptide of 77 amino acids, required for routing of the precursor to the plastids, shows much less identity with the transit peptides of the other waxy preproteins, but resembles the hydropathic distributions of these peptides. Alignment of the amino acid sequences of the four mature starch synthases with the *Escherichia coli glgA* gene product revealed the presence of at least three conserved boxes; there is no homology with previously proposed starch-binding domains of other enzymes involved in starch metabolism. We report the use of chimeric constructs with wild-type and *amf* sequences to localize, via complementation experiments, the region of the *amf* allele in which the mutation resides. Direct sequencing of polymerase chain reaction products confirmed that the *amf* mutation is a deletion of a single AT basepair in the region coding for the transit peptide. Premature termination of translation as a result of this frameshift mutation results in a small peptide. However, a protein reacting with anti-GBSS serum, slightly larger than the wild-type mature GBSS, can be detected in a membrane fraction from amylose-free tubers. A possible explanation for this phenomenon will be discussed.

Key words: Frameshift – Potato – Starch – Transit peptide – Waxy

Offprint requests to: F.R. van der Leij

Introduction

Granule-bound starch synthase (GBSS, starch granule-bound ADP(UDP)glucose:α-1,4-D-glucan 4-α-glucosyltransferase, EC 2.4.1.21), also called the waxy protein (Echt and Schwartz 1981), is a nuclear-encoded enzyme of about 60 kDa which is active in starch synthesizing plastids of plants (Shannon and Garwood 1984; Vos-Scheperkeuter et al. 1986). Although GBSS catalyses in vitro the elongation of both amylose and amylopectin (Leloir et al. 1961), mutations affecting GBSS genes in several plant species result in a loss of amylose only, while the total amounts of starch remain the same (Shannon and Garwood 1984). Recently, Smith (1990) has provided evidence that in pea the waxy protein is not the major starch granule-bound starch synthase. In this plant species a protein of 77 kDa, which can be solubilized from starch granules and which is antigenically only weakly related to the waxy protein of potato, seems to be the major GBSS. However, it is not known whether this enzyme is responsible for amylose synthesis.

Application of the iodine stain for starch to pollen and maize endosperm has made the waxy character one of the most useful characters for genetic studies in plants. After the isolation of the *Waxy* gene from maize (Shure et al. 1983), the sequences of the structural genes from maize (Klösgen et al. 1986), barley (Rohde et al. 1988) and rice (Wang et al. 1990) have been determined, and also partial sequences from the potato gene have been published (Hergersberg 1988; Rohde et al. 1990). In this study we present the complete sequence of the potato GBSS gene. We compare the amino acid sequence deduced from the potato gene with those deduced from the *Waxy* genes from maize, barley and rice and from the gene for glycogen synthase from *Escherichia coli*.

We also present an analysis of the GBSS gene from an amylose-free (amf) mutant of potato. This monogenic recessive mutant was isolated after X-irradiation (Hovenkamp-Hermelink et al. 1987) and lacks amylose in all starch-containing tissues (Jacobsen et al. 1989). Analogous to *waxy* mutations in maize and rice (Wessler and

Varagona 1985; Okagaki and Wessler 1988), the *amf* mutation is located in the structural gene for GBSS. This was shown in a complementation experiment using *Agrobacterium rhizogenes*-mediated transfer of binary constructs containing the wild-type genomic GBSS sequence into the mutant (van der Leij et al. 1991). Subsequent analysis of transformed hairy roots and regenerants showed the presence of active GBSS and amylose in the starch of the complemented mutants. Therefore, in potato, GBSS is the sole enzyme responsible for the presence of amylose in all starch-containing tissues.

Earlier restriction pattern comparison between cloned GBSS genes from the wild-type parent potato clone AM79.7322 (2n=x=12) and the amf mutant 86.040 (2n=x=12) revealed no detectable differences (Visser et al. 1989). The same result was obtained after comparison of the wild-type allele with an *amf* allele isolated from an amylose-free F_2 plant (this report). The presence of a normally sized GBSS mRNA in amylose-free plants indicated that the mutation had no negative effect on the efficiency of transcription (Visser et al. 1989). Therefore, we expected the *amf* mutation to be a small lesion in the transcribed part of the GBSS gene. Here we provide evidence that the *amf* mutation causes a shift in the reading frame which leads to the subsequent termination of translation in the region coding for the transit peptide, the part of the pre-protein needed for efficient transport across the plastid membrane. We deduced the post-translational processing site of the wild-type pre-GBSS by comparison with the N-terminus of a mature GBSS from potato and discuss the putative (re-)initiation of translation in the amf mutant which results in production of an inactive GBSS-like protein.

Materials and methods

Plant material. Clones obtained from homozygous amf plants 1029-32 and 1031-29 (2n=2x=24) from families of $F_1 \times F_1$ crosses (Jacobsen et al. 1989) and the diploidized (2n=2x=24) original mutant 86040 were grown in vitro as previously described (van der Leij et al. 1991). For the construction of libraries, cultivar (cv.) Maris Piper (2n=4x=48) and 1031-29 plants were grown in a greenhouse. Tubers from cv. Doré were used for the isolation of wild-type GBSS protein.

Standard methods and reagents. Manipulations of DNA were performed essentially as described in Maniatis et al. (1982), plasmid isolations were according to Birnboim and Doly (1979). Recombinant plasmids were purified by means of PEG precipitation (0.4 M NaCl, 12.5% polyethyleneglycol 6000, 1 h, 0° C) and extensive washing with 80% ethanol for sequencing by the dideoxy chain termination method (Sanger et al. 1977).

Sequencing of the wild-type GBSS gene. The lambda EMBL4 clone LGBSS^{wt}-6 (Visser et al. 1989) was subcloned in pUC18 (Yanisch-Perron et al. 1985), using the *E. coli* K12 strain JM83 (Vieira and Messing 1982) as host. A combination of sequencing subclones and unidirectional deletion clones and primer walking was used to determine the sequences of both strands of the three *Hin*dIII fragments, except for a part of DNA further than 380 bp downstream of the transcribed region, which was sequenced in one direction only (Fig. 4A).

A lambda gt11 potato cDNA library was constructed with the Amersham cDNA cloning kit, using tubers of cv. Maris Piper as source for RNA. Screening of 120000 plaques yielded 30 positive clones after hybridization with the GBSS cDNA isolated earlier from cv. Granola (Hergersberg 1988). Ten of the positive clones were further characterized. One clone hybridized with the 5′ *Eco*RI fragment of the previously isolated cDNA and was subcloned for sequencing.

Isolation of GBSS protein and sequencing of the N-terminus. Proteins were extracted from potato starch granules as described by Vos-Scheperkeuter et al. (1986). The protein extract was separated on an 8% SDS-polyacrylamide gel, transferred as described by Hovenkamp-Hermelink et al. (1987) to polyvinylidene difluoride (PVDF) membrane (Millipore) and stained with Coomassie Brilliant Blue as described by Matsudaira (1987). Approximately 15 µg of GBSS was excised from the filter and sequenced by automatic Edman degradation (Sequenator 477A, Applied Biosystems). Membrane fractions were isolated from potato tubers as described previously by Lavintman et al. (1974).

Cloning of the amf *allele.* A genomic bank of the amf plant 1031-29 was constructed with *Bam*HI arms of lambda GEM11 (Promega). From 230000 plaques, three positive recombinants were picked up, one hybridizing with probes upstream and downstream of the three *Hin*dIII fragments carrying the gene. This clone, designated G1107, does not contain the downstream *Bgl*II site, which enables asymmetrical subcloning using *Bgl*II and *Sal*I, as lambda GEM11 has a site for *Sal*I near the *Bam*HI site (Frischauf et al. 1983). The *amf* allele was subcloned as a *Bgl*II-*Sal*I fragment in pMTL20 (Chambers et al. 1988), yielding the construct pAMF210. The same fragment from G1107 was ligated between the *Bam*HI and *Sal*I sites of the plant transformation vector pBIN19 (Bevan 1984) to give pAMF110 (Fig. 4C).

Construction of chimeric genes. The binary plasmid pAMF110 was used as the backbone for the construction of pWM114 (Fig. 4C). For replacement of the *Eco*RI fragment carrying the 5′ part of the gene, pAMF110 was cut with *Eco*RI, religated and transformed into *E. coli*. The resulting plasmid, pAMF112, was linearized with *Eco*RI and treated with calf intestinal alkaline phosphatase, before insertion of the corresponding wild-type fragment. Plasmid pMW104 (Fig. 4C) was made the other way round, with pWAM100 (van der Leij et al. 1991) as backbone and pAMF210 as source of the mutant gene fragment.

Removal of the upstream *Hin*dIII site in pWAM10, which contains the wild-type 1.2 kb promoter and 5′ coding region was done by a Klenow reaction on a par-

tial *Hin*dIII digest, enabling the subcloning to be carried out with an *Eco*RI-*Hin*dIII fragment. The adjacent *Hin*dIII-*Eco*RI fragment of the mutant allele was cloned in-frame with the wild-type fragment in the *Eco*RI site of pUC18. The resulting *Eco*RI fragment was inserted in pAMF112 (see above) to construct pWM115 (Fig. 4C).

Triparental mating, plant transformation and screening of hairy roots. Binary plasmid transfer to *A. rhizogenes* and subsequent stem segment transformation were carried out as previously described (van der Leij et al. 1991). In each experiment a minimum of 30 starch-containing kanamycin-resistant hairy roots were screened for the presence of amylose by staining with a fresh mixture of chloral hydrate and Lugol's solution.

Sequencing of polymerase chain reaction (PCR) products. Direct sequencing of PCR fragments was carried out with each of the primers used for the PCR, after removal of the excess of both primers with the Geneclean II kit of Bio101. The primers used were 5′-CGTGTCGACTGAAACCTGCTACA-3′ and 5′-TTGGTCTCGGATCCGGATGCCAT-3′, which anneal at positions −340 to −318 and 177 to 200, respectively, in the wild-type sequence. A detailed description of the PCR procedure will be given elsewhere (van der Leij et al., manuscript in preparation).

Computer analysis. The PCGene software package (Intelligenetics, Geneva) version 6.26 and the alignment programs Fasta (Lipman and Pearson 1985) and Clustal (Higgins and Sharp 1989) were used for sequence analyses. The Swiss Prot database release 15 (August 1990) was used for protein sub-sequence comparison.

Results and discussion

Promoter sequence of the wild-type allele

The DNA sequence of the potato GBSS gene was determined from positions −1073 to 3590 relative to the first nucleotide of the start site of translation (Figs. 1, 4A, B). The promoter region of the wild-type allele from AM79.7322 has features of each of the promoters of two wild-type alleles, G1 and G28, from the tetraploid cv. Granola (Rohde et al. 1990). The promoter part of the sequence shown here is of about the same length as the shorter of the two Granola sequences, i.e. the 1.2 kb *Hin*dIII fragment of allele G1. The CAAT box of the two Granola promoters is in our sequence represented by CTAT at the same position, 32 nucleotides upstream of the putative promoter box TACAAAT. An imperfect inverted repeat of 29–31 nucleotides is positioned between −846 and −738 (Fig. 1). The sixfold direct repeat TCAC around the putative start site of transcription found in the Granola sequences (Rohde et al. 1990) is not so extensive in our sequence, but is present as a ninefold A/TC repeat (positions −298 to −280). The distance between this repeat and the TACAAAT box is different in all three alleles.

Between the CTAT and TACAAAT boxes the palindrome TGACACGTGTCA occurs; it probably carries the recognition sequence for a transcription factor since it shares seven out of eight nucleotides with the transcription factor target sequence, CACGTGGC, known as the abscisic acid (ABA) response element (Guiltinan et al. 1990) and UV-boxII or G-box (Schulze-Lefert et al. 1989). This sequence is found in regulatory regions of many differently expressed plant genes, including *rbcS*, patatin and wound-inducible genes of potato (Schulze-Lefert et al. 1989; Guiltinan et al. 1990) and is often surrounded by degenerate sequences. Two degenerate sequences are present in the GBSS gene as imperfect direct repeats starting 50 bp upstream of the G-box palindrome.

Scanning of the GBSS sequence with some other known regulatory sequences from plants (Buzby et al. 1990; Gilmartin et al. 1990; Jefferson et al. 1990; Liu et al. 1990; Maas et al. 1990; Springer et al. 1990; Takahashi et al. 1990) did not reveal regions with high similarity. It should be noted that sequences with low similarities (25–50% of the maximum values of a weight matrix scan) were found on searching with the two parts of the patatin repeat which is responsible for the sucrose response (Jefferson et al. 1990). These regions are located between positions −613 and −548 (three overlapping matches), −370 and −338 (overlapping the G-box palindrome) and −188 and −155 (in the first intron).

Analysis of the transcribed region

The cDNA used for the isolation of the genomic sequences contained one unspliced intron in the coding part (Hergersberg 1988) and also had an unspliced intron in the leader, as inferred from comparisons with the genes from maize and barley (Rohde et al. 1990). We decided to isolate more cDNAs in order to verify the existence of the latter intron. One cDNA clone with an almost complete leader lacked a part of the sequence present in the genome, showing that the potato GBSS gene indeed contains an intron in the leader sequence, with the intron/exon junction 36 nucleotides upstream of the translational start site. This is similar to the sequences of maize and barley (Klösgen et al. 1986; Rohde et al. 1988). The first introns of many plant genes, including genes encoding enzymes involved in starch metabolism like the sucrose synthase gene *Shrunken-1* of maize (Vasil et al. 1989), are reported to play a regulatory role in gene expression. Genomic sequence information on potato starch metabolism genes in the literature is for the moment restricted to the phosphorylase gene (Camirand et al. 1990), which has been reported not to carry an intron in the untranslated leader. The first intron of the potato GBSS gene shows many short direct repeats, among them an almost perfect 11-fold CTT stretch (positions −106 to −73). The presence of the first two introns in the cDNA of Hergersberg (1988) might indicate regulation at the level of splicing. The

Fig. 1. Complete genomic nucleotide sequence of the potato granule-bound starch synthase (GBSS) gene with the translational start at nucleotide +1. The stop codon at position 2961 is shown in *bold*. Introns are shown in *small italics*, bordered by *bold* type. The putative promoter sequences and polyadenylation signals are *underlined*, poly (A) addition sites are marked with *asterisks*. The start of the largest cDNA is marked with a *black dot*; the sequence around the start of transcription is *underlined* with a *dashed line*. G-box-like sequences are *overlined* with *arrows*, large imperfect inverted repeats with *dashed arrows*

Fig. 2. Alignment of four deduced amino acid sequences of precursors of granule-bound starch synthases from plants, with the similarity shown in the lower row. *Asterisks* are identical residues and *dots* are conservative substitutions. Potato sequences are as deduced from the DNA sequence (1) and as determined by automatic Edman degradation (2). The upper sequence is a translation of the *Escherichia coli glgA* gene which was aligned with all four mature sequences of the starch synthases using the program Clustal. Similarity and identity between the *glgA* sequence and the four plant sequences are given as described above. Gaps in the alignment are indicated by *dashes*; *question marks* are residues which could not be unambigiously determined by Edman degradation. Conserved boxes are indicated

first exon and the first nucleotides of the first intron are missing in the cDNA sequence of Hergersberg (1988), which probably accounts for the negative outcome of attempts to complement the *amf* mutation with constructs containing this cDNA under the control of the CaMV promoter (our unpublished results). The second intron shows only two substitutions when compared with the genomic sequence (data not shown) and the latter was shown to be functionally active in complementation experiments (van der Leij et al. 1991).

The positions of the introns were determined by alignment of the genomic sequence with the cDNA sequences. As in the maize *Waxy* gene (Klösgen et al. 1986), the total number of intervening sequences in the potato GBSS gene is 13. The same number has been reported for the rice gene (Wang et al. 1990) without, however, an intron in the leader. Given the well-conserved intron position in the leader of the genes from potato, maize and barley, it would not be surprising if a 14th intron were present in the leader of the rice gene. The barley gene (Rohde et al. 1988) carries only 11 introns.

All introns in the potato GBSS gene follow the universal GT/AG rule and coincide with AT-rich regions. As is found for other dicot plant genes (Camirand et al. 1990; Hanley and Schuler 1988) and indeed plant genes in general (Brown 1986), the positions near the splice junctions are less fixed, although, in the potato GBSS gene, the third nucleotide of all splice donor sites is an A.

Putative polyadenylation signals have already been found 21 and 11–13 nucleotides upstream of the polyadenylation sites of three 3′ cDNA sequences (Hergersberg 1988).

After processing of the GBSS primary transcript a mature messenger RNA of about 2.4 kb remains, which encodes the precursor protein, using the first AUG sequence as the start codon. The codon usage in the potato GBSS gene complies with that in dicots (Campbell and Gowri 1990) with frequencies of 25% and 34% for A and U as third nucleotide, respectively, where monocots show much lower values for these nucleotides. For example, the maize *Waxy* sequence shows the respective nucleotides in the third position in only 4.4% and 2.8% of the codons used.

Protein sequences

The potato GBSS gene encodes a 607 amino acid preprotein with a calculated molecular weight of 66575 daltons. The amino acid sequence of the N-terminus of the mature GBSS protein, as determined by automated Edman degradation, enabled us to locate the transit peptide processing site (Fig. 2). A transit peptide of 77 amino acids, needed for the routing of the enzyme, is cleaved from the preprotein during transport into plastids. The mature form of potato GBSS, as deduced from the DNA sequence, is a 540 amino acid protein of 58243 kDa. Vos-Scheperkeuter et al. (1986) determined a M_r (rela-

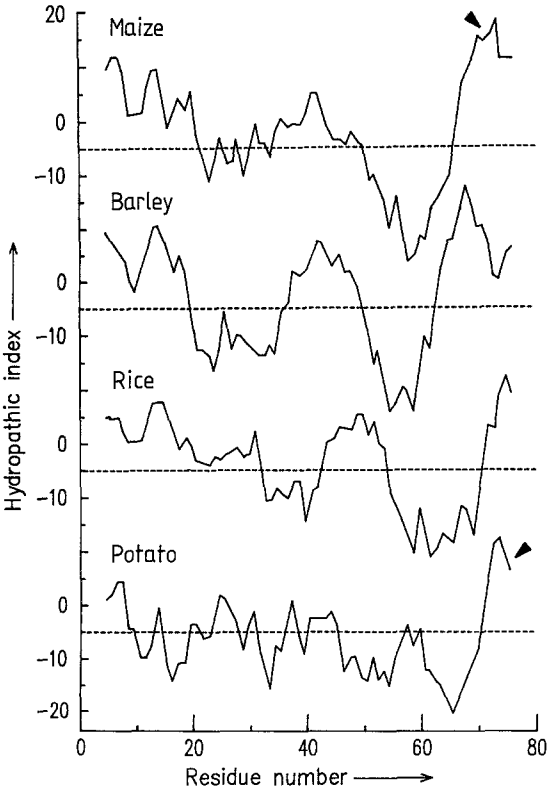

Fig. 3. Hydrophobicity plots according to Kyte and Doolittle (1982) of the 80 N-terminal residues of the four plant GBSS precursors containing the transit peptides. The cleavage sites, which are only known for the maize and the potato protein, are indicated by an *arrowhead*. The interval used for the calculation of the indices was nine amino acid residues

tive molecular weight) of 60 kDa by SDS-polyacrylamide gel electrophoresis.

A comparison of the amino acid sequence of the potato pre-GBSS with the sequences of the three waxy preproteins of the monocots maize, barley and rice reveals a high degree of similarity between the mature forms of all four enzymes, but not between the transit peptides (Fig. 2). Of the 540 amino acid residues of the mature potato GBSS, 336 (62%) are identical and 472 (87%) are similar (identical amino acids plus conservative substitutions) in all four enzymes. Alignment of the plant enzymes with the *E. coli glgA* gene product glycogen synthase (Kumar et al. 1986) shows 26% identity and 63% similarity of the glycogen synthase with the starch synthases and reveals at least three homologous boxes. The conservation of the ADP-glucose binding site, KTGGL (Furukawa et al. 1990), at residues 15–19 of the glycogen synthase has already been discussed by Preiss (1990) in a comparison of the *glgA* sequence with two monocot starch synthases and the N-terminus of the potato GBSS. A second conserved sequence is located at residues 372–384 of the glycogen synthase (478–490 of the potato sequence). Within this box, the sequence VPSRFEPCGL is identical between the potato, rice and *E. coli* sequences and shares one conservative substitution compared with the maize and barley sequences (Fig. 2). A scan of the Swiss Prot databank with this sequence did not reveal similarities with other known proteins.

A third box is separated from box II by 12 amino acids, with a proline 4–5 residues before box III. The sequence TGGLAD in the *E. coli* protein resembles the sequence of the six *E. coli* residues immediately beyond the ADP-glucose-binding lysine residue in box I. In glycogen synthase, the lysine in box III is replaced by an arginine and in the starch synthases by a serine, which makes it unlikely that box III can have the same function as box I. The major parts of the three boxes mentioned here are also present in the deduced protein sequence of the *Bacillus subtilis glgA* gene (J. Kiel, personal communication).

The well-conserved primary structures of the mature proteins of the four plants compared here do not show any similarity with proposed starch granule binding domains of bacterial and fungal amylases, branching enzymes and cyclodextrin glycosyl transferases (Svensson et al. 1989).

The primary structures of the transit peptides of the four plant preproteins, although highly diverged, share the common features of transit peptides (Keegstra et al. 1989) in containing a relatively high amount of alanine, serine and threonine residues, starting with a hydrophobic region followed by one or more hydrophilic parts and ending with a dramatic increase in hydrophobicity near the (putative) processing sites. Plots of the hydropathic indices are shown in Fig. 3; the pattern mentioned above is less pronounced in the case of the potato GBSS transit peptide.

Sequence of the amf *allele*

Since we had indications that the GBSS gene isolated earlier from 86040 was derived from a chimeric plant, with mutant and either wild-type or revertant tissues, we decided not to use clones from the monoploid 86040 for further analysis. The isolation of a genomic clone from a diploid amf plant was done by probing with parts of DNA surrounding the structural GBSS gene. Despite this approach of isolating only those clones which carry upstream as well as downstream sequences, the only complete clone obtained did not carry the distal *Bgl*II site. This did not interfere with our aim to reduce the part of the gene to be sequenced, since a wild-type subclone with even fewer potato sequences at the 3′ end (pWAM150, van der Leij et al. 1991, see also Fig. 4C) was shown to be able to complement the amf mutant.

Transformation experiments using the chimeric constructs pWM114 and pMW104 revealed that the mutation was located in the part of the gene upstream of the central *Eco*RI site (Fig. 4C). Sequencing of the region between the *Nsi*I site in the first intron and the central *Eco*RI site revealed the deletion of one of the three adenines at positions 46–48 of the GBSS sequence. The part of the gene between the *Hin*dIII site and the central *Eco*RI site showed no alterations in a range of single direction sequencing reactions. The absence of a mutation in that part was confirmed by the complemen-

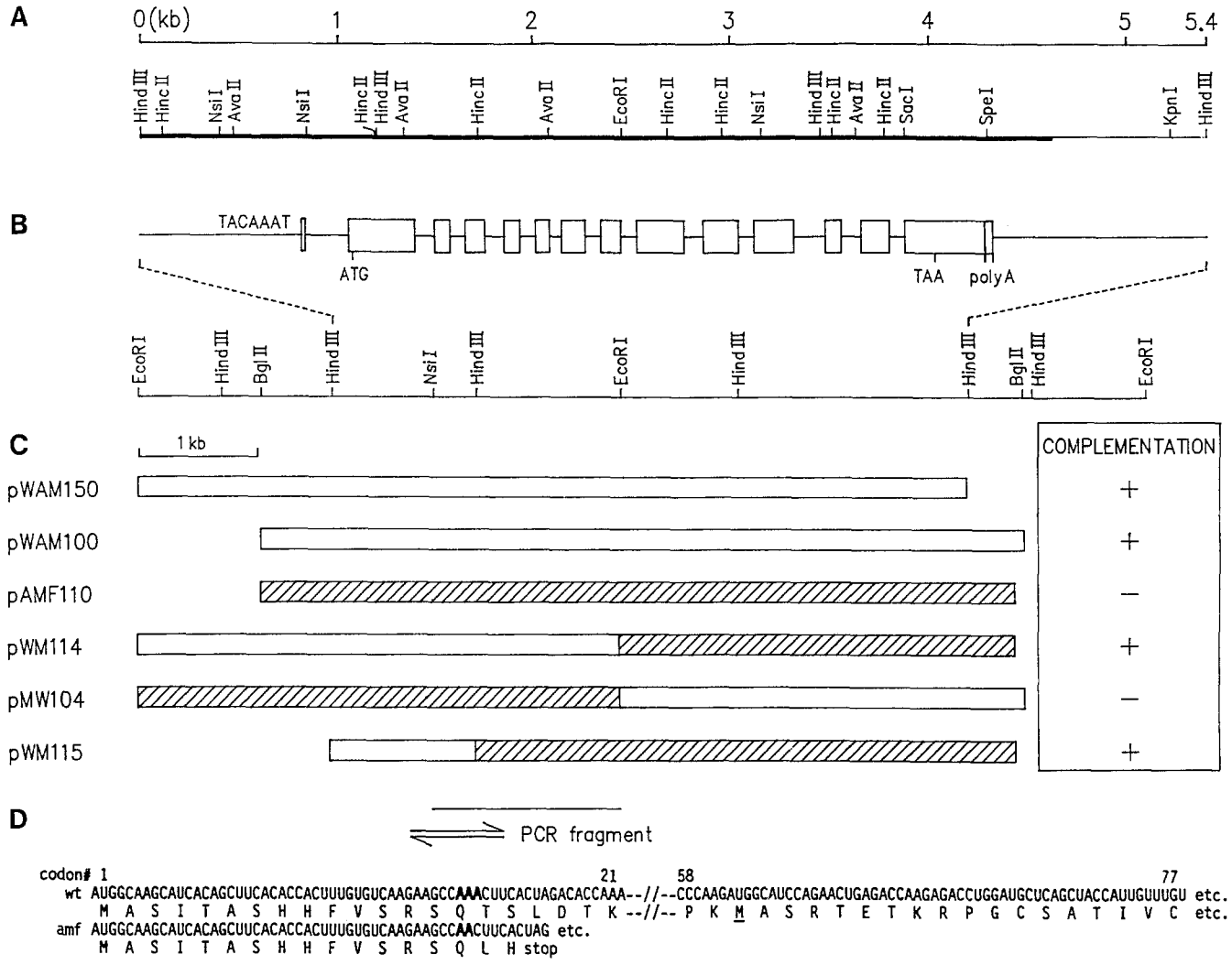

Fig. 4A–D. Schematic overview of the potato GBSS gene and derivatives and the sequence of the *amf* allele. **A** Representation of the sequenced parts and restriction map of the wild-type allele. The *thick line* represents the part of the three *Hin*dIII fragments which has been sequenced in both directions. The sequence of this part is shown in Fig. 1. **B** The GBSS gene as part of two *Eco*RI fragments with the promoter, translational start and stop, and poly(A) sites assigned. *Boxes* represent exons. **C** Chimeric constructs used to localize the *amf* mutation with the corresponding complementation results. *Open boxes* (wild type), *hatched boxes* (mutant). **D** Part of the *amf* allele which has been sequenced in one direction (upper) and both directions (polymerase chain reaction fragment). The relevant parts of the sequences of wild type and mutant are shown below as RNA sequences, together with the deduced protein sequences. The adenine residues in the wild type from which one has been deleted in the mutant are shown in *bold*. The *underlined* methionine is the possible start of (re-)initiation of translation in the mutant. The numbering of codons is indicated

tation of the amf mutant with construct pWM115 (Fig. 4C), which only carries the first *Hin*dIII fragment from the wild-type allele, the rest of the construct being derived from the amf clone G1107. To exclude the possibility of sequence alterations in the transcribed part upstream of the *Nsi*I site, a polymerase chain reaction (PCR) was performed with primers upstream of the promoter TACAAAT box and downstream of the *Hin*dIII site mentioned above. Sequencing of both strands of PCR fragments derived from G1107 and from plant DNA of 1031-29 and the original mutant 86040 confirmed that the only difference between the wild-type and the *amf* allele is the deletion of one of the three adenines of codons 15 and 16 (Fig. 4D). Point deletions, next to larger gene alterations, have been reported for mutant genes which were obtained after ionizing radiation (Grosovsky et al. 1988). However, the mutation may also have arisen spontaneously.

The consequence of this single base deletion is a frameshift and the subsequent termination of translation at the position of codon number 18, an amber (UAG) stop codon, of the *amf* sequence. Therefore, the amf mutant is expected to synthesize only a short peptide of 17 amino acids. However, in Fig. 5 we show the presence of a GBSS-like protein in a membrane fraction from tubers from the amf mutant. The M_r of this protein turned out to be 4 ± 2 kDa larger than a membrane-bound GBSS-like protein from the wild type, the latter having the same M_r (60 kDa) as the wild-type mature GBSS. It seems logical that the mutant GBSS-like protein is the product of the *amf* GBSS sequence. This phenomenon can be explained if it is assumed that (re-)initi-

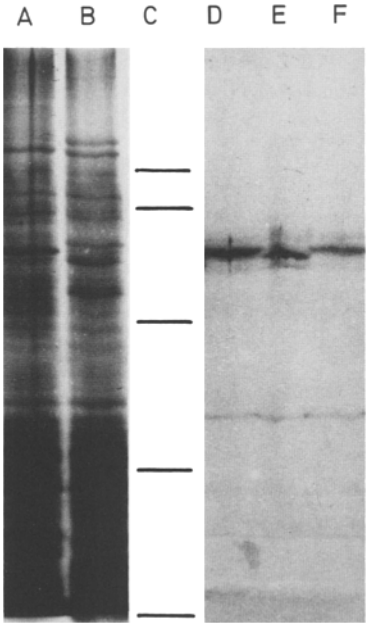

Fig. 5. Proteins cross-reacting with anti-GBSS serum. Membrane-bound proteins were separated in a 10% SDS-polyacrylamide gel either stained with Coomassie Brilliant Blue (lanes A, B) or blotted and immunoprecipitated (lanes D–F). Tuber membrane fractions obtained from the amf mutant (lanes A, D, F) and a wild-type potato (lanes B, E) show the presence of immunoreacting proteins of about 60 kDa, the mutant protein being apparently larger than the band from the wild-type preparation. Lane C represents the molecular weight markers with proteins of 78, 68, 45, 30 and 17 kDa, respectively, from the top downwards

ation of translation in the mutant takes place at internal AUG codons. Two of these AUG codons are out of phase but initiation at codon 60 of the wild-type sequence (Fig. 4D) would result in a protein which is about 2 kDa larger than the mature wild-type form. This protein carries only 17 amino acids of the transit peptide, starting with the amino acid sequence MASRT. Except for the hydrophilic arginine residue, this sequence forms a perfect start for a transit peptide which, however, will probably be too short to allow transport of the precursor across the plastid membranes. The shortest transit peptide known to be functionally active contains 29 amino acids (Keegstra et al. 1989). The amf mutant does not have GBSS attached to starch grains inside plastids (Hovenkamp-Hermelink et al. 1987). Binding to the plastid membrane, but without transport and the accompanying splitting off of the transit peptide, would result in the observed presence of a membrane-bound protein with M_r 62 kDa. Vancanneyt et al. (1990) showed in an analysis of transgenic tobacco plants that steady-state plant mRNA levels dramatically decrease when artificial stop codons prevent further translation. In contrast with their results, slightly increased amounts of GBSS mRNA were found in the amf mutant (Visser et al. 1989). This discrepancy implies either a difference in stability of the mRNAs, which might be related to the translatibility of the part of the gene beyond the stop codon, or a compensating feedback mechanism at the transcriptional level in the amf mutant.

Although the general rule is that the first start codon is used, initiation of translation at internal AUG codons has been shown to occur in plant cells (Putterill and Gardner 1989) and in plant viruses (Tacke et al. 1990).

Acknowledgements. The authors wish to thank H. Nijdam, A. Hesseling-Meinders, H. Mohamedjoenoes, P. Bakker and I. Straatman-Engelen for technical assistance and H. Mulder for preparing the figures. We gratefully acknowledge the assistance of H. Enequist during sequencing and Dr. P. Terpstra for help with the computer analyses. We thank G. van der Steege for helpful discussions. Protein sequencing was performed by Eurosequence, Groningen, the Netherlands.

References

Bevan M (1984) Binary Agrobacterium vectors for plant transformation. Nucleic Acids Res 12:8711–8721

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Brown JWS (1986) A catalogue of splice junction and putative branch point sequences from plant introns. Nucleic Acids Res 14:9549–9559

Buzby JS, Yamada T, Tobin EM (1990) A light-regulated DNA-binding activity interacts with a conserved region of a *Lemna gibba rbcS* promoter. Plant Cell 2:805–814

Camirand A, St-Pierre B, Marineau C, Brisson N (1990) Occurrence of a copia-like transposable element in one of the introns of the potato starch phosphorylase gene. Mol Gen Genet 224:33–39

Campbell WH, Gowri G (1990) Codon usage in higher plants, green algae, and cyanobacteria. Plant Physiol 92:1–11

Chambers SP, Prior SE, Bartsow DA, Minton NP (1988) The pMTL *nic*$^-$ cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. Gene 68:139–149

Echt GS, Schwartz D (1981) Evidence for the inclusion of controlling elements within the structural gene at the waxy locus in maize. Genetics 99:275–284

Frischauf AM, Lehrach H, Poustka A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827–842

Furukawa K, Tagaya M, Inoye M, Preiss J, Fukui T (1990) Identification of lysine 15 at the active site in *Escherichia coli* glycogen synthase. J Biol Chem 265:2086–2090

Gilmartin PM, Sarokin L, Memelink J, Chua N-H (1990) Molecular light switches for plant genes. Plant Cell 2:369–378

Grosovsky AJ, de Boer JG, de Jong PJ, Drobetsky EA, Glickman BW (1988) Base substitutions, frameshifts, and small deletions constitute ionizing radiation-induced point mutations in mammalian cells. Proc Natl Acad Sci USA 85:185–188

Guiltinan MJ, Marcotte WR, Quatrano RS (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. Science 250:267–271

Hanley BA, Schuler MA (1988) Plant intron sequences: evidence for distinct groups of introns. Nucleic Acids Res 16:7159–7176

Hergersberg M (1988) Molekulare Analyse des *waxy* Gens aus *Solanum tuberosum* und Expression von *waxy* antisense RNA in transgenen Kartoffeln. PhD Thesis, University of Cologne

Higgins DG, Sharp PM (1989) Multiple sequence alignment program. Comput Appl Biosci 5:151–153

Hovenkamp-Hermelink JHM, Jacobsen E, Ponstein AS, Visser RGF, Vos-Scheperkeuter GH, Bijmolt EW, Vries JN de, Witholt B, Feenstra WJ (1987) Isolation of an amylose-free starch mutant of the potato (*Solanum tuberosum* L.). Theor Appl Genet 75:217–221

Jacobsen E, Hovenkamp-Hermelink JHM, Krijgsheld HT, Nijdam H, Pijnacker LP, Witholt B, Feenstra WJ (1989) Phenotypic

and genotypic characterization of an amylose-free starch mutant of the potato. Euphytica 44:43–48
Jefferson R, Goldsbrough A, Bevan M (1990) Transcriptional regulation of a patatin-1 gene in potato. Plant Mol Biol 14:995–1006
Keegstra K, Olsen LJ, Theg SM (1989) Chloroplastic precursors and their transport across the envelope membrane. Annu Rev Plant Physiol Plant Mol Biol 40:471–501
Klösgen RB, Gierl A, Schwarz-Sommer Z, Saedler H (1986) Molecular analysis of the *waxy* locus of *Zea mays*. Mol Gen Genet 203:237–244
Kumar A, Larsen CE, Preiss J (1986) Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* ADP-glucose:α-1,4-glucan,4-glucosyltransferase as deduced from the nucleotide sequence of the *glgA* gene. J Biol Chem 261:16256–16259
Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132
Lavintman N, Tandecarz J, Carceller M, Mendiara S, Cardini CE (1974) Role of uridine diphosphate glucose in the biosynthesis of starch. Mechanism of formation and enlargement of a glucoproteic acceptor. Eur J Biochem 50:145–155
van der Leij FR, Visser RGF, Oosterhaven K, van der Kop DAM, Jacobsen E, Feenstra WJ (1991) Complementation of the amylose-free starch mutant of potato (*Solanum tuberosum* L.) by the gene encoding granule-bound starch synthase. Theor Appl Genet, in press
Leloir LF, Rongine de Fekete MA, Cardini CE (1961) Starch and oligosaccharide synthesis from uridine diphosphate glucose. J Biol Chem 236:636–641
Lipman DJ, Pearson WR (1985) Rapid and sensitive protein similarity searches. Science 227:1435–1441
Liu XJ, Prat S, Willmitzer L, Frommer WB (1990) *Cis* regulatory elements directing tuber-specific and sucrose-inducible expression of a chimeric class I patatin promoter/GUS gene fusion. Mol Gen Genet 223:401–406
Maas C, Schaal S, Werr W (1990) A feedback control element near the transcription start site of the maize *Shrunken* gene determines promoter activity. EMBO J 9:3447–3452
Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
Matsudaira P (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J Biol Chem 262:10035–10038
Okagaki RJ, Wessler SR (1988) Comparison of non-mutant and mutant *waxy* genes in rice and maize. Genetics 120:1137–1143
Preiss J (1990) Biology and molecular biology of starch synthesis and its regulation. In: Miflin BJ (ed) Oxford Surveys of Plant Molecular and Cell Biology, vol 7. Oxford University Press, UK, in press
Putterill JJ, Gardner RC (1989) Initiation of translation of the β-glucuronidase reporter gene at internal AUG codons in plant cells. Plant Sci 62:199–205
Rohde W, Becker D, Salamini F (1988) Structural analysis of the *waxy* locus from *Hordeum vulgare*. Nucleic Acids Res 16:7185–7186
Rohde W, Becker D, Kull B, Salamini F (1990) Structural and functional analysis of two *waxy* gene promoters from potato. J Genet Breed 44:311–315
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
Schulze-Lefert P, Dangl JF, Becker-André M, Hahlbrock K, Schulz W (1989) Inducible *in vivo* DNA footprints define sequences necessary for UV light activation of the parsley chalcone synthase gene. EMBO J 8:651–656
Shannon JC, Garwood DL (1984) Genetics and physiology of starch development. In: Whistler RL, BeMiller JN, Paschall JF (eds) Starch: chemistry and technology, 2nd edn. Academic Press, Orlando, pp 25–86
Shure M, Wessler S, Fedoroff N (1983) Molecular identification and isolation of the *Waxy* locus in maize. Cell 35:225–233
Smith AM (1990) Evidence that the "waxy" protein of pea (*Pisum sativum* L.) is not the major starch-granule-bound starch synthase. Planta 182:599–604
Springer B, Bellmann R, Werr W (1990) Formation of composite nucleoprotein complexes near the transcription start of the *Shrunken* gene from maize. Mol Gen Genet 221:94–101
Svensson B, Jespersen H, Sierks MR, MacGregor EA (1989) Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem J 264:309–311
Tacke E, Prüfer D, Salamini F, Rohde W (1990) Characterization of a potato leafroll luteovirus subgenomic RNA: differential expression by internal translation initiation and UAG suppression. J Gen Virol 71:2265–2272
Takahashi Y, Niwa Y, Machida Y, Nagata T (1990) Location of the cis-acting auxin-responsive region in the promoter of the *par* gene from tobacco mesophyll protoplasts. Proc Natl Acad Sci USA 87:8013–8016
Vancanneyt G, Rosahl S, Willmitzer L (1990) Translatability of a plant-mRNA strongly influences its accumulation in transgenic plants. Nucleic Acids Res 18:2917–2921
Vasil V, Clancy M, Ferl RJ, Vasil IK, Hannah LC (1989) Increased gene expression by the first intron of maize *Shrunken-1* locus in grass species. Plant Physiol 91:1575–1579
Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268
Visser RGF, Hergersberg M, van der Leij FR, Jacobsen E, Witholt B, Feenstra WJ (1989) Molecular cloning and partial characterization of the gene for granule-bound starch synthase from a wildtype and an amylose-free potato (*Solanum tuberosum* L.). Plant Sci 64:185–192
Vos-Scheperkeuter GH, de Boer W, Visser RGF, Feenstra WJ, Witholt B (1986) Identification of granule-bound starch synthase in potato tubers. Plant Physiol 82:411–416
Wang Z, Wu Z, Xing Y, Zheng F, Guo X, Zang W, Hong M (1990) Nucleotide sequence of rice waxy gene. Nucleic Acids Res 18:5898
Wessler SR, Varagona MJ (1985) Molecular basis of mutations at the waxy locus of maize: Correlation with the fine structure genetic map. Proc Natl Acad Sci USA 82:4177–4181
Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene 33:103–119

Communicated by H. Saedler

Note added in proof. The sequence data will appear in the EMBL/Gen Bank/DBBJ Nucleotide Sequence Data Bases under accession number X58453