Plant Molecular Biology **22**: 67–82, 1993.

Expression, organisation and structure of the genes encoding the *waxy* protein (granule-bound starch synthase) in wheat

Charles Ainsworth, Joanna Clark [1] and John Balsdon
Department of Biological Sciences, Wye College (University of London), Wye, Kent TN25 5AH, UK
[1] *Present address: Biological Laboratory, University of Kent, Canterbury, Kent CT2 7NZ, UK*

Received 1 October 1992; accepted in revised form 20 January 1993

Key words: granule-bound starch synthase, hexaploid wheat, *waxy* protein

Abstract

A full-length cDNA clone representing the *waxy* protein (GBSSI) isolated from a hexaploid wheat developing grain cDNA library has been used to characterise the organisation and expression of the *waxy* genes in wheat. The genes are organised as a triplicate set of single copy homeoloci on chromosome arms 4AL, 7AS and 7DS. The genes are active throughout grain filling where the main 2.3 kb transcript accumulates to high levels. The 2.3 kb transcript is not expressed in leaves where the presence of a related, but less homologous, transcript of 1.6 kb suggests that a different set of genes operates. Gel analysis and purification of the *waxy* protein isolated from starch granules, followed by N-terminal amino acid sequencing in conjunction with data from hybrid select translation experiments and sequence analysis of the cDNA, shows that the mature protein has a molecular weight of 60kDa (615 amino acids) and that the preprotein includes a chloroplast/amyloplast transit peptide of 7kDa (75 amino acids). Analysis of the derived amino acid sequence and alignment with five other plant *waxy* proteins shows that they exhibit substantial homology. The wheat protein differs from all others in that it contains an 11amino acid insertion towards the N-terminus. The protein contains the conserved motif KTGGL found in other *waxy* proteins and which has been implicated as the active site in glycogen synthase.

Introduction

Starch is the major storage product of the world's most important food crops: the seeds of cereals and some legumes, and tuber and root crops such as potato and yam. Starch biosynthesis in plants is confined to the chloroplast and amyloplast; in the latter, storage starch is deposited as water-insoluble granules. The starch granule usually comprises two different forms of polymer: amylose, an essentially linear chain of α(1–4)-linked α-*D*-glucopyranosyl residues and amylopectin which is the branched form of amylose (by α(1–6) linkages). The ratio of amylose to amylopectin varies between different starches but in most starches the amylopectin content is 67570% of the total [29].

Although the precise pathways leading to the accumulation of starch as granules in storage organs are not fully understood, evidence strongly suggests that the pathway involving ADP-glucose predominates with hexose sugars being imported

into the amyloplast [14]. There is increasing evidence for the fundamental importance of the last three enzymes in the pathway (which have been shown to be confined to the amyloplast and chloroplast [27, 30]) in regulating the type of starch synthesised and the flux through the pathway. These enzymes are ADPG pyrophosphorylase (which catalyses the formation of ADP-glucose from glucose-1-phosphate and ATP), starch synthase (which uses ADP-glucose to extend the amylose chain by $\alpha(1-4)$ linkages) and starch branching enzyme (which cleaves the amylose chain and adds $\alpha(1-6)$-linked branches. The evidence comes from the analysis of the enzymic composition of starch mutants in a wide variety of species and is now being extended by the application of molecular biology and, in particular, by the analysis of transgenic plants.

Starch synthase (ADP-glucose glucosyl transferase; EC 2.4.1.2.1) occurs in two main forms, one bound to the starch grain and the other soluble, located in the stroma of the chloroplast or amyloplast. The *waxy* protein, first identified from the *waxy* mutant of maize, has for some time been thought to be the granule bound starch synthase (GBSS). *Waxy* mutants have been identified in a variety of other plant species including *Amaranthus hypochondriacus*, barley, rice, potato and some millets [6, 11, 12, 28]. In the maize *waxy* mutants, the type of starch in the kernel (and in the pollen and embryo sac) is profoundly affected such that it is completely free of amylose, compared to the normal levels of 25–30% found in the wild-type starches [38]. In addition, although some starch synthase activity has been associated with the *waxy* endosperm, the activity of granule-bound starch synthase is undetectable [25]. Analysis of the proteins associated with the starch granule showed that in the *waxy* mutant, a major protein of 58–60 kDa (the *waxy* protein) and a number of less abundant proteins are absent [6]. The GBSS activity in *waxy* mutants correlates with the abundance of the 60 kDa *waxy* protein [22]. Although these data only provide correlative evidence for the *waxy* protein being a granule-bound starch synthase, there is supporting evidence from a number of other sources. Polyclonal antibodies raised to the purified potato 60kDa *waxy* protein have been shown to cross react with the *waxy* proteins of both maize and *Amaranthus* and starch synthase activity in *Amaranthus* is inhibited by the antibody [42]. A mutant of potato, produced by mutagenesis, is both amylose-free and shows a strongly reduced activity of granule-bound starch synthase and the loss of the 60kDa *waxy* protein [12]. Experiments in which antisense granule-bound starch synthase RNA has been expressed in potato tubers provide the best evidence that the *waxy* protein is a granule-bound starch synthase [41].

Analysis of starch synthases in developing pea cotyledons suggests that starch synthase activity is associated with a 77 kDa protein and not with the antigenically unrelated 59 kDa *waxy* protein [39]. These proteins have been termed GBSS1 and GBSSII respectively [5], and the authors propose that they represent functionally distinct isoforms of starch synthase which make different contributions to starch synthesis.

Purification of the 60 kDa protein from maize has enabled the isolation of cDNA clones and genomic clones [36, 15] and *waxy* protein (GBSSI) genes have now been isolated from several other plant species. These include cDNA clones from pea [5], potato [40], barley [31] and wheat [4] and genomic clones from potato [32, 40], rice [43], barley [31] and wheat (C.C. Ainsworth, unpublished). The GBSSII genes of both pea and potato have also been isolated [5]. In this study, we present an analysis of the nature and expression of a GBSSI gene and its encoded protein from the wheat endosperm. The sequence of the cDNA clone has recently been published [4]. We also compare the available GBSSI sequences and that of *Escherichia coli* glycogen synthase.

Materials and methods

Plant materials

Hexaploid bread wheat (*Triticum aestivum*) cv. Chinese Spring and the relevant nullisomic-

tetrasomic and ditelosomic aneuploid derivatives [33, 34, 35] were grown under glasshouse conditions. Developing grains were dissected from wheat ears, snap-frozen in liquid nitrogen and stored at −70 °C.

Isolation of DNA

Leaves were frozen in liquid nitrogen and rapidly ground to a fine powder in a coffee grinder or pestle and mortar, suspended in 2 volumes of 100 mM Tris-HCl pH 8.5, 100 mM NaCl, 50 mM EDTA, 2% SDS, 0.05 mg/ml proteinase K and incubated at 60 °C for 1 to 2 h. The mixture was extracted with phenol/chloroform and precipitated with 0.6 volumes of isopropanol. The precipitate was dissolved in TE (10 mM Tris-HCl pH8, 1 mM EDTA) and RNA was removed by digestion with 1 μg/ml RNase A for 30 min at 37 °C. After extraction with phenol/chloroform, the DNA was ethanol-precipitated and dissolved in TE.

Southern analysis

Digested wheat DNA (8 μg) was fractionated on 0.8% TAE agarose gels and transferred to nylon membranes (Hybond N, Amersham) as described by Maniatis *et al.* [23]. Blots were prehybridised in 0.6 M NaCl, 20 mM PIPES pH 6.8, 4 mM EDTA, 0.2% gelatin, 0.2% PVP, 0.2% Ficoll 400, 1% SDS containing 500 μg/ml denatured autoclaved salmon sperm DNA for 4 h at 65 °C. The *waxy* cDNA was released from pcSS22 [4] by digestion with *Eco* RI, purified by electrophoresis and spin dialysis through Sepharose CL-6B (Pharmacia) and labelled with ^{32}P-dCTP by primer extension [7]. Unincorporated nucleotides were removed by spin dialysis through Sepharose CL-6B. Prehybridised blots were hybridised with the labelled probe in the buffer described above for 16 h at 65 °C. Blots were washed with 2× SSC, 0.1% SDS (2× 15 min) followed by 0.2× SSC, 0.1% SDS (2× 15 min) and then exposed to Kodak XAR-5 film between intensifying screens at −70 °C.

Isolation of RNA

Frozen developing grains and leaves were ground thoroughly in a pre-cooled coffee grinder and the frozen powder suspended in 3 volumes of 100 mM Tris-HCl pH 9.0, 200 mM NaCl, 5 mM DTT, 1% Sarcosyl, 20 mM EDTA before dispersing by grinding with a Polytron for 30 s. The mixture was extracted once with phenol/chloroform/isoamyl alcohol (50:50:1) and twice with chloroform/isoamyl alcohol (50:1). The solution was adjusted to 2 M with 8 M LiCl solution and the RNA was allowed to precipitate at 4 °C overnight. The RNA was pelleted by centrifugation (10 000 × ***g*** for 10 min at 4 °C), washed twice in 2 M LiCl by resuspension and centrifugation (10 000 × ***g*** for 10 min at 4 °C) before dissolving in TE. Insoluble material was removed by centrifugation before the RNA was ethanol-precipitated and dissolved in water.

Poly(A)$^+$ RNA was selected from total RNA using two rounds of oligo-d(T) cellulose chromatography [22].

Northern analysis

Total or Poly(A)$^+$ RNA was electrophoresed in 1.2% agarose formaldehyde gels and transferred to nylon membranes without pre-treating the gels [23]. Blots were hybridised with labelled probe as for Southern blots (above) except that poly(A) was included in the hybridisation and prehybridisation solutions to prevent non-specific hybridisation.

Hybrid select translation

Circles of DBM paper (Transa-Bind, Schleicher and Schuell) were converted to the active form by washing on sterile foil with ice-cold 1.2 M HCl, immersing in freshly prepared 1.2 M HCl/1% sodium nitrite solution for 45 min on ice, washing twice in ice-cold sterile water followed by a wash in ice-cold 0.2 M sodium acetate pH 4.0. 10 μg of plasmid DNA, dissolved in 0.2 M sodium acetate pH 4.0 was incubated at 80 °C for 1 min

and 40 μl DMSO added before mixing and pipetting onto individual filters contained in sterile siliconised 5 ml bottles. After overnight incubation at room temperature, filters were washed three times in sterile water at 37 °C, four times in 0.4 M NaOH at 37 °C and four times in sterile water. 200 μl storage buffer (50% deionised formamide, 0.1% SDS, 0.2mg/ml poly(A), 0.2mg/ml calf liver tRNA, 100 mM sodium phosphate pH 7.0, 0.6 M NaCl) was added and the filters stored at 4 °C. After removal of the storage buffer, 100 μl hybridisation buffer (50% deionised formamide, 40 mM PIPES pH 6.4, 0.9 M NaCl, 1 mM EDTA, 0.4% SDS, 0.05 mg/ml poly(A)) containing 20 μg total RNA was added and the filters hybridised at 41 °C overnight. The hybridisation solution was removed and the filters washed four times for 20 min at 32 °C in 50% deionised formamide, 8 mM sodium citrate pH 7.6, 20 mM NaCl, 1 mM EDTA, 0.2% SDS. The bound RNA was eluted by washing twice in 90% formamide, 20 mM PIPES pH6.4, 1 mM EDTA, 0.5% SDS at 60 °C for 20 min. The selected RNA was ethanol-precipitated and dissolved in 5 μl sterile water.

2 μl of selected RNAs were translated *in vitro* using a rabbit reticulocyte lysate kindly provided by Dr M.F. Tuite (University of Kent, Canterbury, UK). Translation was carried out for 90 min at 30 °C in a reaction volume of 25 μl containing 12.5 μl of lysate, 100 μM amino acids (minus Met), 1 mM ATP, 200 μM GTP, 12.5 mM creatine phosphate, 5 U creatine phosphokinase, 2 mM magnesium acetate, 75 mM potassium chloride, 20 mM Tris-HCl pH 7.6, 3 mM glucose and 5 μCi [^{35}S] – methionine (> 1000 Ci/mmol, Amersham). After translation, samples were treated with 2.5U RNase A for 10 min at 37 °C.

Labelled polypeptides were resolved on 10% polyacrylamide gels using the Laemmli [19] buffer system and visualised by fluorography.

Isolation of starch granules from wheat grains

Wheat grains (100 g harvested at 15 days after anthesis) were removed from their husks and were homogenised in 200 ml ice-cold buffer (50 mM Tris-acetate pH 7.5, 1 mM EDTA, 1 mM DTT) using a Polytron homogeniser. The homogenised tissue was squeezed through four layers of muslin and the filtrate centrifuged at 10 000 × *g* for 10 min at 4 °C. After discarding the supernatant, loose material was brushed from the surface of the pellet and after removal of the grey layer of gluten, the starch pellet was resuspended in 50 ml of homogenisation buffer and centrifuged as before. This procedure was repeated twice more. The starch preparation was washed four times in cold (−20 °C) acetone, each time the starch being resuspended thoroughly in 100 ml acetone and allowed to settle at −70 °C for 30 min. After the final wash the acetone was removed by aspiration, the starch was allowed to air-dry and was stored at −20 °C.

SDS-PAGE of starch granule-bound proteins

50 mg starch was resuspended in 0.5 ml deionized water. After addition of 2 × sample buffer (100 mM Tris-Cl pH 6.8, 2% SDS, 10% glycerol, 5% 2-mercaptoethanol), the suspension was boiled for 5 minutes and the gelled starch pelleted by centrifugation at 10 000 × *g* for 10 min. SDS-PAGE of the solubilised protein was carried out using 7.5% polyacrylamide gels and the discontinuous Tris-glycine buffer system of Laemmli [19]. 10 mM 2-mercaptoethanol was included in the upper buffer tank as a free-radical scavenger. Coloured molecular weight standards (Amersham) were used to allow identification of the strip of gel containing the *waxy* protein band without the need to stain the gel.

Preparation of protein samples for electroblotting and protein sequencing

Solubilised starch granule-bound proteins were separated by SDS-PAGE. Strips containing the *waxy* protein were cut from the gels and proteins were electroeluted into a dialysis bag at 10 V/cm for 2 h. At the end of that period the current was

reversed for 1 minute. The eluted proteins were concentrated using an Amicon microconcentration system (M_r cut-off 10 000). The concentrated protein was electrophoresed on a 7.5% polyacrylamide gel and proteins were electroblotted onto Immobilon PVDF membrane (Millipore) at 2.5 mA per cm^2 membrane for 30 min. After transfer, the membrane was stained for one min in 0.1% Coomassie blue G250, 40% methanol, 1% acetic acid. The membrane was destained in two changes of 50% methanol, 10% acetic acid for 2 min each. Finally, the membrane was washed in deionized water for 5 min and then allowed to air-dry.

Sequencing of the immobilized *waxy* protein was performed by Dr L. Packman (Department of Biochemistry, University of Cambridge, UK) using an Applied Biosystems 477A protein sequencer.

DNA and protein sequence analysis

Computer analysis of the sequence of the cDNA insert of pcSS22 was carried out using the DNASTAR programs Align, Mapdraw and Protean. Multiple alignments were done using the Intelligenetics program Genalign.

Results and discussion

Cloning the waxy *locus*

A cDNA library of cDNAs synthesised from poly(A)$^+$ RNA isolated from developing wheat grains (cv. Chinese Spring) harvested at 20 days post anthesis (dpa) and cloned into λgt10 was screened with the insert from pcwx27, a cDNA clone representing the barley *waxy* protein mRNA [31]. Of the clones which hybridised strongly to the barley probe, the cDNA inserts ranged in size from a few hundred nucleotides to 2.2 kb. A number of less strongly hybridising clones were isolated which included two with insert sizes of 3.3 kb and 4.4 kb respectively.

Hybridisation of the insert from pcwx27 to northern blots of RNA isolated from developing wheat grains harvested at 20 dpa showed an abundant transcript of 2.3 kb (data not shown). It was assumed that the 2.2 kb cDNA clone represented the full-length wheat *waxy* mRNA and the cDNA insert from this clone was subcloned into pUC18. The resultant plasmid was designated pcSS22. The cDNA insert of 2186 nucleotides appeared to be a full-length copy of the mRNA and had an open reading frame of 1845 nucleotides. The sequence of the GBSSI insert contained in the full-length 2.2 kb cDNA clone, pcSS22, has been presented elsewhere [4].

Organisation of GBSSI genes in the wheat genome

The hexaploid nature of cultivated wheat has enabled the development of a wide range of genetic stocks such as the nullisomic-tetrasomic (NT) and ditelocentric aneuploid lines of Chinese Spring in which the absence of whole chromosomes or chromosome arms can be assessed phenotypically. These lines, developed by E.R. Sears [33, 34, 35] are of great value in enabling the chromosomal locations of genes to be determined. DNA samples were isolated from the set of nullisomic-tetrasomic Chinese Spring lines such that each chromosome was represented as a nullisomic and 8 μg of each sample was digested with *Eco* RI, *Bam* HI and *Hin*d III before electrophoresis. Hybridisation of Southern blots of these gels with the oligolabelled insert from pcSS22 showed that in euploid Chinese Spring the main hybridising fragments can be assigned to chromosomes 4A, 7A or 7D. Figure 1 shows an autoradiograph of a blot of *Eco* RI digested DNA from representative NT lines of the group 4, 6 and 7 chromosomes. The three major hybridising fragments (with sizes of 21 kb, 8.5 kb and 7.5 kb) are absent in the NT lines lacking chromosomes 7A, 4A and 7D respectively. The two weaker bands (16 kb and 9.5 kb), which were unable to be assigned chromosomal locations, may result from the comigration of similar sized fragments from more than one chromosome. Analysis of the ditelosomic lines showed that the loci are located on

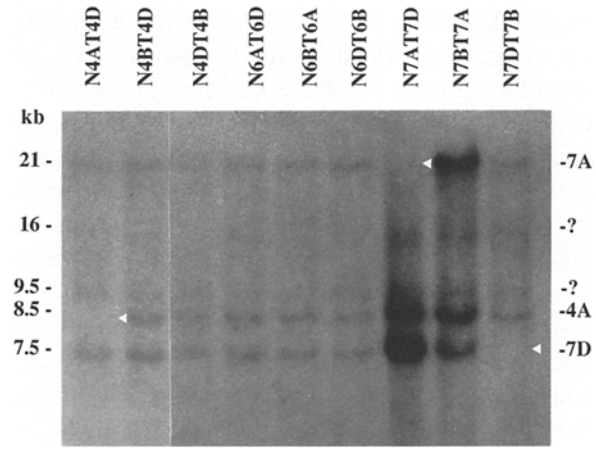


Fig. 1. Waxy gene organisation. Genomic Southern blot of *Eco* RI-digested DNA from euploid CS wheat and chromosome group 4, 6 and 7 nullisomic-tetrasomic (N--T--) genotypes hybridised with the cDNA insert from pcSS22. Fragments missing as a result of the removal of whole chromosomes from chromosome groups 4 and 7 are arrowed. The group 6 NT lines are included as an example of chromosome set in which nullisomy has no effect on the fragments hybridising to the pcSS22 probe.

the long arm of chromosome 4A, the short arm of chromosome 7A and the short arm of chromosome 7D (data not shown). Copy number reconstructions (not shown), in conjunction with the nullisomic data, indicate that there is one gene at each locus. The genes encoding the *waxy* proteins of other monocot plants, for example rice and maize [10, 36], have also been shown to be single-copy.

The fact that the loci are not carried on the three chromosomes of a single homeologous group, but are apparently split between groups 4 and 7 can be explained by the existence of a translocation between chromosomes 4A and 7B which has been identified by the presence of other genes, including isozyme markers [2, 16] and RFLP loci [21], which have translocated positions. Analysis of chromosome pairing also demonstrates that the terminal region of chromosome 4AL is homeologous to 7AS and 7DS [24]. Although we have not used pcSS22 in RFLP studies to map the loci, these data show that the locus on chromosome 4A is located on the terminal segment, which is distal to another translocated region (from 5AL [21]). This translocation has also been identified by the non-homeologous locations of marker genes [1].

GBSSI gene expression

In order to characterise the expression profile of the *waxy* genes in the wheat plant, northern blots of RNA isolated from developing grains and leaves were probed with the cDNA insert from pcSS22. Northern blots of poly(A)$^+$ RNA isolated from wheat grains harvested over the period of grain development (5, 10, 15, 20 and 25 dpa) show that the GBSSI genes are active throughout grain filling where the main 2.3 kb transcripts accumulate to high levels (Fig. 2). It is thought that the slight reduction in transcript levels at 15 dpa (relative to the levels at 10 and 20 dpa) reflects a difference in the efficiency of selection of poly(A)$^+$ RNA rather than a reduction in transcript level. In addition to the 2.3 kb mRNA, three additional, and very much weaker, transcripts are evident. A 4.4 kb transcript appears during early to mid grain development (10 and 15 dpa) but is not evident in the later stages of grain development. Two transcripts, of around 1.6 kb,

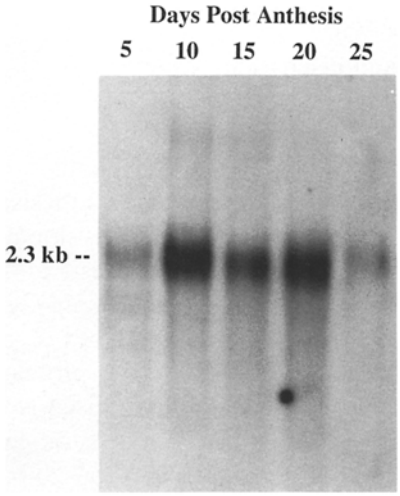


Fig. 2. Northern analysis of *waxy* locus expression during grain development. 5 μg poly(A)$^+$ RNA was run in each lane and the blot hybridised with the insert from pcSS22. Blots were washed at a stringency of 0.2 × SSC, 65 °C.

also appear early in grain development (5 dpa). Since the blots were washed under stringent conditions (0.2 × SSC, 65 °C), these transcripts may result from other less homologous starch synthase genes or may reflect some homology in functionally unrelated genes.

On northern blots containing both total leaf RNA and total grain RNA, hybridised with the insert from pcSS22 and washed under stringent conditions (0.2 × SSC, 65 °C) there is no evidence of the 2.3 kb transcript in leaves, but, instead, there is a very weak transcript of 1.6 kb which is less abundant or less homologous to the endosperm mRNA species (data not shown). In order to investigate this further, duplicate blots containing equal amounts (10 μg) of total RNA from leaves and grains were hybridised as before but washed at a range of stringencies from low (4 × SSC, 65 °C) to high stringency (0.2 × SSC, 65 °C). It is clear from this analysis (Fig. 3) that, whereas the 2.3 kb grain transcript does not change in intensity with washing stringency, the 1.6 kb leaf transcript most certainly does. At low stringency, the 1.6 kb band is clearly evident but the intensity is markedly reduced at washing stringencies below 2 × SSC. This is indicative that the mRNA in leaves is produced by separate genes and does not result from differential splicing of the RNA which produces the endosperm *waxy* mRNA.

Comparison of the expression profiles of the *waxy* (GBSSI) genes in the plants from which the genes have been cloned produces an interesting dichotomy between dicotyledonous and monocotyledonous plants. In maize and rice, in northern experiments where the blots were washed at high stringency, no mRNA homologous to the cloned *waxy* gene has been shown in leaves or other green tissues [15, 10]. Since no lower stringency experiments were performed it is not possible to say whether, as in wheat, there is a smaller homologous leaf specific transcript in maize and rice. In pea, GBSS1 transcripts accumulate to high levels in the embryo, which is the site for storage starch synthesis, and mRNAs of the same size are also detected in pods and leaves [5]. Analysis of GBSSI transcripts in potato shows that the GBSSI genes are expressed in a number of different tissues including tubers and leaves [40], the transcript size being the same in tubers and leaves. These data indicate that the enzymology of starch synthesis in the leaves of wheat, maize and rice differs from that in leaves of pea and potato. In the latter group, the same GBSSI genes appear to be involved in the synthesis of both storage starch and transient starch (for example, in leaves). In the monocot group, unrelated proteins or different isoforms of GBSSI must operate in the synthesis of transient starch in the leaves. It is tempting to speculate that the 1.6 kb leaf transcript encodes a starch synthase which is involved in the synthesis of starch in the leaf. However, no proteins of this size (the encoded polypeptide could not be larger than 50

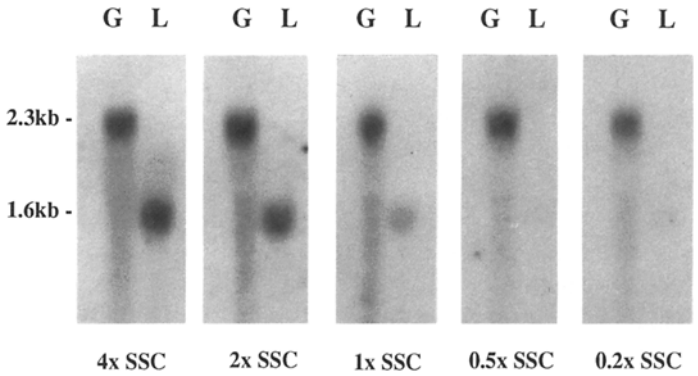


Fig. 3. Northern analysis of *waxy* locus expression in grains and leaves. 10 μg total RNA from 15 dpa grains (G) and from mature leaves (L) was run in each lane and the blots hybridised with the insert from pcSS22. Blots were washed at five different stringencies, ranging from 4 × SSC to 0.2 × SSC.

kDa) have been shown to have starch synthase activity, although in pea cotyledons, a minor 45 kDa protein, which is related immunologically to the recently described 77 kDa granule-bound starch synthase (GBSSII) protein present in pea and potato [5], has been reported [37]. Without further experiments, the function of the protein encoded by it must remain an enigma.

In pea and potato, the sets of genes encoding the GBSSII starch synthase [5] are expressed in all organs, although the level of expression in pea is much higher in the embryo than in other tissues. In addition, a smaller GBSSII transcript is present in roots and leaves as well as the normal transcript. In pea, the GBSSI and GBSSII proteins are only weakly related antigenically and the genes are expressed at different times during embryo development, the GBSSI genes being expressed earlier than the GBSSII genes [5].

The existence of another granule-bound starch synthase, in addition to the *waxy* protein, was proposed by Nelson *et al.* [26], whose analysis of starch synthase activity in extracts of *waxy* maize endosperm showed that the residual activity (10% relative to the wild type) was attributable to a protein with a lower K_m for ADP-glucose. This enzyme was most active in endosperm tissue from younger kernels. This enzyme would seem likely to be the GBSSII type.

Preliminary attempts at FPLC purification of granule-bound starch synthases from wheat granule-associated proteins (data not shown) implicate two proteins of 77 kDa and 95 kDa as starch synthases. Although activity was not associated with the 60 kDa *waxy* protein, this may simply reflect that the *waxy* protein has lost activity in the purification method used. In pea, it has similarly been difficult to demonstrate starch synthase activity associated with the 59 kDa *waxy* protein [5]. The 77 kDa protein in wheat may well be a GBSSII type. The 95 kDa protein may be homologous to the 92 kDa granule-bound protein found in maize which is immunologically related to the 60 kDa *waxy* protein but is suggested to be less abundant [22].

In summary, it seems that there are at least three starch synthases which operate in cereal endosperms, the 60 kDa *waxy* protein (GBSS1), a 77 kDa protein (probably homologous to GBSSII found in pea and potato) and a 95 kDa protein (probably homologous to the 92 kDa and protein of maize.

The wheat waxy *protein*

mRNA isolated from 20 dpa Chinese Spring grains was used in hybrid selection experiments using pcSS22. Filters carrying pUC18 DNA alone were used as a control. The selected mRNA was translated using rabbit reticulocyte lysate and the translation products electrophoresed on 10% polyacrylamide gels. Total RNA from wheat grains was also translated and the translation

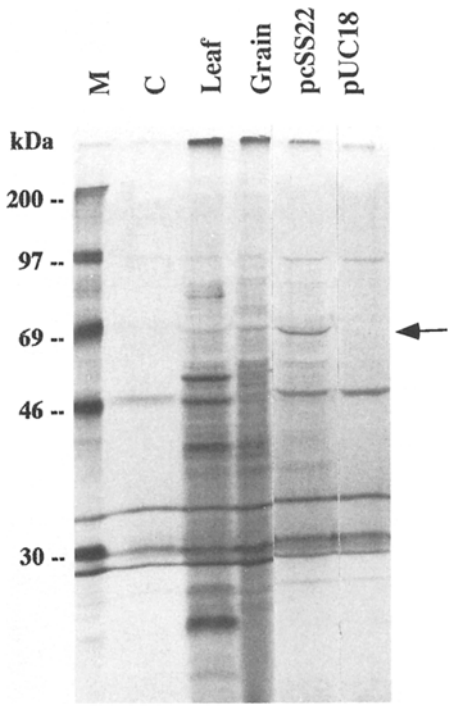


Fig. 4. Hybrid-select translation analysis of the *waxy* locus (pcSS22) cDNA clone. Translation products from mRNA selected using pcSS22 DNA were electrophoresed on a 10% polyacrylamide gel together with translation products from total wheat leaf RNA, total RNA from 15 dpa grains, selected RNA from the pUC18 plasmid control and selected RNA from control (no DNA) filters. The resultant fluorograph is shown. Lane M represents size markers. The 67 kDa selection product of pcSS22 is arrowed.

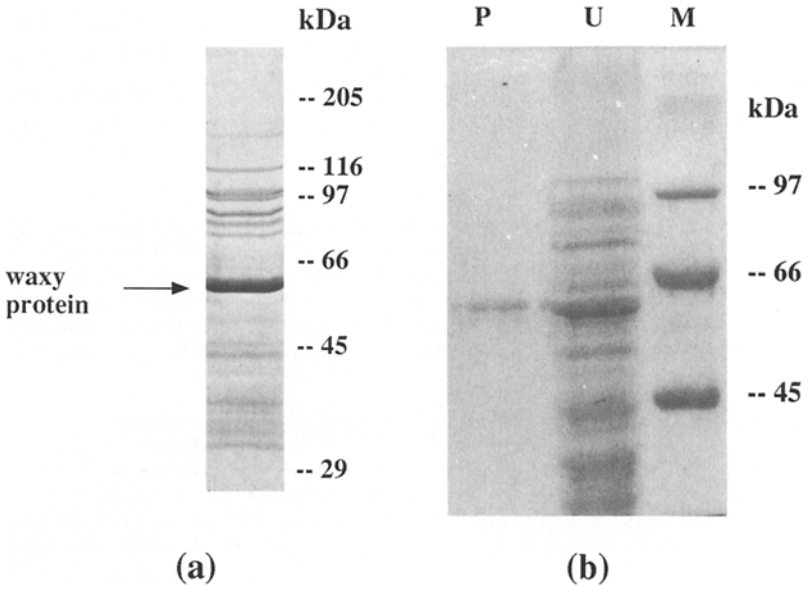


Fig. 5. (a) SDS-PAGE of proteins associated with starch granules in developing wheat grain. The 60 kDa *waxy* protein is arrowed. (b) Purification of the *waxy* protein. SDS-PAGE of unpurified proteins isolated from starch granules (U) and the purified protein (P). Lane M represents size markers.

products included on the gels together with ^{14}C-labelled molecular weight markers. The resultant fluorograph shows that a translation product with a molecular weight of approximately 67 kDa is produced specifically with mRNA selected by pcSS22 (Fig. 4). The 67 kDa protein band may correlate with a band of similar size in the translation products of total grain RNA.

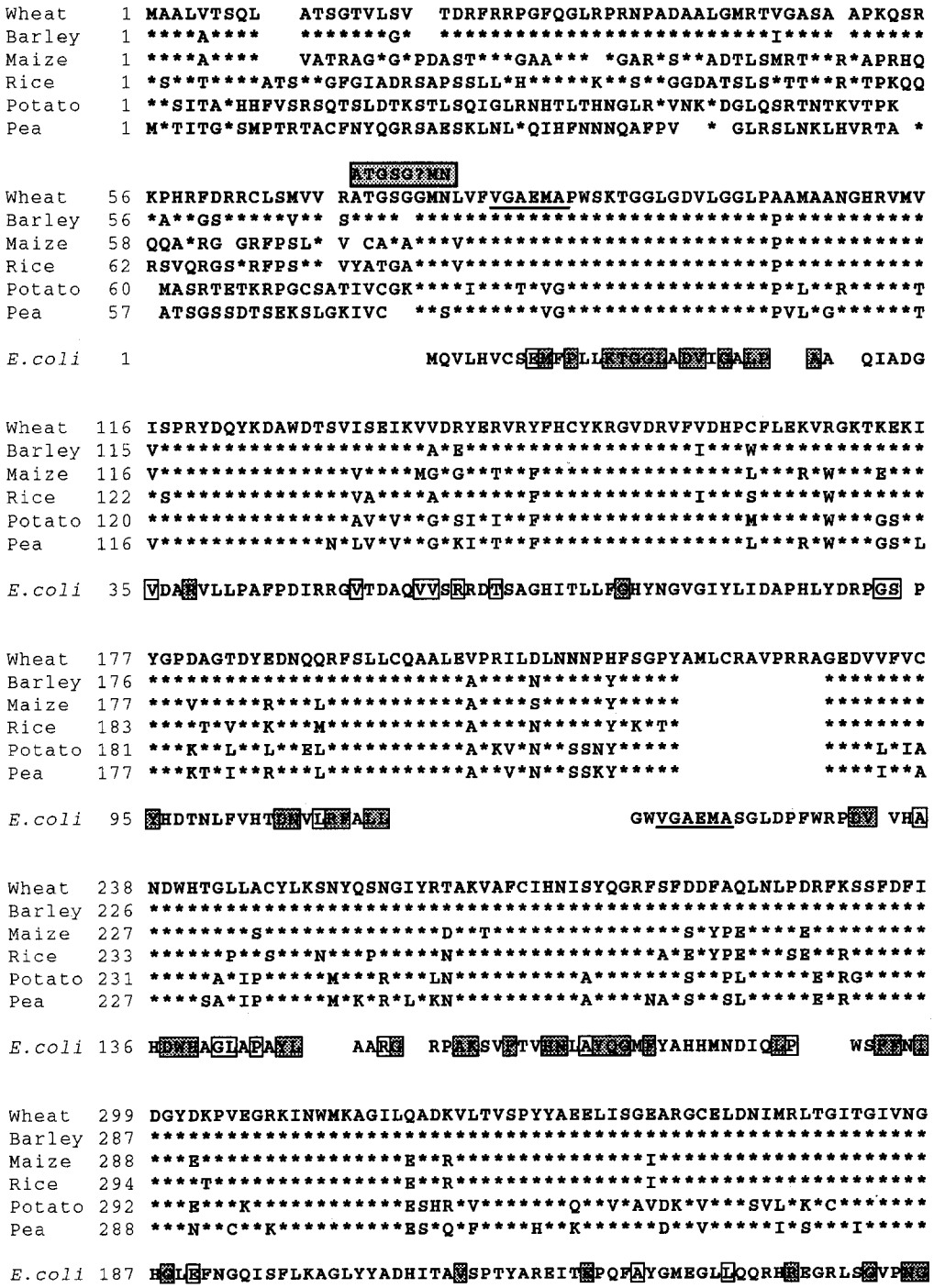
Translation of the 1845 nucleotide open reading frame of pcSS22, from the ATG (nucleotides 64–66) to the termination codon, TGA (nucleotides 1909–1911), gives a protein containing 615 amino acids with a calculated molecular weight of 67.7 kDa.

The relative abundance of the *waxy* protein and its molecular weight *in planta* was determined by SDS-polyacrylamide gel electrophoretic analysis of proteins isolated from starch granules. Starch granules were isolated from 15 dpa wheat grains and proteins solubilised from the starch. Granule-bound proteins were separated on 7.5% polyacrylamide gels. Approximately twenty protein bands were observed on the stained gels, with the most abundant protein having a molecular mass of 60 kDa (Fig. 4a). This protein was shown to be the *waxy* protein by using a polyclonal antibody raised to the potato *waxy* protein [42] to probe western blots of wheat granule associated proteins (data not shown). The discrepancy of 7 kDa between the molecular weight of the mature protein as observed on SDS gels and that given by hybrid select translation or translation of the open reading frame of pcSS22 (M_r 67 758) is presumed to represent the transit peptide. In all other plant *waxy* proteins for which clones are available, a transit peptide sequence is present with a calculated molecular weight of around 7 kDa.

Purification of the waxy *protein and the N-terminal sequence*

In order to define the transit peptide cleavage point, the *waxy* protein was purified from solubilised starch granule-bound proteins by cutting the protein band from SDS- polyacrylamide gels and concentration of the protein following electroelution (Fig. 5b). The proteins were electroblotted onto PVDF membrane and 8 residues of amino acid sequence at the N-terminus were identified

by microsequencing. The N-terminal sequence obtained was Ala-Thr-Gly-Ser-Gly-unknown-Met-Asn, where the sixth amino acid was not clear-cut, but was most likely to be glycine. Inspection of the deduced amino acid sequence shows that the identical sequence, Ala-Thr-Gly-Ser-Gly-Gly-Met-Asn, occurs, with the alanine at position 71 (Fig. 6). The position of the cleavage



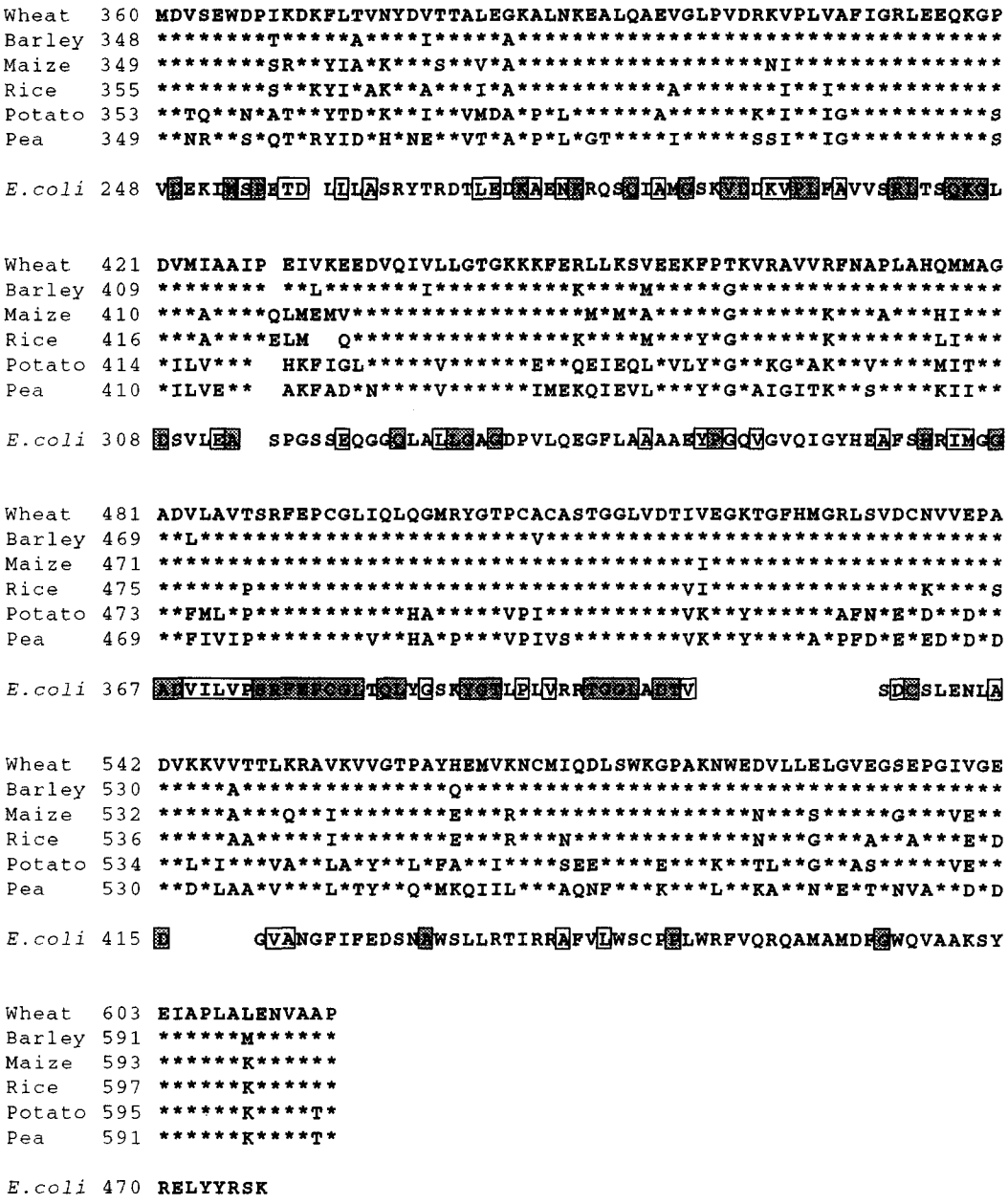


Fig. 6. Alignment of the derived amino acid sequences of the plant *waxy* proteins and *E. coli* glycogen synthase. Identical residues (relative to the wheat sequence) are represented by asterisks. Residues which are conserved between all plant sequences shown and the *E. coli* sequences are shown as filled boxes in the *E. coli* sequence. Residues which are conserved between the *E. coli* sequences and at least one plant sequence are shown as open boxes. The conserved sequence VGAEMA is underlined in the wheat and *E. coli* sequences. The point of cleavage of the transit peptide from the mature *waxy* protein from wheat was identified by comparison of 8 residues of N-terminal amino acid sequence (shown as a filled box over the wheat sequence) obtained from the purified *waxy* protein.

of the transit peptide is, therefore, between Arg_{70} and Ala_{71}, the length of the transit peptide being 70 amino acids. The predicted molecular weight of the transit peptide is 7.62 kDa, which is close to the size difference between the size of the protein on SDS gels and the deduced molecular weight.

Comparison of the transit pepides of plant *waxy* proteins for which clones are available shows that there is homology between the transit peptides of wheat, barley, maize and rice but they are significantly different from the potato and pea transit peptides. The transit peptide sequences of the proteins from wheat and barley are very similar (89%) with only 8 of the 75 residues differing. The primary structures of the four cereal transit peptides exhibit some of the features which are commonly asssociated with transit sequences (reviewed in [13]) as does the potato transit peptide, though to a lesser extent [39]. Examination of the hydrophilicity plot of the derived amino acid sequence from pcSS22 (Fig. 7) shows that there is a hydrophobic region encompassing the N-terminal twenty amino acids and two strongly hydrophilic regions (residues 20–37 and 50–63). The sequence includes a high proportion of hydroxylated amino acids, threonine and serine (totalling 17%) and of small hydrophobic amino acids, alanine and valine (totalling 21%) and has a net positive charge (Fig. 7).

Table 1. Similarity comparison (%) of the mature protein sequences of the plant GBSSI proteins and *E. coli* glycogen synthase. Amino acid sequences were compared using the Lipman-Pearson method with a gap penalty of 4 and a length penalty of 12 [19]. Amino acid sequences were deduced by translation of the open reading frames from the cDNA and genomic clones if no protein sequence was given in the reference. Transit peptide sequences were excluded from the comparison.

	Barley	Rice	Maize	Potato	Pea	*E. coli* [8]
Wheat [4]	93.3	84.2	83.4	67.3	63.3	32.4
Barley [31]		86.6	86.4	69.6	69.6	30.6
Rice [43]			88.1	70.1	67.0	32.3
Maize [15]				69.9	67.6	31.5
Potato [39]					75.4	29.6
Pea [5]						29.5

The position of the site of cleavage of the transit peptide in the preprotein, between Arg_{70} and Ala_{71} (VVR↓A) fits the consensus cleavage site motif of I/VXA/C↓A, proposed for chloroplast transit peptides [9] and is similar to the sequence around the cleavage site (identified by N-terminal sequencing) of the maize *waxy* gene (VVC↓A; [15]). Inspection of the sequences of barley and rice deduced *waxy* proteins shows that alanine residues, likely to represent the start of the mature protein, occur with similar position and sequence context.

The mature waxy *protein sequence*

Comparison of the deduced amino acid sequences (Table 1) shows that there is a large degree of similarity between the mature peptides of the six plant GBSSI proteins examined, with no two sequences showing less than 63% identity. The proteins fall into two distinct classes; the cereal proteins share in excess of 83% identity but are less similar to the two dicot proteins (potato and pea), showing between 60% and 70% identity. The pea and potato proteins share 75% identity. The proteins of wheat and barley are 93.3% similar which reflects the relatively short genetic distance between these species.

Examination of the multiple alignment of the six plant genes and the bacterial homologue, the glycogen synthase protein from *E. coli*, gives more clue as to the divergence between the proteins (Fig. 6). The primary sequences of the plant mature *waxy* proteins are conserved to a large degree as is indicated by the identity scores. The most striking difference amongst the plant sequences is that the wheat protein contains an insertion of 11 amino acids (AMLCRAVPRRA) between positions 219 and 230. A search of the SwissProt database (release 21) did not detect other proteins which contained this or very closely related peptides. The significance of this peptide is, therefore, impossible to ascertain. It is noteworthy, however, that this region has a high probability of forming an alpha helix in a surrounding area that is predicted to be predominantly beta sheet or

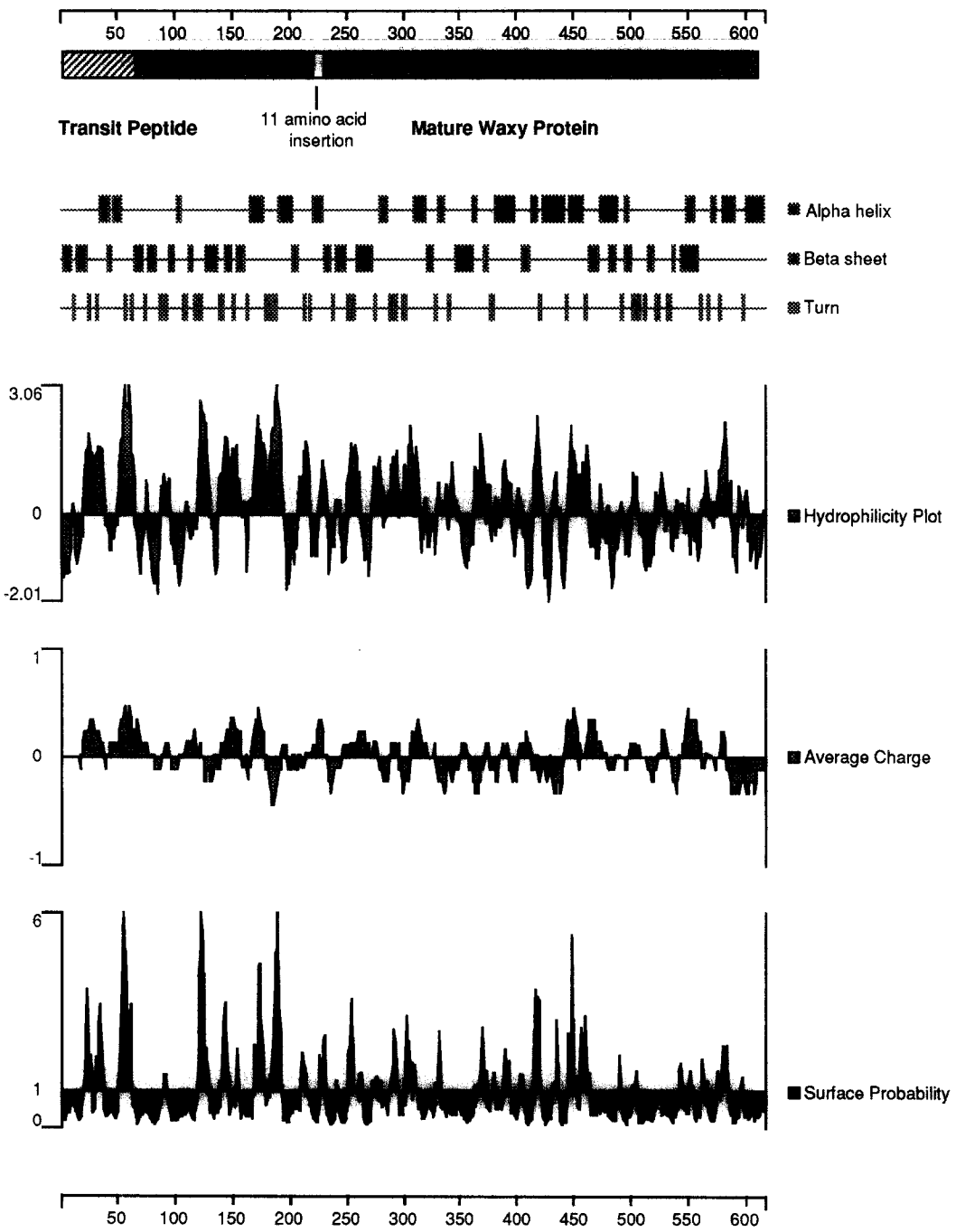


Fig. 7. Characteristics of the deduced protein from pcSS22. Hydrophilicity and charge density are calculated according to Kyte and Doolittle [17], secondary structure probabilities according to Chou and Fasman [3].

turn (Fig. 7). The peptide has a low probability of being on the surface of the protein (Fig. 7) and is consequently unlikely to be intimately involved in interactions with the substrate or other molecules. It is not known whether this insertion is confined to this particular gene or is found in all the homeoloci in the wheat genome.

The longest region of identical and contiguous

amino acid residues in the plant *waxy* proteins tested is the 16 residues towards the N-terminus of the mature protein (PWSKTGGLGDVLGGLP, residues 88–103 in the wheat sequence; Fig. 6). This conserved glycine-rich region contains the motif KTGGL which is also found in the glycogen synthase proteins of *E. coli* and animal muscle and which has been implicated in the binding site for the substrates, ADP-glucose and UDP-glucose, respectively [8]. This has previously been discussed by Preiss [29] and van der Leij *et al.* [39] and the sequence motif has also been identified in the GBSSII gene from pea [5]. The region of homology around the motif is reasonably extensive, and in the *E. coli* enzyme, the 19 amino acid region which includes the KTGGL motif contains 14 residues which are identical in one or more of the plant sequences. Two other regions of homology between the plant synthases and bacterial glycogen synthase have been noted towards the carboxy termini [39]. A region of 19 residues (GADVILVPSRFEPCGLTQL) in the *E. coli* protein occurs in the plant sequences (positions 480–498 in the wheat sequence) with 13 residues identical in all plant synthases, 5 are present in one or more of the plant sequences and only one is different (the threonine is replaced by valine in pea and by isoleucine in the others). The second region of 8 residues in the *E. coli* protein (TGGLADTV) has a counterpart in the plant sequences (positions 513–519 in the wheat sequence) where 6 residues are identical, one (valine) is present in three of the plant proteins and one, the alanine, is replaced by valine in all the plant proteins. It has been noted [39] that the part of the last conserved domain, TGGL, occurs in the substrate binding domain at the N-termini of the proteins (KTGGL) but as it lacks the lysine, it is unlikely to represent an active site.

Apart from these three domains, the plant GBSSI proteins share only around 30% identity with the *E. coli* glycogen synthase (Table 1), the remaining identical residues being spread throughout the protein sequence. A feature of interest is the sequence VGAEMA that occurs in all the plant proteins (residues 82–87 in wheat), immediately preceding the putative active site, and which occurs in the *E. coli* enzyme between residues 116 and 121. The significance of this is unknown and screening the SwissProt data base revealed no homologous sequences in other proteins.

There is now a significant body of information which provides evidence for the involvement of the *waxy* protein in the synthesis of storage starch. The correlative evidence from analysis of *waxy* mutants has been difficult to substantiate by *in vitro* experiments where enzyme activity is measured. This, however, may simply be due to the loss of activity as a result of the protein isolation procedures employed as has been suggested to be the case for the pea *waxy* protein [5]. The best evidence for the *waxy* protein being a granule-bound starch synthase comes from experiments in which antisense GBSSI RNA expressed in potato tubers was demonstrated to reduce the levels of both amylose and GBSS to zero [41]. The similarity between the amino acid sequences of the plant *waxy* proteins and the glycogen synthases, particularly in the respect that the plant proteins contain a sequence analagous to the active site of glycogen synthase is additional evidence for the role of plant *waxy* proteins as starch synthase. The relationship between GBSSI and GBSSII, which exhibit substantial sequence homology [5] and the nature and role of the starch syntheses which are active in leaves remain to be elucidated.

Acknowledgements

This work was supported by a grant from the Agriculture and Food Research Council. We are very grateful to Dr Len Packman (Department of Biochemistry, Cambridge, UK) for performing the N-terminal microsequencing, to Dr Alison Smith (John Innes Institute, Norwich, UK) for help in protein purification and to Dr G Scheperkeuter (Department of Biochemistry, University of Groningen, Netherlands) for kindly donating the antibodies raised to the potato *waxy* protein. We are very grateful to Vicky Buchanan-Wollaston for discussions and critical reading of the manuscript.

References

1. Ainsworth CC, Gale MD, Baird S: The genetics of β-amylase isozymes in wheat. I. Allelic variation among hexaploid varieties and intrachromosomal gene locations. Theor Appl Genet 66: 39–49 (1983).
2. Benito C, Perez de la Vega M: The chromosomal location of peroxidase isozymes of the wheat kernel. Theor Appl Genet 55: 73–76 (1979)
3. Chou PY, Fasman GD: Prediction of the secondary structure of of proteins from their amino acid sequence. Adv Enzymol 47: 45–147 (1978)
4. Clark JR, Robertson M, Ainsworth CC: Nucleotide sequence of a wheat (*Triticum aestivum* L.) cDNA clone encoding the *waxy* protein. Plant Mol Biol 16: 1099–1101 (1991)
5. Dry I, Smith AM, Edwards E, Bhattacharya M, Dunn P, Martin C: Characterisation of cDNAs encoding two isoforms of granule-bound starch synthase which show differential expression in developing storage organs of pea and potato. Plant J 2: 193–202 (1992)
6. Echt CS, Schwartz D: Evidence for the inclusion of controlling elements within the structural gene at the *waxy* locus in maize. Genetics 99: 275–284 (1981)
7. Feinberg AP, Vogelstein B: A technique for radio-labelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137: 266–267 (1984).
8. Furukawa K, Tagaya M, Preiss J, Fukui Y: Identification of lysine 15 at the active site in E.coli glycogen synthase. J Biol Chem 265: 2086–2090 (1990).
9. Gavel Y, von Heine G: A conserved cleavage-site motif in chloroplast transit peptides. FEBS Lett 261: 455–458 (1990).
10. Hirano H, Sano Y: Molecular characterisation of the *waxy* locus of rice (*Oryza sativa*). Plant Cell Physiol 32: 989–997 (1991).
11. Hixon RM, Brimhall B: The waxy cereals and starches which stain red with iodine. In: Radley JA (ed) Starch and its Derivatives, pp. 252–290. Chapman and Hall, London (1953).
12. Hovenkamp-Hermelink JHM, Jacobsen E, Ponstein AS, Visser RGF, Vos-Scheperkeuter GH, Bijmolt EW, de Vries JN, Witholt B, Feenstra WJ: Isolation of an amylose-free mutant of the potato (*Solanum tuberosum*). Theor Appl Genet 75: 217–221 (1987).
13. Keegstra K, Olsen LJ: Chloroplastic precursors and their transport across the envelope membranes. Annu Rev Plant Physiol Plant Mol Biol 40: 471–501 (1989)
14. Keeling PL, Wood JR, Tyson RH, Bridges IG: Starch synthesis in the developing wheat grain. Evidence against the direct involvement of triose phosphates in the metabolic pathway. Plant Physiol 87: 311–311 (1988).
15. Klösgen RB, Gierl A, Schwarz-Sommer Z, Saedler H: Molecular analysis of the *waxy* locus of maize. Mol Gen Genet 203: 237–244 (1986).
16. Kobrehel K, Feillet P: Identification of genomes and chromosomes involved in peroxidase synthesis of wheat seeds. Can J Bot 53: 2336–2344 (1975).
17. Kumar A, Larsen CE, Preiss J: Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* ADP-glucose: a 1,4-glucan, 4-glycosyltransferase as deduced from the nucleotide sequence of the glgA gene. J Biol Chem 261: 16256–16259 (1986).
18. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132 (1982).
19. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685 (1970).
20. Lipman DJ, Pearson WR: Rapid and sensitive protein similarity searches. Science 227: 1435–1441 (1985).
21. Lui CJ, Atkinson CN, Chinoy CN, Devos KM, Gale MD: Nonhomoeologous translocations between group 4, 5 and 7 chromosomes within wheat and rye. Theor Appl Genet 83: 305–312 (1992).
22. MacDonald FD, Preiss J: Partial purification and characterisation of granule-bound starch synthases from normal and *waxy* maize. Plant Physiol 78: 849–852 (1985).
23. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1982).
24. Naranjo T, Roca A, Goicoecha PG, Giraldz R: Arm homoeology of wheat and rye chromosomes. Genome 29: 873–882 (1987).
25. Nelson OE, Rhines HW: The enzymatic deficiency in the *waxy* mutant of maize. Biochem Biophys Res Commun 9: 297–300 (1962).
26. Nelson OE, Chourey PS, Chang MT: Nucleoside diphosphate sugar-starch glucosyl transferase activity of *wx* starch granules. Plant Physiol 62: 383–386 (1978).
27. Okita TW, Greenberg E, Kuhn DN, Preiss J: Subcellular localisation of the starch degradative and biosynthetic enzymes of spinach leaves. Plant Physiol 64: 187–192 (1979).
28. Okuno K, Sakaguchi S: Inheritance of starch characteristics in endosperm of *Amaranthus hypochondriacus*. J Hered 73: 467 (1982).
29. Preiss J: Biology and molecular biology of starch synthesis and its regulation. Oxford Surv Plant Mol Cell Biol 7: 59–114 (1990).
30. Robinson N, Preiss J: Localisation of carbohydrate metabolising enzymes in guard cells of *Commelina communis*. Plant Physiol 85: 360–364 (1987).
31. Rohde W, Becker D, Salamini F: Structural analysis of the *waxy* locus from *Hordeum vulgare*. Nucl Acids Res 16: 7185 (1988).
32. Rohde W, Becker D, Kull B, Salamini F: Structural and functional analysis of two *waxy* gene promoters from potato. J Genet Breed 44: 311–315 (1990).
33. Sears ER: The aneuploids of common wheat. Mo Agric Extn Bull 572: 1–58 (1954).

34. Sears ER: Nullisomic-tetrasomic combinations in hexaploid wheat. In: Riley R, Lewis KR (eds) Chromosome Manipulation and Plant Genetics, pp. 29–45. Oliver and Boyd, London (1966).
35. Sears ER: Chromosome mapping with the aid of telocentrics. Proc 2nd Int Wheat Genet Symp. Heriditas (suppl) 2: 370–381 (1966).
36. Shure M, Wessler SR, Federoff N: Molecular identification and isolation of the *waxy* locus in maize. Cell 35: 225–233 (1983).
37. Smith AM: Evidence that the *waxy* protein of pea (*Pisum sativum* L.) is not the major starch-granule-bound starch synthase. Planta 182: 599–604 (1990).
38. Sprague GF, Brimhall B, Hoxon RM: Some effects of the waxy gene in corn on properties of endosperm starch. J Am Soc Agron 35: 817–822 (1943).
39. van der Leij FR, Visser RGF, Ponstein AS, Jacobsen E, Feenstra WJ: Sequence of the structural gene for granule bound starch synthase of potato (*Solanum tuberosum* L.). Mol Gen Genet 228: 240–248 (1991).
40. Visser RGF, Hergersberg M, van der Leij FR, Jacobsen E, Witholt B, Feenstra WJ: Molecular cloning and partial characterisation of the gene for granule-bound starch synthase from a wild-type and an amylose-free potato (*Solanum tuberosum* L.). Plant Sci 64: 185 (1989).
41. Visser RGF, Somhorst I, Kuipers GJ, Ruys NJ, Feenstra WJ, Jacobsen E: Inhibition of the expression of the gene for granule-bound starch synthase in potato by antisense constructs. Mol Gen Genet 225: 289–296 (1991).
42. Vos-Scheperkeuter GH, de Boer W, Visser RGF, Feenstra WJ, Witholt B: Identification of granule-bound starch synthase in potato tubers. Plant Physiol 82: 411–416 (1986).
43. Wang Z, Wu Z, Zheng F, Guo X, Zhang W, Hong M: Nucleotide sequence of the rice *waxy* gene. Nucl Acids Res 18: 5898 (1990).