Theor Appl Genet (2001) 102:431–439

F.R. Murray · J.H. Skerritt · R. Appels

A gene from the *Sec2* (*Gli-R2*) locus of a wheat 2RS.2BL chromosomal translocation line

Received: 31 February 2000 / Accepted: 21 May 2000

Abstract A genomic DNA clone coding for a rye secalin gene (gSec2A) was isolated from a wheat translocation line carrying the 2RS.2BL chromosome, using a previously identified partial secalin (Sec2) cDNA clone as a probe. The predicted N-terminal amino-acid sequence of the gSec2A gene was identical to the N-terminal sequence obtained for Sec2 polypeptide bands isolated from SDS-PAGE gels. Bacterially expressed gSec2A protein was identical in size to that of the smallest Sec2 polypeptide band observed on SDS PAGE gels and is recognized by a monoclonal antibody specific for Mr 75000 2RS γ-secalins. Overall, the predicted protein sequence of gSec2A was most similar (50%) to the family of γ-gliadins and consists of a short N-terminal region containing one cysteine residue followed by a glutamine/proline-rich repetitive domain and a long C-terminal domain containing eight cysteine residues. The repetitive domain can be divided into two regions. One region coded for 15 units, each consisting of eight amino acids similar in sequence to that found in the ω-secalins and C-hordeins. The second region coded for 17 units each consisting of a sequence of 7–10 amino acids similar to that observed in γ-gliadins.

Keywords Secalin · Sec2 gene · 2RS.2BL translocation · Seed storage proteins

Introduction

Secalin proteins are an alcohol-soluble (prolamin) group of seed storage proteins present in rye grain and can be classified into four major groups; the high-molecular-weight (HMW) secalins, the Mr 40000 γ-secalins, the Mr 75000 γ-secalins (2RS γ-secalins) and the ω-secalins (reviewed by Shewry et al. 1994). Purification and characterization of some of these polypeptides have indicated that they are structurally related to the prolamin proteins of wheat (gliadins and glutenins) and barley (hordeins). The 2RS γ-secalins are the most abundant group of secalins, accounting for approximately 50% of the total secalins (Shewry et al. 1983a) and are of particular interest because the 2RS.2BL translocation can produce additional polymeric glutenin providing potentially superior dough properties (Gupta et al. 1989).

The Sec2 (Gli-R2) locus encoding the 2RS γ-secalin proteins is located relatively close (2.9 cM, Alonso-Blanco et al. 1993) to the telomeric C-band on the short arm of chromosome 2R. Several 2RS.2BL translocation lines have allowed the Sec2 locus from Imperial rye to be introduced into several wheat backgrounds (Jagannath and Bhatia 1972; May and Appels 1980; Gupta et al. 1989) and the analysis of grain from these lines suggested that this locus, or linked genes, may contribute to an increased amount of protein in the mature grain (Feldman 1988; Gupta et al. 1989; Stoddard et al. 1991). However, these lines have reduced fertility and/or grain yield, limiting their use in breeding programs. Between 3–6 genes are believed to be present at the 2RS secalin locus (Hull et al. 1992) and 3–4 polypeptide bands are assigned to this locus using SDS-PAGE. The effects of the proteins from the Sec2 locus on wheat-flour quality have not been clearly determined and, in order to carry out detailed studies, a full-length genomic clone of one of the genes at the locus was isolated. This gene when expressed in bacteria, produced a protein equal in size to that of the smallest Sec2-encoded polypeptide. Analysis of the predicted protein sequence provides some indication of how it may act to modify wheat flour properties. As this gene has limited homology to other wheat seed-storage proteins it is an excellent candidate for introduction into wheat. Transgenic wheat highly expressing this protein could be readily identified and a functional analysis performed to gain a greater understanding of dough properties.

F.R. Murray (✉) · J.H. Skerritt · R. Appels
CSIRO Plant Industry, P.O. Box 1600, Canberra City ACT 2601, Australia
e-mail: fmurray@pi.csiro.au
Fax: +61-2 62465000

Materials and methods

Plant material

Triticum aestivum L. cultivars Hartog and Gabo, as well as a translocation line, 2RS.2BL (May and Appels 1984), were used in this study. The translocation line 2RS.2BL was isolated in the wheat cv Gabo background. Plants were cultivated in 15-cm plastic pots in a glasshouse without supplementary lighting.

Amino acid-sequencing

Proteins were extracted from 2RS.2BL wholemeal in 50% isopropanol for 1 h at 21–25°C. Following centrifugation (10 min, 15000 g) the supernatant was extracted with 125 mM Tris-HCl, pH 7.5, containing 1% (w/v) DTT for 30 min at 37°C. After centrifugation (as above), the polypeptides were alkylated and fractionated on a 12% SDS-PAGE gel as described in Sissons et al. (1998). The gel was stained with 0.1% Coomassie Blue R-250 in 40% methanol/10% acetic acid, de-stained in 40% methanol/10% acetic acid and the desired bands excised with a razor blade, placed in separate tubes and washed 4–5 times in distilled water until no colour was observed. Protein was eluted from gel fragments and processed as described in Sissons et al. (1998).

cDNA library construction and screening

Total RNA was isolated from 16 to 20 day-old 2RS.2BL developing grains according to the method of Chandler et al. (1983). To remove contaminating starch, RNA was re-suspended in 0.4 M sodium acetate/0.5 vol of ethanol, centrifuged for 15 min (15000 g), and RNA in the supernatant precipitated with 1.5 vol of ethanol. Polyadenylated mRNA was separated from mRNA using magnetic beads (Dynal), first and second cDNA strand synthesis was performed following the manufacturer's instructions (Stratagene) and the cDNA cloned into a Lambda Zap II vector. Approximately 30000 plaques were hybridized with a pSc503 (Kreis et al. 1985b) DNA probe labelled with ^{32}P-dCTP by random primer extension using an Amersham Multiprime DNA labelling kit. Filters were pre-hybridized and hybridized at 42°C using the solutions described in Murray et al. (1997). Post-hybridization washes were performed at 65°C with 2× SSC/0.1% SDS and 0.1× SSC/0.1% SDS (1× SSC: 0.15 M NaCl plus 0.015 M sodium citrate). Over one-hundred hybridizing plaques were identified in the primary screen and lambda DNA purified (Sambrook et al. 1989) from 15 plaques containing cloned DNA of at least 800 bp in size. Insert DNA was subcloned into the NotI restriction site of pBluescript SK$^-$ (Stratagene) instead of using phagmid excision as this methodology gave rise to plasmids with shortened inserts. Sequences were determined using the automated ABI sequencing system with dye terminators, following the manufacturer's instructions. Analysis of DNA sequences was performed using GCG software (Devereaux et al. 1984).

Genomic DNA gel blot analysis

Plant genomic DNA was extracted from leaf tissue according to the method of Lagudah et al. (1991). DNA gel-blot analysis was carried out as described in Witrzens et al. (1998) except that following hybridization, filters were washed twice with 2× SSC/0.1% SDS at 55°C then twice with 0.1× SSC/0.1% SDS at 65°C. Filters were hybridized with a 354-bp probe corresponding to the 3′ end of cSec2. The probe was labelled with digoxigenin by polymerase chain reaction (PCR)-amplification using the 0.5-kb *Pst*I/*Hin*dIII fragment from cSec2 as a template. This DNA fragment had previously been purified from a 0.8% Seaplaque agarose gel according to the method of Thuring et al. (1975). PCR was carried out using the conditions described in Witrzens et al. (1998) with the following primers: S1, 5′ GAGTGCCAGGTGATGCAAC 3′ and S3, 5′ CCTTGACTTTTCAGTGGCCA 3′.

RNA gel-blot analysis

RNA was extracted from the developing grain as indicated above. RNA gel-blot analysis was carried out as described in Dolferus et al. (1994) loading 10 µg of total RNA per lane. Filters were hybridized at 42°C with the ^{32}P-labelled pSc503 DNA probe described above.

Partial genomic library construction and screening

2RS.2BL plant genomic DNA, extracted as indicated above, was digested with *Eco*R1 before being fractionated on a glycerol gradient as described in Sambrook et al. (1989). An aliquot from each fraction was analysed by gel-blot analysis, as above, using the 354-bp digoxigenin-labelled cSec2 probe to identify fractions containing hybridizing DNA fragments. The desired fractions were pooled, extracted with phenol/chloroform and ethanol-precipitated. DNA was ligated into Lambda Zap II and approximately 100000 plaques screened with the ^{32}P-labelled cSec2 probe. Two hybridizing plaques were identified; however, only one plaque hybridized strongly following secondary screening. DNA from this plaque was excised with *Eco*R1, subcloned into pBluescipt SK$^-$ (Stratagene) and sequenced.

Bacterial expression of gSec2A

To create the expression construct pet-gSec2 A, PCR was used to introduce a *Bsp*HI and *Bam*HI restriction site into gSec2A at the end of the secretion-signal peptide and 15 nucleotides downstream from the TGA stop codon respectively. PCR was carried out using 1 pg of template DNA, and 1×10^{-5} µM each of the primer and 'supermix' reaction mix (GibcoBRL). The primers used were; *Bsp*HI-A: 5′GCGACTACCATCGCCATCATGAA-CATGCAAGTCAACCC-TAGTGG3′ and Sbam2: 5′AGTGGATCCCTTGACTTT-TCAGT-GGCCAA3′. Cycling parameters were: four cycles of 95°C for 2 min, 64°C 1 min, 72°C 1 min, then 28 cycles of 95°C for 30 s, 66°C 1 min, 72°C 1 min. Following PCR, the reactions were passed through a 'QIAquick PCR purification column' (Qiagen) to remove unincorporated nucleotides before being digested with *Bsp*HI and *Bam*HI. The digested DNA was extracted twice with phenol/chloroform and ethanol-precipitated before being subcloned into the *Nco*I/*Bam*HI restriction sites of expression vector pet11D (New England Biolabs). The correct DNA sequence was confirmed by sequencing. Expression vectors were transformed into cell line AD494 and grown at 30°C on a rotary shaker (250 rpm) to an OD_{600} of 0.5–0.6. Cells were induced with 1 mM of IPTG and grown further for 5–6 h before harvesting by centrifugation (15 min at 5000 g). After washing once in TE (10:1), cells were re-suspended in a small volume of water and then ethanol was added to a final concentration (v/v) of 70%. Suspensions were incubated at 65°C for 1 h, mixing occasionally, before being centrifuged (15 min at 25000 g), and the supernatant collected. Samples were dried under vacuum and re-suspended in 50% (v/v) isopropanol. Proteins were extracted from Gabo and 2RS.2BL wholemeal as for the amino-acid sequencing but were not alkylated. All samples were fractionated on 12% SDS-PAGE gels with 2% cross-linking under either reducing or non-reducing conditions as described by Skerritt and Underwood (1986). Fractionated proteins were either stained with colloidal Coomassie G-250 (Neuhoff et al. 1991) or blotted and probed with antibody (Skerritt and Lew 1990).

Results

N-terminal protein sequence

To determine the N-terminal protein sequence of the 2RS polypeptides visualized on SDS-PAGE gels, protein was extracted from mature 2RS.2BL (translocation of Imperi-

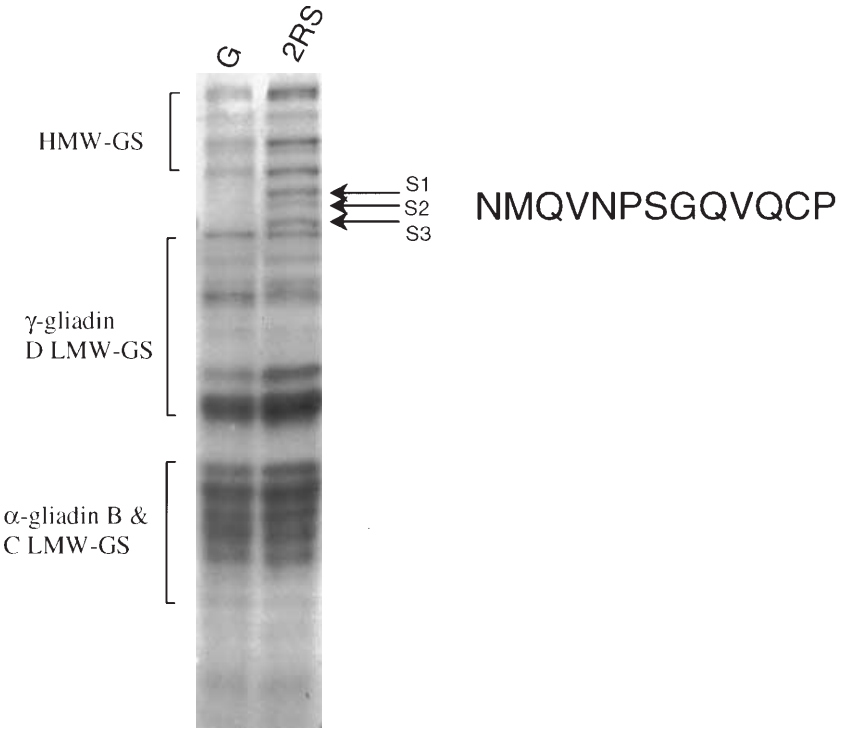

Fig. 1 SDS-PAGE of reduced protein extracted from Gabo (*G*) and 2RS.2BL (*2RS*). Secalin bands (*S1–S3*) and the N-terminal sequence are indicated by *arrowheads*. Protein groups are *bracketed*: high-molecular-weight glutenin subunits (*HMW-GS*); *B*, *C* and *D* low-molecular-weight glutenin subunits (*LMW-GS*) and α and γ-gliadins

al rye; May and Appels 1984) and Gabo (parent) wheat seeds and fractionated using SDS-PAGE. The three bands shown in Fig. 1 (S1–S3) were separately excised and sequenced. The N-terminal sequence for the first 13 residues of all three bands (identified with confidence) was NMQVNPSGQVQCP (Fig. 1). This sequence is similar to that published by Shewry and Field (1982) for unfractionated Mr 75000 secalin (NMQVNPSGQVQWP) except that our sequence contains a cysteine residue at position 12 instead of a tryptophan residue. Both samples were alkylated prior to sequencing although the parental rye source was different (Imperial and Rheidol). This cultivar difference may account for the difference in the N-terminal sequence.

cDNA isolation

Previously Kreis et al. (1985b) had isolated a partial cDNA clone, pSc503, believed to be a 2RS γ-secalin. Northern analysis in which the non-repetitive region of this clone was hybridized to RNA isolated from either the 2RS.2BL or the Gabo wheat cultivars was performed. A transcript was only detected in RNA isolated from the 2RS.2BL cultivar (Fig. 2A) suggesting that this clone was indeed part of a 2RS γ-secalin gene. This DNA sequence was subsequently used to screen a cDNA library constructed from 16–20-day-old 2RS. 2BL developing wheat grain and 15 positive plaques purified. From these, two cDNAs of similar size, whose 5′ sequence encoded an N-terminal amino-acid sequence identical to that determined by protein sequencing, were identified. The other 13 clones were smaller and sequencing indicated that they were partial cDNAs. The two putatively full-length cDNA clones had an identical nucleotide sequence at the 5′ and 3′ ends so only one cDNA, cSec2, was completely sequenced.

Analysis of the cSec2 sequence revealed a 1071-bp open reading frame (ORF) encoding a protein with a predicted length of 357 residues and molecular weight of Mr 40000. This is considerably smaller than the molecular weight predicted from SDS-PAGE gels (Mr 75000) possibly because, as with the prolamins of wheat and barley, the proline-rich repetitive region inhibits the migration of the protein through polyacrylamide gels (Shewry and Field 1982). However, it is also less than the molecular weight of Mr 54000 determined for 2RS γ-secalins by equilibrium centrifugation (Shewry and Field 1982). The predicted amino-acid sequences for cSec2, and the partial sequence for pSc503 are compared in Fig. 3. The sequence for pSc503, is different from that previously published (Kreis et al. 1985b) as re-sequencing of the clone identified errors in the original sequence. Both sequences were very similar at the nucleotide level with greater than 99% identity (4-bp changes in the 537-bp sequenced). At the amino-acid level this translated into 99.2% identity. Only one amino acid was different; in cSec2 there is a valine instead of a glutamic acid residue at amino-acid position 449.

Northern analysis of Gabo and 2RS.2BL RNA using a 0.5-kb non-repetitive region of cSec2 as a probe gave a result identical to that seen when using the pSc503 probe (Fig. 2A) confirming that the isolated cDNA is specific to chromosome 2 of rye (data not shown). Genomic Southern analysis using the cSec2 probe, following digestion of Gabo or 2RS.2BL DNA with different restriction en-

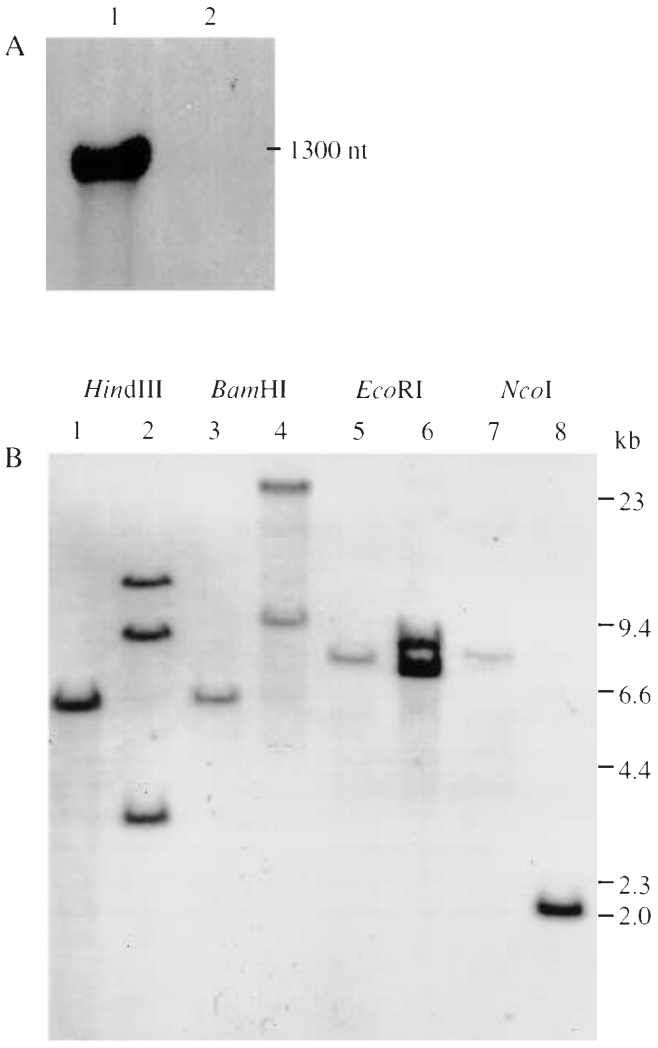

Fig. 2A Analysis of total RNA isolated from developing wheat seeds (16–20 dpa) of 2RS.2BL (*lane 1*) and Gabo (*lane 2*) wheat lines. RNA was hybridized with a ^{32}P-labelled DNA probe corresponding to the *Hin*dIII/*Pst*I 0.5-kb fragment from the 3′ end of the partial cDNA clone, pSc503 (Kreis et al. 1985b). The size of RNA is given in nucleotides (*nt*). **B** Genomic DNA analysis of the Mr 75000 2RS γ-secalin genes. Genomic DNA from 2RS.2BL (*lanes 2*, *4*, *6* and *8*) and Gabo (*lanes 1*, *3*, *5* and *7*) wheat lines digested with the restriction enzymes: *Hin*dIII, *Bam*HI, *Eco*RI and *Nco*I, and hybridized with a 354-bp digoxigenin-labelled DNA probe corresponding to the 3′ end of cSec2. The sizes of markers are given in kilobases (kb)

zymes, resulted in at-most three hybridizing bands (*Hin*dIII digest) (Fig. 2B). As only a single *Hin*dIII restriction-enzyme site is present approximately 150 bp downstream from the cSec2 stop codon, and the probe used hybridizes to the 3′ region of the cSec2 ORF, this indicates that at least three secalin-like genes are present in the genome. This is similar to the 3–6 copies of the 2RS γ-secalin genes per genome predicted by Hull et al. (1992). More than three genes may be present if the hybridizing bands observed here represent more than one gene copy.

A single hybridizing band was also seen in digested Gabo DNA (background of the 2RS.2BL line), although the observed band was always a different size to that seen in similarly digested 2RS.2BL DNA. This suggested that the sequence was present on the short arm of chromosome 2B and hence was removed following the translocation which created the 2RS.2BL line. Absence of a signal in Gabo RNA indicates that the sequence is transcribed only at low levels at this stage of development, or else not at all, in this cultivar. No hybridizing band was detected in DNA from unrelated wheat cultivars such as Hartog and Bob White (data not shown) indicating that the hybridizing sequence present in Gabo is absent in these cultivars.

Genomic clone isolation

As the isolated cDNA encodes for a protein significantly smaller than previously predicted, and because the presence of some repetitive domains in prolamin sequences can lead to problems in stability during the cloning procedure (O. Anderson, L. Tamas, pers. communication), the genomic Sec2 sequence was isolated. Genomic gel-blot analysis of 2RS.2BL DNA indicates that cSec2 hybridizes to at least two *Eco*RI fragments of 6–8 kb in size (Fig. 2B). Further DNA analysis with other restriction enzymes indicated that the Sec2 ORF was present within the central region of these *Eco*RI fragments (results not shown). In order to isolate the genomic clone of cSec2, a partial 2RS.2BL genomic library composed of 6–8-kb *Eco*RI fragments was made in Lambda Zap II. Screening with the cSec2 probe used previously identified one strongly hybridizing plaque, gSec2A. Gel-blot analysis indicated that the cloned DNA hybridized to the larger (8-kb) *Eco*R1 band observed on the genomic 2RS.2BL gel blot (data not shown).

Within the gSec2A sequence (genebank accession number AF201084), a 1365-bp ORF was identified which is identical at the nucleotide level to that of cSec2 except that it contains an additional 294-bp approximately 220-bp downstream from the initiating ATG start codon. No intron splice sites were present adjacent to this extra nucleotide sequence and no introns have been identified in cereal seed-storage protein genes to-date. Within the promoter upstream of the proposed ATG start codon are the putative recognition sequences important in transcriptional regulation, i.e. a TATA box at –101 and the '–300 element' conserved in many cereal storage protein genes and implicated in seed-specific expression (Kreis

Fig. 3 Alignment of predicted gSec2A amino-acid sequence (*bold*) ▶ with pSc503 (Kreis et al. 1985b), cSec2, and other closely related γ-gliadins and LMW glutenins (γ-type). γ-gliadins: whtgligp, Scheets et al. (1985); whtglgb, Rafalski (1986); whtggmpa, Maruyama et al. (1998); whtggmpb, Maruyama et al. EMBL/Genbank/DDBJ accession number D78184. LMW glutenins: whtggln, Scheets and Hedgcoth (1988) (sequence corrected by Shewry and Tatham 1997); tdgag1, D'Ovidio et al. (1991). *Numbers 1–8* indicate conserved cysteine residues, * indicates stop codons, – indicates sequence unavailable, . indicates spaces introduced to maximise alignment (not necessarily reflective of evolutionary relationship)

```
pSc503   (γ-secalin)      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
cSec2         "           MKTLLMLAIL AMATTIATAN MQVNPSGQVQ CPQQQPFPQP QQSSPQQPQQ PFPQQSQQPF PQQPQQSSPQ PQQPY.....
gSec2A        "           MKTLLMLAIL AMATTIATAN MQVNPSGQVQ CPQQQPFPQP QQSSPQQPQQ PFPQQSQQPF PQQPQQSSPQ PQQPYPQQPF
whtglig  (γ-gliadin)      MKTLLILTIL AMAITIGTAN IQVDPSGQVQ WLQQQLVPQL QQPS...... .......... .......... ..........
whtgligp      "           MKTLLILTIL AMAITIGTAN MQVDPSSQVQ WPQQQPVPQP HQPS...... .......... .......... ..........
whtglgb       "           MKTLLILTIL AMAITIATAN MQADPSGQVQ WPQQQPFLQP HQPS...... .......... .......... ..........
whtggmpb      "           ---------- ---------N IQVDPSGQVQ WPQQQPFPQP HQPF...... .......... .......... ..........
whtggmpa      "           ---------- ---------N IQVDPSGQVQ WPQQQPFPQP HQPF...... .......... .......... ..........
whtggln (LMW-glutenin)    MKTLLILTIL AMATTIATAN MQVDPSGQVQ WPQQQPFPQP QQPC...... .......... .......... ..........
tdgagl        "           MKTLLILTIL AMATTIATAN MQVDPSGQVQ WPQQQPFPQP QQPC...... .......... .......... ..........

pSc503   (γ-secalin)      ---------- ---------- ---------- ---------- ---------- ---------- --------- ----------
cSec2         "           .......... .......... .......... .......... .......... .......... .......... ..........
gSec2A        "           PQQPQQPYPQ QPQQPFPQQP QQPYPQQPQQ QFPQQPQQPV PQQPLQQFPQ QPQQPFPQQP LQQFPQQPQQ PFPQQPQQPV
whtglig  (γ-gliadin)      .......... .......... .......... .......... .......... .......... .......... ..........
whtgligp      "           .......... .......... .......... .......... .......... .......... .......... ..........
whtgligb      "           .......... .......... .......... .......... .......... .......... .......... ..........
whtggmpb      "           .......... .......... .......... .......... .......... .......... .......... ..........
whtggmpa      "           .......... .......... .......... .......... .......... .......... .......... ..........
whtggln (LMW-glutenin)    .......... .......... .......... .......... .......... .......... .......... ..........
tdgag1        "           .......... .......... .......... .......... .......... .......... .......... ..........

pSc503   (γ-secalin)      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
cSec2         "           .......... ...PQQPFPQ PQQPQQLFPQ TQQSSPQQPQ QVTSQPQQPF PQAQPPQQSS PQS.QQPYPQ EPQQLFPQSQ
gSec2A        "           PQQSQQPFPQ TQQPQQPFPQ PQQPQQLFPQ TQQSSPQQPQ QVTSQPQQPF PQAQPPQQSS PQS.QQPYPQ EPQQLFPQSQ
whtglig  (γ-gliadin)      .......... .......... .QQPQQTFPQ PQQTFPHQPQ QQVPQPQQPQ QPFLQPQQPF PQQPQQPFPQ TQQPQQPFPQ
whtgligp      "           .......... .......... .......... ......QQPQ QTFPQPQQTF PHQ..PQQQF PQ......PQ QPQQQFLQ..
whtglgb       "           .......... .......... .......... ......QQPQ QIFPQPQQTF PHQ..PQQQF PQ......PQ QPQQQFLQ..
whtggmpb      "           .......... .......... .......... .....SQQPQ QTFPQPQQTF PHQ..QQQQF SQ......PQ QPQQQFIQ..
whtggmpa      "           .......... .......... .......... .....SQQPQ QTFPQPQQTF PHQ...QQQF SQ......PQ QPQQQFIQ..
whtggln (LMW-glutenin)    .......... .......... .QQPQRTIPQ PHQTFHHQPQ QTFPQPQQTY PHQ..PQQQF PQ......TQ QPQQPFPQ..
tdgagl        "           .......... .......... .EQPQRTIPQ PHQTFHHQPQ QTFPQPEQTY PHQ..PQQQF PQ......TQ QPQQPFPQ..

                                                                                                          1
pSc503   (γ-secalin)      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
cSec2         "           QPQQPFPQPQ QPQQPFPQPQ PQTQQSIPQP QQPFPQ...P QQPFPQSQEQ FPQVHQPQQP SPQQQQPSIQ LSLQQQLNPC
gSec2A        "           QPQQPFPQPQ QPQQPFPQPQ PQTQQSIPQP QQPFPQ...P QQPFPQSQEQ FPQVHQPQQP SPQQQQPSIQ LSLQQQLNPC
whtglig  (γ-gliadin)      QPQQPFPQTQ QPQQPFPQQP QQPFPQTQQP QQPFPQLQQP QQPFPQPQQQ LPQPQQPQQS FPQQQRPFIQ PSLQQQLNPC
whtgligp      "           .PQQPFPQ.. QPQQPYPQQP QQPFPQTQQP QQLFPQSQQP QQQFSQPQQQ FPQPQQPQQS FPQQQPPFIQ PSLQQQVNPC
whtglgb       "           .PRQPFPQ.. QPQQPYPQQP QQPFPQTQQP QQPFPQSKQP QQPFPQ.... ...PQQPQQS FPQQQPSLIQ QSLQQQLNPC
whtggmpb      "           .PQQPFPQ.. QPQQTYPQRP QQPFPQTQQP QQPFPQSQQP QQPFPQPQQQ FPQPQQPQQS FPQQQPSLIQ QSLQQQLNPC
whtggmpa      "           .PQQPFPQ.. QPQQTYPQRP QQPFPQTQQP QQPFPQSQQP QQPFPQPQQQ FPQPQQPQQS FPQQQPSLIQ QSLQQQLNPC
whtggln (LMW-glutenin)    .PQQTFPQ.. QPQLPFPQQP QQPFPQPQQP QQPFPQSQQP QQPFPQPQQQ FPQPQQPQQS FPQQQQPAIQ SFLQQQMNPC
tdgagl        "           .PQQTFPQ.. .PQLPFPQQP QQPFPQPQQP QQPFPQSQQP QQPFPQPQQQ FPQPQQPQQS FPQQQQPAIQ SFLQQQMNPC

                                 2                   3       45           6
pSc503   (γ-secalin)      ----LQQCSP VALVSSLRSK IFPQSECQVM QQQCCQQLAQ IPQQLQCAAI HSVVHAIIM. ...QQEQREG VQILLPQSHQ
cSec2         "           KNVLLQQCSP VALVSSLRSK IFPQSECQVM QQQCCQQLAQ IPQQLQCAAI HSVVHAIIM. ...QQEQREG VQILLPQSHQ
gSec2A        "           KNVLLQQCSP VALVSSLRSK IFPQSECQVM QQQCCQQLAQ IPQQLQCAAI HSVVHAIIM. ...QQEQREG VQILLPQSHQ
whtglig  (γ-gliadin)      KNILLQQCKP ASLVSSLWSI IWPQSDCQVM RQQCCQQLAQ IPQQLQCAAI HSVVHSIIMQ QQQQQQQQQG IDIFLPLSQH
whtgligp      "           KNFLLQQCKP VSLVSSLWSM IWPQSDCQVM RQQCCQQLAQ IPQQLQCAAI HTIIHSIIM. QQEQQEQQQG MHILLPLYQQ
whtglgb       "           KNFLLQQCKP VSLVSSLWSI ILPPSDCQVM RQQCCQQLAQ IPQQLQCAAI HSVVHSIIM. QQEQQEQLQG VQILVPLSQQ
whtggmpb      "           KNFLLQQCKP VSLVSSLWSM ILPRSDCQVM RQQCCQQLAQ IPQQLQCAAI HGIVHSIIM. QQEQQEQRQG VQILVPLSQQ
whtggmpa      "           KNFLLQQCKP VSLVSSLWSM ILPRSDCQVM RQQCCQQLAQ IPQQLQCAAI HSIVHSIIM. QQEQQEQRQG VQILVPLSQQ
whtggln (LMW-glutenin)    KNFLLQQCNH VSLVSSLVSI ILPRSDCQVM QQQCCQQLAQ IPQQLQCAAI HSVAHSIIM. ...QQEQQQG VPILRPLFQ.
tdgagl        "           KNFLLQQCNH VSLVSSLVSI ILPRSDCQVM QQQCCQQLAQ IPQQLQCAAI HSVAHSIIM. ...QQEQQQG VPILRPLFQ.

                                                                  7        8
pSc503   (γ-secalin)      QHVGQGALAQ VQGIIQPQQL SQLEVVRSLV LQNLPTMCNV YVPRQCSTIQ APFASIETGI VGH*
cSec2         "           QHVGQGALAQ VQGIIQPQQL SQLEVVRSLV LQNLPTMCNV YVPRQCSTIQ APFASIVTGI VGH*
gSec2A        "           QHVGQGALAQ VQGIIQPQQL SQLEVVRSLV LQNLPTMCNV YVPRQCSTIQ APFASIVTGI VGH*
whtglig  (γ-gliadin)      EQVGQGSLVQ GQGIIQPQQP AQLEAIRSLV LQTLPSMCNV YVPPECSIMR APFASIVAGI GGQ*
whtgligp      "           QQVGQGTLVQ GQGIIQ---- ---------- ---------- ---------- ---------- ----
whtglgb       "           QQVGQGILVQ GQGIIQPQQP AQLEVIRSLV LQTLPTMCNV YVPPYCSTIR APFASIVASI GGQ*
whtggmpb      "           QQVGQGTLVQ GQGIIQPQQP AQLEVIRSSV LQTLATMCNV YVPPYCSTIR APFASIVAGI GGQ*
whtggmpa      "           QQVGQGTLVQ GQGIIQPQQP AQLEVIRSSV LQTLATMCNV YVPPYCSTIR APFASIVAGI GGQ*
whtggln (LMW-glutenin)    .......LAQ GLGIIQPQQP AQLEGIRSLV LKTLPTMCNV YVPPDCSTIN VPYANIDAGI GGQ*
tdgagl        "           .......LAQ GLGIIQPQQP AQLEGIRSLV LKTLPTMCNV Y--------- ---------- ----
```

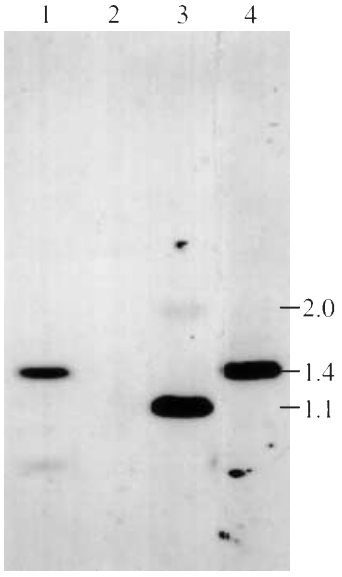

Fig. 4 Genomic gel-blot analysis of *Rsa*I/*Hind*III-digested 2RS.2BL DNA hybridized with the 3′ cSec2 probe. *Lane 1* 2RS.2BL; *lane 2* Hartog; *lane 3* Hartog DNA spiked with cSec2 DNA; *lane 4* Hartog DNA spiked with gSec2A DNA. The sizes of hybridizing bands are in kilobases

et al. 1985a) at –316. The gSec2A gene contains two consensus polyadenylation signals (AATAAA) 62 and 119-bp downstream from the stop codon. The sequence of cSec2 (data not shown) indicates that the second polyadenylation signal (119-bp downstream) is utilized. Multiple polyadenylation signals have been observed in other prolamin genes such as the Bx7 HMW glutenin gene (Anderson and Greene 1989).

To determine if both the genomic and cDNA sequences were present in their cloned form in the 2RS.2BL genome, DNA analysis were carried out in which 2RS.2BL genomic DNA digested with *Rsa*I and *Hin*dIII was hybridized with the 0.5-kb 3′ cSec2 probe used previously. Sequence analysis indicated that hybridization would occur to two bands of 1.4-kb and 1.1-kb (gSec2A and cSec2 respectively) if both the cDNA and the genomic Sec2 sequences were present in the 2RS.2BL genome. Figure 4 shows the presence of the 1.4-kb hybridizing band but no 1.1-kb band, indicating that the cSec2 sequence probably does not exist in the 2RS.2BL genome and that its sequence is incorrect as a result of a deletion which occurred during the cDNA cloning process.

Translation of the gSec2 A ORF predicts that it encodes a 455-aa protein with a molecular weight of Mr

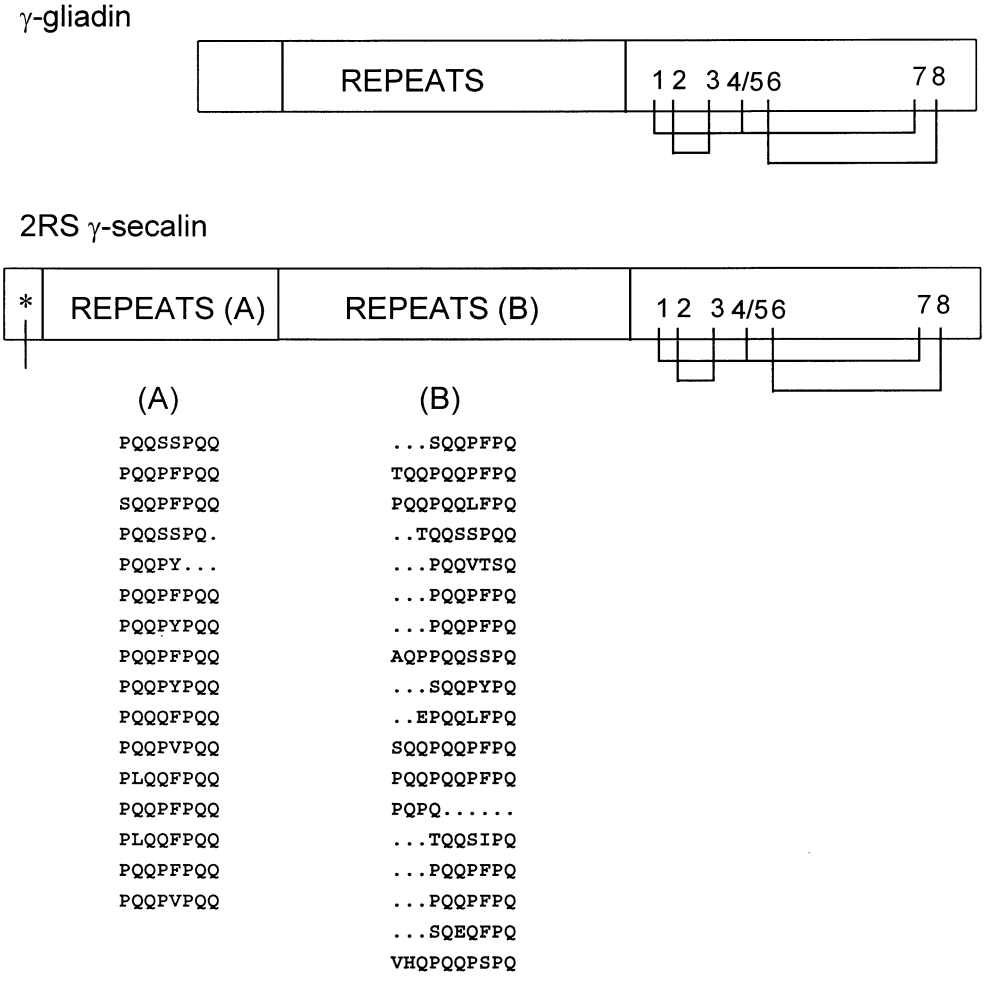

Fig. 5 Schematic diagram showing the predicted disulphide structures of γ-type gliadins (Müller and Wieser 1997) and gSec2A. Numbers relate to the cysteine residues shown in Fig. 3, * indicates the cysteine at the N-terminus of gSec2A. *Lines* indicate predicted disulphide linkages. The repeat regions (*A*) and (*B*) identified in gSec2A are listed beneath the appropriate regions

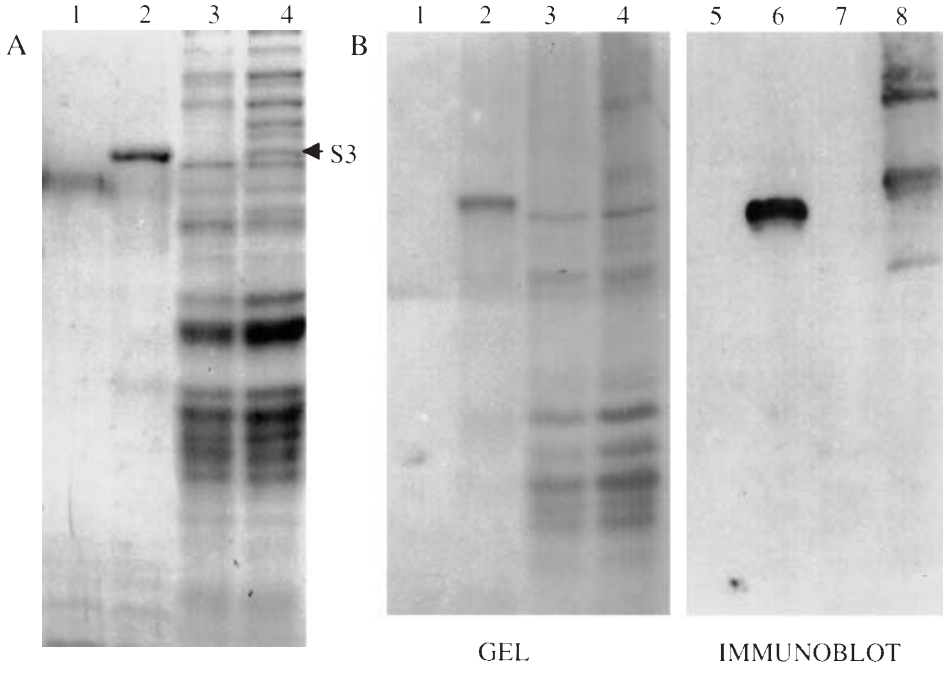

Fig. 6A SDS-PAGE of reduced bacterially expressed gSec2A protein. *Lane 1* pet11D; *lane 2* gSec2A; *lane 3* Gabo extract; *lane 4* 2RS.2BL extract. *S3* indicates the smallest secalin protein band in the 2RS.2BL extract. **B** SDS-PAGE and the corresponding immunoblot of unreduced bacterially expressed gSec2A protein. *Lanes 1* and *5* pet11D; *lanes 2* and *6*, gSec2A; *lanes 3* and *7* Gabo extract; *lanes 4* and *8* 2RS.2BL extract

52 000. Compared to the predicted cSec2 protein, the gSec2A protein has an extra 98 amino acids of a proline/glutamine-rich repetitive domain (Fig. 3). Analysis of the gSec2A protein sequence reveals that it consists of a 19-aa signal peptide prior to three domains; a short N-terminal domain (20-aa), followed by a proline/glutamine-rich repetitive region (247-aa) and a non-repetitive C-terminal domain (156-aa). The repetitive region consists of two types of repeats, approximately 15 eight-aa repeats (consensus PQQPFPQQ, Fig. 5) followed by 17 seven-aa repeats interspersed with a tripeptide repeat [consensus $(PQQ)_{1-2}PFPQ$, Fig. 5].

A blast search (Altschul et al. 1990) of the data base indicates that gSec2A has highest homology to the γ-gliadin gene family showing approximately 50% identity to members of this family at the amino-acid level (Fig. 3). The 2RS γ-Sec2 N-terminal domain (mature peptide) matches closely the equivalent regions in γ-gliadins and LMW glutenins (γ-type) but is quite distinct from the α-gliadins (consensus VRVPVPQLQPQNPSQ). Compared to the γ-gliadins and the γ-type LMW glutenins, the 2RS γ-Sec2 N-terminal domain has an asparagine at position 5 instead of aspartic acid, and a cysteine at position 12 instead of tryptophan. In addition, the γ-type LMW glutenins have a further mutation, the serine at position 25 in gSec2A has been mutated to a cysteine in these LMW glutenins. The identity of gSec2A with the C-terminal domain of γ-gliadins and LMW glutenins is also high (75%) and in all cases the eight cysteines align at the same positions (Fig. 3). In the repetitive region, the second part of the 2RS γ-Sec2 repetitive domain (pentapeptide with interspersed tripeptide repeats) is very similar to that seen in γ-gliadins but the octapeptide repeat seen in gSec2A is completely absent from the γ-gliadins. However, this octapeptide repeat region is seen in ω-secalins and C-hordeins (Entwistle et al. 1991; Hull et al. 1991). The octapeptide repeat region of gSec2A is most similar to that of the rye ω-secalins (70% identity).

Bacterial expression of the gSec2A protein

To determine if the isolated gSec2A sequence coded for a 2RS γ-secalin protein of an appropriate size, the sequence was engineered into the bacterial expression vector Pet11d, and the expressed protein isolated. When reduced and electrophoresed on a SDS-PAGE gel, the gSec2A protein migrated at the same rate as the smallest 2RS.2BL γ-secalin polypeptide band (S3) (Fig. 6A). Immunoblot analysis of a SDS-PAGE gel electrophoresed under non-reducing conditions using a 2RS γ-secalin-specific antibody (Skerritt et al. 1996) indicates that the gSec2A protein is recognised by the antibody 80709 (Fig. 6B). Non-reducing conditions were used as this antibody only recognizes the Sec2 protein under non-reducing conditions. Under these conditions 2RS γ-secalins migrate as dimers and small oligomers, resulting in the appearance of several bands on the immunoblot (Skerritt et al. 1996).

Discussion

We have successfully isolated a gene, gSec2A, coding for a 2RS γ-secalin protein. The N-terminal amino-acid sequence of the predicted protein encoded by this gene is identical to the N-terminal protein sequence obtained from the 2RS γ-secalin protein bands isolated from SDS-PAGE gels. When reduced, the bacterially expressed gSec2A protein has the same mobility on SDS-PAGE gels as the smallest Sec2 polypeptide band. Under non-reducing conditions this protein is recognized by a

monoclonal antibody specific for 2RS γ-secalins. Also isolated was a cDNA clone, cSec2, which is identical to the gSec2A nucleotide sequence except that it does not have 294 bp of internal sequence. Genomic DNA analysis indicated that the cSec2 gene sequence does not exist in the 2RS2BL wheat line from which it was isolated, indicating that this sequence has probably had part of its sequence deleted during the cloning process.

gSec2A encodes a protein which has broad similarities to other prolamin genes, in particular the γ-gliadins and LMW glutenins, although it is substantially larger due to the inclusion of an additional proline/glutamine repetitive domain close to the N-terminus. This repetitive region is predicted to form an octapeptide repeat similar to that seen in the ω-secalins and C-hordeins (Shewry et al. 1994), although in gSec2A the repeats have significantly less degeneracy. The repetitive region is almost completely absent from cSec2 and other cDNA clones (data not shown) suggesting that structurally this DNA is unstable when not flanked by non-repetitive DNA as it is in the genomic clone. Such instability may account for the difficulty in obtaining cDNA clones of C-hordeins, ω-secalins and ω-gliadins.

The protein coded for by gSec2A consists of three domains, a short N-terminal domain followed by a proline/glutamine-rich repetitive domain and a C-terminal domain. Given the amino-acid sequence similarity of these domains to other proteins, (ω-secalins and γ-gliadins), some predictions concerning protein secondary structure can be made. The repetitive domain comprising an octapeptide repeat region followed by a pentapeptide interspersed tripeptide repeat region is predicted to form β-reverse turns and poly-L-proline II structures, whereas the non-repetitive sequence in the C-terminal domain is predicted to have a globular structure rich in α-helix (Shewry et al. 1994).

As disulphide bonds play a key role in determining the structure and properties of wheat gluten proteins, the numbers and locations of cysteine residues in the gSec2A protein are of considerable interest. Alignment of the gSec2A C-terminal domain aa sequence with that of other γ-gliadins (Fig. 3) indicates that in all cases the cysteines from both proteins occur in the same position. In γ-gliadins, it is predicted that these eight cysteines form intramolecular disulphide bonds with each other (Müller and Wieser 1997) and the same is likely to occur in the gSec2A protein (Fig. 5). The remaining N-terminal unpaired cysteine in gSec2A is presumably able to form a disulphide bond with other unpaired cysteines in the gluten polymer. This is consistent with the observation that Sec2 proteins often form oligomers which are dissociated by reducing agents (Shewry et al. 1983b; Skerritt et al. 1996). When bacterially expressed gSec2A protein was electrophoresed on native SDS-PAGE gels, dimer formation was not observed, possibly due to the isolation of protein without sonication, leaving most of the expressed protein present in the inclusion bodies.

As gSec2A is predicted to have effectively only one cysteine with which to bind to the gluten polymer, it is likely that the inclusion of this protein into wheat would promote 'chain terminating' (Kasarda 1989), and stop lengthening of the gluten polymer resulting in a reduction in dough strength. This is consistent with the fact that in the absence of reducing agents, 2RS γ-secalins form small oligomers as opposed to the HMW glutenin subunits which form large polymers. These considerations predict that gSec2A would reduce dough strength. This is in contradiction to the findings of Gupta et al. (1989) that 2RS.2BL lines have increased dough strength (R_{max} 298 compared to 271) but the observed change in R_{max} is marginal and probably due to the accompanying change in protein content (13.8% compared to 11.8%). However, the work of Gupta et al. was performed with translocation lines containing several secalin genes and we have only isolated a single member of this gene family. N-terminal sequencing suggests that all Sec2 members have a single cysteine at the N-terminus but it is possible that different numbers of cysteine residues occur in the C-terminal domain of other family members. The absence or presence of an additional cysteine in the C-terminus would significantly alter the way in which the protein binds to gluten polymers and may promote larger polymer formation and an increase in dough strength consistent with the work of Gupta et al.

The presence of the large proline/glutamine repetitive domain in the gSec2A protein may also effect dough properties. In C-hordeins and γ-gliadins, these regions, because of their abundance of glutamine residues, are capable of forming numerous hydrogen bonds both within and between protein subunits. These and other non-covalent interactions are thought to contribute to dough viscosity (Shewry and Tatham 1997). As a result of the presence of the repetitive domain, the viscosity (extensibility) of dough containing gSec2A protein could be significantly altered. In the future, incorporation of bacterially expressed purified protein into small-scale mixing experiments (Tamas et al. 1997), and/or the generation of transgenic plants expressing gSec2A, should allow the effect of this gene on dough properties and grain protein content to be specifically investigated.

Acknowledgements We thank Ram Gupta, Rob Solomon, Lazlo Tamas, Judy Flanigan and Richard Brettell for their assistance with this work. We are grateful to the Grains Research and Development Corporation (Australia) for their financial support throughout this research.

References

Alonso-Blanco C, Goicoechea PG, Roca A, Giraldez R (1993) Genetic linkage between cytological markers and the seed storage protein loci sec 2 (*Gli-R2*) and sec 3 (*Glu-R1*) in rye. Theor Appl Genet 87:321–327

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Anderson OD, Greene FC (1989) The characterization and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat. Theor Appl Genet 77:689–700

Chandler PM, Higgens TJV, Randell PJ, Spencer D (1983) Regulation of legumin levels in developing pea seeds under conditions of sulphur deficiency. Plant Physiol 71:47–54

Devereaux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

D'Ovidio R, Tanzarella OA, Porceddu E (1991) Cloning and sequencing of a PCR-amplified gamma-gliadin gene from durum wheat [Triticum turgidum (L) Thell. Conv. *durum* (Desf.) MK]. Plant Sci 75:229–236

Dolferus R, Jacobs M, Peacock WJ, Dennis ES (1994) Differential interactions of promoter elements in stress responses of the *Arabidopsis Adh* gene. Plant Physiol 105:1075–1087

Entwistle J, Knudsen S, Muller M, Cameron-Mills V (1991) Amber codon suppression: the in vivo and in vitro analysis of two C-hordein genes from barley. Plant Mol Biol 17:1217–1231

Feldman M (1988) Cytogenetic and molecular approaches to alien gene transfer in wheat. In: Miller TE, Koebner RMD (eds) Proc 7th Int Wheat Genet Symp, 13–19 July 1988, Cambridge, UK. Agricultural and Food Research Council, Institute of Plant Science Research, Cambridge Laboratory, Trumpington, UK, pp 23–32

Gupta RB, Shepherd KW, MaCritchie F (1989) Effect of rye chromosome arm 2RS on flour properties and physical dough properties of wheat. J Cereal Sci 10:169–173

Hull GA, Halford NG, Kreis M, Shewry PR (1991) Isolation and characterisation of genes encoding rye prolamins containing a highly repetitive sequence motif. Plant Mol Biol 17:1111–1115

Hull G, Sabelli PA, Shewry PR (1992) Restriction fragment analysis of the secalin loci of rye. Biochem Genet 30:85–97

Jagannath DR, Bhatia CR (1972) Effect of rye chromosome-2 substitution on kernal protein content of wheat. Theor Appl Genet 42:89–92

Kreis M, Shewry PR, Forde BG, Forde J, Miflin BJ (1985a) Structure and evolution of seed storage proteins and their genes with particular reference to those for wheat, barley and rye. Oxford Surveys Plant Mol Cell Biol 2:253–317

Kreis M, Forde BG, Rahman S, Miflin GJ, Shewry PR (1985b) Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol 183:499–502

Lagudah ES, Appels R (1991) The *Nor-D3* locus of *Triticum tauschii*: natural variation and genetic linkage to markers on chromosome 5. Genome 34:387–395

Maruyama N, Ichise K, Katsube T, Kawase S, Matsumura Y, Takeuchi Y, Sawada T, Utsumi S (1998) Identification of major wheat allergens by means of the *Escherichia coli* expression system. Eur J Biochem 255:739–745

Kasarda DD (1989) Glutenin structure in relation to wheat quality. In: Pomeranz Y (ed) Wheat is unique. Am Assoc Cereal Chem. St Paul, Minnesota, pp 277–302

May CE, Appels R (1980) Rye chromosome translocations in hexaploid wheat: a re-evaluation of the loss of heterochromatin from rye chromosomes. Theor Appl Genet 56:17–23

May CE, Appels, R (1984) Seedling lethality in wheat: a novel phenotype associated with a 2RS/2BL translocation chromosome. Theor Appl Genet 68:163–168

Müller S, Wieser H (1997) The location of disulphide bonds in monomeric γ-type gliadins. J Cereal Sci 26:169–176

Murray FR, Llewellyn DJ, Peacock WJ, Dennis ES (1997) Isolation of the glucose oxidase gene from *Talaromyces flavus* and characterisation of its role in the biocontrol of *Verticillium dahliae.* Curr Genet 32:367–375

Neuhoff V, Arnold N, Taube D, Ehrhardt W (1991) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with a clear background at nanogram sensitivity, using Coomassie Brilliant Blue G-250 and R-250. Electro 9:255–262

Rafalski JA (1986) Structure of wheat gamma-gliadin genes. Gene 43:221–229

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Scheets K, Hedgcoth C (1988) Nucleotide sequence of a γ-gliadin gene: comparison with other γ-gliadin sequences show the structure of the γ-gliadin genes and the general primary structure of γ-gliadins. Plant Sci 57:141–150

Scheets K, Rafalski JA, Hedgcoth C, Soell DG (1985) Heptapeptide repeat structure of a wheat gamma-gliadin. Plant Sci Lett 37:221–225

Shewry PR, Field M (1982) The purification and characterization of two groups of storage proteins (secalins) from rye (Secale cereale L.). J Exp Bot 33:261–268

Shewry PR, Tatham AS (1997) Disulphide bonds in wheat gluten proteins. J Cereal Sci 25:207–227

Shewry PR, Kreis M, Burgess SR, Parmar S, Miflin BJ (1983a) The synthesis and deposition of the prolamin storage proteins (secalins) of rye. Planta 159:439–445

Shewry PR, Parmar S, Miflin BJ (1983b) Extraction, separation and polymorphism of the prolamin storage proteins (secalins) of rye. Cereal Chem 60:1–6

Shewry PR, Miles MJ, Tatham AS (1994) The prolamin storage proteins of wheat and related cereals. Prog Biophys Mol Biol 61:37–59

Sissons MJ, Bekes F, Skerritt JH (1998) Isolation and functionality testing of low-molecular-weight glutenin subunits. Cereal Chem 75:30–36

Skerritt JH, Lew PY (1990) Homologies between grain storage proteins of different cereal species. 1. Monoclonal antibody reaction with total protein extracts. J Cereal Sci 11:103–121

Skerritt JH, Underwood PA (1986) Specificity characteristics of monoclonal antibodies to wheat grain storage proteins. Biochem Biophys Acta 874:245–254

Skerritt JH, Andrews JL, Stoddard FL, Gupta RB, Howes NK (1996) A rapid antibody based test for *Sec-2*, a marker for the short arm of chromosome 2 of rye (2RS). Genome 39: 1006–1012

Stoddard FL, Moody DB, Gupta RB, Marshall DR (1991) Effects of seed weight and rye germplasm on grain protein concentration in wheat. In: Bushuk W, Tkachuk R (eds) Gluten Proteins, 1990. Proc 4th Int Wkshp on Gluten Proteins, 27–29 June 1990, Winnipeg, Mannitoba, Canada. Eagan Press, St Paul, Minnesota, USA, pp 707–714

Sugiyama T, Rafalski A, Soell D (1986) The nucleotide sequence of a wheat gamma-gliadin genomic clone. Plant Sci 44:205–209

Tamas L, Bekes F, Greenfield J, Tatham A, Shewry P, Appels R (1997) Heterologous expression and mixing studies of wild-type and mutant C hordeins. J Cereal Sci 27:15–22

Thuring RWJ, Sanders JPM, Borst P (1975) A freeze-squeeze method for recovering long DNA from agarose gels Anal Biochem 66:213–220

Witrzens B, Brettell RIS, Murray FR, McElroy D, Li Z, Dennis ES (1998) Comparison of three selectable marker genes for transformation of wheat by microprojectile bombardment. Aust J Plant Physiol 25:39–44