

Improved protein quality in transgenic soybean expressing a de novo synthetic protein, MB-16

Yunfang Zhang · Johann Schernthaner ·
Natalie Labbé · Mary A. Hefford · Jiping Zhao ·
Daina H. Simmonds

Received: 23 May 2013 / Accepted: 16 December 2013 / Published online: 17 January 2014
© Her Majesty the Queen in Right of Canada 2014

Abstract To improve soybean [*Glycine max* (L.) Merrill] seed nutritional quality, a synthetic gene, MB-16 was introduced into the soybean genome to boost seed methionine content. MB-16, an 11 kDa de novo protein enriched in the essential amino acids (EAAs) methionine, threonine, lysine and leucine, was originally developed for expression in rumen bacteria. For efficient seed expression, constructs were designed using the soybean codon bias, with and without the KDEL ER retention sequence, and β -conglycinin or cruciferin seed specific protein storage promoters. Homozygous lines, with single locus integrations, were identified for several transgenic events. Transgene transmission and MB-16 protein expression were confirmed to the T5 and T7 generations, respectively. Quantitative RT-PCR analysis of developing seed showed that the transcript peaked in

growing seed, 5–6 mm long, remained at this peak level to the full-sized green seed and then was significantly reduced in maturing yellow seed. Transformed events carrying constructs with the rumen bacteria codon preference showed the same transcription pattern as those with the soybean codon preference, but the transcript levels were lower at each developmental stage. MB-16 protein levels, as determined by immunoblots, were highest in full-sized green seed but the protein virtually disappeared in mature seed. However, amino acid analysis of mature seed, in the best transgenic line, showed a significant increase of 16.2 and 65.9 % in methionine and cysteine, respectively, as compared to the parent. This indicates that MB-16 elevated the sulfur amino acids, improved the EAA seed profile and confirms that a de novo synthetic gene can enhance the nutritional quality of soybean.

Electronic supplementary material The online version of this article (doi:10.1007/s11248-013-9777-5) contains supplementary material, which is available to authorized users.

Y. Zhang · J. Schernthaner · N. Labbé ·
D. H. Simmonds (✉)
Eastern Cereals and Oilseed Research Centre, Agriculture
and Agri-Food Canada, Ottawa, ON K1A0C6, Canada
e-mail: simmondsdh@agr.gc.ca;
daina.simmonds@agr.gc.ca

Present Address:

Y. Zhang
Shanghai Institute of Biological Science, China Academic
Science, 320 Yueyang Road, Life Science Building,
Shanghai 200032, China

Keywords Transgenic soybean · Milk bundle
protein · MB-16 · β -Conglycinin · Cruciferin ·
Sulfur amino acids · Essential amino acids

M. A. Hefford
Centre for Biologics Research, Biologics and Genetic
Therapies Directorate, Health Canada, Ottawa,
ON K1A 0L2, Canada

J. Zhao
Ball Helix, Ball Horticultural Company, 622 Town Road,
West Chicago, IL 60185-2698, USA

Introduction

Soybean [*Glycine max* (L.) Merr.], an important legume for food and feed, is rich in isoflavones, the secondary metabolites that have been cited for enhancing numerous health benefits including reducing the risk of cancer, cardiovascular disease, osteoporosis and in controlling weight (Messina 1999; Messina and Wood 2008; Wong et al. 2008; Cederroth and Nef 2009; Taylor et al. 2009). The nutritional quality of soybean is considered to be equivalent to casein and egg white (Krishnan 2005) except for its deficiency in sulfur amino acids (SAAs) and especially in methionine (Met). Although the SAA level of 22–23 mg/g protein in soybean meets the WAO/WHO (2007) recommendation for adults, an increase of approx. 20 % would be required to meet the needs of young children. However, more than double the current SAA level is needed in feed of poultry and weaning swine. These industries augment feed with synthetic Met, spending an estimated \$100 million annually (Imsande 2001).

Over the past 50 years, considerable effort has been directed at improving protein quality in soybean. The approaches used traditional breeding as well as biotechnological methods, including, protein modification, and introduction of heterologous or synthetic SAA rich proteins (Krishnan 2005). Breeding methods were used to alter the ratio of β -conglycinins and glycinins in seed to reduce the 7S β -conglycinins that are very low in SAAs (Paek et al. 1997, 2000). Spontaneous or induced soybean mutants with reduced β -conglycinin composition were identified but were also accompanied with developmental abnormalities. Ethyl methanesulfonate mutagenesis was used to identify a methionine overproducing line with an SAA concentration approximately 20 % higher than the parent line (Imsande 2001), but further work on this line has not been reported. One spontaneous mutant extremely low in the 7S globulins was identified in a Japanese wild soybean collection that could be incorporated into soybean cultivars without obvious effects on agronomic performance (Hajika et al. 1996). The single dominant gene, *Scg-1*, was shown to suppress the three subunits of β -conglycinin (Teraishi et al. 2001) and was thus a potential method to control soybean protein content. However, this approach may not produce the desired results as integrating soybean mutants with deficiencies in both

β -conglycinin and glycinin did not change the overall amino acid composition and did not alter the SAA levels (Takahashi et al. 2003).

Elevation of SAAs by means of modification of the glycinins has been attempted as these five endogenous proteins (Gy1–Gy5) have a highly variable region suitable for Met residue insertion. Oligonucleotides encoding Met residues have been inserted into different regions of the Gy1 gene to identify optimal folding and the assembly of the new protein in potato (Utsumi et al. 1994). A modified Gy1 proglycinin gene with a synthetic DNA encoding four continuous Met was integrated into the soybean genome (El-Shemy et al. 2007). Although the accumulation of these Met-enriched glycinins in potato and soybean indicate the possibility of introducing additional Met residues in glycinin, the overall effect on the Met content is small.

Transformation of soybean with several heterologous genes encoding proteins rich in SAAs has produced transgenic lines with elevated SAAs. Brazil nut 2S albumin (BNA) containing 18 % Met and 8 % cysteine (Cys) (Altenbach et al. 1987) was successfully integrated into soybean with an increase of SAA content as high as 40 % (Townsend and Thomas 1994), but commercial lines were not developed because BNA is an allergen (Nordlee et al. 1996). Several zein proteins, also known to be high in SAAs, have been expressed in soybean. Two transgenic soybean lines carrying the 15-kDa delta zein showed an increase of 12–20 and 15–30 % in Met and Cys, respectively (Dinkins et al. 2001). An 11-kDa delta zein driven by the β -conglycinin promoter showed a 1.5–1.7 fold Met increase in the alcohol-soluble protein fraction but the overall Met content in seed was not increased (Kim and Krishnan 2004). The 7-kDa gamma-zein, also introduced into soybean, showed an increase of nearly 30 and 19 % for Cys and Met, respectively (Li et al. 2005); however, recently it has been reported that this protein has allergenic properties (Krishnan et al. 2010). Soybean transformation with a gene coding for the hypoallergenic protein, bovine β -casein, also high in Met, has shown a maximum protein accumulation of 0.4 % of the total soluble seed protein, but an improvement of SAAs has not been reported (Maughan and Philip 1999; Philip et al. 2001).

Progress in understanding the structure, function, folding, and topology of proteins allows the design and synthesis of genes encoding novel proteins with

desired amino acids. Such a gene, (ASP1), has been designed and delivered to tobacco, sweet potato, and rice (Kim et al. 1992; Potrykus 2003; Zhang et al. 2003), resulting in an increase in lysine in these species. Another de novo protein, CP3-5, containing 31 % lysine and 20 % Met, was expressed in transgenic tobacco and resulted in a significant increase in the total lysine content in the seed (Keeler et al. 1997).

Substantial progress has been achieved with genetic engineering in improving soybean protein quality, but these improvements are not sufficient to adequately meet the nutritional requirements for food and feed. The objective of this work was to boost soybean sulfur amino acid content by expressing the de novo Milk Bundle proteins, MB-1 and MB-16, in transgenic soybean seed. MB-1, an artificial 11 kDa protein containing 67 % essential amino acids (EAAs), methionine, threonine, lysine and leucine, was designed for a high complement of EAA, tailored for dairy animal nutrition (Beauregard et al. 1995); however it was found lacking in folding stability as assessed by urea and thermal denaturation (Hefford et al. 1999). The variant MB-16, has a flexible, five-glycine residue sequence in the second loop that results in a tighter folded structure than MB-1, resulting in higher stability to thermal denaturation and protease degradation (Simons et al. 2001). This is the first attempt of using a synthetic protein to increase soybean nutritional quality. MB-1Trp, also more resistant to thermal and proteolytic degradation than its parent molecule, MB-1 (Gagnon et al. 2000), was introduced into maize (Kirihaara et al. 2001) and alfalfa (Khouidi and Beauregard 2005); the mRNA was expressed but the de novo protein could not be detected. In this report, we present data of MB-1 and MB-16 mRNA and protein expression in soybean using three different promoters and different codons with or without the KDEL ER retention sequences.

Materials and methods

Vector construction

pFFMB-1: The MB-1 coding region was cloned as a 300 bp *Hind* III (Klenow-blunted) *Bam*H I fragment from pMal-MB-1 (Beauregard et al. 1995) into the *Sal* I (Klenow-blunted)—*Bam*H I site from pFF19

(Timmermans et al. 1990) to produce pFFMB-1 containing the CaMV 35S promoter, with the enhancer duplicated, MB-1 and the 35S polyadenylation signal.

pGMCP3: The promoter GMCP3, which drives an α' subunit of β -conglycinin, was amplified from *Glycine max* cv. Harosoy 63 genomic DNA using forward primer 5'ACACaagcttGTCTCGAATGCCA TAACACTGACTAG3' and reverse primer 5'ACACccatggTATATCTTAAATTCTTTAATACGG3'. The promoter sequence corresponds to sequence positions 454–962 of GenBank accession M13759 (Doyle et al. 1986). The PCR fragment was cloned into the *Eco*R V site of pBluescript II-KS (Stratagene) generating vector pGMCP3.

β -con-MB-16: To improve the stability of MB-1, MB-16 was generated by replacing ACTAAGAACA AA (encoding AKNK) with GGTGGAGGAGG TGGG (encoding GGGGG) (Simons et al. 2001). The MB-16 gene was amplified from plasmid pMal-2 with PCR primers 5'-GATCGAGGGAACCATG GCTACGGAAGAC-3' and 5'-TGGCGAAAGGGGG ATGTGCT-3'. The PCR product was ligated into the *Nco* I site of vector pGMCP3 behind the β -conglycinin promoter. A *Hind* III fragment comprising the β -conglycinin promoter with MB-16 gene was sub-cloned into the *Hind* III site of pFFMB-1, in front of the 35S polyadenylation signal, generating β -con-MB-16.

MB-16sc: The MB-16 gene, originally designed for bacterial expression, was optimized in order to eliminate rare codons in soybean. A MB-16 gene with soybean codon bias was generated by mixing overlapping oligonucleotides to obtain a PCR fragment from the combined overlaps. The final sequence was verified (Supplemental Fig. 1S).

MB-16scKDEL: The ER-retention signal, KDEL coding sequence, AAAGATGAGCTA, was fused to the C-terminal of MB-16sc to generate MB-16scKDEL.

β -con-MB-16sc: in this vector, the MB-16 fragment in β -con-MB-16 was replaced by the codon-optimized MB-16sc.

Cru-MB-16scKDEL: The *Brassica napus* cruciferin promoter (cru, sequence corresponding to pos. 3,318–3,390 of GenBank accession number AF319771) was inserted into the β -con-MB-16sc, vector, replacing the β -conglycinin promoter.

The vectors used for transformation are listed in Table 1.

Table 1 Transgenic events recovered from four different *MB-1/MB-16* constructs

Construct	Gene of Interest ^a		Transgenic Events	
	Promoter	Milk bundle gene	Original name	Simplified name
<i>pFFMB-1</i>	<i>CaMV 35S</i>	<i>MB-1</i> bacterial codon	2B6-2 ^b	35S-bc
<i>β-con-MB-16</i>	<i>β-conglycinin</i>	<i>MB-16</i> bacterial codon	2B21-2A20	Con-bc
<i>β-con-MB-16sc</i>	<i>β-conglycinin</i>	<i>MB-16</i> soybean codon	2B12-4A2	Con-sc
<i>Cru-MB-16scKDEL^c</i>	<i>Cruciferin</i>	<i>MB-16</i> soybean codon	5B9-10A2	Cru-sc-KDEL

^a The CaMV 35S terminator was used for all constructs

^b Figures in front of hyphen refer to experiment number; figures following hyphen refer to one transgenic event

^c The ER retention sequence KDEL was included in the construct *Cru-MB-16scKDEL*

Plant material, tissue culture and transformation

MB-1 or MB-16 constructs were introduced into soybean via biolistics (PDS-1000/He Particle Delivery System, Bio-Rad Laboratories, Richmond, CA). The transformation, selection and plant regeneration protocols were previously described (Simmonds 2003) and are similar to the methods using proliferative embryogenic cultures described by Parrott and Clemente (2004). Briefly, proliferative embryogenic cultures of soybean cv. X5 (X2650-7-2-3) (Simmonds and Donaldson 2000) and We (OAC Westag 95) were co-bombarded with the various MB-1 or MB-16 constructs and pHygR constructs (encoding aphIV driven by CaMV 35S promoter and 3' Nos terminator). Transgenic events were selected and maintained on 55 mg L⁻¹ hygromycin. Embryos were matured on antibiotic-free medium, air-desiccated and converted on B5 solid medium (Gamborg et al. 1968); tissue cultures and regenerating plantlets were maintained at 20 °C and 20 h photoperiod. Upon transfer to soil, plants designated the T0 generation were grown in a controlled environment chamber. T0 plants carrying the transgene were identified by PCR and brought to seed (T1). Homozygous T1 plants were identified by screening the T2 progeny. Homozygous T2–T6 transgenic plants were planted in 5 inch pots in 3:1:1:1 mix of ASB Original grower Mix (a peat based mix from Greenworld, ON, Canada). They were grown in controlled environment chambers at 26/24 °C (day/night), 13 h photoperiod provided by cool white fluorescent lights and incandescent bulbs, at a light intensity of 300–400 μ moles m⁻² s⁻¹, watered every day and fertilized with a 20-20-20 (N–P–K) nutrient solution.

Seed stages

Seeds were collected for analysis at different stages of development and categorized according to length and DAP (days after pollination) into six stages (Supplemental Table 1S),

- Stage 1: 3–3.5 mm long (green translucent, 10–12 DAP),
- Stage 2: 5–6 mm long (green opaque, 14–16 DAP),
- Stage 3: full size (green opaque, 20–24 DAP),
- Stage 4: full size (turning yellow, 28–31 DAP),
- Stage 5: full size (yellow/dry, 31–35 DAP),
- Stage 6: mature dry seed, imbibed overnight)

PCR and southern analysis

Genomic DNA isolation was based on Doyle and Doyle CTAB procedure (Doyle 1991). Identification of transgenic lines and T2 progeny segregation analysis was carried out by PCR. Primers for detecting *β-con-MB-16sc* were 5'-CCCAAAAGCCATGCACAACAAC-3' and 5'-TCCACCTCCCCCACCCTTCT-3'); for *Cru-MB-16scKDEL*, the primers used were 5'-CCTCTTAA CATGCAACCGAACA-3' and 5'-ATCTCCGCATAC TCCAAACCAC. Gene copy number and integration patterns were examined by Southern blotting using the Roche DIG hybridization system. For the probe, a 240 bp MB-16 coding region was amplified with primers 5'-CCGATATGATGACCACCCTCTTTA-3' and 5'-GATTTTTTCATGGTTGTAGCAGTCT-3' using the DIG DNA labeling kit (Roche, Cat # 1 277 065).

Real time RT-PCR and northern analysis

Total mRNA was isolated from transgenic soybean seeds at different developmental stages using TRizol Reagent (Invitrogen Cat # 15596-018) and MaXtract High Density tubes (Qiagen Cat # 129065). RNA was treated with RNase-free DNase enzyme (Ambion, Cat. # AM1906) followed by purification with RNeasy spin column (Qiagen, Cat. # 74104). For real time RT-PCR, 1 µg of total mRNA was used for cDNA synthesis (Bio-Rad, iScript cDNA synthesis kit, 170-8891). Amplifications were performed in a total volume of 20 µL containing iQ SYBR Green Supermix (Bio-Rad), primer pairs (0.4 µM), cDNA (amount corresponding to 5 ng starting mRNA). Samples were cycled 45 times, starting with an initial activation step of 10 min at 95 °C, followed by denaturation for 15 s at 95 °C, annealing for 15 s at 60 °C and extension for 15 s at 72 °C. Data were acquired for 2 s after extension. At the end of the PCR cycles, the temperature was increased from 50 to 99 °C at a rate of 0.2 °C s⁻¹, and the fluorescence was measured continuously to construct the melting curve. Soybean β-actin gene was used as a reference to normalize the amounts of MB-16 gene transcription in each sample. MB-16 specific primers 5'-GCGATGGATGAGGCCACTAAG-3' and 5'-GCCATGTCTGTCATATCTTCTCCA-3' were used for target gene amplification; β-actin primers 5'-TCCAA GGGGACCTAACGGAGA-3' and 5'-TGGGTCAAGA GCTGGATGGTG were used for reference gene amplification. In addition, the same MB-16 and β-actin primer pairs were used in PCR with serially diluted cDNA samples to establish the efficiency of amplification. The reactions were run on Bio-Rad Chromo4 Real-Time PCR Detection system and the output data were analyzed by REST software (Pfaffl et al. 2002). Northern blotting was used to confirm the real time RT-PCR results. Total RNA samples (15 µg) were separated on denaturing gel and transferred to a nylon membrane; the Roche DIG hybridization instructions were followed using the same probe as that for Southern blotting.

MB-16 protein detection by immunoblotting

Total soluble proteins were extracted from seeds at different developmental stages. The protein extraction buffer (62 µM Tris, pH 6.8, 2.3 % SDS, 10 % glycerol, 0.1 M DTT) was supplemented with a protease inhibitor cocktail (Sigma, Cat. # P9599), to

prevent proteolysis of MB-16 during extraction. Soluble proteins were quantified with the Bradford assay (Sigma, Cat. # P6916). For western blotting, 120 µg protein was loaded into each well and separated by sodium dodecyl sulfate poly acrylamide gel electrophoresis (SDS-PAGE) (15 % tricine gels) and transferred onto PVDF membranes (Bio-Rad Cat. # 162-0177). The membranes were blocked overnight with 1 % blocking reagent (Roche, Cat # 11 096 176), and treated for 1 h with 1:1,000 dilution of rabbit polyclonal antibody raised against *E. coli*-derived MB-16 (Alpha Diagnostic International). A goat anti-rabbit antibody conjugated with horseradish peroxidase (Roche Diagnostics) was used as the secondary antibody. Immunoblot detection was performed by chemiluminescence according to the manufacturer's instructions (Roche Diagnostics).

MB-16 protein production

MB-16 was expressed, produced and purified as described elsewhere (Simons et al. 2001). After the final purification step, protein was exhaustively dialyzed against ammonium bicarbonate (10 mM) and then lyophilized to constant weight. Lyophilized protein was stored at -20 °C and then resuspended in appropriate buffer immediately before use. Analysis by SDS-PAGE followed by silver staining confirmed that each preparation was at least 95 % pure.

MB-16 proteolysis in soybean seed extract

Total soluble protein was prepared from non-transgenic soybean seed at the immature stage (stage 2 or 3) and the mature stage (stage 5 or 6). The seed was macerated in double-distilled water (400 µL) instead of protein extraction buffer to extract total protein. 5 µg of purified MB-16 protein was added to 20 µL of the extract, and incubated at room temperature for 0, 10, 30, 120 and 180 min. Reactions were stopped by adding 4 µL of 6× concentrated sample loading buffer and boiling for 5 min. Samples were separated by SDS-PAGE and immunoblotted as described above.

Amino acid profile analysis

Dry seed of three replicate samples, randomly picked, were pulverized in a standard electrically driven end

runner mill (Cyclone Sample Mill, U.D.Corp., Fort Collins, CO.), and defatted with hexane. Amino acid profiles were determined at the Advanced Protein Technology Centre of the Hospital for Sick Children, Toronto, Ontario, with the Waters Pico-Tag System. The defatted samples were hydrolyzed in 6 M HCl with 1 % phenol in a vacuum chamber under nitrogen for 24 h at 110 °C; the samples were then derivatized using phenylisothiocyanate, followed by reverse phase HPLC. Cysteine and tryptophan amounts were analyzed separately (Hirs 1967). Non-hydrolyzed samples were analyzed for free amino acids.

Total protein content was evaluated using an Infratec 1241 analyzer (FOSS NIR Systems, Inc., Eden Prairie, MN). Briefly, readings from three samples of a minimum of 100 g of seed were taken and averaged.

1-D SDS-PAGE

For 1-D SDS-PAGE, 20 stage-5 seeds per sample were ground in liquid nitrogen; the powder was added to 30 mL extraction buffer [20 mM Tris pH 6.8 containing Protease Inhibitor Cocktail Tablet (Roche 04 693 124 001) prepared according to the manufacturer guidelines]. Samples were further processed using a handheld homogenizer, vortexed for 15 min and cleared by centrifugation at 13,900×g for 15 min at 10 °C. Protein concentration was determined on two aliquots using the Bio-Rad Protein Assay Reagent (cat. 500-0006) and BSA as a standard.

Ten µg protein of each, parent and transgenic line, were loaded in triplicate on 13.5 % gels and run for 45 min at 200 V. Gels were stained with Coomassie Blue and imaged using the Bio-Rad Image Lab 3.0 software. Band intensities were quantified using the ‘quantity tool’ of Bio-Rad Image Lab 3.0 software using the rolling ball method of subtraction. Values were averaged and plotted.

Results

Integration of synthetic MB-16 into soybean

The first milk bundle protein to be introduced into soybean was MB-1 driven by the constitutive promoter *CaMV 35S* (Table 1, *pFFMB-1*). MB-1 protein in the transgenic soybean line, 35S-bc, was expressed

in leaves but was undetectable in seed on western blots. To improve seed expression, constructs were designed using the soybean codon bias, with and without the KDEL ER retention sequence, and β-conglycinin or cruciferin seed specific promoters to drive the MB-16 gene (Table 1).

More than ten events, per construct, were regenerated from hygromycin selection as determined by PCR screening (data not shown). Lines with good fertility and stable expression of the MB-16 protein were advanced. Southern blot analyses of genomic DNA of selected T1 plants indicated that most lines carried multiple copies of the transgene and showed different integration patterns (Supplemental Fig. 2aS). *Hind* III digestion that drops out the expression cassettes (1 kb for *Cru-MB-16scKDEL* and 0.8 kb for *β-con-MB-16sc*), produced fragments both larger and smaller than the expected size, indicating possible rearrangements or fragmentations of the MB-16 gene. The three top expressing lines, Con-bc, Con-sc and Cru-sc-KDEL, each have 5 to 6 copies of the MB-16 gene (data not shown). Mendelian segregation (3:1) of T1 progeny ($\chi^2 = 0.11$, $p > 0.75$, data not shown) indicated that the multiple copies of the transgenes were within one genetic locus. This is not surprising as DNA copies delivered by particle bombardment generally integrate into one genetic locus and segregate 3:1 (Li et al. 2011). Homozygous T1 lines were identified by PCR screening of T2 progeny; Con-bc and Cru-sc-KDEL were advanced to the T3 generation and Con-sc to T5. Stable integration of the MB-16 gene was confirmed by Southern blot analysis that showed no change in the integration pattern for each line over multiple generations (data shown for Con-sc, Supplemental Fig. 2bS).

Transcription profiles of *MB-16* driven by β-conglycinin or cruciferin promoters

Transcription profiles of MB-16 during seed development showed a different pattern for each of the promoters (Fig. 1a). Con-sc showed the strongest transcription at stages 2 and 3, a lower level at stage 1 and 6 and almost no transcription at stage 4 (see Supplemental Table 1S for stage definition). Transcription levels of the Cru-sc-KDEL line steadily increased from stage 1 to stage 3, dropped at stage 4 and was undetectable at stage 6. Quantitative RT-PCR was conducted for pair-wise statistical comparisons between different seed development stages in the same lines as

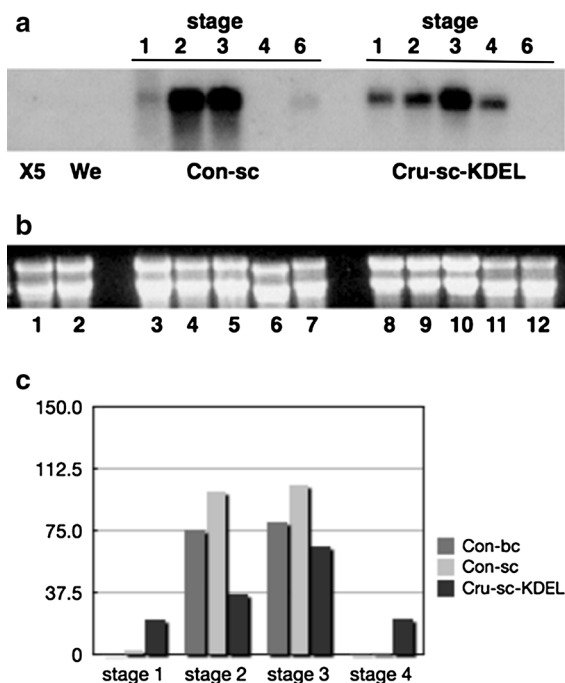


Fig. 1 **a** Northern blot analysis and **b** corresponding RNA gel, showing transcription levels of MB-16 driven by two different promoters at different stages of seed development. Seed stages (see Supplemental Table 1S) are indicated in **(a)**. Lanes 1,2—negative control, wild type *parental lines* X5 and Westag 97 (We); 3,4,5,6,7—RNA from Con-sc seed; 8,9,10,11,12—RNA from Cru-sc-KDEL seed. **c** Transcription profiles from developing seed of three MB-16 transgenic lines, Con-bc, Con-sc and Cru-sc-KDEL. Quantitative RT-PCR data, shown as comparative $9C(t)$ values, was generated from three biological replicates (each with 3 RT-PCR replicates) and analyzed by REST software

well as corresponding stages between different lines (Fig. 1c). Lines Con-sc and Con-bc, driven by the same promoter, had a similar expression pattern where stage 2 and stage 3 were at least 25 times stronger than stage 1 and stage 4 ($p = 0.001$). However, Con-sc had significantly higher transcription levels than Con-bc ($p = 0.001$) at stage 1 and stage 2, indicating improved expression with soybean codon bias. Cru-sc-KDEL had significantly higher transcription levels than Con-sc line at stage 1 and stage 4 ($p = 0.001$), and lower levels at stage 2 and 3 ($p = 0.05$).

MB-16 protein expression

Western blot analysis data of line Con-sc (Fig. 2a) agreed with the results of the quantitative RT-PCR

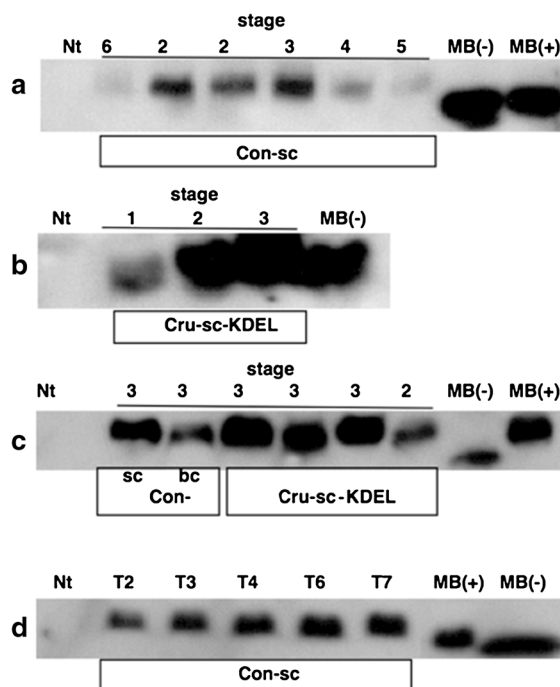


Fig. 2 Western blot of MB-16 protein of seed at different developmental stages from three transgenic lines. **a** Transgenic line Con-sc: Nt, negative control wild type parental line X5; MB(–), purified MB-16 protein; MB(+), purified MB-16 mixed with X5 seed extract. **b** Transgenic line Cru-sc-KDEL. **c** Comparison of MB-16 protein levels from different transgenic lines at seed stage 2 or 3. **(d)** MB-16 protein levels of transgenic line Con-sc over generations T2 to T7

(Fig. 1c). MB-16 protein was not detectable in stage 1 developing seed (data not shown), reached the highest levels at stages 2 and 3, then decreased at stage 4, and finally was almost undetectable at stage 5 and 6. Line Con-bc had a similar protein profile but lower expression. In transgenic line Cru-sc-KDEL, MB-16 began to accumulate at stage 1, reached a peak at stage 3 and declined thereafter (Fig. 2b). Western blot data showed negligible protein expression in mature seed in all three lines. However, expression during the early developmental stages was consistent through multiple generations, as shown in line Con-sc (Fig. 2d). Comparison of the MB-16 protein levels at stage 3 in the three transgenic lines showed that line Cru-sc-KDEL had the highest levels, and line Con-bc had the lowest levels (Fig. 2c). The higher protein levels of Cru-sc-KDEL at stage 3 coupled with a lower transcript levels, as compared to Con-sc and Con-bc (Fig. 1c) suggests that the KDEL peptide may be extending the residency time

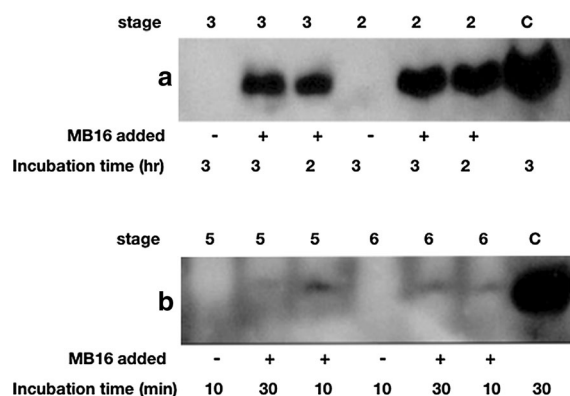


Fig. 3 Western blot of purified MB-16 protein incubated with extracts of wild type seed from different development stages. **a** MB-16 protein incubated with extracts from stage 2 or 3 wild type seed for 2 or 3 h. **b** MB-16 protein incubated with extracts from stage 5 or 6 wild type seed for 10 or 30 min. Seed extract without (–) or with (+) MB-16 protein (50 ng). Extraction buffer with MB-16 protein (C)

in the developing seed, thereby slowing the degradation of the de novo protein (Herman et al. 1990).

Proteolysis of MB-16 protein in mature soybean seeds

The cause for the disappearance of MB-16 protein from maturing seed, as indicated by western blots, was further investigated. The degradation capacity of proteolytic enzymes of mature seed was compared to that of immature seed (Fig. 3). Pure MB-16 protein was added to wild type seed extract and incubated at room temperature for different time periods. The MB-16 protein was not degraded after a 30 min incubation with extracts from stage 2 or 3 seed (data not shown) but was reduced by 30–50 % after 2–3 h of incubation (Fig. 3a). However, almost all MB-16 protein was degraded after only 10 min incubation with stage 5 or 6 seed extracts (Fig. 3b).

Amino acid analysis

Total protein dry mass in Con-sc (39.9 %) is lower than in its X5 parental line (43.3 %) and higher in Cru-sc-KDEL (42.9 %) than in its Westag parental line (38.7 %). The amino acid balance of both transgenic lines was improved. Met and Cys increased significantly by 16.2 and 65.9 %, respectively, resulting in a 42 % increase in total SAA content for Con-sc

(Table 2 and Supplemental Table 2S); the total SAA increase of 19 % was less dramatic for Cru-sc-KDEL as Met and Cys increased by 6.0 and 31.4 %, respectively (Table 2 and Supplemental Table 2S). The levels of the EAAs, threonine, lysine and leucine, the other three major components of MB-16 protein also increased (Table 2 and Supplemental Table 2S). The free amino acid pool was very small, indeed, several amino acids were undetectable (Supplemental Table 3S). The level of free Met was similar in the transgenic and parental lines, while Cys was undetectable in the multiple samples of all the lines.

Protein quality for food and feed is generally evaluated on the profile of its EAAs and digestibility. The gold standard is that of the requirements of a 2–5 year-old child (Young and Steinke 1992). Judged against this standard, the transformed lines surpass the requirements for all the EAAs except Lysine (Supplemental Table 2S). However, Lysine too, increased in the two transgenic lines by 18 and 26 % as compared to the WT lines. Furthermore, both transgenic lines have an increased total EAA, another indicator of improved protein quality (Supplemental Table 2S).

Seed protein analysis

The protein content of mature, stage 5 seed was analyzed on 1-D gels to determine whether an increased pool of SAAs affects seed protein composition. Expression differences of the major storage proteins, the conglycinins, were detected. The Cru-sc-KDEL line showed a marked increase of conglycinin compared to the parental line, while the same fraction was lowered in the Con-sc line (Fig. 4). These changes might account for the observation that the total protein contents of the two transgenic lines differ from that of their parental lines with Cru-sc-KDEL having a higher, and Con-sc having a lower total protein content.

Discussion

This report describes the integration of a synthetic gene, Milk Bundle (MB-16), into soybean in order to improve seed protein quality by elevating the SAAs, Met and Cys. The efficacy of β -conglycinin and cruciferin seed specific protein storage promoters was examined and the MB-16 bacterial and soybean codon preferences were compared. Southern blot analysis of

Table 2 Amino acid profiles of Con-sc and Cru-sc-KDEL transgenic lines and their WT parental lines, X5 and Westag, respectively, expressed as mg/g total protein

Amino Acid	X5 ^a	Con-sc	% Change	Westag ^a	Cru-sc-KDEL	% Change
Asx (Asp + Asn)	125.10 ± 0.49	121.93 ± 0.08	−2.5	128.10 ± 0.68	122.55 ± 0.33	−4.3
Glx (Glu + Gln)	207.48 ± 0.85	195.51 ± 0.06	−5.8	209.99 ± 1.06	196.82 ± 0.50	−6.3
Ser	48.80 ± 0.18	48.71 ± 0.10	−0.2	49.59 ± 0.26	48.44 ± 0.12	−2.3
Gly	37.18 ± 0.16	36.90 ± 0.08	−0.7	36.45 ± 0.23	36.48 ± 0.12	+0.1
His	28.82 ± 0.11	28.41 ± 0.01	−1.4	27.62 ± 0.12	27.75 ± 0.06	+0.5
Arg	93.12 ± 0.56	83.83 ± 0.02	−10.0 ^b	81.35 ± 0.27	77.61 ± 0.23	−4.6
Thr	37.17 ± 0.17	38.10 ± 0.09	+2.5	36.08 ± 0.25	36.73 ± 0.14	+1.8
Ala	42.06 ± 0.20	42.57 ± 0.10	+1.2	41.02 ± 0.29	40.85 ± 0.15	−0.4
Pro	52.30 ± 0.18	51.44 ± 0.03	−1.6	51.17 ± 0.25	49.81 ± 0.10	−2.7
Tyr	39.67 ± 0.08	40.86 ± 0.18	+3.0	41.33 ± 0.07	41.18 ± 0.09	−0.4
Val	51.74 ± 0.12	51.41 ± 0.06	−0.6	50.04 ± 0.14	50.84 ± 0.05	+1.6
Met	11.11 ± 0.04	12.91 ± 0.03	+16.2 ^b	11.47 ± 0.09	12.16 ± 0.02	+6.0
Ile	45.64 ± 0.15	47.34 ± 0.17	+3.7	47.64 ± 0.16	48.80 ± 0.16	+2.4
Leu	75.29 ± 0.32	77.97 ± 0.37	+3.6	76.64 ± 0.35	79.62 ± 0.33	+3.9
Phe	42.42 ± 0.83	45.81 ± 0.25	+8.0	44.88 ± 0.96	49.82 ± 0.50	+11.0
Lys	36.93 ± 1.31	43.81 ± 0.39	+18.6	41.02 ± 1.62	51.67 ± 0.82	+26.0
Cys A	11.72 ± 0.17	19.44 ± 0.10	+65.9 ^b	13.00 ± 0.13	17.08 ± 0.34	+31.4 ^c
Trp	13.44 ± 0.04	13.04 ± 0.35	−0.3	12.60 ± 0.15	11.77 ± 0.08	−6.7

^a The value of each amino acid is the average of three replicates

T-tests: ^b two tail significant ($\alpha = 0.05$) and ^c one tail significant ($\alpha = 0.05$) between the transgenic line and its parent

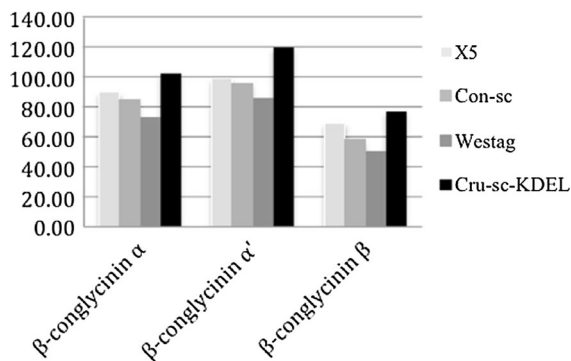


Fig. 4 β -conglycinin subunit band intensities as measured from 1-D gel bands. X5 (parent), Con-sc (transgenic); Westag (parent). Cru-sc-KDEL (transgenic). The non-quantitative intensity values are generated by the analytical software and are relative to an arbitrarily set background value

recovered T1 plants showed that most transgenic lines carried multiple copies of MB-16 genes. Transgenic lines with very high copy numbers (Supplemental Fig. 2aS, lanes 3, 4, 7 and 8) had low expression or no expression (data not shown). High copy numbers of a transgene most likely resulted in co-suppression and

gene silencing (Iyer et al. 2000; Sparks and Jones 2009). However, the low copy number events (Supplemental Fig. 2aS, lanes 5, 6, 19 and 20) also had poor expression (data not shown). The three best expressing lines, Con-bc, Cru-sc-KDEL and Con-sc (e.g. Supplemental Fig. 2bS), carry 5–6 copies of the transgene. Bradeen et al. (2009) reported that transgenic lines with higher copy numbers had enhanced transcription and expression levels, and suggested that the gene transcript levels may have to surpass a threshold before silencing is triggered (Bradeen et al. 2009); this agrees with our findings of poor expression in our low and very high copy number lines and high expression in the mid copy number lines.

The spatial and temporal expression pattern of the α' subunit promoter of β -conglycinin has been described in detail (Chen et al. 1986; Ladin et al. 1987; Imoto et al. 2008). The transcription profiles of MB-16 driven by the β -conglycinin promoter (Fig. 1) agreed with the previous studies (Walling et al. 1986; Harada et al. 1989), showing gene activation in early seed development, high expression by mid-maturation, and repression of transcription in drying seed.

The transcription levels of MB-16 with the soybean codon were higher than the bacterial codon at every developmental stage. On the other hand, transcription of MB-16 driven by the cruciferin promoter rose steadily from early seed development (12 DAP) to mid-maturity (24 DAP), dropping at maturity (31 DAP) and reaching zero in dry seed. This was somewhat different from a previous study in *Brassica* and tobacco that showed accumulation of cruciferin mRNA beginning at 23 DAP, peaking at 38 DAP, and thereafter declining but showing very low mRNA levels in the dry seed (Sjödahl et al. 1993, 1995).

Improvement of storage protein in plants by the introduction of foreign genes has been a challenge. MB-1, the antecedent of MB-16, was introduced to maize to improve its nutritional content but expression data was absent (Kirihara et al. 2001). Alfalfa was transformed with MB-1Trp, a stable mutant of MB-1, and in spite of the high efficiency of transformation and the large amounts of mRNA, the protein was not detected (Khouidi and Beauregard 2005). The expression of foreign proteins in seeds frequently results in post-translational degradation of the newly synthesized proteins. As early as 1988, attempts to improve the amino acid balance of seed storage proteins were reported when Hoffman et al. (1988) modified phaseolin to increase the SAA content. The modified protein however proved to be unstable when expressed in tobacco (Hoffman et al. 1988). The addition of the ER retention signal HDEL somewhat alleviated the problem but degradation still occurred (Pueyo et al. 1995). Other examples of degradation of foreign proteins in transgenic plants include zeins in tobacco seed (Scherthaner et al. 1988; Williamson et al. 1988; Ohtani et al. 1991), sea anemone equistatin in potato (Outchkourov et al. 2003) and hepatitis B in potato (Richter et al. 2000). In the study presented here, the expected 11-kDa band, unique to the transgenic lines, was detected on western blots, indicating that this *de novo* synthetic protein was expressed and accumulated in immature seed. However, the protein band was diminished during seed maturation, and was no longer detectable in mature seed. Philip et al. (2001) also reported a great reduction of a foreign protein, bovine β -casein, in mature soybean seed (Philip et al. 2001). These findings may indicate that foreign proteins are turned over rapidly in developing soybean seed and that their storage is not protected. This is supported by the Cru-sc-KDEL line which, in spite of lower

transcription during seed mid-maturation, showed a higher accumulation of the protein than Con-sc (no KDEL); it appears that the KDEL peptide protected MB-16 transiently, prolonging its native state. In contrast, Cunha et al. (2011) reported the successful expression of a functional recombinant human growth hormone in soybean seeds. In this study, the hGH gene was placed under the control of the β -conglycinin α' subunit promoter and terminator. In addition, a α -Coixin signal peptide was placed in front of the hGH gene. The recombinant hGH protein was found to be properly processed and stored in seed protein storage vesicles (Cunha et al. 2011).

The disappearance of MB-16 at the end of seed development may be explained by findings that the ER is generally degraded at the end of seed maturation and replaced by newly formed ER after germination (Gilkes et al. 1979). In addition to this mechanism, proteolytic enzymes appear to have had a role in MB-16 degradation. The digestion experiment (Fig. 3), comparing purified MB-16 protein degradation in extracts from immature or mature seed, showed that MB-16 remained stable in immature seed extract but was degraded rapidly in mature seed extract. This demonstrates that endogenous proteases present during the later stages of seed development may be responsible for the disappearance of MB-16 from the immunoblots. However, it remains to be determined whether MB-16 was present in mature seed but degraded during processing or whether endogenous enzymes *in situ* degraded it. As the small free amino acid pool was similar in the transgenic and parental lines, we can infer that MB-16 underwent partial degradation into peptides. A breakdown of the native MB-16 protein structure inevitably results in the loss of its antigenic specificity and may explain its disappearance from immunoblots. Furthermore, the conversion of MB-16 into peptides also explains the elevation of SAAs when MB-16 is no longer detectable.

The change in the relative amounts of EAAs is an important contribution of the transgenic lines. The Met content in seed was increased by 6.0 and 16.2 % in Cru-sc-KDEL and Con-sc lines, respectively. Previous studies have shown that Met had a negative effect on the transcription of the β -conglycinin gene, a major soybean storage protein (Creason et al. 1983; Ohkama et al. 2002; Ohkama-Ohtsu et al. 2008). It is possible that Met, released from MB-16 turnover depressed the production of the β -conglycinins in the Con-sc lines (Fig. 4), moreover it may also have enhanced the rapid

decrease of transcription of MB-16 driven by the β -conglycinin promoter. Although both transgenic lines displayed changes in conglycinin expression compared to their parental lines (Fig. 4), it is not clear whether this is a direct consequence of MB-16 expression itself or caused by genetic side effects stemming from the transformation process. Altogether due to the large proportion of conglycinins in the overall seed protein content, these changes might explain why the total protein contents of the two transgenic lines differ from their parental lines with Cru-sc-KDEL having a higher, and Con-sc having a lower total protein content. Further studies will reveal the cause of this differential conglycinin expression.

Cysteine (Cys), also a sulfur rich amino acid, was significantly higher in both Cru-sc-KDEL and Con-sc lines than the parental line, by 31.4 and 65.9 %, respectively. A major increase might be expected for Met rather than Cys because MB-16 is rich in Met and does not contain Cys. The biosynthetic pathway of Met and Cys helps to explain this result (Hesse and Hoefgen 2003). Two biosynthetic pathways involved with Met biosynthesis consist of i) a branch of the aspartate amino acid family and ii) the sulfur assimilation and biosynthesis of Cys, a precursor of cystathionine. Cystathionine is an important intermediate for Met synthesis that is regulated by cystathionine gamma-synthase (CgS). Studies of Arabidopsis plants over-expressing CgS revealed that CgS is transcriptionally and post-transcriptionally regulated by Met or one of its metabolites and that unregulated expression of CgS leads to an increase in Met (Gakière et al. 2000; Kim and Leustek 2000). Elevated levels of Met resulted in a reduction in CgS activities and RNA abundance (Inaba et al. 1994), (Chiba et al. 1999) showing that Met may be a negative feed-back regulator of CgS. In the case presented here, the Met from degraded MB-16 may have reduced CgS activity resulting in accumulation of Cys.

In conclusion, the sulfur amino acid deficiency of soybean seed has been addressed by the elevation of methionine and cysteine in lines carrying the de novo MB-16 gene. This study has shown that it is feasible to alter the amino acid ratio in soybean seed with a novel nutritive protein. Future investigation will focus on best approaches to enhance the storage of foreign proteins in soybean seed.

Acknowledgments We thank Sheryl Hubbard for her technical assistance and are grateful to the financial support from the

Ontario Soybean Board and Eastern Cereal and Oilseed Research Centre. Agriculture and Agri-Food Canada Contribution # 13-468.

References

- Altenbach SB, Pearson KW, Leung FW, Sun SSM (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. *Plant Mol Biol* 8:239–250
- Beauregard M, Dupont C, Teather RM, Hefford MA (1995) Design, expression, and initial characterization of MB-1, a de novo protein enriched in essential amino acids. *Bio-technology* 13:974–981 (NY)
- Bradeen JM, Iorizzo M, Molloy DS, Raasch J, Kramer LC, Millett BP, Austin-Phillips S, Jiang J, Carpato D (2009) Higher copy numbers of the potato RB transgene correspond to enhanced transcript and late blight resistance levels. *Mol Plant Microbe Interact* 22:437–446
- Cederroth CR, Nef S (2009) Soy, phytoestrogens and metabolism: a review. *Mol Cell Endocrinol* 304:30–42
- Chen Z-L, Schuler MA, Beachy RN (1986) Functional analysis of regulatory elements in a plant embryo-specific gene. *Proc Natl Acad Sci* 83(22):8560–8564
- Chiba Y, Ishikawa M, Kijima F, Tyson RH, Kim J, Yamamoto A, Nambara E, Leustek T, Wallsgrove RM, Naito S (1999) Evidence for autoregulation of cystathionine gamma-synthase mRNA stability in Arabidopsis. *Science* 286:1371–1374
- Creason GL, Holowach LP, Thompson JF, Madison JT (1983) Exogenous methionine depresses level of mRNA for a soybean storage protein. *Biochem Biophys Res Commun* 117:658–662
- Cunha NB, Murad AM, Cipriano TM, Araújo ACG, Aragão FJL, Leite A, Vianna GR, McPhee TR, Souza GHMF, Waters MJ, Rech EL (2011) Expression of functional recombinant human growth hormone in transgenic soybean seeds. *Transgenic Res* 20(4):811–826
- Dinkins RD, Reddy MSS, Meurer CA, Yan B, Trick H, Thibaud-Nissen F, Finer JJ, Parrott WA, Collins GB (2001) Increased sulfur amino acids in soybean plants over-expressing the maize 15 kDa zein protein. *In vitro cellular and developmental biology. Plant* 37:742–747
- Doyle JJ (1991) DNA isolation from small amounts of plant tissue. *Phytochem Bull* 19:11–15
- Doyle JJ, Schuler MA, Godette WD, Zenger V, Beachy RN, Slightom JL (1986) The glycosylated seed storage proteins of *Glycine max* and *Phaseolus Vulgaris*. Structural homologies of genes and proteins. *J Biol Chem* 261(20):9228–9238
- El-Shemy HA, Khalafalla MM, Fujita K, Ishimoto M (2007) Improvement of protein quality in transgenic soybean plants. *Biol Plant* 51:277–284
- Gagnon MC, Williams M, Doucet A, Beauregard M (2000) Replacement of tyr62 by trp in the designer protein milk bundle-1 results in significant improvement of conformational stability. *FEBS Lett* 484(2):144–148
- Gakière B, Ravanel S, Droux M, Douce R, Job D (2000) Mechanisms to account for maintenance of the soluble

- methionine pool in transgenic *Arabidopsis* plants expressing antisense cystathionine gamma-synthase cDNA. *C R Acad Sci III* 323:841–851
- Gamborg OL, Miller RA, Ojima K (1968) Nutrient requirements of suspension cultures of soybean root cells. *Exp Cell Res* 50:151–158
- Gilkes NR, Herman EM, Chrispeels MJ (1979) Rapid degradation and limited synthesis of phospholipids in the cotyledons of mung bean seedlings. *Plant Physiol* 64:38–42
- Hajika M, Takahashi M, Sakai S, Igita K (1996) A new genotype of 7S globulin (β -conglycinin) detected in wild soybean (*Glycine soja* Sied. et. Zucc). *Breed Sci* 44:65–66
- Harada JJ, Barker SJ, Goldberg RB (1989) Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell* 1:415–425
- Hefford MA, Dupont C, MacCallum J, Parker MH, Beauregard M (1999) Characterization of MB-1. A dimeric helical protein with a compact core. *Eur J Biochem* 262(2):467–474
- Herman EM, Tague BW, Hoffman LM, Kjettrup SE, Chrispeels MJ (1990) Retention of phytohemagglutinin with carboxylterminal tetrapeptide KDEL in the nuclear envelope and the endoplasmic reticulum. *Planta* 182:305–312
- Hesse H, Hoefgen R (2003) Molecular aspects of methionine biosynthesis. *Trends Plant Sci* 8:259–262
- Hirs, CHW (1967) Book chapter in: methods of enzymology. Elsevier
- Hoffman LM, Donaldson DD, Herman EM (1988) A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants. *Plant Mol Biol* 11:717–729
- Imoto Y, Yamada T, Kitamura K, Kanazawa A (2008) Spatial and temporal control of transcription of the soybean β -conglycinin alpha subunit gene is conferred by its proximal promoter region and accounts for the unequal distribution of the protein during embryogenesis. *Genes Genet Sys* 83(6):469–476
- Imsande J (2001) Selection of soybean mutants with increased concentrations of seed methionine and cysteine. *Crop Sci* 41:510–515
- Inaba K, Fujiwara T, Hayashi H, Chino M, Komeda Y, Naito S (1994) Isolation of an *Arabidopsis thaliana* Mutant, mto1, that over accumulates soluble methionine (temporal and spatial patterns of soluble methionine accumulation). *Plant Physiol* 104:881–887
- Iyer LM, Kumpatla SP, Chandrasekharan MB, Hall TC (2000) Transgene silencing in monocots. *Plant Mol Biol* 43:323–346
- Keeler SJ, Maloney CL, Webber PY, Patterson C, Hirata LT, Falco SC, Rice JA (1997) Expression of de novo high-lysine alpha-helical coiled-coil proteins may significantly increase the accumulated levels of lysine in mature seeds of transgenic tobacco plants. *Plant Mol Biol* 34:15–29
- Khoudi H, Beauregard M (2005) The de novo designed nutritive protein MB-1Trp does not resist proteolytic degradation in alfalfa leaves. *Plant Physiol Biochem* 43:1039–1043
- Kim WS, Krishnan HB (2004) Expression of an 11 kDa methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. *Plant Biotechnol J* 2:199–210
- Kim J, Leustek T (2000) Repression of cystathionine [gamma]-synthase in *Arabidopsis thaliana* produces partial methionine auxotrophy and developmental abnormalities. *Plant Sci* 151:9–18
- Kim JH, Cetiner S, Jaynes JM (1992) Enhancing the nutritional quality of crop plants: design, construction and expression of an artificial plant storage protein gene. In *Molecular Approaches to Improving Food Quality and Safety*, An Avi Book, Van Nostrand Reinhold, New York (
- Kirihara, JA, Hibberd, KA, and Anthony, J (2001) Method for altering the nutritional content of plant seed. US Patent 6326527
- Krishnan HB (2005) Engineering soybean for enhanced sulfur amino acid content. *Crop Sci* 45:454–461
- Krishnan HB, Kerley MS, Allee GL, Jang S, Kim WS, Fu CJ (2010) Maize 27 kDa gamma-zein is a potential allergen for early weaned pigs. *J Agric Food Chem* 58:7323–7328
- Ladin BF, Tierney ML, Meinke DW, Hosangadi P, Veith M, Beachy RN (1987) Developmental regulation of β -conglycinin in soybean axes and cotyledons. *Plant Physiol* 84(1):35–41
- Li Z, Meyer S, Essig JS, Liu Y, Schapaugh MA, Muthukrishnan S, Hainline BE, Trick HN (2005) High-level expression of maize γ -zein protein in transgenic soybean (*Glycine max*). *Mol Breed* 16:11–20
- Li R, Bancroft B, Hutcheon C, Zhao S, Zheng S, Blahut-Beatty L, De Rocher J, Simmonds Daina (2011) Multiple inserts of gene of interest and selectable marker gene are CO-integrated and stably transmitted as a single genetic locus in transgenic soybean plants. *In Vitro Cell Dev Biol Plant* 47(2):274–281
- Maughan PJ, Philip R (1999) Biolistic transformation, expression, and inheritance of bovine beta-casein in soybean (*Glycine max*). *In Vitro Cell Dev Biol Plant* 35:344–349
- Messina MJ (1999) Legumes and soybeans: overview of their nutritional profiles and health effects. *Am J Clin Nutr* 70:439S–450S
- Messina MJ, Wood CE (2008) Soy isoflavones, estrogen therapy, and breast cancer risk: analysis and commentary. *Nutr J* 7:17
- Nordlee JA, Taylor SL, Townsend JA, Thomas LA, Bush RK (1996) Identification of a Brazil-nut allergen in transgenic soybeans. *N Engl J Med* 334:688–698
- Ohkama N, Goto DB, Fujiwara T, Naito S (2002) Differential tissue-specific response to sulfate and methionine of a soybean seed storage protein promoter region in transgenic *Arabidopsis*. *Plant Cell Physiol* 43:1266–1275
- Ohkama-Ohtsu N, Kezuka A, Onouchi H, Fujiwara T, Naito S (2008) Promoter region of the β subunit gene of β -conglycinin responds to methionine and glutathione in transient assays using *Arabidopsis* protoplasts. *Soil Sci Plant Nutr* 54:128–132
- Ohtani T, Galili G, Wallace JC, Thompson GA, Larkins BA (1991) Normal and lysine-containing zeins are unstable in transgenic tobacco seeds. *Plant Mol Biol* 16:117–128
- Outchkourov NS, Rogelj B, Strukelj B, Jongsma MA (2003) Expression of sea anemone equistatin in potato. Effects of plant proteases on heterologous protein production. *Plant Physiol* 133:379–390
- Paek NC, Imsande J, Shoemaker RC, Shibles R (1997) Nutritional control of soybean seed storage protein. *Crop Sci* 37:498–503
- Paek NC, Sexton PJ, Naeve SL, Shibles R (2000) Differential accumulation of soybean seed storage protein subunits in

- response to sulfur and nitrogen nutritional sources. *Plant Prod Sci Tokyo* 3:266–272
- Parrott, WA and Clemente, TE (2004) Transgenic Soybean. In: Soybeans: Improvement, Production, And Uses. H. Roger Boerma and James E. Specht (ed.) ed. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America 10, Dec 2012, p. 265–302. ISBN: 978-0-89118-266-5
- Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* 30:e36
- Philip R, Darnowski DW, Maughan PJ, Vodkin LO (2001) Processing and localization of bovine beta-casein expressed in transgenic soybean seeds under control of a soybean lectin expression cassette. *Plant Sci* 161:323–335
- Potrykus I (2003) Nutritionally enhanced rice to combat malnutrition disorders of the poor. *Nutr Rev* 61:S101–S104
- Pueyo JJ, Chrispeels MJ, Herman EM (1995) Degradation of transport-competent destabilized phaseolin with a signal for retention in the endoplasmic reticulum occurs in the vacuole. *Planta* 196:586–596
- Richter LJ, Thanavala Y, Arntzen CJ, Mason HS (2000) Production of hepatitis B surface antigen in transgenic plants for oral immunization. *Nat Biotechnol* 18:1167–1171
- Schermerhaner JP, Matzke MA, Matzke AJ (1988) Endosperm-specific activity of a zein gene promoter in transgenic tobacco plants. *EMBO J* 7:1249–1255
- Simmonds D (2003) Genetic transformation of soybean with biolistics. *Mol Methods Plant Anal* 23:159–174
- Simmonds DH, Donaldson PA (2000) Genotype screening for proliferative embryogenesis and biolistic transformation of short-season soybean genotypes. *Plant Cell Rep* 19:485–490
- Simons B, Scholl D, Cyr T, Alice HM (2001) Effects of increased loop flexibility on the structure and stability of a de novo designed helical protein. *Protein Pept Lett* 8:89–96
- Sjödahl S, Gustavsson HO, Rödin J, Lenman M, Höglund AS, Rask L (1993) Cruciferin gene families are expressed coordinately but with tissue-specific differences during *Brassica napus* seed development. *Plant Mol Biol* 23:1165–1176
- Sjödahl S, Gustavsson HO, Rödin J, Rask L (1995) Deletion analysis of the *Brassica napus* cruciferin gene cru 1 promoter in transformed tobacco: promoter activity during early and late stages of embryogenesis is influenced by cis-acting elements in partially separate regions. *Planta* 197:264–271
- Sparks CA, Jones HD (2009) Biolistics transformation of wheat. *Methods Mol Biol* 478:71–92
- Takahashi M, Uematsu Y, Kashiwaba K, Yagasaki K, Hajika M, Matsunaga R, Komatsu K, Ishimoto M (2003) Accumulation of high levels of free amino acids in soybean seeds through integration of mutations conferring seed protein deficiency. *Planta* 217(4):577–586
- Taylor CK, Levy RM, Elliott JC, Burnett BP (2009) The effect of genistein aglycone on cancer and scancer risk: a review of in vitro, preclinical, and clinical studies. *Nutr Rev* 67:398–415
- Teraishi M, Takahashi M, Hajika M, Matsunaga R, Uematsu Y, Ishimoto M (2001) Suppression of soybean β -conglycinin genes by a dominant gene, *Scg-1*. *Theor Appl Genet* 103:1266–1272
- Timmermans MC, Maliga P, Vieira J, Messing J (1990) The pFF plasmids: cassettes utilising CaMV sequences for expression of foreign genes in plants. *J Biotechnol* 14:333–344
- Townsend JA, Thomas LA (1994) Factors which influence the *Agrobacterium*-mediated transformation of soybean. *J Cell Biochem* 18:78–88
- Utsumi S, Kitagawa S, Katsube T, Higasa T, Kito M, Takaiwa F, Ishige T (1994) Expression and accumulation of normal and modified soybean glycinins in potato tubers. *Plant Sci* 102:181–188
- Walling L, Drews GN, Goldberg RB (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc Natl Acad Sci USA* 83:2123–2127
- Williamson JD, Galili G, Larkins BA, Gelvin SB (1988) The synthesis of a 19 kDa zein protein in transgenic petunia plants. *Plant Physiol* 88:1002–1007
- Wong MC, Emery PW, Preedy VR, Wiseman H (2008) Health benefits of isoflavones in functional foods? Proteomic and metabonomic advances. *Inflammopharmacology* 16:235–239
- Young, VR, and Steinke, FH (1992) Protein and amino acid requirements in relation to dietary food protein needs. *News protein food in human health: Nutrition, prevention and therapy*. Boca Raton, FL: CRC, 9–31
- Zhang P, Jaynes JM, Potrykus I, Gruissem W, Puonti-Kaerlas J (2003) Transfer and expression of an artificial storage protein (ASP1) gene in cassava (*manihot esculenta* crantz). *Transgenic Res* 12:243–250

Improved Protein Quality in Transgenic Soybean Expressing a De Novo Synthetic Protein, MB-16

Transgenic Research

Yunfang Zhang, Johann Scherthaner, Natalie Labbé, Mary A. Hefford, Jiping Zhao, Daina H. Simmonds

Corresponding author: Daina Simmonds, Agriculture and Agr-Food Canada, Daina.Simmonds@agr.gc.ca

Supplemental Table 1S Description of seeds during and following development

Stage	Seed Length (mm)	DAP*	Appearance
1	3-3.5	10-12	Green Translucent
2	5-6	14-16	Green opaque
3	Full Size	20-24	Green opaque
4	Full Size	28-31	Turning Yellow
5	Full Size	31-35	Yellow/Dry
6	Full Size	-	Imbibed overnight

*DAP – days after pollination

Improved Protein Quality in Transgenic Soybean Expressing a De Novo Synthetic Protein, MB-16

Transgenic Research

Yunfang Zhang, Johann Scherthaner, Natalie Labbé, Mary A. Hefford, Jiping Zhao, Daina H. Simmonds

Corresponding author: Daina Simmonds, Agriculture and Agr-Food Canada, Daina.Simmonds@agr.gc.ca

Supplemental Table 2S Essential Amino Acid (EAA or IAA^a) profiles of transgenic lines, the WT parental lines and the established requirement of children and adults, expressed as mg/g total protein.

EAA	EAA Reference Pattern ^a				X5 line		Westag line	
	Child (yrs)			Adult	WT Parent	Con-sc	WT Parent	Cru-sc-KDEL
	0.5	1-2	3-10					
Histidine	20	18	16	15	28.8	28.4	27.6	27.8
Isoleucine	32	31	31	30	45.6	47.3	47.6	48.8
Leucine	66	63	61	59	75.3	78.0	76.6	79.6
Lysine	57	52	48	45	37.0	43.8	41.0	51.7
Met+cys ^b	28	26	24	22	22.8	32.4	24.5	29.2
Phe + tyr	52	46	41	38	82.1	86.7	86.2	91.0
Threonine	31	27	25	23	37.2	38.1	36.1	36.7
Tryptophan	8.5	7.4	6.6	6	13.4	13.0	12.6	11.8
Valine	43	42	40	39	51.7	51.4	50.0	50.8
% total protein	33.8	31.2	29.3	27.7	39.4	41.9	40.2	42.7

^a Data for Essential Amino Acids (EAA), also referred to as Indispensible Amino Acids (IAA); the minimum requirements of infants, children and adults from WAO/FAO/WHO (2007).

^b Met+Cys is increased by more than 42% and 19% for Con-sc and Cru-sc-KDEL, respectively, relative to their WT parents.

Improved Protein Quality in Transgenic Soybean Expressing a De Novo Synthetic Protein, MB-16

Transgenic Research

Yunfang Zhang, Johann Scherthaner, Natalie Labbé, Mary A. Hefford, Jiping Zhao, Daina H. Simmonds

Corresponding author: Daina Simmonds, Agriculture and Agr-Food Canada, Daina.Simmonds@agr.gc.ca

Supplemental Table 3S Free amino acid profiles of Con-sc and Cru-sc-KDEL transgenic lines and their parental lines, X5 and Westag, respectively, expressed as mg/g total protein.

Amino Acid	X5 ^a	Con-sc	Westag	Cru-sc-KDEL
Asx (Asp+Asn)	3.98±2.49	1.86±0.55	3.25±1.05	1.81±0.55
Glx (Glu+Gln)	3.53±4.41	1.05±0.15	2.39±0.99	1.24±0.35
Ser	3.32±1.54	- ^b	0.22±0.11	-
Gly	1.13±0.92	0.14 ^c	0.20±0.09	-
His	1.11±0.79	0.77	0.29±0.16	-
Arg	9.63±0.44 ^{*d}	4.74±1.31 [*]	2.42±1.38	1.15±0.52
Thr	1.52±0.62	-	0.25±0.12	-
Ala	1.27±0.98	0.42±0.14	0.62±0.18 [*]	0.18±0.03 [*]
Pro	1.30±1.23	0.47±0.11	0.63±0.22	0.29±0.02
Tyr	0.92±0.26	-	0.34	0.26
Val	1.10±1.14	0.19±0.04	0.36±0.15	0.12±0.11
Met	0.60±0.12	0.72±0.05	0.78±0.24	0.72±0.06
Ile	1.33±1.15	0.54±0.03	0.82±0.21	0.75±0.08
Leu	2.22±2.20	0.24	0.42±0.21	-
Phe	1.15±1.27	0.44±0.12	0.48±0.19	-
Lys	1.23±1.40	0.34±0.15	0.48±0.22	0.17±0.02
Cys A	-	-	-	-
Trp	0.66±0.39	0.65±0.15	0.60±0.25	0.64±0.09

^a Each free amino acid value is shown as mg/g of total protein and is the average of three replicates, except where three replicate values were unavailable because the amino acid was below the level of detection, as follows: ^b all three samples contained an undetectable amino acid, or ^c the amino acid was not detectable in one or two samples resulting in insufficient sample numbers for standard deviation determination.

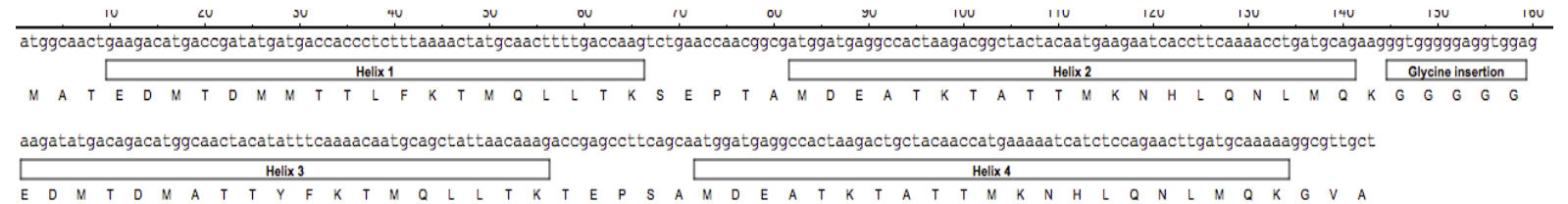
^dT-test; * two tail significant ($\alpha=0.05$)

Improved Protein Quality in Transgenic Soybean Expressing a De Novo Synthetic Protein, MB-16

Transgenic Research

Yunfang Zhang, Johann Schernthaner, Natalie Labbé, Mary A. Hefford, Jiping Zhao, Daina H. Simmonds

Corresponding author: Daina Simmonds, Agriculture and Agr-Food Canada, Daina.Simmonds@agr.gc.ca



Supplementary Fig. 1S

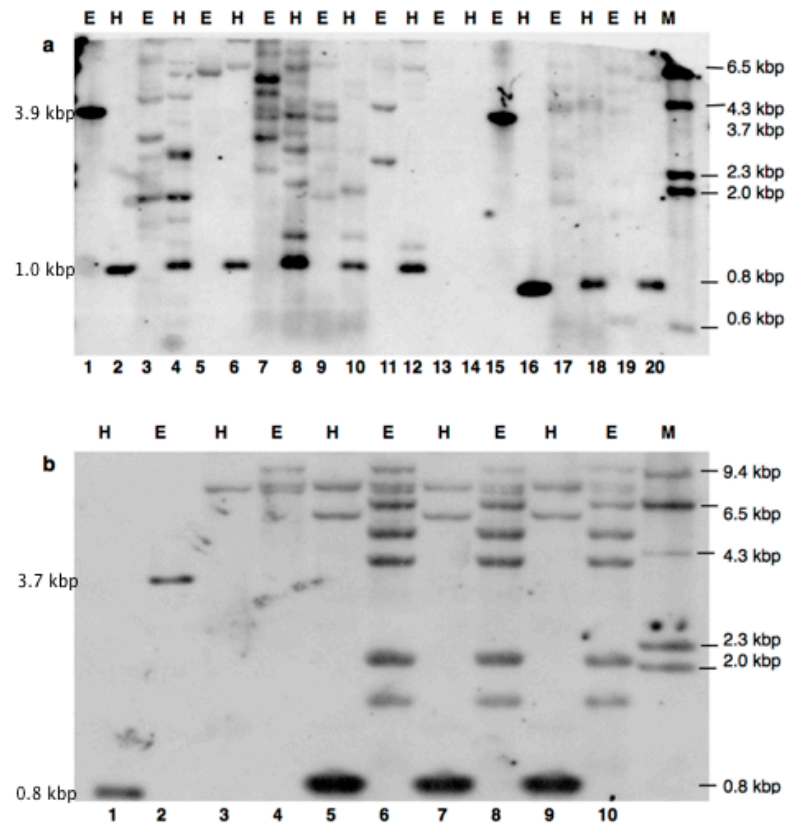
MB-16 DNA and amino acid sequences. The boxes indicate the positions of the four helixes and the glycine insertions into loop 2.

Improved Protein Quality in Transgenic Soybean Expressing a De Novo Synthetic Protein, MB-16

Transgenic Research

Yunfang Zhang, Johann Scherthaner, Natalie Labbé, Mary A. Hefford, Jiping Zhao, Daina H. Simmonds

Corresponding author: Daina Simmonds, Agriculture and Agr-Food Canada, Daina.Simmonds@agr.gc.ca



Supplemental Fig. 2S

Southern blot analysis of soybean lines carrying either the *β-con-MB-16sc* or the *Cru-MB-16scKDEL* transgenes.

Restriction digest with *Eco RI* (E) linearizes the vector and produces a 3.9 kbp fragment from *Cru-MB-16scKDEL* and a 3.7 kb fragment from *β-con-MB-16sc*. *Hind III* (H) digestion drops out the *MB-16* gene with the promoter and generates a 1.0 kb fragment from *Cru-MB-16scKDEL*, and a 0.8 kb fragment from *β-con-MB-16sc*. (a) Transgenic lines probed with 286 bp *MB-16* fragment: 1, 2 - positive control, plasmid *Cru-MB-16scKDEL*; 3, 4 - 5B9-2C5; 5, 6 - 5B9-7B6; 7, 8 - 5B9-9B6; 9, 10 - 5B9-10A2 (*Cru-sc-KDEL*); 11, 12 - 5B9-5E2; 13, 14 - negative control, wild type parental line; 15, 16 - positive control, plasmid *β-con-MB-16sc*; 17, 18 - 2B12-1B11; 19, 20 - 2B12-2A20; M - DNA Marker. (b) A stable line with good expression, 2B12-4A2 (Con-sc), maintained the same integration pattern over several generations: 1, 2 - positive control, plasmid *β-con-MB-16sc*; 3, 4 - negative control, parental line; 5, 6 - T2 plant; 7, 8 - T3 plant; 9, 10 - T5 plant; M - DNA marker.