FOOD STANDARDS
Australia New Zealand
Te Mana Kounga Kai – Ahitereiria me Aotearoa

Supporting document 1

Safety Assessment Report – Application A1110

Food derived from Insect-protected Soybean Line MON87751

Summary and conclusions

Background

A genetically modified (GM) soybean line with OECD Unique Identifier MON-87751-7, hereafter referred to as MON87751, has been developed by Monsanto Company (Monsanto).

The soybean has been modified such that it is protected against lepidopteran pests including the key soybean pests bean shoot moth, sunflower looper and fall armyworm. Protection against these pests is achieved through expression of two Cry proteins (Cry1A.105 and Cry2Ab2) encoded by the *cry1A.105* and *cry 2Ab2* genes derived from the common soil bacterium *Bacillus thuringiensis.* The Cry1A.105 protein is chimeric and comprises functional domains derived from the Cry1Ab, Cry1F and Cry1Ac proteins. The safety of the Cry1A.105 and Cry2Ab2 proteins has previously been assessed by FSANZ.

In conducting a safety assessment of food derived from MON87751, a number of criteria have been addressed including: a characterisation of the transferred gene sequences, their origin, function and stability in the corn genome; the changes at the level of DNA and protein in the whole food; compositional analyses; and evaluation of intended and unintended changes.

This safety assessment report addresses only food safety and nutritional issues. It therefore does not address:

- environmental risks related to the environmental release of GM plants used in food production
- the safety of animal feed, or animals fed with feed, derived from GM plants
- the safety of food derived from the non-GM (conventional) plant.

History of Use

Soybean (*Glycine max*) is grown as a crop in over 93 countries worldwide. Soybean-derived products have a range of food and feed as well as industrial uses and have a long history of safe use for both humans and livestock. Oil, in one form or another, accounts for the major food use of soybean and is incorporated in salad and cooking oil, bakery shortening, and frying fat as well as processed products such as margarine.

Molecular Characterisation

Comprehensive molecular analyses of MON87751 indicate there is a single insertion site comprising a single, complete copy of each of the two expression cassettes. The introduced genetic elements are stably inherited from one generation to the next. There are no antibiotic resistance marker genes present in the line and directed sequencing analysis shows no plasmid backbone has been incorporated into the transgenic locus.

Characterisation and Safety assessment of New Substances

Newly expressed proteins

Soybean MON87751 contains two newly expressed proteins, Cry1A.105 and Cry2Ab2. Mean levels of Cry1A.105 were highest in the R6 leaf (790 µg/g dry weight) and lowest in the roots and pollen where the level was below the limit of quantitation (LOQ). For Cry2Ab2, mean protein levels were highest in the R2 – R3 leaf (32 µg/g dry weight) and were lowest in the pollen (<LOQ). In the seed, from which most food products are derived, Cry1A.105 was present at a mean level of 4 µg/g dry weight and Cry2Ab2 was present at 2.4 µg/g dry weight.

A range of characterisation studies confirmed the identity of the Cry1A.105 and Cry2Ab2 proteins produced in MON87751 and also their equivalence with the corresponding proteins produced in a bacterial expression system. It was found that incomplete cleavage of the chloroplast targeting sequence associated with Cry1A.105 results in the expression of a protein that is four amino acids longer than predicted. Conversely, cleavage of the chloroplast targeting sequence associated with the Cry2Ab2 protein has also resulted in the removal of the first 15 amino acids such that the protein expressed in MON87751 is 15 amino acids shorter than predicted. The variant proteins have the expected molecular weights, immunoreactivity, lack of glycosylation and functional activity.

For both plant-expressed proteins, bioinformatic studies confirmed the lack of any significant amino acid sequence similarity to know protein toxins or allergens; digestibility studies suggest the proteins would be rapidly degraded in the gastro-intestinal tract following ingestion; and thermolability studies indicate both proteins are functionally inactivated following heating. Taken together, the evidence indicates that neither Cry1A.105 nor Cry2Ab2 are likely to be toxic or allergenic in humans.

Compositional Analyses

Detailed compositional analyses were done to establish the nutritional adequacy of seed from MON87751 and to characterise any unintended compositional changes. Analyses were done of proximates, fibre, fatty acids, amino acids, minerals, vitamins, anti-nutrients and isoflavones. The levels were compared to levels in a) the non-GM parental cultivar A3555 b) a tolerance interval compiled from results taken for 19 non-GM lines grown under the same conditions and c) levels recorded in the literature. Only 6 of the 44 reported analytes deviated from the control in a statistically significant manner. However, the mean levels of all of these analytes fell within both the tolerance interval and the historical range from the literature. It can therefore be concluded that seed from MON87751 is compositionally equivalent to seed from conventional soybean varieties.

Conclusion

No potential public health and safety concerns have been identified in the assessment of insect-protected soybean line MON87751. On the basis of the data provided in the present Application, and other available information, food derived from MON87751 is considered to be as safe for human consumption as food derived from conventional soybean varieties.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADF	acid detergent fibre
BLAST	Basic Local Alignment Search Tool
bp	base pairs
Bt	*Bacillus thuringiensis*
Cry	crystal
CTP	chloroplast transit peptide
kDa	kilo Dalton
DNA	deoxyribonucleic acid
T-DNA	transferred DNA
dw	dry weight
ELISA	enzyme linked immunosorbent assay
FAO	Food and Agriculture Organization of the United Nations
FARRP	Food Allergy Research and Resource Program
FASTA	Fast Alignment Search Tool - All
FSANZ	Food Standards Australia New Zealand
GM	genetically modified
HRP	horseradish peroxidase
JSA	junction sequence analysis
kDa	kilo Dalton
LB	Left Border of T-DNA
LOQ	Limit of quantitation
MALDI-TOF MS	matrix-assisted laser desorption/ionisation–time of flight mass spectrometry
MMT	Million metric tons
MRL	Maximum Residue Limit
NDF	neutral detergent fibre
NGS	next generation sequencing
OECD	Organisation for Economic Co-operation and Development
ORF	open reading frame
PCR	polymerase chain reaction
P-value	probability value
RB	Right Border of T-DNA
mRNA	messenger ribonucleic acid
SAS	Statistical Analysis Software
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGF	simulated gastric fluid
SIF	simulated intestinal fluid
U.S.	United States of America
WHO	World Health Organization

1 Introduction

Monsanto Australia Limited has submitted an application to FSANZ to vary Standard 1.5.2 – Food produced using Gene Technology – in the *Australia New Zealand Food Standards Code* (the Code) to include food from a new genetically modified (GM) soybean line MON-87751-7 (referred to as MON87751). The soybean has been modified such that it is protected against lepidopteran pests including the key soybean pests bean shoot moth (*Crocidosema aporema*), sunflower looper (*Rachiplusia nu*) and fall armyworm (*Spodoptera frugiperda*).

Protection against these pests is achieved through expression of two Cry proteins (Cry1A.105 and Cry2Ab2) encoded by the *cry1A.105* and *cry 2Ab2* genes derived from the common soil bacterium *Bacillus thuringiensis.* The Cry1A.105 protein is chimeric and comprises functional domains derived from the Cry1Ab, Cry1F and Cry1Ac proteins. The safety of the Cry1A.105 and Cry2Ab2 proteins has previously been assessed by FSANZ (2008).

According to the Applicant, MON87751 will be combined, through traditional breeding, with other approved GM soybean lines (a process known as 'stacking') to provide additional protection against lepidopteran pests as well as tolerance to various herbicides. The product is designed primarily for commercial growing in South America (e.g. Brazil and Argentina) and approval for cultivation in Australia or New Zealand is not being sought. Therefore, if approved, food derived from this line may enter the Australian and New Zealand food supply as imported food products.

MON 87751 is intended for use as a broad-acre commodity (or field) soybean and not for vegetable, garden, or food-grade soybean that is generally used to produce food items such as tofu, soybean sprouts, soymilk, or green soybean (e.g. edamame). Vegetable and food-grade soybean generally have different characteristics (e.g. size, flavour, texture) from field soybean, and are more easily cooked.

2 History of Use

2.1 Host organism

The host organism is a conventional soybean (*Glycine max* (L.) Merr.), belonging to the family Leguminosae. The soybean cultivar A3555 was used as the parental variety for the genetic modification described in this application, and thus is regarded as the near-isogenic line for the purposes of comparative assessment with MON87751. A3555 is a mid−maturity group III soybean variety developed by Asgrow Seed Company[1].

Soybean is grown as a commercial food and feed crop in many countries worldwide, with some 93 countries listed as producers in 2013 (FAOSTAT3 2015), and has a long history of safe use for both humans and livestock. The major producers of soybean seed, accounting for some 90% of world production, are the U.S. (89 MMT), Brazil (81 MMT), Argentina (49 MMT), China (11.9 MMT) and India (11.9 MMT) (FAOSTAT3 2015). Australia, while a net importer of soybean seed, grows crops in latitudes extending from the tropics (16^{o} S) to temperate regions (37^{o} S), mainly in the eastern states and as a rotational crop (James and Rose 2004). The seed is used mainly to produce meal for use in animal feed (Grey 2006).

[1] Asgrow/ Dekalb 2012 Seed Resource Guide - http://www.scribd.com/doc/93230394/Seed-Resource-Guide-North#scribd

In many soybean producing countries, GM soybean (mainly with a herbicide tolerant trait) accounts for a significant proportion of the total soybean grown e.g. U.S. (93%); Argentina (99%); Brazil (89%); South Africa (90%); Uruguay (99%), Canada (78%) (Brookes and Barfoot 2014). Australia does not currently grow any commercial GM soybean lines[2].

Soybean food products are derived either from whole or cracked soybeans:

- Whole soybeans are used to produce soy sprouts, baked soybeans, roasted soybeans, full fat soy flour, non-fermented traditional soy foods (e.g. tofu, soy milk) and fermented traditional soy foods (e.g. soy sauce, miso, natto and tempeh).

- Cracked soybeans have the hull (seed coat) removed and are then rolled into flakes which undergo solvent extraction to remove the oil. The oil is further refined to produce cooking oil, shortening and lecithin as well as being incorporated into a variety of edible and technical/industrial products. Glycerol, fatty acids and sterols are also derived from soybean oil.

  The defatted flakes are dried and undergo further processing to form products such as toasted, defatted meal (for use in livestock, pet and poultry food), protein concentrate and isolate (for use in both edible and technical/industrial products), and textured flour (for edible uses). Soybean protein is also added to a number of meat, dairy, bakery and cereal products. The hulls are used in mill feed as well as being incorporated into dietary fibre products such as fibre bran breads, cereals and snacks.

Unprocessed (raw) soybeans are not suitable for food use, and have only limited feed uses, as they contain toxicants and anti-nutritional factors, such as lectins and trypsin inhibitors (OECD 2012). Appropriate heat processing inactivates these compounds.

Soybean oil constitutes approximately 27% of global consumption of edible fats and oils and is currently the second largest source of vegetable oil worldwide behind palm oil (American Soybean Association 2015). Oil, in one form or another, accounts for the major food use of soybean (Shurtleff and Aoyagi 2007) and is incorporated in salad and cooking oil, bakery shortening, and frying fat as well as processed products such as margarine.

Another possible food product that can be derived from the soybean plant is bee pollen. (Krell 1996).

2.2 Donor organisms

2.2.1 *Bacillus thuringiensis*

Many different subspecies of *Bacillus thuringiensis (Bt)* have been isolated from dead or dying insects, mostly from the orders Coleoptera, Diptera and Lepidoptera, but many subspecies have also been found in the soil, aquatic environments and other habitats (WHO 1999). The source of the *cry2Ab2* gene used in MON87751 is the *Bt* subsp. *kurstaki.* The *cry1A.105* gene comprises sequences from *Bt* subspecies *kurstaki* and *aizawai* both of which are spore-forming, gram-positive bacteria that are primarily associated with the soil and leaf surfaces.

[2] See information on approved commercial; releases of GM crops in Australia on the website of the Office of the Gene Technology Regulator http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/Content/ir-1

Studies on mammals, particularly laboratory animals, demonstrate that *B. thuringiensis* is mostly non-pathogenic and non-toxic. *B. thuringiensis* has been demonstrated to be highly specific in its insecticidal activity and has demonstrated little, if any, direct toxicity to non-target insects (see NPTN 2000; OECD 2007 and references therein). Infection in humans is unusual although there have been at least two clinical reports, one in the wounds of a soldier (Hernandez et al. 1998) and one in burn wounds (Damgaard et al. 1997), and in both cases impaired immunosuppression was implicated in the cause of the infection. *B. thuringiensis* has also been rarely associated with gastroenteritis (see eg Jackson et al. 1995) but generally, *B. thuringiensis* present in drinking water or food has not been reported to cause adverse effects on human health (WHO 1999; NPTN 2000; OECD 2007).

Sporulated *B. thuringiensis* has a long history of safe use for pest control in agriculture, especially in organic farming. The effect of such *Bt*-containing products on human health and the environment was the subject of a critical review by the WHO International Programme on Chemical Safety (WHO 1999). The review concluded that '*B. thuringiensis* products are unlikely to pose any hazard to humans or other vertebrates or the great majority of non-target invertebrates provided that they are free from non-*B. thuringiensis* microorganisms and biologically active products other than the insecticidal proteins'. Products containing *Bt* are approved for use on crops in Australia[3] and New Zealand[4] and in both countries there is an exemption from maximum residue limits (MRLs) when *Bt* is used as an insecticide[5].

2.2.2 Other organisms

Genetic elements from several other organisms have been used in the genetic modification of MON87751 (refer to Table 1). These non-coding sequences are used to drive, enhance, target or terminate expression of the novel genetic material. None of the sources of these genetic elements is associated with toxic or allergenic responses in humans.

3 Molecular Characterisation

Molecular characterisation is necessary to provide an understanding of the genetic material introduced into the host genome and helps to frame the subsequent parts of the safety assessment. The molecular characterisation addresses three main aspects:

- the transformation method together with a detailed description of the DNA sequences introduced to the host genome
- a characterisation of the inserted DNA including any rearrangements that may have occurred as a consequence of the transformation
- the genetic stability of the inserted DNA and any accompanying expressed traits.

Studies submitted:

2014. Amended Report for MSL0025106: Segregation Analysis of the Coding Sequences Present in Insect Protected Soybean MON 87751 Across Multiple Generations. MSL0025712. Monsanto Company (unpublished)

2014 Amended Report for MSL0025312: Molecular Characterization of Insect Protected Soybean MON 87751. MSL0025901. Monsanto Company.

2013. Bioinformatics Evaluation of DNA Sequences Flanking the 5' and 3' Junctions of Inserted DNA in MON 87751: Assessment of Putative Polypeptides. MSL0024956. Monsanto Company (unpublished)

[3] https://portal.apvma.gov.au/pubcris
[4] http://www.biosecurity.govt.nz/pests-diseases/forests/white-spotted-tussock-moth/about-btk.htm
[5] New Zealand: http://www.foodsafety.govt.nz/elibrary/industry/nz-mrl-agricultural-compounds-food-standards-07-2014.pdf ; Australia: http://www.comlaw.gov.au/Details/F2014C00821

3.1 Method used in the genetic modification

Soybean cultivar A3555 was transformed with binary plasmid vector PV-GMIR13196 containing two T-DNA inserts (Figure 1) using an *Agrobacterium*-mediated method (Martinell et al. 2011). Each T-DNA insert is delineated by a Right and Left border (RB and LB). Insert T-DNA I contains the two *cry* genes while insert T-DNA II contains the selectable marker genes *splA* and *aadA*. Expression of *splA* produces an enzyme with sucrose phosphorylase-like activity which interferes with sucrose metabolism, leading to a recognisable seed phenotype (Piper et al. 1999). The *aadA* gene encodes an aminoglycoside-modifying enzyme that confers resistance to the antibiotics spectinomycin and streptomycin (Fling et al. 1985). Although following transformation both T-DNAs were inserted into the genome, subsequent conventional breeding and segregation (refer to Section 3.3) were used to isolate those plants containing only T-DNA I.

Briefly, the transformation procedure used germinated seeds of A3555 from which shoot meristem tissues were excised. After co-culturing with the disarmed *Agrobacterium* strain AB30 carrying the vector, the meristems were placed on selection medium containing spectinomycin, carbenicillin, cefotaxime, and ticarcillin/clavulanate acid mixture, to inhibit the growth of untransformed plant cells and excess *Agrobacterium*. The meristems were then placed on a medium that supported shoot and root development. Rooted plants (R_0) with normal phenotypic characteristics were selected and transferred to soil for growth and further assessment.

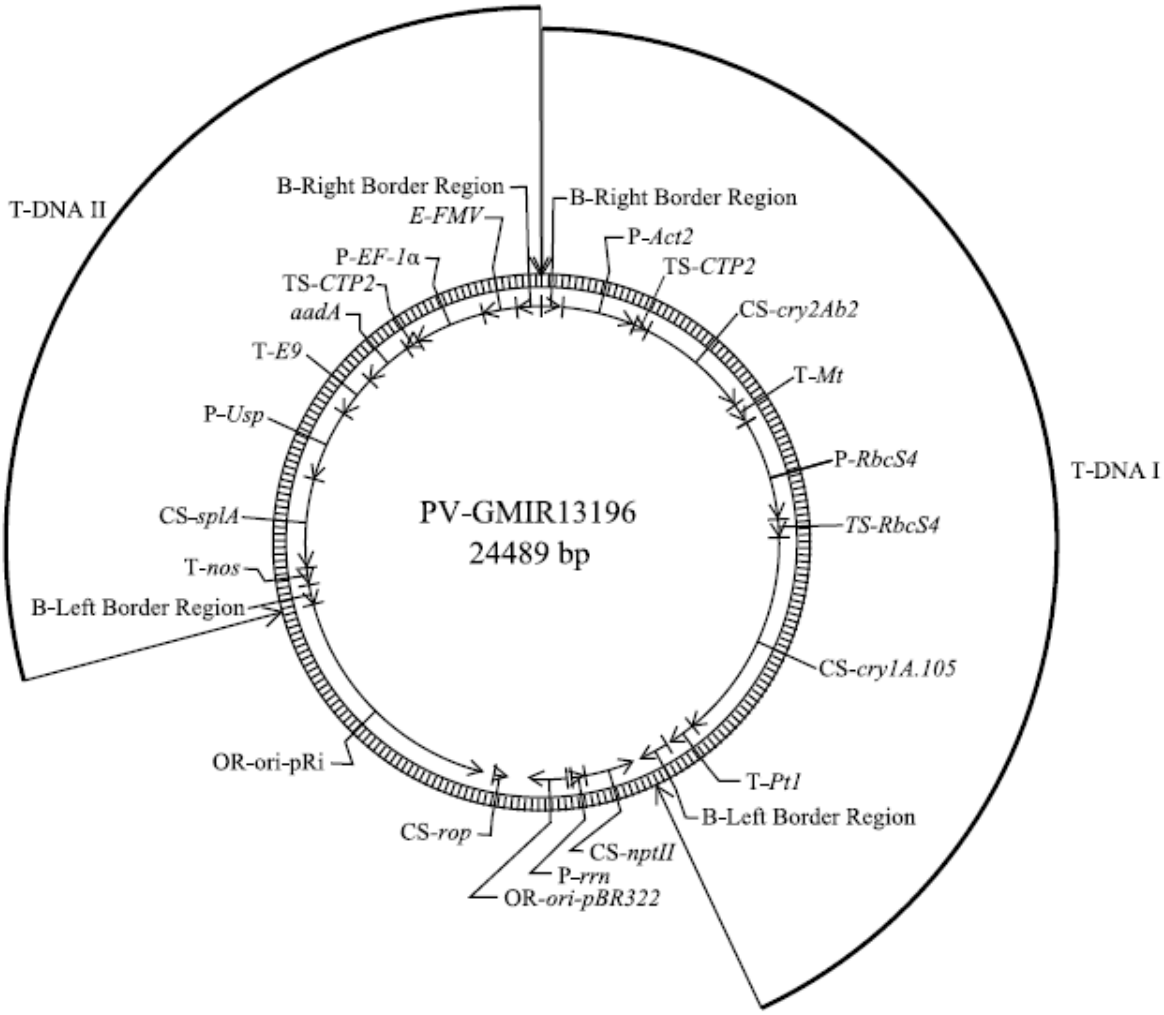


Figure 1: Genes and regulatory elements contained in plasmid PV-GMIR13196

3.2 Function and regulation of introduced genetic material

The complete PV-GMIR13196 plasmid is 24,489 bp in size and comprises:

- 10,530 bp of T-DNA I
- 7,122 bp of T-DNA II
- 6,837 bp of backbone

Information on the genetic elements in the T-DNA I of PV-GMIR13196 is summarised in Table 1. There are two cassettes comprising a total of 9, 803 bp located between a 442 bp Left Border (LB) and a 285 bp Right Border (RB).

Table 1: Description of the genetic elements contained in the T-DNA I of PV-GMIR13196

Genetic element	bp location on plasmid	Size (bp)	Source	Orient.	Description & Function	References
RIGHT BORDER	1 - 285	285				
***cry2Ab2* cassette**						
Intervening sequence	286 - 337	52				
Act2	338 - 1545	1208	*Arabidopsis thaliana*	Clockwise	• Promoter, leader and intron from the *act2* gene • Direct transcription of *cry2Ab2*	An *et al.* (1996)
Intervening sequence	1546 - 1555	10				
CTP2	1556 - 1783	228	*Arabidopsis thaliana*	Clockwise	• Targeting sequence of the *ShkG* gene encoding the EPSPS transit peptide • Directs transport of the Cry2Ab2 protein to the chloroplast	Herrmann (1995); Klee *et al.* (1987)
Intervening sequence	1784 - 1792	9				
cry2Ab2	1793 - 3697	1905	*Bacillus thuringiensis* subsp. *kurstaki*	Clockwise	• Codon-optimised coding sequence of the *cry2Ab2* gene	Donovan (1991)
Intervening sequence	3698 - 3700	3				
Mt	3701 - 4000	300	*Oryza sativa* (rice)	Clockwise	• 3' untranslated region sequence from the *Mt* gene encoding metallothionein-like protein • Directs polyadenylation of the *cry2Ab2* gene	Hunt (1994)
***cry1A.105* cassette**						
Intervening sequence	4001 - 4045	45				
RbcS4	4046 - 5768	1723	*Arabidopsis thaliana*	Clockwise	• Promoter and leader sequence from the *rbcs* gene family encoding small subunit *ats1A* • Directs transcription of the *cry1A.105* gene	De Almeida *et al.* (De Almeida et al. 1989); Krebbers *et al*. (1988)
RbcS4	5769 - 6032	264	*Arabidopsis thaliana*	Clockwise	• Targeting sequence from the *rbcs* gene family encoding small subunit *ats1A* • Directs transport of the Cry1A.105 protein to the chloroplast	Wong *et al*. (1992)
cry1A.105	6033 - 9566	3534	*Bacillus thuringiensis* subsp. *kurstaki* and subsp. *aizawai*	Clockwise	• A codon-optimised chimeric gene comprising coding sequences for the Cry1Ab, Cry1F and Cry1Ac proteins.	Monsanto (unpublished)
Intervening sequence	9567 - 9569	3				
Pt1	9570 - 9969	400	*Medicago truncatula* (barrel medic)	Clockwise	• 3' untranslated region from the gene encoding phosphate transporter PT1 • Directs polyadenylation of the *cry1A.105* mRNA	Liu *et al.* (1998)
Intervening sequence	9970 - 10088	119				
LEFT BORDER	10089 - 10530	442				

3.2.1 *cry2Ab2* expression cassette

The *cry2Ab2* gene encodes a 79 kDa Cry2Ab2 insecticidal protein, which is a variant of the wild-type Cry2Ab2 protein from *B. thuringiensis* subsp. *kurstaki*. The *cry2Ab2* coding sequence has been modified to change codon usage for optimised expression in plants.

The sequence is driven by the constitutive *Act2* promoter (with leader sequence and intron that help enhance expression) from the plant *Arabidopsis thaliana*. A chloroplast transit peptide (CTP), CTPT2, derived from elements from the *ShkG* gene from *A. thaliana,* targets the Cry2Ab2 protein to the chloroplasts. The CTP is typically cleaved on uptake of the mature protein into the chloroplast, and is subsequently rapidly degraded. A sequence from the 3' untranslated region of the *Mt* gene from *Oryza sativa* (rice) functions to terminate transcription of the *cry2Ab2* gene and direct polyadenylation of the mRNA.

3.2.2 cry1A.105 expression cassette

The *cry1A.105* gene encodes a 142 kDa Cry1A.105 insecticidal protein, which is a chimeric protein (see Section 4.1.3). The *cry1A.105* coding sequence was optimised for expression in plant cells. The sequence is driven by the green tissue-specific *RbcS4* promoter from *A. thaliana*, and a sequence from the *RbcS4* gene encodes the small subunit *ats1A* that directs transport of the Cry1A.105 protein to the chloroplasts. A sequence from the 3' untranslated region of the *Ptl* gene from *Medicago truncatula* (barrel medic) functions to terminate transcription of the *cry1A.105* gene and direct polyadenylation of the mRNA.

3.3 Breeding of MON87751

The breeding pedigree for the various generations is given in Figure 2.

Many hundreds of R_0 plants generated through the transformation process described in Section 3.1 were self-pollinated to produce R_1 seed. Any seeds containing the *splA* scorable phenotype and *aadA* sequence (detected by polymerase chain reaction – PCR) of T-DNA II were then eliminated. R_1 plants homozygous for the T-DNA I insert were then taken for further rounds of self-pollination. At each generation, the progeny were evaluated for desirable molecular and phenotypic characteristics. MON87751 was selected as the lead event, based on its superior agronomic, phenotypic and molecular characteristics.

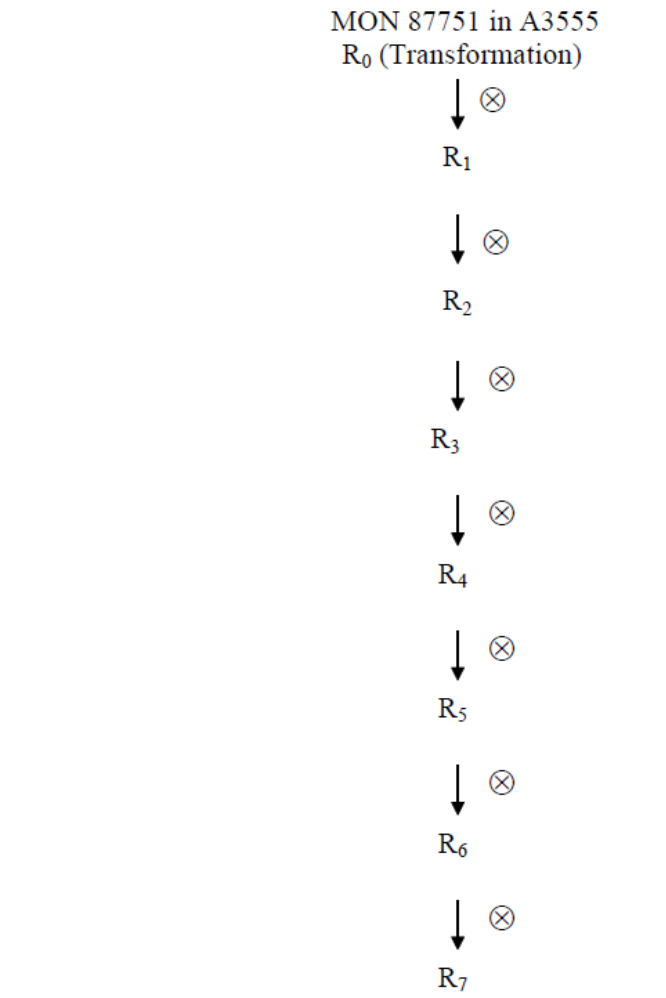


Figure 2: Breeding diagram for MON87751

Table 2 indicates the generations and controls that were used in the various studies characterising MON87751.

Table 2: MON87751 generations used for various analyses

Analysis	MON87751 Generation used	Control(s) used
Molecular characterisation (Section 3.4)	R_3	A3555
Genetic stability (Section 3.5.1)	R_3, R_4, R_5, R_6, R_7	N/A
Mendelian inheritance (Section 3.5.2)	F_2, F_3, and F_4 (see Figure 6)	
Protein expression levels in plant parts (Section 4.1.2)	R_7	N/A
Protein characterisation (Section 4.1.3)	R_7	*E.coli*-produced
Compositional analyses (Section 5)	R_7	A3555 (+ 19 non-GM reference lines)

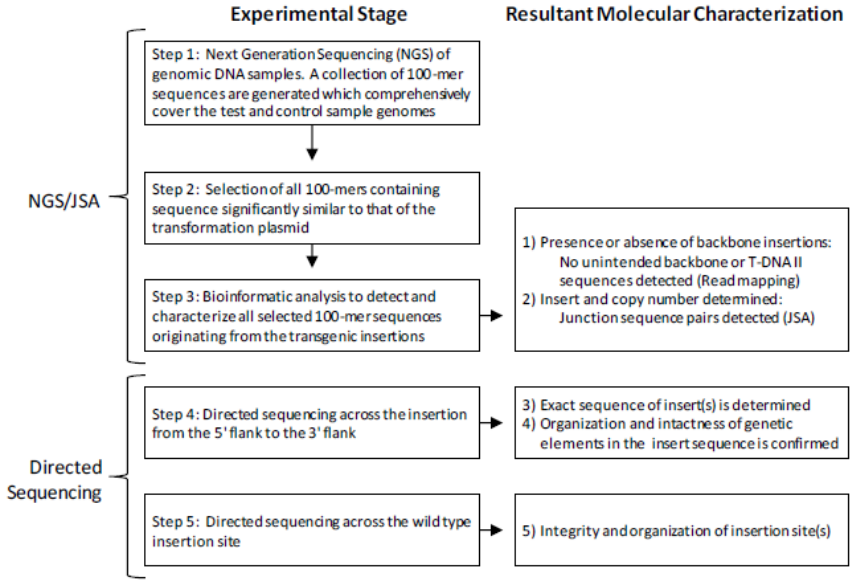
3.4 Characterisation of the genetic modification in the plant

A range of analyses were undertaken to characterise the genetic modification in line MON87751. These analyses focussed on the nature of the insertion of the introduced genetic elements and whether any unintended genetic re-arrangements may have occurred as a consequence of the transformation procedure.

The molecular characterisation of MON87751 incorporated a relatively new approach (Kovalic et al. 2012; DuBose et al. 2013) that applies Next Generation Sequencing (NGS) and Junction Sequence Analysis (JSA) together with bioinformatics to determine the number of inserts. In the past, this has been determined by Southern blot analysis. The organisation and sequence of the insert and adjacent flanking DNA, and the sequencing of the insertion site were all determined by methods employing directed sequencing (locus-specific PCR and DNA sequence analysis).

The rationale for junction sequence analysis is that, since junctions are characteristic of DNA insertion, it follows that each insertion will produce two (i.e. 5' and 3') unique junction sequences. By evaluating the number and the sequences of unique junctions detected, the number of plasmid sequence insertions and the T-DNA copy number can be determined.

The Applicant supplied the flow-diagram shown in Figure 3 to illustrate this approach to molecular characterisation.



Genomic DNA from the test and the conventional control was sequenced using technology that produces a set of short, randomly distributed sequence reads (each approximately 100 bp long) that comprehensively cover both genomes (Step 1). Utilizing these genomic sequence reads, bioinformatics search tools were used to select all sequence reads that were significantly similar to the transformation plasmid (Step 2) for use in read mapping to determine the presence/absence of backbone and T–DNA II sequences, and Junction Sequence Analysis (JSA) bioinformatics to determine the insert and copy number (Step 3). Overlapping PCR products are also produced which span any inserts and their wild type locus (Step 4 and Step 5 respectively); these overlapping PCR products are sequenced to allow for detailed characterization of the inserted DNA and insertion site(s).

Figure 3: Steps in the molecular characterisation of MON87751

3.4.1 Insert number and backbone presence

Total genomic DNA from seed of verified MON87751 (generation R_3) and the untransformed parent (A3555) was sequenced using Illumina®[6] NGS technology. Reference DNA was also used from the plasmid vector PV-GMIR13196. As a positive control, plasmid DNA was spiked into A3555 DNA at a single copy genome equivalent ratio and 1/10 copy genome equivalent ratio. It was noted from the subsequent positive control results, that any portion of the plasmid could be detected at both single copy and 1/10 copy; this indicated there was adequate sensitivity to be able to observe any inserted fragment.

The DNA was sheared, using a Covaris S-220 ultrasonicator, into approximately 325 bp fragments, processed for deep sequencing (end-repaired, A-tailed and ligated to adapters), enriched through ten cycles of PCR and then sequenced using Illumina HiSeq® technology that produces short-sequence reads approximately 100 bp long. To confirm sufficient sequence coverage in the samples, the 100-mer sequence reads from all samples were analysed to determine the effective depth of coverage (i.e. the average number of times any base of the genome is expected to be independently sequenced) by mapping all reads to a known single-copy endogenous soybean lectin-like gene (*Le1* - GenBank accession: K00821.1). The analysis showed that *Le1* was covered by the 100-mers at >75x for each sample, a coverage that is considered to be comprehensive (Kovalic et al. 2012).

An *in silico* analysis using the BLAST[7] algorithm then followed, in which only those 100-mer reads containing sequence similarity to the plasmid PV-GMIR13196 were selected i.e. this analysis found all 100-mer reads that were either fully matched to the insert plasmid sequences or contained both plasmid sequences and junction sequences. The analysis collected all sequencing reads with an e-Value[8] of less than 1e-5 and at least 30 bases match of greater than 96.7% identity to the transformation plasmid (Kovalic et al. 2012). Using Bowtie[9] short sequence alignment software, good-quality non-duplicated reads of ≥30 bp were collected. Following *in silico* adapter removal (Novoalign software[10]) and *in silico* removal of low quality read ends (Phred score[11] ≤ 12) the remaining reads were then aligned to the whole plasmid PV-GMIR13196 sequence in order to find junction region sequences (Figure 4). Reads were also aligned against the control genome in order to remove those reads sourced from endogenous homologues.

Figure 4 shows a map of the junction sequences (illustrated as stacked bars) that were detected. Each detected junction sequence read is shown trimmed to include only 30 bp of plasmid sequence. Only two unique junction sequence classes, both containing portions of T-DNA and flanking sequence were detected. This indicates that MON87751 contains a single DNA insert. No junction sequences were found in the DNA from A3555.

Through mapping the sequence reads obtained from MON87751 and the control, to plasmid PV-GMIR13196 ,the presence/absence of backbone sequences and T-DNA II can also be determined. For a single insert of T-DNA I at a single locus few, if any, reads aligning with plasmid backbone or T-DNA II would be expected in the JSA analysis.

---

[6] http://www.illumina.com/technology/next-generation-sequencing/sequencing-technology.ilmn

[7] BLAST is the acronym for Basic Local Alignment Search Tool (Altschul et al. 1990), a computer algorithm that can rapidly align and compare a query DNA sequence with other DNA sequences..

[8] Comparisons between highly homologous proteins yield *E*-values approaching zero, indicating the very low probability that such matches would occur by chance. A larger *E*-value indicates a lower degree of similarity.

[9] http://bowtie-bio.sourceforge.net/index.shtml

[10] http://www.novocraft.com/main/index.php

[11] Phred is a base calling programme for DNA sequences. Phred quality scores have become widely accepted to characterise the quality of sequences (http://www.phrap.com/phred/) .

In fact, 143 of the total of over 10,000 reads obtained from MON87751 and 30 reads from the conventional control did align with sequences in the backbone. Of the 143 reads, 126 of these aligned to the CTP2 sequence that is present in both the T-DNA I and T-DNA II (see Figure 1). The fact that a) full sequence analysis (Section 3.4.2) indicated no CTP2 sequences were present in the junctions and b) over 10,000 reads in total aligned with T-DNA I (compared to the 30 reads from the control and 17 from MON87751 that were scattered across the length of the backbone) leads to the weight-of-evidence conclusion that MON87751 contains neither plasmid backbone nor T-DNA II sequences.

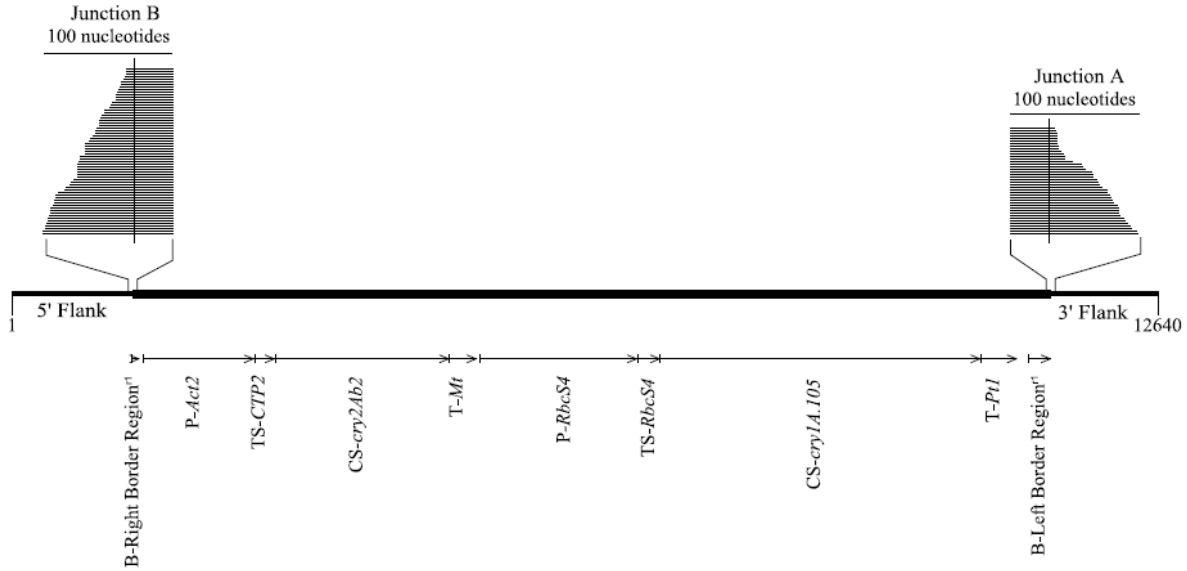


Figure 4: Schematic representation of the junction sequences detected in MON87751

3.4.2 Insert organisation and sequence

PCR primers were designed to amplify four overlapping regions of MON87751 genomic DNA incorporating the insert and flanking regions [no products were obtained for DNA from A3555]. The products were used to determine the nucleotide sequence of the insert and flanking regions using BigDye® Terminator chemistry[12]. A consensus sequence was generated by compiling sequences from multiple sequencing reactions performed on the overlapping PCR products. The consensus sequence was then aligned to the PV-GMIR13196 sequence to determine the integrity and organisation of the insert and flanking regions.

The results showed that the insert is 10,119 bp in length and comprises the identical 9,803 bp sequence found within the T-DNA of plasmid PV-GMIR13196 together with terminally truncated Right and Left Border regions (RB missing 215 bp and LB missing 196 bp – see Table 1 for details of full sequence length). This analysis also showed that no plasmid backbone sequences or T-DNA II sequences are present in MON87751, and confirmed the conclusion from the NGS/JSA analysis that a single copy of the T-DNA I has been inserted.

In addition to the insert, 1,334 bp flanking the 5' end of the insert and 1,187 bp flanking the 3' end of the insert were sequenced.

[12] http://www.appliedbiosystems.com.au/

3.4.3 The insertion site

In order to identify any changes to the genomic DNA as a result of the insertion event, two primers (one specific to the 5' flanking sequence of MON87751 and one specific to the 3' sequence) were used for PCR of genomic DNA isolated from the untransformed parent (A3555). The product (approximately 2,600 bp) was then sequenced and the sequence was compared with the sequences obtained for the 5' and 3' flanking regions of MON87751. The results showed that a 1 bp insertion and a 7 bp deletion had occurred at the insertion site during transformation and also that a 16 bp deletion in the 5' flanking region had occurred. Changes such as this are not uncommon and most likely result from double-stranded break repair in the plant during the *Agrobacterium*-mediated transformation process (Salomon and Puchta 1998).

3.4.4 Open reading frame (ORF) analysis

Sequences spanning the 5' and 3' junctions of the MON87551 insert were translated using DNAStar software[13] from stop codon to stop codon (TGA, TAG, TAA) in all six reading frames. A total of 12 ORFs (six in the 5' junction and six in the 3' junction) were identified that encode putative polypeptides ranging in size from 17 – 90 amino acids. No analysis was done to determine whether any potential regulatory elements were associated with the polypeptides.

The putative polypeptides encoded by the 12 identified ORFs were then analysed using a bioinformatic strategy to determine similarity to known protein toxins or allergens (refer to Section 4.1.5).

3.5 Stability of the genetic changes in MON87751

The concept of stability encompasses both the genetic and phenotypic stability of the introduced trait over a number of generations. Genetic stability refers to maintenance of the modification (as produced in the initial transformation event) over successive generations. It is best assessed by molecular techniques, such as Southern blot analysis or NGS/JSA. Phenotypic stability refers to the expressed trait remaining unchanged over successive generations. It is often quantified by a trait inheritance analysis to determine Mendelian heritability via assay techniques (chemical, molecular, visual).

3.5.1 Genetic stability

The genetic stability of event MON87751 was evaluated by NGS/JSA (as described in Section 3.4.1) in verified genomic DNA isolated from the seed of plants of five generations (refer to Table 2). Control genomic DNA was isolated from the non-GM parental line (A3555). The depth of coverage for each generation sample was ≥ 75x.

No junction sequences were detected in DNA obtained from A3555. Analysis of the MON87751 DNA from all generations showed the presence of the same two junction sequences identified in Section 3.4.1. No other junction sequences were present. The consistency of this junction sequence data across all generations tested, demonstrates that the single insert is stably maintained in MON87751.

[13] http://www.dnastar.com/

3.5.2 Phenotypic stability

Since it was demonstrated that the insert resides at a single locus within the MON87751 genome, the expectation would be the genetic material within it would be inherited according to Mendelian principles.

Chi-square (**X^2**) analysis was undertaken over several generations to confirm the segregation and stability of the complete T-DNA sequence within the insert. The breeding path followed for this analysis was different from that represented in Figure 2 and is shown in Figure 5. Basically, at each stage, a Real-Time TaqMan® PCR assay and Invader® analysis was used to select plants containing the T-DNA I insert. From the R_1 segregating population an individual plant homozygous for the *cry1A.105* and *cry2Ab2* genes was identified. This plant was then self-pollinated to give rise to a population of R_2 plants, which were in turn self-pollinated to obtain the R_3 generation. The R_3 plants were crossed to a Monsanto proprietary soybean line (MonSoy8329) that did not contain the *cry1A.105* and *cry2Ab2* cassettes, in order to generate hemizygous generations. Ultimately, the inheritance of the T-DNA I was assessed in the F_2 F_3 and F_4 generations for which it was predicted to segregate at a 1:2:1 ratio (homozygous positive: hemizygous: homozygous negative) according to Mendelian inheritance principles.

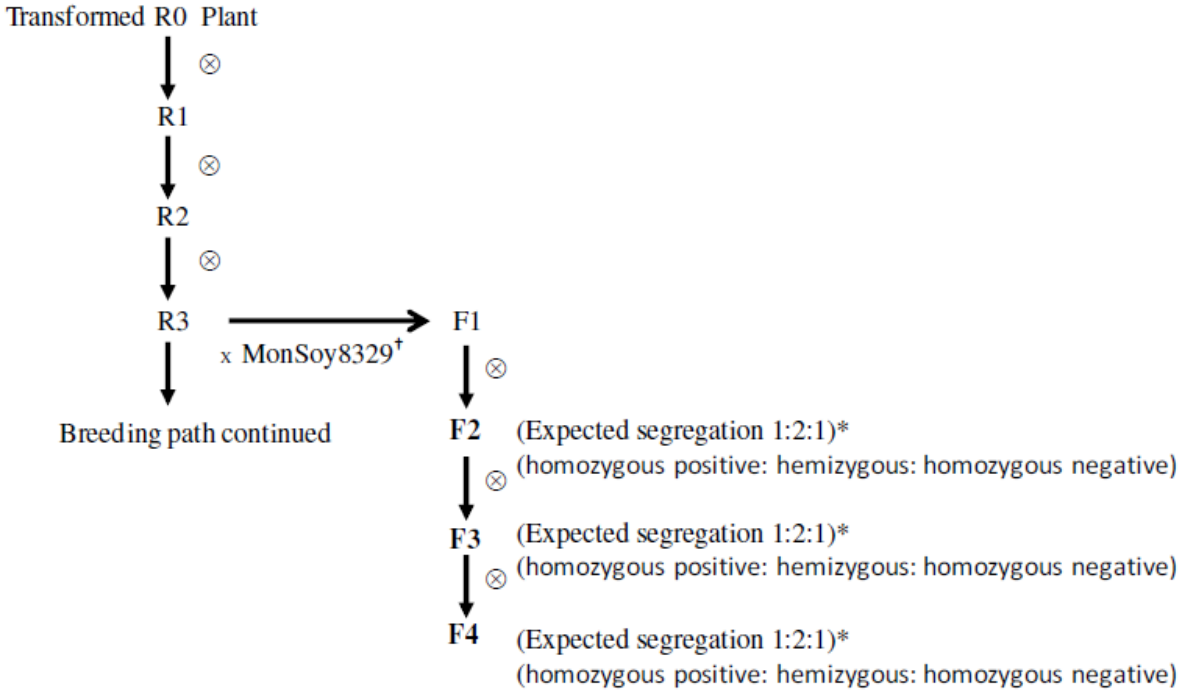


Figure 5: Breeding path for generating segregation data for MON87751

The results (Table 3) indicated there were no significant differences between the observed and expected segregation ratios in any of the generations. This supported the conclusion that the T-DNA I resides at a single locus and showed that it is inherited according to Mendelian principles.

Table 3: Segregation of the MON87411 T-DNA sequences over three generations

Generation	Total plants	Ratio[1]		X^2	Probability (P)[2]
		Observed	Expected		
F_2	152	1:1.8:1.05	1:2:1	0.47	0.79
F_3	214	1:2.32:1.04	1:2:1	0.95	0.62
F_4	204	1:1.81:0.7	1:2:1	3.01	0.22

[1] The ratio is homozygous positive:hemizygous:homozygous negative for each of observed and expected

[2] Statistical significance is when P≤0.05

3.6 Antibiotic resistance marker genes

No antibiotic marker genes are present in MON87751. The insert sequence analysis (Section 3.4.2) showed no T-DNA II or plasmid backbone sequences have been integrated into the MON87751 genome during transformation, i.e. the *aadA* gene (in T-DNA II) used as a plant selectable marker gene and the *nptII* (in the plasmid backbone) used as a bacterial selectable marker gene, are not present in MON87751.

3.7 Conclusion

Soybean line MON87751 contains two expression cassettes: the *cry2Ab2* cassette and the *cry1A.105* cassette.

Comprehensive molecular analyses of MON87751 indicate there is a single insertion site comprising a single, complete copy of each of the two expression cassettes. The introduced genetic elements are stably inherited from one generation to the next. There are no antibiotic resistance marker genes present in the line and directed sequencing analysis shows no plasmid backbone has been incorporated into the transgenic locus.

The results from both the insert sequence analysis (Section 3.4.2) and the phenotypic stability analysis (Section 3.5.2) are consistent with the results from NGS/JSA which was used to determine the number of inserts.

4 Characterisation and safety assessment of new substances

4.1 Newly expressed proteins

In considering the safety of newly expressed proteins it is important to note that a large and diverse range of proteins are ingested as part of the normal human diet without any adverse effects, although a small number have the potential to impair health, e.g., because they are allergens or anti-nutrients (Delaney et al. 2008). As proteins perform a wide variety of functions, different possible effects have to be considered during the safety assessment including potential toxic, anti-nutritional and allergenic effects. To effectively identify any potential hazards requires knowledge of the characteristics, concentration and localisation of all newly expressed proteins in the organism as well as a detailed understanding of their biochemical function and phenotypic effects. It is also important to determine if the newly expressed protein is expressed as expected, including whether any post-translational modifications have occurred.

Two types of proteins were considered:

- Those that were expected to be directly produced as a result of the translation of the introduced genes. A number of different analyses were done to characterise these proteins and determine *in planta* expression.
- Those that may be potentially translated as a result of the creation of ORFs during the transformation process (see Section 3.4.4).

4.1.1 Cry proteins

Crystal (Cry) proteins produced by *Bacillus thuringiensis* (*Bt*) are classified by their primary amino acid sequence and more than 500 different *cry* gene sequences have been classified into 73 groups (Cry1–Cry73)[14]. The largest family is the 3D-Cry group and contains proteins subdivided into further groups based on their specificity for target insects; Cry1 (such as Cry1A.105) and Cry2 (such as Cry2Ab2) proteins act specifically on Lepidopterans while Cry 3 proteins act specifically on Coleopterans (Höfte and Whiteley 1989).

The primary action of all Cry toxins is to lyse midgut epithelial cells in the target insect by forming pores in the apical microvilli membrane of the cells, which subsequently leads to ion leakage and cell lysis. The crystal inclusions ingested by susceptible larvae dissolve in the alkaline environment of the gut, and the solubilised inactive protoxins are cleaved by midgut proteases yielding 60-70 kDa protease resistant core toxins (Bravo et al. 2007). Toxin activation involves the proteolytic removal of an N-terminal peptide. The activated toxin then binds to specific receptors on the brush border membrane of the midgut epithelium columnar cells (Hofmann et al. 1988; Aronson and Shai 2001) before inserting into the membrane. Toxin insertion leads to formation of lytic pores in microvilli apical membranes (de Maagd et al. 2001; Aronson and Shai 2001) and eventually to cell lysis and disruption of the gut epithelium. The septicaemia that inevitably follows may be mediated by an influx of enteric bacteria into the haemocoel (Broderick et al. 2006).

Approved GM crops incorporating *Bt* Cry proteins have been available for a number of years and have not raised any food safety concerns (Koch et al. 2015).

4.1.2 Cry1A.105 and Cry2Ab2 expression in MON87751 tissues

Study submitted:

2013. Amended Report for MSL0024805: Assessment of Cry1A.105 and Cry2Ab2 Protein Levels in Soybean Tissues Collected from MON 87751 Produced in U.S. Field Trials during 2012. **MSL0025199**. Monsanto Company (unpublished)

Plants of MON87751 (generation R_7) and A3555 were grown from verified seed lots at five field sites in the U.S.[15] during the 2012 growing season. These plantings overlapped with the eight sites used for the compositional analyses described in Section 5. There were four replicated plots at each site planted in a randomised complete-block design. Flowers for the collection of pollen/anther tissue were also taken in 2012 from a single non-randomised plot at a field site in Champaign County, Illinois

Samples were taken at various stages of growth (Table 4) and the levels of Cry 1A.105 and Cry2Ab2 proteins were determined for each sample type using a validated enzyme-linked immunosorbent assay (ELISA).

[14] http://www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/toxins2.html
[15] Jackson County, Arkansas; Jefferson County, Iowa; Pawnee County, Kansas; Perquimans County, North Carolina; and Lehigh County, Pennsylvania.

Both proteins were first captured using an affinity-purified, protein-specific, monoclonal mouse capture antibody and were then detected using an affinity-purified, protein-specific, biotinylated polyclonal goat antibody and a NeutrAvidin (Thermo Fisher Scientific) detection reagent conjugated to horseradish peroxidase. Plates were analysed on a commercial microplate spectrophotometer (SPECTRAmax Plus 384, Molecular Devices). Quantification of each protein was accomplished by interpolation on a protein standard curve.

The results, averaged over all sites, are given in Table 4.

Table 4: Cry2Ab2 and Cry1A.105 protein content in MON87751 soybean parts at different growth stages (averaged across 5 sites except for pollen)

Tissue/Growth stage[1]	Cry2Ab2 µg/g dry weight		Cry1A.105 µg/g dry weight	
	Mean (SD)	Range	Mean (SD)	Range
Leaf / V3 – V4	24 (5.9)	17 - 37	580 (250)	260 - 1100
Leaf / V5 – V7	26 (3.1)	20 - 33	590 (270)	68 - 1100
Leaf / R2 – R3	32 (5.2)	25 - 43	400 (220)	50 - 780
Leaf / R6	24 (2.7)	18 - 29	790 (280)	430 - 1600
Root / R6	15 (2.7)	11 - 22	< LOQ	
Forage / R6[2]	14 (2.2)	11 - 18	230 (91)	110 - 440
Seed / R8	4 (0.77)	2.6 – 5.1	2.4 (0.5)	1.7 – 3.2
Pollen; anther / R2[3]	<LOQ		<LOQ	

[1]For information on soybean growth stages see e.g. Iowa State University (2009).
[2] Forage is the above ground plant parts used for animal feed.
[3]Flowers from which pollen and anther tissue was collected were grown only at one site in Champaign, Illinois

Mean levels of Cry1A.105, while low, were generally much higher than those of Cry2Ab2 and were highest in the R6 leaf (790 µg/g dry weight), although unquantifiable in the roots and the pollen and very low in the seed. This is consistent with the use of the *RbcS4* promoter to drive expression of *cry1A.105*; this promoter is generally regarded as being stronger than the constitutive 35S promoter used to drive *cry2Ab2* expression but is specific to green parts of the plant. For Cry2Ab2, mean protein levels were extremely low in all tissues but were highest in the R2 – R3 leaf (32 µg/g dry weight) and were so low as to not be quantifiable in the pollen

In the seed, Cry1A.105 was present at a mean level of 4 µg/g dry weight and Cry2Ab2 was present at 2.4 µg/g dry weight.

4.1.3 Characterisation of the Cry1A.105 and Cry2Ab2 proteins

Cry1A.105 is a full-length Cry protein consisting of 1,181 amino acids with a predicted molecular weight of approximately 133 kDa. It is a chimeric protein consisting of domains I and II from Cry1Ab or Cry1Ac[16], domain III from Cry1F, and the C-terminal domain from Cry1Ac (Figure 6).

[16] Cry1Ab and Cry1Ac share 100% amino acid sequence identity in domains I and II.

Cry1A.105 was designed using a domain exchange strategy to achieve high levels of activity against the target insect pests. Domain exchange is a naturally occurring mechanism that increases protein diversity in *B. thuringiensis* (de Maagd et al. 2001; Masson et al. 2002; de Maagd et al. 2003). A domain exchange strategy has been used to switch the functional domains of Cry1 proteins to develop a commercial biopesticide with improved specificity to lepidopteran insect pests.

Cry1A.105 is most closely related to Cry1Ac (93.6% amino acid sequence identity), but it also has a high degree of amino acid identity (90%) with Cry1Ab. Identity with Cry1F is 76.7%.

Cry1A.105 is targeted to the chloroplast using CTP from *A. thaliana* which is fused to the N-terminus of Cry1A.105. CTPs are typically cleaved from the mature protein upon translocation into the chloroplast, and then rapidly degraded (Bruce 2000).

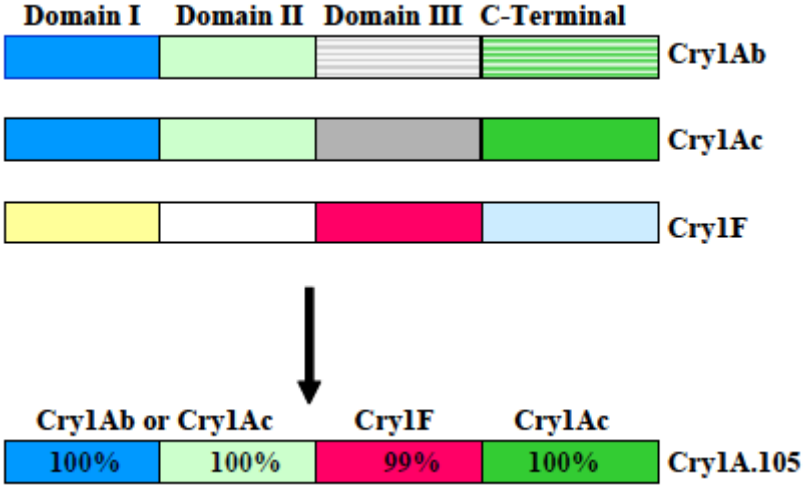


Figure 6: Schematic representation of the origin of Cry1A.105 protein domains

The amino acid sequence for the Cry2Ab2 expressed in MON87751 should be identical to the Cry2Ab2 coding sequence present in MON89034 and is a variant of the wild-type Cry2Ab2 protein from *B. thuringiensis* subsp. *kurstaki*. It is predicted to have a molecular weight of approximately 61 kDa and comprise 637 amino acids. This variant protein has 88% amino acid sequence identity to Cry2Aa, which is present in a number of registered pest control products in Australia and New Zealand.

As with Cry1A.105, a CTP targets the protein to the chloroplasts before it is cleaved away.

It was necessary to confirm that the proteins expressed in MON87751 have the expected biochemical characteristics. Accordingly, the Applicant used a number of analytical techniques to characterise the Cry1A.105 and Cry2Ab2 proteins and compare them with previously characterised microbially-derived proteins.

The techniques used were:

- Sodium dodecyl polyacrylamide gel electrophoresis (SDS - PAGE)
- Western blot analysis
- N-terminal sequencing
- Peptide mass mapping
- Glycosylation analysis
- Enzyme bioactivity assay

Studies submitted:

2013. Characterization of the Cry1A.105 Protein Purified from the Soybean Seed of MON 87751 and Comparison of the Physicochemical and Functional Properties of the Plant-Produced and *Escherichia coli*-Produced Cry1A.105 Proteins. **MSL0025088**. Monsanto Company (unpublished).

2013. Amended from MSL0024792: Characterization of the Cry2Ab2 Protein Purified from the Soybean Seed of MON 87751 and Comparison of the Physicochemical and Functional Properties of the Plant-Produced and *Escherichia coli*-Produced Cry2Ab2 Proteins. **MSL0025096**. Monsanto Company (unpublished).

Plant-produced Cry2Ab2 and Cry1A.105 proteins were purified from defatted enzyme active seed flour of MON87751 (generation R_7). Several batches of each protein were prepared in order to generate sufficient amounts of pooled protein. Microbially-derived Cry2Ab2 and Cry1A.105 were produced and characterised from *Escherichia coli* and reflected the same variant amino acid sequences shown in the plant-derived proteins (see 4.1.3.2 and 4.1.3.3 below).

4.1.3.1 Molecular weight and immunoreactivity of Cry1A.105 and Cry2Ab2

The molecular weights of plant- and microbially-derived proteins were estimated from SDS-PAGE. Following electrophoresis, gels were stained with Brilliant Blue G-Colloidal stain and analysed by densitometry. Apparent molecular weight was reported as an average of six lanes containing the MON87551-produced protein.

Cry1A.105

The *E.coli*-derived protein showed, on the stained SDS-PAGE gel, a prominent band with an apparent molecular weight of 130.8 kDa. Several faint bands were also observed at lower molecular weights and these were also observed in the Western blot (see below).

For protein obtained from MON87751, three bands were observed. The most prominent of these had an apparent molecular weight of 132.9 kDa and migrated to the same position on the gel as the *E. coli*-derived protein. The occurrence of the other, much fainter, bands is most likely due to contaminating proteins (the purity of the samples was estimated to be 60%) not completely removed during the purification process. This conclusion is supported by the fact that in the Western blot analysis (next paragraph) none of these additional bands was shown to be immunoreactive. The apparent molecular weight estimates for the *E.coli*- and plant-derived proteins is in good agreement with the calculated molecular weight of 133 kDa .

The Western blot analysis used a polyclonal goat anti-Cry1A.105 primary antibody and a commercial (Vector Lab, Burlingame) anti-goat horseradish peroxidase (HRP) linked secondary antibody. The blot showed a single immunoreactive band, increasing in intensity with protein load, that had co-migrated in separate extracts from MON87411 and *E. coli*. at the expected apparent molecular weight (approximately 133kDa). The three bands that occurred below approximately 75 kDa are likely to represent Cry1A.105 degradation products due to storage and/or sample processing; it has been shown that C-terminal regions of Cry1 proteins can be degraded by proteolytic enzymes present in the cells or released during protein purification (Gao et al. 2006).

Cry2Ab2

For the microbially-derived protein, a single band was observed in the SDS-PAGE gel. For the plant-derived protein four bands were observed. The highest and most prominent of these corresponded to the single band from the microbially-derived protein. The other bands are likely to be due to contaminating proteins given they were not immunoreactive.

The apparent molecular weight of the *E. coli*-derived protein was estimated to be 60.1 kDa while that of the prominent plant-derived protein band was 61.4 kDa. This molecular weight estimate is in good agreement with the calculated molecular weight of 61 kDa.

Immunoreactivity was detected on the Western blots using a monoclonal mouse anti-Cry2Ab2 primary antibody and a commercial (Thermo Scientific) anti mouse horseradish peroxidase-linked secondary antibody. The Western blot analysis showed a single immunoreactive band at approximately 60 kDa, increasing in intensity with protein load, that had co-migrated in separate extracts from MON87751 and *E. coli*.

Conclusion

The Cry2Ab2 and Cry1A.105 proteins expressed in MON87751 had the expected size and immunoreactivity.

4.1.3.2 N-terminal sequence analysis

Automated Edman degradation chemistry was performed on the 133 kDa band of the plant derived Cry1A.105 protein and the 61 kDa band of the plant-derived Cry2Ab2 protein, that had been eluted from SDS-PAGE gels.

While the Cry1A.105 protein in MON87751 should have the same amino acid sequence as that in the previously-assessed corn line MON89034 (FSANZ 2008), the N-terminal sequencing indicated that the Cry1A.105 protein in MON87751 contains an additional four amino acids at the N-terminus. These are derived from the *RbcS4* targeting sequence (see Section 3.2). The additional four amino acids are cysteine (C), methionine (M), glutamine (Q), and alanine (A) (Figure 7). While the identities of methionine, glutamine, and alanine were clearly determined by N-terminal sequencing, the identity of the first amino acid, cysteine, was inferred based on the *RbcS*4 targeting sequence in MON 87751. The chemistry employed in N-terminal sequencing is known to degrade cysteine, preventing its clear identification (Speicher et al. 2009). With the exception of the four additional CTP-derived amino acids, the deduced sequence of the Cry1A.105 protein that accumulates in MON 87751 shares 100% amino acid identity with the deduced sequence of the Cry1A.105 protein present in MON 89034.

Amino acid residue # from the N-terminus →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Expected Sequence →	C	M	Q	A	M	D	N	N	P	N	I	N	E	C	I
Experimental Sequence →	X	M	Q	A	M	D	N	N	(P)	X	X	X	X	X	X

X = residue not identified
( ) = tenuous designation

Figure 7: N-terminal sequence of the Cy1A.105 protein in MON87751

The Cry2Ab2 protein in MON87411 should have the same sequence as the wild type Cry2Ab2 which is the same as the protein in the previously-assessed corn line MON89034 (FSANZ 2008), except that the latter is three amino acids longer than the wild type Cry2Ab2. However, the N-terminal sequencing results (Figure 8) indicated that the N-terminus of the MON87751 protein starts at position 16 relative to the first methionine of the predicted Cry2Ab2.

Accumulation of the Cry2Ab2 protein in MON 87751 is targeted to the chloroplasts due to the addition of a *CTP2* sequence at the 5′ end of the coding sequence (see Section 3.2) and it is evident that cleavage of the CTP2 on uptake of the mature Cry2Ab2 protein into the chloroplast, has occurred at a position 15 amino acids within the Cry2Ab2 protein sequence. This deletion results in an amino acid sequence for the MON 87751-derived Cry2Ab2 protein that is 18 amino acids shorter than the MON 89034-derived Cry2Ab2 protein (bearing in mind that the MON 89034-derived protein is three amino acids longer than wild type Cry2Ab2).

Amino acid residue # from the N-terminus →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Expected Sequence →	A	Y	N	V	A	A	H	D	P	F	S	F	Q	H	K
Experimental Sequence →	A	Y	N	V	A	A	H	D	P	F	S	X	Q	X	X

X = residue not identified

Figure 8: N-terminal sequence of the Cry2Ab2 protein in MON87751

4.1.3.3 MALDI-TOF tryptic mass fingerprint and intact mass analyses

A protein identification made by peptide mass fingerprinting is considered to be reliable if the measured coverage of the sequence is 15% or higher with a minimum of five unique peptide matches with those expected from a specified enzyme digestion of the theoretical protein (Jensen et al. 1997).

For Cry2Ab2 and Cry1A.105, mass spectral analysis using matrix-assisted laser desorption/ionisation–time of flight mass spectrometry (MALDI-TOF MS) was performed on the trypsin-digested Cry2Ab2 60 kDa excised band and Cry1A.105 133 kDa band obtained by running the MON87751 protein samples on SDS-PAGE.

The peptide mapping of the Cry1A.105 protein identified 60 unique peptides corresponding to the masses expected to be produced by trypsin digestion, with a coverage of 53.5% of the amino acids expected in Cry1A.105. This was adequate to confirm the identity of the protein. Based on the evidence from the sequence of the *cry1A.105* gene together with N-terminal sequencing and MALDI-TOF MS, the sequence of the MON87751-derived Cry1A.105 protein is as given in Figure 9 and gives rise to a protein with a calculated molecular weight of 133 kDa and comprising 1,177 amino acids.

```
                          CMQAMD NNPNINECIP YNCLSNPEVE VLGGERIETG
YTPIDISLSL TQFLLSEFVP GAGFVLGLVD IIWGIFGPSQ WDAFLVQIEQ LINQRIEEFA
RNQAISRLEG LSNLYQIYAE SFREWEADPT NPALREEMRI QFNDMNSALT TAIPLFAVQN
YQVPLLSVYV QAANLHLSVL RDVSVFGQRW GFDAATINSR YNDLTRLIGN YTDHAVRWYN
TGLERVWGPD SRDWIRYNQF RRELTLTVLD IVSLFPNYDS RTYPIRTVSQ LTREIYTNPV
LENFDGSFRG SAQGIEGSIR SPHLMDILNS ITIYTDAHRG EYYWSGHQIM ASPVGFSGPE
FTFPLYGTMG NAAPQQRIVA QLGQGVYRTL SSTLYRRPFN IGINNQQLSV LDGTEFAYGT
SSNLPSAVYR KSGTVDSLDE IPPQNNNVPP RQGFSHRLSH VSMFRSGFSN SSVSIIRAPM
FSWIHRSAEF NNIIASDSIT QIPLVKAHTL QSGTTVVRGP GFTGGDILRR TSGGPFAYTI
VNINGQLPQR YRARIRYAST TNLRIYVTVA GERIFAGQFN KTMDTGDPLT FQSFSYATIN
TAFTFPMSQS SFTVGADTFS SGNEVYIDRF ELIPVTATLE AEYNLERAQK AVNALFTSTN
QLGLKTNVTD YHIDQVSNLV TYLSDEFCLD EKRELSEKVK HAKRLSDERN LLQDSNFKDI
NRQPERGWGG STGITIQGGD DVFKENYVTL SGTFDECYPT YLYQKIDESK LKAFTRYQLR
GYIEDSQDLE IYSIRYNAKH ETVNVPGTGS LWPLSAQSPI GKCGEPNRCA PHLEWNPDLD
CSCRDGEKCA HHSHHFSLDI DVGCTDLNED LGVWVIFKIK TQDGHARLGN LEFLEEKPLV
GEALARVKRA EKKWRDKREK LEWETNIVYK EAKESVDALF VNSQYDQLQA DTNIAMIHAA
DKRVHSIREA YLPELSVIPG VNAAIFEELE GRIFTAFSLY DARNVIKNGD FNNGLSCWNV
KGHVDVEEQN NQRSVLVVPE WEAEVSQEVR VCPGRGYILR VTAYKEGYGE GCVTIHEIEN
NTDELKFSNC VEEEIYPNNT VTCNDYTVNQ EEYGGAYTSR NRGYNEAPSV PADYASVYEE
KSYTDGRREN PCEFNRGYRD YTPLPVGYVT KELEYFPETD KVWIEIGETE GTFIVDSVEL
LLMEE
```

Figure 9: Sequence of the Cry1A.105 protein present in MON87751 (four N-terminus amino acids from the RbcS4 targeting sequence are highlighted).

The peptide mapping of the Cry2Ab2 protein identified 37 unique peptides corresponding to the masses expected to be produced by trypsin digestion, with a coverage of 66% of the amino acids expected in Cry2Ab2, which was adequate to confirm the identity of the protein.

Based on the evidence from the sequence of the *cry2Ab2* gene together with N-terminal sequencing and MALDI-TOF MS, the sequence of the MON87751-derived Cry2Ab2 protein is as shown in Figure 10 and gives rise to a protein with a calculated molecular weight of 62 kDa and comprising 619 amino acids.

```
                          M DNSVLNSGRT TICDAYNVAA HDPFSFQHKS LDTVQKEWTE
WKKNNHSLYL DPIVGTVASF LLKKVGSLVG KRILSELRNL IFPSGSTNLM QDILRETEKF
LNQRLNTDTL ARVNAELTGL QANVEEFNRQ VDNFLNPNRN AVPLSITSSV NTMQQLFLNR
LPQFQMQGYQ LLLLPLFAQA ANLHLSFIRD VILNADEWGI SAATLRTYRD YLKNYTRDYS
NYCINTYQSA FKGLNTRLHD MLEFRTYMFL NVFEYVSIWS LFKYQSLLVS SGANLYASGS
GPQQTQSFTS QDWPFLYSLF QVNSNYVLNG FSGARLSNTF PNIVGLPGST TTHALLAARV
NYSGGISSGD IGASPFNQNF NCSTFLPPLL TPFVRSWLDS GSDREGVATV TNWQTESFET
TLGLRSGAFT ARGNSNYFPD YFIRNISGVP LVVRNEDLRR PLHYNEIRNI ASPSGTPGGA
RAYMVSVHNR KNNIHAVHEN GSMIHLAPND YTGFTISPIH ATQVNNQTRT FISEKFGNQG
DSLRFEQNNT TARYTLRGNG NSYNLYLRVS SIGNSTIRVT INGRVYTATN VNTTTNNDGV
NDNGARFSDI NIGNVVASSN SDVPLDINVT LNSGTQFDLM NIMLVPTNIS PLY
```

Figure 10: Sequence of the full-length Cry2Ab2 protein present in MON87751 (the 15 N-terminal amino acids removed during cleavage of the CTP2 sequence are shown in highlight).

4.1.3.4 Glycosylation status

Many eukaryotic proteins are glycoproteins that have been post-translationally modified by the addition of carbohydrate moieties (glycans) covalently linked to the polypeptide backbone.

N-glycosylated proteins are glycosylated on an asparagine residue and commonly contain an asparagine-X-serine/threonine sequence (N-X~(P)-[S/T), where X~(P) indicates any amino acid except proline (Orlando and Yang 1998). Although rare, the sequence asparagine-X-Cysteine (N-X-C) can also be an N-glycosylation site (Miletich and Broze Jr. 1990). A basic search using NetNGlyc[17] predicted five potential sites in the Cry1A.105 protein and nine potential N-glycosylation sites in the Cry2Ab2 protein.

Analysis of purified plant- and microbially-derived Cry1A.105 and Cry2Ab2 proteins was done using an ECL™ Glycoprotein Detection Kit (GE Healthcare) following SDS-PAGE. The kit detects carbohydrates by covalently linking biotin and visualizing for its presence using a streptavidin/peroxidase system. A glycosylated protein (transferrin) was applied to each gel as a positive control.

A visible band was obtained for transferrin. No glycosylation signal was observed in either the molecular weight range of 130 kDa in the lanes containing the *E.coli*- or plant-produced Cry1A.105 protein, or the molecular weight range of 60 kDa in the lanes containing the *E.coli*- or plant-produced Cry2Ab2 protein.

A signal of ~40 kDa in the lanes of plant-produced Cry1A.105 protein is probably due to a glycosylated soy protein that was co-purified with Cry1A.105 protein, since no Cry1A.105 protein of this molecular weight was identified in the Western blot analysis.

These results support the conclusion that neither microbially- nor MON87751-derived Cry1A.105 or Cry2Ab2 or proteins are glycosylated.

4.1.3.5 Functional activity

Corn ear worm (*Helicoverpa zea*) larvae (≤ 30 h old) were used to measure biological activity of the Cry1A.105 and Cry2Ab2 proteins produced in MON87751 and *E. coli*. For each protein the bioassay was replicated (three and two times respectively) on separate days, each with a separate batch of larvae. For Cry1A.105, each bioassay replicate consisted of a series of six dilutions yielding a dose series ranging from 0.0008 – 0.025 µg Cry1A.105 protein/ml diet for the *E. coli*-produced and MON 87751-produced protein and a single buffer control. For Cry2Ab2, each bioassay replicate consisted of a series of seven dilutions yielding a dose series ranging from 0.0016 – 1.0 µg Cry2Ab2 protein/ml diet for the *E. coli*-produced and MON 87751-produced protein and a single buffer control.

The larvae (16/dose level) were allowed to feed individually in separate wells on an agar-based insect diet containing one of the dilutions for 7 days, after which the number of survivors at each dose level was recorded. Activity was expressed as EC_{50}[18] µg/ml diet..

[17] http://www.cbs.dtu.dk/services/NetNGlyc/

[18] The median effective concentration (EC50) is the statistically derived concentration of a substance in an environmental medium expected to produce a certain effect in 50% of test organisms in a given population under a defined set of conditions (IUPAC Compendium of Chemical Terminology, 1997, *Median effective concentration (EC50)*. [online] Available at: http://old.iupac.org/goldbook/M03807.pdf)

The EC_{50} of the MON 87751- and *E. coli*-produced Cry1A.105 proteins was determined to be 0.0035 µg/ml diet and 0.0032 µg/ml diet respectively and indicated that the proteins from the two sources have equivalent bioactivity. For the plant- and microbially derived Cry2Ab2 proteins the EC_{50} was determined to be 0.0734 and 0.1145 µg/ml diet respectively and this was indicative of equivalent functional activity of the proteins from the two sources.

4.1.3.6 Discussion

The results indicate that the Cry1A.105 and Cry2Ab2 proteins expressed in MON 87751 have greater than 99% and 97% amino acid identity, respectively, to the Cry1A.105 and Cry2Ab2 expressed proteins in MON 89034. Significantly, the protease-resistant 55 kDa core domains (Schnepf et al. 1998) of the modified Cry1A.105 and Cry2Ab2 proteins expressed in MON 87751 share 100% deduced amino acid identity to the core domains of the Cry1A.105 and Cry2Ab2 expressed in MON89034; these domains are responsible for insecticidal activity and specificity (Widner and Whiteley 1990; Gill et al. 1992).

4.1.4 Safety of the Cry1A.105 and Cry2Ab2 proteins

Given the slight differences in amino acid sequence between the proteins in MON89034 and MON87751, data were provided to assess the potential toxicity and allergenicity of the proteins expressed in MON87751. The characterisation described in Section 4.1.3 indicated that the modified proteins produced in an *E.coli* system and in MON87751 are equivalent and therefore it is appropriate to use the *E.coli*-produced proteins in safety assessment studies where insufficient quantities of MON87751-produced protein can be obtained.

4.1.4.1 Amino acid sequence similarity to known protein toxins and allergens

Bioinformatic analyses are useful for assessing whether introduced proteins share any amino acid sequence similarity with known protein toxins or allergens.

Studies submitted

2013. Bioinformatics Evaluation of the Cry1A.105 Protein in MON 87751 Utilizing the AD_2013, TOX_2013 and PRT_2013 Databases. **MSL0024925**. Monsanto Company (unpublished).
2013. Bioinformatics Evaluation of the Cry2Ab2 Protein in MON 87751 Utilizing the AD_2013, TOX_2013 and PRT_2013 Databases. **MSL0024926**. Monsanto Company (unpublished).

A Genbank protein database[19] (PRT_2013) containing 27,998,271 sequences was searched for all proteins with similarity to the two Cry proteins. A subset of sequences derived from PRT_2013 was compiled to contain only toxin proteins (TOX_2013) and contained 8,881 sequences.

For sequence similarities towards allergens, searches were done to compare sequences of the two Cry proteins with known allergens in the Allergen, Gliadin and Glutenin sequence database (AD_2013) residing in the 2013 FARRP (Food Allergy Research and Resource Program) dataset within AllergenOnline[20]. The database contained 1,630 sequences.

[19] http://www.ncbi.nlm.nih.gov/genbank/
[20] University of Nebraska; http:www.allergenonline.org/

The Fast Alignment Search Tool - All (FASTA) algorithm (Pearson and Lipman 1988), version 3.4t 26, was used to search all databases using the BLOSUM50[21] scoring matrix (Henikoff and Henikoff 1992).

The search of the PRT_2013 database indicated there was no biologically relevant structural similarity between either Cry1A.105 or Cry2Ab2 and any proteins of concern. The TOX_2013 database search showed that no alignments with either of the protein sequences generated an E-score[22] of ≤1e^{-5}. Similarly, the search of the AD_2013 database produced no alignments with either protein sequence displaying an E-score of ≤1e^{-5}. Additionally, no alignment met or exceeded the Codex Alimentarius (Codex 2009) FASTA alignment threshold for potential allergenicity (35% identity over 80 amino acids) and no alignments of eight or more consecutive identical amino acids (Metcalfe et al. 1996) were found.

These results demonstrated the lack of both structurally and immunologically relevant similarities between the Cry1A.105 or Cry2Ab2 proteins in MON 87751and known allergens, toxins, or other biologically active proteins that could be harmful to human or animal health.

4.1.4.2 In vitro *digestibility*

Typically, food proteins that are allergenic tend to be stable to enzymes such as pepsin and the acidic conditions of the digestive system, exposing them to the intestinal mucosa and leading to an allergic response (Astwood and Fuchs 1996; Metcalfe et al. 1996; Kimber et al. 1999). Therefore some correlation exists between resistance to digestion by pepsin and potential allergenicity although it does not necessarily follow that resistance to digestion is always an indicator of an allergenic protein (Thomas et al. 2004; Herman et al. 2007). As a consequence, one of the criteria for assessing potential allergenicity is to examine the stability of newly expressed proteins in conditions mimicking human digestion. Proteins that are rapidly degraded in such conditions are considered less likely to be involved in eliciting an allergic response. However, evidence of slow or limited protein digestibility does not necessarily indicate that the protein is allergenic.

Studies submitted

2013. Assessment of the in vitro Digestibility of Cry1A.105 Protein in Simulated Gastric and Simulated Intestinal Fluids. **MSL0024977**. Monsanto Company (unpublished).
2013. Assessment of the in vitro Digestibility of Cry2Ab2 Protein in Simulated Gastric and Simulated Intestinal Fluids. **MSL0025099**. Monsanto Company.

For both proteins, analyses using simulated gastric fluid – SGF (containing pepsin) and simulated intestinal fluid - SIF (containing pancreatin, a mixture of enzymes including amylase, trypsin, lipase, ribonuclease and protease) were done. The SIF study by itself may not be entirely informative because ordinarily an ingested protein would first be exposed to pepsin-mediated hydrolysis in the acidic environment of the stomach before being subject to further digestion in the small intestine.

[21] The BLOSUM series of matrices tabulate the frequency with which different substitutions occur in conserved blocks of protein sequences and are effective in identifying distant relationships.
[22] Comparisons between highly homologous proteins yield E-values approaching zero, indicating the very low probability that such matches would occur by chance. A larger E-value indicates a lower degree of similarity. Typically, alignments between two sequences will need to have an *E*-score of 1e-5 (1×10-5) or smaller to be considered to have significant homology.

SGF and SIF digestibility assays were performed independently but because a 5 kDa protein fragment attributable to the Cry1A.105 protein was identified after 10 min of digestion in SGF, a sequential digestion of the Cry1A.105 protein in SGF followed by SIF was also performed.

Simulated gastric fluid (SGF)

A pepsin digestibility assay (Thomas et al. 2004) was conducted to determine the digestive stability of the Cry1A.105 and Cry2Ab2 proteins purified from *E.coli* using SGF (U.S.Pharmacopeia 2000). For both proteins the test material was evaluated following incubation in SGF at 37º for 0, 0.5, 2, 5, 10, 20, 30 and 60 minutes.

The samples were then run on SDS-PAGE. Proteins were visualised by Colloidal Brilliant Blue G staining of the resulting gels. Western blotting of the SDS gels was also performed using a protein-specific goat primary antibody and an HRP-conjugated anti-goat secondary antibody.

In the SDS-PAGE gel for Cry1A.105, it was estimated that 98.7% of the protein was digested within 0.5 min of incubation in SGF. A peptide fragment of ≈ 60 kDa was present at 0.5 min but had disappeared by 2 min, and a fragment of ≈ 5 kDa appeared at 5 min but had disappeared by 30 min. The presence of the latter fragment suggested that a sequential SGF/SIF digestion should be done (see below) to determine whether the full digestive process resulted in complete degradation of the protein. The Western blot for Cry1A.105 indicated that 98.4% of the protein was digested within 0.5 min; no peptide fragments were detected at any time point.

For Cry2Ab2, SDS-PAGE results indicated that more than 99% of the protein was digested within 0.5 min. Peptide fragments of 4 – 5 kDa were observed in the 0.5 min time point but were gone by 2 min. The Western blot analysis indicated that more than 98% of the protein was digested within 0.5 min; no peptide fragments were detected at any time point.

Simulated intestinal fluid (SIF)

As for the SGF studies, Cry1A.105 and Cry2Ab2 proteins derived from *E. coli* were used. For both proteins the test material was evaluated following incubation in SIF (U.S. Pharmacopeia, 1990) at 37º for 0, 5, 15, and 30 minutes and 1, 2, 4, 8, and 24 h. Western blotting (using the same antibodies as for the SGF analyses) was then used to visualise the protein bands.

The Western blot of the Cry1A.105 protein indicated that more than 98% of the protein was digested within 5 min. As expected an immunoreactive band corresponding to the size of the ≈ 55 kDa protease-resistant core (see Section 4.1.3.6) was observed throughput the course of the SIF digestion.

For Cry2Ab2, the Western blot analysis indicated more than 96.8% of the protein was digested within 0.5 min. As for the Cry1A.105 protein, a protease-resistant core (≈ 50 kDa) remained during the course of the SIF digestion. Other immunoreactive bands (≈ 10, 15 and 57 kDa) were noted to be transiently present but disappeared over time. An immunoreactive band of ≈180 kDa was observed intermittently and it was concluded that it was due to aggregation of the intact Cry2Ab2 and/or the tryptic core under the conditions necessary to conduct the SIF digestion assay.

Sequential SGF/SIF digestion of Cry1A.105

For this procedure, the Cry1A.105 protein was digested in SGF for 2 min, the SGF digestion was quenched, and digestion was then continued in SIF for 0, 0.5, 2, 5, 10, 30, 60 and 120 min. The SDS-PAGE gel showed the intact Cry1A.105 protein was digested within 2 min of incubation in SGF and the 5 kDa fragment was completely digested within 0.5 min of exposure to SIF. In the Western blot, no bands were detected after the SGF 0 min time point.

Conclusion

Rapid digestion of the full-length Cry1A.105 and Cry2Ab2 proteins in SGF and SIF, together with rapid digestion in sequential digestion of a transiently stable ≈5 kDa fragment associated with Cry1A.105 indicates that both proteins are readily digestible in the mammalian digestive tract.

4.1.4.3 Heat Stability

The thermolability of a protein provides an indication of its stability under cooking/processing conditions.

Studies submitted

2013. Effect of Heat Treatment on the Functional Activity of *Escherichia coli* (*E. coli*)-Produced MON 87751 Cry1A.105 Protein. **MSL0025092**. Monsanto Company (unpublished).
2013. Effect of Heat Treatment on the Functional Activity of *Escherichia coli* (*E. coli*)-Produced MON 87751 Cry2Ab2 Protein. **MSL0025102**. Monsanto Company (unpublished).

Cry1A.105 and Cry2Ab2 proteins derived from *E. coli* were incubated at 25^{o}, 37^{o}, 55^{o}, 75^{o} or 95^{o} C for 15 and 30 min. The samples were then analysed for functional activity using a corn earworm assay, as described in principle in Section 4.1.3.5. The assay basically entailed feeding 16 larvae per treatment for a period of 7 d and then recording the number of survivors and the weight of surviving insects.

Samples were also analysed by SDS-PAGE where the intensity of the major protein band obtained for each heat-treated sample after staining in Brilliant Blue Colloidal G was compared to the same band in a control sample maintained on wet ice.

Results showed that the Cry1A.105 protein is functionally active following heat treatment at 25 ºC, 37 ºC and 55 ºC for 15 or 30 min. Cry1A.105 protein heated to 75 ºC and 95 ºC demonstrated >97% loss in activity at the highest dose level tested in the insect bioassay. SDS-PAGE analysis demonstrated that treatment at 25, 37, 55, 75 and 95 ºC for 15 minutes or 25, 37, 55 and 75 ºC for 30 minutes did not significantly change the band intensity of the Cry1A.105 protein. When incubated at 95 ºC for 30 minutes, a slight reduction in band intensity of the Cry1A.105 protein was observed.

For Cry2Ab2, the results showed the protein is functionally active following heat treatment at 25 ºC and 37 ºC for 15 or 30 minutes. Heating to a temperature of 55 ºC for 15 minutes, leads to a 96% reduction of functional activity while there is a 99% reduction when it is heated to 55 ºC for 30 minutes. Heating to 75 and 95 ºC for 15 or 30 minutes, leads to a >99% loss of functional activity. SDS-PAGE analysis demonstrated that treatment at 25, 37, or 55 ºC for 15 or 30 minutes did not change the band intensity. Heating to 75 ºC and above leads to an increasing reduction in band intensity, with a 90% reduction when the protein is heated to 95 ºC for 30 minutes.

Overall, the results for both proteins indicate there is a loss of functional activity and a tendency towards protein degradation at elevated temperatures.

4.1.5 Bioinformatic analysis of additional ORFs created by the transformation procedure

Studies submitted:

2013. Bioinformatics Evaluation of DNA Sequences Flanking the 5' and 3' Junctions of Inserted DNA in MON 87751: Assessment of Putative Polypeptides. **MSL0024956**. Monsanto Company (unpublished)

2013. Bioinformatics Evaluation of the Transfer DNA Insert in MON87551 Utilizing the AD-2013, TOX_2013 and PRT_2013 Databases. **MSL0024971**. Monsanto Company (unpublished)

This analysis, which was done using an *in silico* bioinformatic approach, was divided into two parts that considered a) the inserted T-DNA sequence and b) the junction sequence ORFs. The T-DNA analysis was done using query sequences generated by translation of the entire T-DNA, using DNAStar, EditSeq software in all six reading frames. For the junction sequence ORFs the bioinformatics analysis was performed to assess the similarity to known allergens and toxins of the putative polypeptides encoded by the 12 sequences obtained from the ORF analysis (refer to Section 3.4.4).

To evaluate the similarity to known allergens of proteins that might potentially be produced by translation of either the T-DNA or 12 junction sequence ORFs, an epitope search was carried out to identify any short sequences of amino acids that might represent an isolated shared allergenic epitope. This search compared the sequences with known allergens in the AD_2013 database (see Section 4.1.4.1) using the FASTA algorithm. No alignments with any of the eight junction sequence ORFs or the T-DNA query sequences generated an E-score of $\leq 1e^{-5}$, no alignment met or exceeded the Codex Alimentarius (Codex 2009) FASTA alignment threshold for potential allergenicity (35% identity over 80 amino acids) and, for the 12 junction regions ORFs, no alignments of eight or more consecutive identical amino acids (Metcalfe et al. 1996) were found.

For the T-DNA, using the eight amino acid sliding window algorithm, the frame 4 translation identified one alignment with tropomyosin from the bivalve mollusc *Solen strictus*. However, since the aligned region in the query sequence is on the reverse compliment strand of the promoter from the *act2* gene, it is unlikely that a mRNA would be produced as this would require downstream CTP or Cry2Ab2 coding sequence to act as a promoter. Additionally, even in the unlikely event of this occurring, the result would be the production of dsRNA which is unstable and is not translated.

The sequences corresponding to the T-DNA translation and the 12 junction sequence ORFs were also compared with sequences present in the PROT_2013 and TOX_2013 (see section 4.1.4) databases using the FASTA algorithm. No significant similarities to any known toxin sequences in the databases were found.

4.1.6 Conclusion

Soybean MON87751 contains two newly expressed proteins, Cry1A.105 and Cry2Ab2. Mean levels of Cry1A.105 were highest in the R6 leaf (790 µg/g dry weight) and lowest in the roots and pollen where the level was below the LOQ. For Cry2Ab2, mean protein levels were highest in the R2 – R3 leaf (32 µg/g dry weight) and were lowest in the pollen (<LOQ). In the seed, from which most food products are derived, Cry1A.105 was present at a mean level of 4 µg/g dry weight and Cry2Ab2 was present at 2.4 µg/g dry weight.

A range of characterisation studies confirmed the identity of the Cry1A.105 and Cry2Ab2 proteins produced in MON87751 and also their equivalence with the corresponding proteins produced in a bacterial expression system. It was found that incomplete cleavage of the chloroplast targeting sequence associated with Cry1A.105 results in the expression of a protein that is four amino acids longer than predicted. Conversely, cleavage of the chloroplast targeting sequence associated with the Cry2Ab2 protein has also resulted in the removal of the first 15 amino acids such that the protein expressed in MON87751 is 15 amino acids shorter than predicted. The variant proteins have the expected molecular weights, immunoreactivity, lack of glycosylation and functional activity.

For both plant-expressed proteins, bioinformatic studies confirmed the lack of any significant amino acid sequence similarity to know protein toxins or allergens; digestibility studies suggest the proteins would be rapidly degraded in the gastro-intestinal tract following ingestion; and thermolability studies indicate both proteins are functionally inactivated following heating. Taken together, the evidence indicates that neither Cry1A.105 nor Cry2Ab2 are likely to be toxic or allergenic in humans.

The results of bioinformatic analyses of putative ORFs created by the transformation process indicate a) there is no evidence to indicate that the T-DNA insert in MON 87751 results in the production of any protein that shares relevant similarities between known allergens, toxins or biologically active proteins of concern, and b) no structurally relevant sequence similarities were observed between the 12 putative junction ORFs and any allergens, toxins or biologically active proteins.

5 Compositional Analyses

The main purpose of compositional analyses is to determine if, as a result of the genetic modification, any unexpected changes have occurred to the food. These changes could take the form of alterations in the composition of the plant and its tissues and thus its nutritional adequacy. Compositional analyses can also be important for evaluating the intended effect where there has been a deliberate change to the composition of the food.

The classic approach to the compositional analyses of GM food is a targeted one. Rather than analysing every single constituent, the aim is to analyse only those constituents most relevant to the safety of the food or that may have an impact on the whole diet. Important analytes therefore include the key nutrients, toxicants and anti-nutrients for the food in question. The key nutrients and anti-nutrients are those components in a particular food that may have a substantial impact in the overall diet. They may be major constituents (fats, proteins, carbohydrates or enzyme inhibitors such as anti-nutrients) or minor constituents (minerals, vitamins). Key toxicants are those toxicologically significant compounds known to be inherently present in an organism, such as compounds whose toxic potency and level may be significant to health (e.g. solanine in potatoes).

5.1 Key components

For soybean intended for human food use, the key components considered important for compositional analysis include the proximates (moisture, crude protein, fat, ash), fibre, amino acids, fatty acids, minerals, vitamins, isoflavones, phospholipids, sterols, saponins and the anti-nutrients phytic acid, trypsin inhibitors, stachyose, raffinose and lectins, (OECD 2012).

Analyses for key components were done on seed and forage. In general, soybean is cultivated for the production of seed, which is used as a source of both human food and animal feed, and is only infrequently used as a forage crop for livestock.

As there are no human food products derived from forage, only the results of the compositional analyses for seed and its processed fractions are presented in this report. The compositional analyses for forage (collected at the R6 stage) focussed only on proximates and fibre. The analyses showed that the only significant differences ($p < 0.05$) between MON87751 and A3555 were for mean levels of total fat and neutral detergent fibre (NDF). For total fat, the mean value was 6.03% dw for MON 87751 and 6.43% dw for the conventional control. For NDF, the mean value was 36.77% dwt for MON 87751 and 34.08% dwt for the conventional control. In both cases, the means for MON87751 were within a 99% tolerance interval and within values observed in the literature (see Section 5.3 below for an explanation of these values).

5.2 Study design and conduct for key components

Study submitted:

2014. Amended Report for MSL0024609: Compositional Analyses of Soybean Forage and Seed Collected from MON 87751 Grown in the United States during the 2012 Season. **MSL0026251**. Monsanto Company (unpublished).

The test (MON87751 seed of generation R_7) and control (A3555) lines were grown from PCR-verified seed lots at eight replicated field sites in the U.S.[23] during the 2012 growing season. Additionally, a total of 19 non-GM cultivars were grown as reference lines with four different lines being planted at each site. The purpose of these was to generate tolerance ranges for each analyte. There were four replicated plots at each site planted in a randomised complete-block design. Production was conducted under normal agronomic conditions for the geographic regions represented by the field sites.

Seed was harvested at physiological maturity and samples were analysed for proximates, fibre (acid detergent fibre – ADF; NDF; total dietary fibre), fatty acids, amino acids, minerals, vitamins, antinutrients and isoflavones. Chain-of-custody documentation supplied with the harvested seed was used to maintain sample identity.

Methods of composition analysis were based on internationally recognised procedures (e.g. those of the Association of Official Analytical Chemists), methods specified by the manufacturer of the equipment used for analysis, or other published methods.

5.3 Analyses of key components in grain

For each analyte 'descriptive statistics' were generated i.e. a mean (least-square mean) and standard error, averaged over all sites (combined-site analysis). The values thus calculated are presented in Tables 5 – 11.

In total, 57 analyte levels were measured for statistical analysis and carbohydrate was calculated rather than being measured i.e. a total of 58 analytes were considered. Moisture values were measured for conversion of components to dry weight, but were not statistically analysed. Fourteen analytes had more than half of the observations below the LOQ and were excluded from the statistical analysis. The data for 44 analytes were therefore analysed. This analysis used a mixed model analysis of variance. Data were transformed into Statistical Analysis Software[24] (SAS) data sets and analysed using SAS® software (SAS MIXED, version 9.2).

[23] Jackson County, Arkansas; Story County, Iowa; Jefferson County, Iowa; Champaign County, Illinois; Pawnee County, Kansas; Perquimans County, North Carolina; Merrick County, Nebraska; Lehigh County, Pennsylvania.
[24] SAS website - http://www.sas.com/technologies/analytics/statistics/stat/index.html

The SAS GLM procedure was applied to all data (test, control and reference) to detect potential outliers in the dataset by screening studentised PRESS residuals[25]. The four replicated plots at each site were analysed using a combined-site analysis and the data are presented in Tables 5 – 11. In assessing the significance of any difference between means, a P-value of 0.05 was used (i.e. a P-value of ≥0.05 was not significant).

Any statistically significant differences between MON87751 and the A3555 control were compared to the 99% tolerance interval compiled from the results for each analyte of the 19 non-GM reference lines combined across all sites, to assess whether the differences are likely to be biologically meaningful. These tolerance intervals contain, with 95% confidence, 99% of the quantities expressed in the population of commercial substances. Additionally, the results for MON87751 and A3555 were compared to a combined literature range for each analyte, compiled from published literature for commercially available soybean[26]. It is noted, however, that information in the published literature is limited and is unlikely to provide a broad reflection of the natural diversity that occurs within soybean (Harrigan et al. 2010). Therefore, even if means fall outside the published range, this is not necessarily a concern.

5.3.1 Proximates and fibre

Results of the proximate and fibre analysis are shown in Table 5. One MON87751 analyte mean showed a significant difference from the control – the mean protein level in MON87751 is significantly higher than the level in the A3555 control. However, the MON87751 mean was within both the tolerance interval and the literature range.

Table 5: Mean (±standard error) percentage dry weight (%dw) of proximates and fibre in seed from MON87751 and A3555

Analyte	MON87751[2] (%dw)	A3555 (%dw)	Overall treat effect (P-value)	Tolerance interval (%dw)	Combined literature range (%dw)
Ash	4.91±0.10	4.89±0.10	0.789	4.32, 5.74	3.8 – 6.9
Protein	40.58±0.5	40.12±0.5	***0.023***	34.33, 45.17	32.2 – 45.4
Total Fat	19.21±0.43	19.48±0.43	0.088	17.12, 24.20	8.1 – 23.56
Carbohydrate[1]	35.3±0.58	35.63±0.58	0.131	29.43, 39.69	29.6 – 50.2
ADF	13.99±0.44	13.98±0.44	0.986	10.71, 17.25	7.81 – 26.2
NDF	15.56±0.33	15.57±0.33	0.982	12.66, 18.44	8.53 – 23.9

[1] Carbohydrate calculated as 100% - (protein %dw+ fat %dw + ash %dw)

[2] orange shading represents MON87751 mean that is significantly higher than the control mean.

5.3.2 Fatty Acids

The levels of 22 fatty acids were measured. Of these, the following had 50% of observations below the limit of quantitation (LOQ) and were therefore excluded from analysis - C8:0 caprylic, C10:0 capric, C12:0 lauric, C14:0 myristic, C14:1 myristoleic, C15:0 pentadecanoic, C15:1 pentadecenoic, C16:1 palmitoleic , C17:0 heptadecanoic, C17:1 heptadecenoic, C18:3 gamma linolenic, C20:2 eicosadienoic, C20:3 eicosatrienoic, and C20:4 arachidonic.

[25] A PRESS (predicted residual sum of squares) statistic provides a comparison of the predicted marginal mean and the observed mean when the predicted value is calculated without the deleted observation in question (Schabenberger 2004).

[26] References included: Kakade et al (1972); Iskander (1987); Hartwig & Kilen (1991); Padgette et al (1996); Taylor et al (1999); OECD (2001); McCann et al (2005); Harrigan et al (2007); Bilyeu et al (2008); Lundry et al (2008); Berman et al (2009); Berman et al (2010); Harrigan et al (2010); ILSI (2011)

Results for the remaining eight fatty acids are given in Table 6 and show there was no significant difference between any of the means of MON87751 and those of the control.

Table 6: Mean (±standard error) percentage composition, relative to total fat, of major fatty acids in seed from MON87751 and A3555

Analyte	MON87751 (%total)	A3555 (%total)	Overall treat effect (P-value)	Tolerance interval (%total)	Combined literature range (%total)
Palmitic acid (C16:0)	11.69±0.19	11.59±0.19	0.603	8.39, 13.35	1.4 – 15.7
Stearic acid (C18:0)	4.22±0.12	4.29±0.12	0.106	2.12, 6.45	2.59 – 5.88
Oleic acid (C18:1)	20.81±0.94	21.50±0.94	0.211	13.27, 32.25	2.6 – 45.6
Linoleic acid (C18:2)	54.57±0.73	53.93±0.73	0.074	44.95, 62.64	7.58 – 58.8
Linolenic acid (C18:3)	7.89±0.38	7.86±0.38	0.740	4.29, 10.55	1.27 – 12.52
Arachidic acid (C20:0)	0.32±0.0087	0.33±0.0087	0.133	0.17, 0.50	0.038 – 0.57
Eicosenoic acid (C20:1)	0.19±0.0084	0.19±0.0084	0.953	0.12, 0.26	<LOQ – 0.35
Behenic acid (C22:0)	0.31±0.0090	0.32±0.0090	0.354	0.20, 0.48	0.043 – 0.65

5.3.3 Amino acids

Levels of 18 amino acids were measured. Since asparagine and glutamine are converted to aspartate and glutamate respectively during the analysis, levels for aspartate include both aspartate and asparagine, while glutamate levels include both glutamate and glutamine.

The results in Table 7 show the means for glycine and proline in MON87751 were significantly higher than the means in A3555 but that these means fell within both the tolerance interval and the literature range. None of the other amino acid means in MON87751 were significantly different from those in A3555.

Table 7: Mean % dw, relative to total dw, of amino acids in seed from MON87751 and A3555

Analyte	MON87751[1] (%dw)	A3555 (%dw)	Overall treat effect (P-value)	Tolerance interval (%dw)	Combined literature range (%dw)
Alanine	1.75±0.018	1.75±0.018	0.611	1.53, 1.93	1.51 – 2.10
Arginine	3.03±0.047	3.00±0.047	0.317	2.42, 3.52	2.15 – 3.40
Aspartate	4.51±0.048	4.48±0.048	0.185	3.84, 5.07	3.81 – 5.72
Cystine	0.56±0.017	0.57±0.017	0.903	0.43, 0.67	0.37 – 0.81
Glutamate	6.89±0.085	6.85±0.085	0.360	5.60, 7.96	5.49 – 8.72
Glycine	1.73±0.017	1.71±0.017	***0.014***	1.47, 1.89	1.41 – 2.00
Histidine	1.01±0.011	1.01±0.011	0.469	0.90, 1.11	0.88 – 1.24
Isoleucine	1.92±0.022	1.90±0.022	0.296	1.63, 2.13	1.41 – 2.08
Leucine	3.04±0.032	3.02±0.032	0.186	2.68, 3.36	2.39 – 3.62
Lysine	2.51±0.025	2.49±0.025	0.096	2.16, 2.77	2.19 – 3.15

Analyte	MON87751[1] (%dw)	A3555 (%dw)	Overall treat effect (P-value)	Tolerance interval (%dw)	Combined literature range (%dw)
Methionine	0.56±0.0092	0.55±0.0092	0.135	0.45, 0.64	0.39 – 0.68
Phenylalanine	2.02±0.024	2.01±0.024	0.290	1.75, 2.27	1.62 – 2.44
Proline	2.07±0.025	2.03±0.025	***0.025***	1.72, 2.32	1.63 – 2.28
Serine	1.90±0.022	1.89±0.022	0.734	1.60, 2.17	1.11 – 2.48
Threonine	1.58±0.011	1.57±0.011	0.127	1.39, 1.72	1.14 – 1.86
Tryptophan	0.55±0.0085	0.54±0.0085	0.131	0.47, 0.60	0.30 – 0.56
Tyrosine	1.55±0.013	1.54±0.013	0.066	1.35, 1.71	0.74 – 1.61
Valine	1.87±0.017	1.87±0.017	0.460	1.66, 2.05	1.50 – 2.20

[1] orange shading represents MON87411 means that are significantly higher than the control means.

5.3.4 Minerals

The levels of two minerals in seed from MON87751 and A3555 were measured. There was no significant difference between the means for calcium in MON87751 and A3555 (Table 8). The mean for phosphorus was significantly higher in MON87751 compared to the control but was within the tolerance interval and literature range.

Table 8: Mean levels (g/100 g dw) of two minerals in the seed of MON87751 and A3555

Analyte	MON87751[1]	A3555	Overall treat effect (P-value)	Tolerance interval	Combined literature range
Calcium	0.29±0.010	00.29± 0.010	0.376	0.20, 0.41	0.12 – 0.31
Phosphorus	0.54±0.020	0.53±0.020	***0.02***	0.40, 0.70	0.50 – 0.94

[1] orange shading represents MON87751 mean that is significantly higher than the control mean.

5.3.5 Vitamins

Levels of two vitamins were measured. There was no significant difference between the means for Vitamin K in MON87751 and A3555 (Table 9). The mean for Vitamin E was significantly lower in MON87751 compared to the control but was within the tolerance interval and literature range.

Table 9: Mean levels (mg/100 g dw) of two vitamins in seed from MON87751 and A3555

Analyte	MON87751[1] (mg/100 g dw)	A3555 (mg/100 g dw)	Overall treat effect (P-value)	Tolerance range (mg/kg dw)	Combined literature range (mg/kg dw)
Vitamin E (α-tocopherol)	2.59±0.29	2.78±0.29	***<0.001***	0, 5.12	0.19 – 6.17
Vitamin K (phylloquinone)	0.64±0.067	0.67±0.067	0.111	0.036, 1.10	N/A

[1] mauve shading represents MON87751 mean that is significantly lower than the control mean.

5.3.6 Anti-nutrients

Levels of five key anti-nutrients were measured. Results in Table 10 show that neither the mean phytic acid nor raffinose level differed significantly between MON87411 and the control.

Table 10: Mean levels of anti-nutrients in seed from MON87751 and A3555

Analyte	Unit	MON87751[1]	A3555	Overall treat effect (P-value)	Tolerance interval	Combined literature range
Lectin	H.U./mg dw	3.42±0.23	3.18±0.23	0.353	0.59, 6.30	0.105 – 9.038
Phytic acid	%dw	1.22±0.067	1.19±0.067	0.234	0.80, 1.68	0.41 – 2.66
Raffinose	%dw	0.88±0.087	0.95±0.087	***0.007***	0.44, 1.27	0.21 – 1.85
Stachyose	%dw	4.09±0.079	4.06±0.079	0.628	3.15, 4.80	1.21 – 6.65
Trypsin inhibitor	TIU/mg dw	26.21±1.97	26.82±1.97	0.693	12.17, 38.77	18.14 – 118.68

[1] mauve shading represents MON87751 means that is significantly lower than the control mean.

5.3.7 Isoflavones

In total, there are 12 different soybean isoflavone isomers, namely three parent isoflavones (genistein, daidzein and glycitein), their respective β-glucosides (genistin, daidzin, and glycitin), and three β-glucosides each esterified with either malonic or acetic acid (Messina 2005). The parent isoflavones are also referred to as free or aglycon isoflavones, while the glucosides and their esters are also referred to as conjugated isoflavones.

The Applicant used an acid hydrolysis method to extract the isoflavones. This method results in the hydrolysis of all isoflavones to aglycons and therefore the results in Table 11 are expressed as total aglycon equivalents. There was no significant difference between any of the means of MON87751 and those of the control.

Table 11: Mean weight (µg/g dw) of two secondary metabolites in seed from MON87751 and A3555

Analyte (aglycon equivalents)	MON87751 (µg/g dw)	A3555 (µg/g dw)	Overall treat effect (P-value)	Tolerance range (µg/g dw)	Combined literature range (µg/g dw)
Total Daidzein	901.33±98.11	893.90±98.11	0.867	0, 1494.46	199 - 2453
Total Genistein	756.78±77.73	755.11±77.73	0.925	150.41, 1437.69	148 - 2837
Total Glycitein	77.67±5.37	84.03±5.37	0.129	8.04, 211.36	15 - 310

5.3.8 Summary of analysis of key components

A summary of the statistically significant differences in the analyte levels found between seed of MON87751 and the control A3555 is provided in Table 12.

Table 12: Summary of analyte levels found in seed of MON87751 that are significantly (P < 0.05) different from those found in seed of the control A3555

Analyte	Unit	MON87751 mean[1]	A3555 mean	diff between MON87751 & A3555 means	diff between max and min in A3555	MON87751 within tolerance interval?	MON87751 within literature range?
Protein	% dw	40.58	40.12	0.46	5.09	yes	yes
Glycine	% dw	1.73	1.71	0.02	0.20	yes	yes
Proline	% dw	2.07	2.03	0.04	0.28	yes	yes
Phosphorus	g/100 g dw	0.54	0.53	0.01	0.23	yes	yes
Vitamin E	mg/100 g dw	2.59	2.78	0.19	3.03	yes	yes
Raffinose	%dw	0.88	0.95	0.07	0.88	yes	yes

[1] mauve shading represents MON87751 means that are significantly lower than the control means while orange shading represents MON87751 means that are significantly higher.

5.4 Conclusion from compositional analyses

Detailed compositional analyses were done to establish the nutritional adequacy of seed from MON87751 and to characterise any unintended compositional changes. Analyses were done of proximates, fibre, fatty acids, amino acids, minerals, vitamins, anti-nutrients and isoflavones. The levels were compared to levels in a) the non-GM parental cultivar A3555 b) a tolerance interval compiled from results taken for 19 non-GM lines grown under the same conditions and c) levels recorded in the literature. Only 6 of the 44 analytes reported in Tables 5 – 11 deviated from the control in a statistically significant manner. However, the mean levels of all of these analytes fell within both the tolerance interval and the historical range from the literature. It is also noted that the differences between these statistically significant analyte means of MON87751 and the control means were smaller than the variation within the control. It can therefore be concluded that seed from MON87751 is compositionally equivalent to seed from conventional soybean varieties.

6 Nutritional impact

In assessing the safety of a GM food, a key factor is the need to establish that the food is nutritionally adequate and will support typical growth and well-being. In most cases, this can be achieved through an understanding of the genetic modification and its consequences, together with extensive compositional analyses of the food.

If the compositional analyses indicate biologically significant changes to the levels of certain nutrients in the GM food, additional nutritional assessment should be undertaken to assess the consequences of the changes and determine whether nutrient intakes are likely to be altered by the introduction of such foods into the food supply.

Where a GM food has been shown to be compositionally equivalent to conventional varieties, the evidence to date indicates that feeding studies using target livestock species will add little to the safety assessment and generally are not warranted (see e.g. OECD 2003; Bartholomaeus et al. 2013; Herman and Ekmay 2014). Soybean line MON87751 is the result of a genetic modification designed to confer protection against lepidopteran pests with no intention to significantly alter nutritional parameters in the food. In addition, the extensive compositional analyses of seed that have been undertaken to demonstrate the nutritional adequacy of MON87751 indicate it is equivalent in composition to conventional soybean cultivars.

The introduction of food from soybean line MON87751 into the food supply is therefore expected to have little nutritional impact and as such no additional studies, including animal feeding studies, are required.

References[27]

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215(3):403–410

American Soybean Association (2015) Soy Stats 2015. St Louis, Missouri. http://soystats.com/

An AQ, McDowell JM, Huang S, McKinney EC, Chambliss S, Meagher RB (1996) Strong constitutive expression of the Arabidopsis ACT2/ACT8 actin subclass in vegetative tissue. The Plant Journal 10(1)(107):121

Aronson AI, Shai Y (2001) Why *Bacillus thuringiensis* insecticidal toxins are so effective: unique features of their mode of action. FEMS Microbiology Letters 195:1–8

Astwood JD, Fuchs RL (1996) Allergenicity of foods derived from transgenic plants. Monographs in Allergy 32 (Highlights in Food Allergy):105–120

Bartholomaeus A, Parrott W, Bondy G, Walker K (2013) The use of whole food animal studies in the safety assessment of genetically modified crops: Limitations and recommendations. Critical Reviews in Toxicology 43(S2):1–24

Berman KH, Harrigan GG, Riordan SG, Nemeth MA, Hanson C, Smith M, Sorbet R, Zhu E, Ridley WP (2009) Compositions of seed, forage, and processed fractions from insect-protected soybean MON 87701 are equivalent to those of conventional soybean. Journal of Agricultural and Food Chemistry 57(23):11360–11369

Berman KH, Harrigan GG, Riordan SG, Nemeth MA, Hanson C, Smith M, Sorbet R, Zhu E, Ridley WP (2010) Compositions of forage and seed from second-generation glyphosate-tolerant soybean MON 89788 and insect-protected soybean MON 87701 from Brazil are equivalent to those of conventional soybean (*Glycine max*). Journal of Agricultural and Food Chemistry

Bilyeu KD, Zeng P, Coello P, Zhang ZJ, Krishnan HB, Bailey A, Beuselinck PR, Polacco JC (2008) Quantitative conversion of phytate to inorganic phosphorus in soybean seeds expressing a bacterial phytase. Plant Physiol 146(2):468–477

Bravo A, Gill SS, Soberón M (2007) Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. Toxicon 49:423–435

Broderick NA, Raffa KF, Handelsman J (2006) Midgut bacteria required for *Bacillus thuringiensis* insecticidal activity. Proceedings of the National Academy of Sciences 103:15196–15199

Brookes G, Barfoot P (2014) GM crops: global socio-economic and environmental impacts 1996 - 2012. PG Economics Ltd, Dorchester, U.K. http://www.pgeconomics.co.uk/page/36/-gm-crop-use-continues-to-benefit-the-environment-and-farmers

[27] All website references were current as at 18 June 2015

Bruce BD (2000) Chloroplast transit peptide: structure, function and evolution. Trends in Cell Biology 10(10):440–447

Codex (2009) Foods derived from modern biotechnology, 2nd edition. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organization of the United Nations, Rome.
http://www.codexalimentarius.org/standards/thematic-publications/

Damgaard PH, Granum PE, Bresciani J, Torregrossa MV, Eilenberg J, Valentino L (1997) Characterization of *Bacillus thuringiensis* isolated from infections in burn wounds. FEMS Immunology & Medical Microbiology 18:47–53

De Almeida ERP, Gossele V, Muller CG, Dockx J, Reynaerts A, Botterman J, Krebbers E, Timko MP (1989) Transgenic expression of two marker genes under the control of an *Arabidopsis rbcS* promoter: sequences encoding the Rubisco transit peptide increase expression levels. Molecular and General Genetics 218:78–86

de Maagd RA, Bravo A, Crickmore N (2001) How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. Trends in Genetics 17:193–199

de Maagd RA, Bravo A, Berry C, Crickmore N, Schnepf HE (2003) Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annual Review of Genetics 37:409–433

Delaney B, Astwood JD, Cunny H, Eichen Conn R, Herouet-Guicheney C, MacIntosh S, Meyer LS, Privalle LS, Gao Y, Mattsson J, Levine M, ILSI (2008) Evaluation of protein safety in the context of agricultural biotechnology. Food and Chemical Toxicology 46:S71–S97

Donovan, W. P (1991) CryIIb crystal protein gene from *Bacillus thuringiensis*. Patent US 5073632 A.
http://www.lens.org/lens/patent/US_5073632_A

DuBose AJ, Lichtenstein ST, Narisu N, Bonnycastle LL, Swift AJ, Chines PS, Collins FS (2013) Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene. Nucleic Acids Research 41(6):e70

FAOSTAT3 (2015) Online database of the Food and Agriculture Organization of the United Nations.
http://faostat3.fao.org/browse/T/TP/E

Fling ME, Kopf J, Richards C (1985) Nucleotide sequence of the transposon Tn7 gene encoding an aminoglycoside-modifying enzyme, 3"(9)-O-nucleotidyltransferase. Nucleic Acids Research 13(19):7095–7106

FSANZ (2008) Application A595 - Food Derived from Insect-Protected Corn Line MON89034. Report Prepared by Food Standards Australia New Zealand, available online at.
http://www.foodstandards.gov.au/code/applications/pages/applicationa595foodd3492.aspx

Gao Y, Fencil KJ, Xu X, Schwedler DA, Gilbert JR, Herman RA (2006) Purification and characterization of a chimeric Cry1F d-endotoxin expressed in transgenic cotton **plants**. Journal of Agricultural and Food Chemistry 54:829–835

Gill SS, Cowles EA, Pietrantonio PV (1992) The mode of action of *Bacillus thuringiensis* endotoxins. Annual Review of Entomology 37:615–636

Grey D (2006) Growing soybeans in northern Victoria. AG1138, State of Victoria, Department of Primary Industries. http://aof.clients.squiz.net/__data/assets/pdf_file/0008/7667/Growing_Soybeans_in_Northern_Victoria.pdf

Harrigan GG, Glenn KC, Ridley WP (2010) Assessing the natural variability in crop composition. Regulatory Toxicology and Pharmacology 58:S13–S20; doi:10.1016/j.yrtph.2010.08.023

Harrigan GG, Ridley WP, Riordan S, Nemeth MA, Sorbet R, Trujillo WA, Breeze M, Schneider RW (2007) Chemical composition of glyphosate-tolerant soybean 40-3-2 grown in Europe remains equivalent with that of conventional soybean (*Glycine max* L.). Journal of Agricultural and Food Chemistry 55(15):6160–6168

Hartwig EE, Kilen TC (1991) Yield and composition of soybean seed from parents with different protein, similar yield. Crop Science 31(2):290–292

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences 89:10915–10919

Herman RA, Ekmay R (2014) Do whole-food animal feeding studies have any value in the safety assessment of GM crops? Regulatory Toxicology and Pharmacology 68:171–174

Herman RA, Woolhiser MM, Ladics GS, Korjagin VA, Schafer BW, Storer NP, Green SB, Kan L (2007) Stability of a set of allergens and non-allergens in simulated gastric fluid. International Journal of Food Sciences and Nutrition 58:125–141

Hernandez E, Ramisse F, Ducoureau J-P, Cruel T, Cavallo J-D (1998) *Bacillus thuringiensis* subsp. *konkukian* (serotype H34) superinfection: case report and experimental evidence of pathogenicity in immunosuppressed mice. Journal of Clinical Microbiology 36:2138–2139

Herrmann KM (1995) The shikimate pathway: Early steps in the biosynthesis of aromatic compounds. Plant Cell 7:907–919

Hofmann C, Vanderbruggen H, Höfte H, Van Rie J, Jansens S, Van Mellaert H (1988) Specificity of *Bacillus thuringiensis* d-endotoxins is correlated with the presence of high-affinity binding sites in the brush border membrane of target inscet midguts. Proceedings of the National Academy of Sciences 85:7844–7848

Höfte H, Whiteley HR (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiological Reviews 53:242–255

Hunt AG (1994) Messenger RNA 3' end formation in plants. Annual Review of Plant Physiology and Plant Molecular Biology 45:47–60

ILSI (2011) International Life Sciences Institute Crop Composition Database Version 4.0. http://www.cropcomposition.org/query/index.html

Iowa State University (2009) Soybean growth stages. Soybean Extension and Research Program, Department of Agronomy, Iowa State University, Ames, IA. http://extension.agron.iastate.edu/soybean/production_growthstages.html

Iskander FY (1987) Maturation stage and mineral content in soybeans. Food Chemistry 24:29–35

Jackson SG, Goodbrand RB, Ahmed R, Kasatiya S (1995) *Bacillus cereus* and *Bacillus thuringiensis* isolated in a gastroeneritis outbreak investigation. Letters in Applied Microbiology 21:103–105

James AT, Rose IA (2004) Integrating crop physiology into the Australian soybean improvement program.
http://www.cropscience.org.au/icsc2004/poster/3/4/6/327_jamesat.htm#TopOfPage.

Jensen ON, Podtelejnikov AV, Mann M (1997) Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. Annals of Chemistry 69(23):4741–4750

Kakade ML, Simons NR, Liener IE, Lambert JW (1972) Biochemical and nutritional assessment of different varieties of soybeans. Journal of Agricultural and Food Chemistry 20(1):87–90

Kimber I, Kerkvliet NI, Taylor SL, Astwood JD, Sarlo K, Dearman RJ (1999) Toxicology of protein allergenicity: prediction and characterization. Toxicological Sciences 48(2):157–162

Klee HJ, Muskopf YM, Gasser CS (1987) Cloning of an *Arabidopsis thaliana* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. Molecular and General Genetics 210(3):437–442

Koch MS, Ward JM, Levine SL, Baum JA, Vicini JL, Hammond BG (2015) The food and environmental safety of Bt crops. Frontiers in Plant Science 6:283

Kovalic D, Garnaat C, Guo L, Yan Y, Groat J, Silvanovich A, Ralston L, Huang M, Tian Q, Christian A, Cheikh N, Hjelle J, Padgette S, Bannon G (2012) The use of next generation sequencing and junction sequence analysis bioinformatics to achieve molecular characterization of crops improved through modern biotechnology. The Plant Genome 5(3):149–163

Krebbers E, Seurinck J, Herdies L, Cashmore AR, Timko MP (1988) Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. Plant Molecular Biology 11:745–749

Krell R (1996) Value-Added Products from BeeKeeping. FAO Agricultural Services Bulletin No. 124, Food and Agriculture Organization of the United Nations, Rome

Liu H, Trieu AT, Blaylock LA, Harrison MJ (1998) Cloning and characterization of two phosphate transporters from *Medicago truncatula* roots: regulation in response to phosphate and to colonization by arbuscular mycorrhizal (AM) fungi. Molecular Plant-Microbe Interactions 11(1):14–22

Lundry DR, Ridley WP, Meyer JJ, Riordan SG, Nemeth MA, Trujillo WA, Breeze ML, Sorbet R (2008) Composition of grain, forage, and processed fractions from second-generation glyphosate-tolerant soybean, MON 89788, is equivalent to that of conventional soybean (*Glycine max* L.). Journal of Agricultural and Food Chemistry 56(12):4611–4622

Martinell, B., Petersen, P., Somers, D., Wan, Y., Williams, E., Ye, X (2011) Methods for transformation using spectinomycin selection. Patent US 8030544 B2.
http://www.lens.org/lens/patent/US_8030544_B2

Masson L, Tabashnik BE, Mazza A, Prefontaine G, Potvin L, Brousseau R, Schwartz J-L (2002) Mutagenic analysis of a conserved region of Domain III in the Cry1Ac toxin of *Bacillus thuringiensis*. Applied and Environmental Microbiology 68(1):194–200

McCann MC, Liu K, Trujillo WA, Dobert RC (2005) Glyphosate-tolerant soybeans remain compositionally equivalent to conventional soybeans (Glycine max L.) during three years of field testing. Journal of Agricultural and Food Chemistry 53:5331–5335

Messina M (2005) Isoflavones. In: Coates P, Blackman MR, Cragg G, Levine M, Moss J, and White J (eds) Encyclopedia of Dietary Supplements. Informa Healthcare,

Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL, Fuchs RL (1996) Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Critical Reviews in Food Science and Nutrition 36 Suppl:S165–S186

Miletich JP, Broze Jr. GJ (1990) b Protein C is not glycosylated at asparagine 329. The Journal of Biological Chemistry 265:11397–11404

NPTN (2000) *Bacillus thuringiensis* (Technical Fact Sheet)., National Pesticide Telecommunications Network, Oregon State University & U.S. Environmental Protection Agency. http://npic.orst.edu/factsheets/BTtech.pdf

OECD (2001) Consensus document on compositional considerations for new varieties of soybean: key food and feed nutrients and anti-nutrients. ENV/JM/MONO(2001)15. Organisation for Economic Co-operation and Development, Paris

OECD (2003) Considerations for the safety assessment of animal feedstuffs derived from genetically modified plants. ENV/JM/MONO(2003)10. Organisation for Economic Co-operation and Development, Paris. http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=ENV/JM/MONO(2003)10&doclanguage=en

OECD (2007) Consensus document on safety information on transgenic plants expressing *Bacillus thuringiensis* - derived insect control proteins. ENV/JM/MONO(2007)14. Organisation for Economic Co-operation and Development. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2007)14&docLanguage=En

OECD (2012) Revised consensus document on compositional considerations for new varieties of soybean [*Glycine max* (L.) Merr.]: Key food and feed nutrients, antinutrients, toxicants and allergens. ENV/JM/MONO(2012)24. Organisation for Economic Co-operation and Development. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2012)24&doclanguage=en

Orlando R, Yang Y (1998) Analysis of Glycoproteins. Ch 9 In: Larsen BS, McEwen CN (eds) Mass Spectrometry of Biological Materials. 2nd ed, Marcel Dekker, p. 216–246

Padgette SR, Taylor NB, Nida DL, Bailey MR, MacDonald J, Holden LR, Fuchs RL (1996) The composition of glyphosate-tolerant soybean seeds is equivalent to that of conventional soybeans. Journal of Nutrition 126(3):702–716

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences 85(8):2444–2448

Piper KR, Beck von Bodman S, Hwang I, Farrand SK (1999) Hierarchical gene regulatory systems arising from fortuitous gene associations: controlling quorum sensing by the opine regulon in *Agrobacterium*. Molecular Microbiology 32(5):1077–1089

Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. The EMBO Journal 17(20):6086–6095

Schabenberger O (2004) Mixed model influence diagnostics. Paper 189-29. Proceedings of the 29th Annual SAS Users Group International Conference (SUGI 29), May 9 - 12, 2004, Montreal. http://www.stat.purdue.edu/~bacraig/notes582/SUGI_Paper_Schabenberger.pdf

Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH (1998) *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiology and Molecular Biology Reviews 62(3):775–806

Shurtleff W, Aoyagi A (2007) History of soybean crushing: Soy oil and soybean meal. In: History of soybeans and soyfoods: 1100 B.C. to the 1980s. SoyInfo Center, available online at http://www.soyinfocenter.com/HSS/soybean_crushing1.php, Lafayette, California,

Speicher KD, Gorman N, Speicher DW (2009) N-terminal sequence analysis of proteins and peptides. Current Protocols in Protein Science 11:11.10.1–11.10.31

Taylor NB, Fuchs RL, MacDonald J, Shariff AR, Padgette SR (1999) Compositional analysis of glyphosate-tolerant soybeans treated with glyphosate. Journal of Agricultural and Food Chemistry 47(10):4469–4473

Thomas K, Aalbers M, Bannon GA, Bartels M, Dearman RJ, Esdaile DJ, Fu T-J, Glatt CM, Hadfield N, Hatzos C, Hefle SL, Heylings JR, Goodman RE, Henry B, Herouet C, Holsapple M, Ladics GS, Landry TD, MacIntosh SC, Rice EA, Privalle LS, Steiner HY, Teshima R, Van Ree R, Woolhiser M, Zawodny J (2004) A multi-laboratory evaluation of a common *in vitro* pepsin digestion assay protocol used in assessing the safety of novel proteins. Regulatory Toxicology and Pharmacology 39:87–98

U.S.Pharmacopeia (2000) The United States Pharmacopeia 24. The National Formulary 19th ed. United States Pharmacopeia Convention, Inc, Rockville, MD

WHO. Microbial pest control agent *Bacillus thuringiensis*. (217). 1999. Geneva, World Health Organization. Environmental Health Criteria.

Widner WR, Whiteley HR (1990) Location of the dipteran specificity region in a lepidopteran-dipteran crystal protein from *Bacillus thuringiensis*. Journal of Bacteriology 172:2826–2832

Wong EY, Hironaka CM, Fischhoff DA (1992) *Arabidopsis thaliana* small subunit leader and transit peptide enhance the expression of *Bacillus thuringiensis* proteins in plant cells. Plant Molecular Biology 20(1):81–93