Application to Food Standards Australia New Zealand for the inclusion of insect-protected soybean MON 87701 in Standard 1.5.2 - Food Derived from Gene Technology

Submitted by:
Monsanto Australia Limited
Level 12 / 600 St Kilda Road,
Melbourne Victoria 3004

August 2009

© 2009 Monsanto Company. All Rights Reserved.
This document is protected under copyright law. This document is for use only by the regulatory authority to which this has been submitted by Monsanto Company, and only in support of actions requested by Monsanto Company. Any other use of this material, without prior written consent of Monsanto, is strictly prohibited. By submitting this document, Monsanto dose not grant any party or entity any right or license to the information or intellectual property described in this document.

TABLE OF CONTENTS

PART 1 GENERAL INFORMATION 1

1.1 APPLICANT DETAILS 1
1.2 NATURE OF APPLICATION 1

PART 2 SPECIFIC DATA REQUIREMENTS FOR SAFETY ASSESSMENT . 2

2.1 GENERAL DETAILS 2
a) Description of the nature and purpose of the introduced trait 2
i) Intended function of the genetic modification 2
ii) Mode of action of the introduced trait 3
b) Proposed name the product will be marketed under 4
c) Product list 4
2.2 HISTORY OF USE 6
a) Donor organism 6
b) Host organism 6
i) Taxonomic classification 6
ii) Characteristics of the recipient soybean 8
iii) History of soybean 8
2.3 NATURE OF GENETIC MODIFICATION 10
a) Transformation method 10
b) Gene construct and transformation event 13
i) T-DNA I 17
ii) T-DNA II 19
iii) Genetic elements outside of the T-DNA borders 20
c) Molecular characterization 20
i) Insert and copy number 24
ii) Southern Blot Analysis to Determine the Presence or Absence of Plasmid PV-GMIR9 Backbone 30
iii) Southern Blot Analysis to Determine the Presence or Absence of T-DNA II 33
iv) Organization and Sequence of the Insert and Adjacent Genomic DNA in MON 87701 35
v) Insert junction open reading frame analysis 38
vi) Assessment of open reading frames contained in the cry1Ac coding sequence 38
d) Derivation of line and generational stability 39
i) Southern blot analysis to examine insert stability in multiple generations of MON 87701 39
ii) Inheritance of the genetic insert in MON87701 43
2.4 ANTIBIOTIC RESISTANCE MARKER GENES 47
a) Clinical / veterinary importance 47
b) Viability 47
c) Presence in food 47
2.5 CHARACTERISATION OF NOVEL PROTEIN 48
a) Description 48
i) Cry1Ac molecular weight equivalence 51
ii) Cry1Ac immunoreactivity equivalence 51
iii) N-terminal sequence analysis 52

iv) MALDI-TOF mass spectrum analysis 53
v) Cry1Ac functional activity equivalence 53
vi) Cry1Ac glycosylation equivalence 54
b) Site of expression *60*
c) Non-expression *61*
d) History of human consumption *61*
i) Safety of the donor organism : Bacillus thuringiensis 61
ii) History of safe use of Cry proteins produced in crop plants 62
e) Oral toxicological studies *62*
f) Amino acid sequence *63*
g) Known allergenicity of source *64*
h) Unknown allergenicity information *64*
i) The Cry1Ac protein as a proportion of total protein 65
ii) Bioinformatics analysis of sequence similarity of the Cry1Ac protein produced in MON 87701 to allergens 65
iii) Stability of the Cry1Ac protein in simulated gastric fluid 66
iv) Stability of the Cry1Ac protein in SGA followed by SIF 67
v) Assessment of human IgE binding to MON87701, control, and conventional soybean extracts 67
2.6 CHARACTERISATION OF OTHER NOVEL PROTEIN 74
a) Identification *74*
b) Toxicity *74*
2.7 COMPARATIVE ANALYSIS 74
a) Key nutrients etc *75*
i) Level of nutrients in soybean forage 75
ii) Level of nutrients in soybean seed 76
b) Other constitutes *77*
c) Allergenic proteins *79*
2.8 NUTRITIONAL IMPACT 93
2.9 OTHER TECHINCAL INFORMATION 93
a) Detection methodology *93*
b) Projected market penetration *93*

PART 3 REGULATORY / LEGISLATIVE IMPLICATIONS 95

3.1 OTHER APPROVALS 95
a) Relevant overseas approvals *95*
b) Approval refusal *95*
3.2 REGULATORY IMPACT STATEMENT 95

PART 4 STATUTORY DECLARATION – AUSTRALIA 96

PART 5 REFERENCES 97

APPENDIX 1 : CONFIDENTIAL ATTACHMENT - DNA Sequence of the Insert and Adjacent Genomic Regions in MON 87701

LIST OF FIGURES

Figure 1. Schematic of the Development of MON 87701 12
Figure 2. Genetic Elements and Restriction Sites of Vector PV-GMIR9 Used in Southern Blot Analysis (Probes 1-11) 14
Figure 3. Deduced Amino Acid Sequence of the CTP1 Targeting Sequence and the Full- Length Cry1Ac Protein Produced in MON 87701 18
Figure 4. Schematic Representation of the Insert and Genomic Flanking Sequences in MON 87701 22
Figure 5. Southern Blot Analysis of MON 87701: T-DNA I Copy Number Analysis (Probe 7) 27
Figure 6. Southern Blot Analysis of MON 87701: T-DNA I Copy Number Analysis (Probes 8 and 10) 28
Figure 7. Southern Blot Analysis of MON 87701: T-DNA I Copy Number Analysis (Probes 9 and 11) 29
Figure 8. Southern Blot Analysis of MON 87701: PV-GMIR9 Backbone Sequence (Probes 1, 2, and 3) 31
Figure 9. Southern Blot Analysis of MON 87701: PV-GMIR9 Backbone Sequence (Probe 4) 32
Figure 10. Southern Blot Analysis of MON 87701: T-DNA II (Probes 5 and 6) 34
Figure 11. Overlapping PCR Analysis across the Insert in MON 87701 37
Figure 12. MON 87701 Breeding Diagram 41
Figure 13. Insert Stability of MON 87701: T-DNA I (Probe 8) 42
Figure 14. Breeding Path for Generating Segregation Data for MON 87701 45
Figure 15. Amino Acid Sequence Alignment for Cry1Ac Proteins 49
Figure 16. SDS-PAGE of MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins 55
Figure 17. Western Blot Analysis of MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins 56
Figure 18. N-terminal Amino Acid Sequence Analysis of the Cry1Ac Protein Purified from Harvested Seed of MON 87701 57
Figure 19. MALDI-TOF MS Coverage Map of the Cry1Ac Protein Isolated from MON 87701 58
Figure 20. Glycosylation Analysis of the MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins 59
Figure 21. Colloidal Brilliant Blue G stained SDS-PAGE Gels of Cry1Ac Protein Digestion in SGF 69
Figure 22. Western Blot Analysis of the Digestion of Cry1Ac Protein in SGF 70
Figure 23. Cry1Ac and Cry1Ab Protein Sequences Identified in the ~4 kDa Stable Fragment from SGF 71
Figure 24. Colloidal Brilliant Blue G stained SDS-gel of the Cry1Ac Protein in SGF Followed by SIF 72
Figure 25. Serum IgE Binding Values for MON 87701, Conventional Control (A5547), and the Tolerance Limits for 17 Conventional References. 73

LIST OF TABLES

Table 1. List of Species in the Genus Glycine Willd., 2n Chromosome Number, Genome Symbol and Distribution 7
Table 2. Summary of Genetic Elements in Plasmid Vector PV-GMIR9 15
Table 3. Summary of Genetic Elements in MON 87701 23
Table 4. Segregation of the *cry1Ac* Gene during the Development of MON 87701 ..46
Table 5. Segregation of the *cry1Ac* Gene in F_2 and F_3 Progeny from a Cross of MON 87701 with a Soybean Variety that did not Contain the *cry1Ac* Gene 46
Table 6. Molecular Weight Difference between the Full-Length MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins Based on SDS-PAGE ..51
Table 7. Comparison of Immunoreactive Signals Between Full-Length MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins 52
Table 8. EC_{50} Values of *E. coli*-Produced and MON 87701-Produced Cry1Ac Proteins in a Corn Earworm Diet-Incorporation Bioassay.....54
Table 9. Summary of Cry1Ac Protein Levels in Tissues Collected from MON 87701 Produced across Five Sites during the U.S. 2007 Growing Season 61
Table 10. Statistical Summary of Combined Site Soybean Forage Fiber and Proximate Content for MON 87701 vs. the Conventional Control (A5547) 80
Table 11. Statistical Summary of Combined-Site Soybean Seed Proximate, Amino Acid, Fatty Acid, Fiber and Vitamin Content for MON 87701 vs. the Conventional Control (A5547) 81
Table 12. Statistical Summary of Combined-Site Soybean Seed Antinutrient and Isoflavone Content for MON 87701 vs. the Conventional Control (A5547) 85
Table 13. Summary of Differences ($p<0.05$) for the Comparison of Soybean Component Levels for MON 87701 vs. the Conventional Control (A5547) 86
Table 14. Literature and ILSI Ranges for Components in Soybean Forage and Seed.91

UNPUBLISHED REPORTS BEING SUBMITTED

Sophia M. Arackal, Kim R. Lawry, Zihong Song, Jeanna R. Groat, James F. Rice, James D. Masucci and Qing Tian. Amended Report for MSL0021960: Molecular Analysis of Insect-Protected Soybean MON 87701. Monsanto Study Report MSL0022176.

Andre Silvanovich and Renee Girault. Bioinformatics Evaluation of DNA Sequences Flanking the 5' and 3' Junctions of the Inserted DNA in MON 87701: Assessment of Putative Polypeptides. Monsanto Study Report MSL0021816.

Erin Bell, Kathleen S. Crowley, Joshua P. Uffman, and Elena A. Rice. Characterization of the Cry1Ac Protein Purified from the Harvested Seed of MON 87701 Soybean and Comparison of the Physicochemical and Functional Properties of the MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins. Monsanto Study Report MSL0021146.

Katherine E. Niemeyer and Andre Silvanovich. Assessment of the Cry1Ac Protein Levels in Soybean Tissues Collected from MON 87701 Produced in U.S. Field Trials During 2007. Monsanto Study Report MSL0021531.

Jason W. Smedley. An Acute Toxicity Study of Cry1Ac Protein Administered by the Oral (Gavage) Route to Mice. Monsanto Study Report CRO-2007-325.

Renee Girault and J. Scott McClain. Bioinformatics Evaluation of the Cry1Ac Protein Present in MON 87701 Soybean Utilizing the AD8, TOXIN6, and PROTEIN Databases. Monsanto Study Report MSL0021658.

Brian E. Goertz, Erin Bell and Elena A. Rice. Assessment of the *In Vitro* Digestibility of the Cry1Ac Protein in Simulated Gastric and Simulated Intestinal Fluids. Monsanto Study Report MSL0021376.

Kristina H. Berman, Susan G. Riordan, Michelle N. Smith and Roy Sorbet. Compositional Analyses of Forage and Seed Collected from MON 87701 Grown in United States during 2007. Monsanto Study Report MSL0021413.

Scott McClain, Elena Rice, Chen Meng and Gary Bannon. Quantitative ELISA Assessment of Human IgE Binding to MON 87701, Control, and Reference Soybean Using Sera from Soybean-Allergic Subjects. Monsanto Study Report MSL0022043.

GM CHECKLIST

	Data Provided	Part No.	Data Not Provided	Omission Explained
Executive Summary				
Separately bound document	☒		☐	☐
Part 1: General Information				
1.1 Applicant				
(a) Company / Organisation Name	☒	1.1	☐	☐
(b) Address	☒	1.1	☐	☐
(c) Contact	☒	1.1	☐	☐
(d) Nature of business	☒	1.1	☐	☐
(e) Sole or joint application	☒	1.2	☐	☐
(f) Co-applicants	☐	1.2	☐	☒
Part 2: Specific Data Requirements				
2.1 General Details				
(a) Description of GM organism	☒	2.1.a.	☐	☐
(b) Name / Number of new line / strain	☒	2.1.a.	☐	☐
(c) Marketing name	☒	2.1.b.	☐	☐
(d) Product list	☒	2.1.c.	☐	☐
2.2 History of Use				
(a) Donor	☒	2.2.a.	☐	☐
(b) Host	☒	2.2.b.	☐	☐
2.3 Nature of Genetic Modification				
(a) Method used	☒	2.3.a.	☐	☐
(b) Bacteria used	☒	2.3.a.	☐	☐
(c) Gene Construct and transformation event	☒	2.3.b.	☐	☐
(d) Molecular characterisation	☒	2.3.c.	☐	☐
(e) Derivation of line or strain	☒	2.3.d.	☐	☐
(f) Evidence of stability	☒	2.3.d.	☐	☐
2.4 Antibiotic Resistance Genes				
(a) Clinical / veterinary importance	☐	2.4.a.(NA)	☒	☐
(b) Viability	☐	2.4.b.(NA)	☒	☐
(c) Presence in food	☐	2.4.c.(NA)	☒	☐
2.5 Characterisation of Novel Protein				
(a) Description	☒	2.5.a.	☐	☐
(b) Site of expression	☒	2.5.b.	☐	☐
(c) Non-expression	☐	2.5.c.(NA)	☒	☐
(d) History of human consumption	☒	2.5.d.	☐	☐
(e) Oral toxicological studies	☒	2.5.e.	☐	☐
(f) Amino acid sequence	☒	2.5.f.	☐	☐
(g) Known allergenicity of source	☒	2.5.g.	☐	☐
(h) Unknown allergenicity information	☒	2.5.h.	☐	☐

GM CHECKLIST (cont'd.)

	Data Provided	Part No.	Data Not Provided	Omission Explained
2.6 Characterisation of Other Novel Substances				
(a) Identification	☐	2.6.a.(NA)	☒	☐
(b) Toxicity	☐	2.6.b.(NA)	☒	☐
2.7 Comparative Analyses				
(a) Key nutrients etc.	☒	2.7.a.	☐	☐
(b) Other constituents	☒	2.7.b.	☐	☐
(c) Allergenic proteins	☐	2.7.c.(NA)	☒	☐
2.8 Nutritional Impact				
(a) Animal feeding studies	☐	2.8	☐	☒
(b) Nutritional changes	☐	2.8	☐	☒
2.9 Other Technical Information				
(a) Detection methodology	☒	2.9.a.	☐	☐
(b) Market penetration	☒	2.9.b.	☐	☐
Part 3: Regulatory / Legislative Implications				
3.1 Other approvals				
(a) Relevant overseas approvals	☒	3.1.a.	☐	☐
(b) Approval refusal	☒	3.1.b.	☐	☐
Part 4: Statutory Declaration	☒		☐	☐

PART 1 GENERAL INFORMATION

1.1 Applicant Details

Company Name: Monsanto Australia Limited

Address: 12 / 600 St Kilda Road,
Melbourne, Victoria, 3004

Postal Address: PO Box 6051,
St Kilda Road Central, Victoria, 8008

Contact: Amanda Forster, +61 3 9522 7109

Nature of Your Business: Technology Provider to the Agricultural and Food Industries.

1.2 Nature of Application

This application is submitted to Food Standards Australia New Zealand by Monsanto Australia Limited and is not made on behalf of any other party.

The purpose of this submission is to make an application to vary **Standard 1.5.2** – *Food Produced Using Gene Technology* to seek the addition of insect-protected soybean MON 87701 and products containing insect-protected soybean MON 87701 (hereafter referred to as MON 87701) to the Table to Clause 2 (see below).

Column 1 **Food derived from gene technology**	**Column 2** **Special requirements**
Food derived from insect-protected soybean MON 87701	None

PART 2 SPECIFIC DATA REQUIREMENTS FOR SAFETY ASSESSMENT

2.1 General Details

a) Description of the nature and purpose of the introduced trait

Soybean is the most commonly grown oilseed in the world. In 2007/08, approximately 218.8 million metric tons (MMT) of harvested soybean seed were produced, representing 56% of the world's oilseed production. The impact and severity of insect pest infestations vary greatly across global soybean production regions, primarily due to the different climate and weather conditions, the distribution and environmental tolerance of insect species, and agricultural practices.

Monsanto Company has developed biotechnology-derived, insect-protected soybean MON 87701 that produces the Cry1Ac insecticidal crystal (δ-endotoxin) protein[1] derived from *Bacillus thuringiensis* (Bt) subsp. *kurstaki*. Cry1Ac provides protection from feeding damage caused by targeted lepidopteran pests and will reduce or replace current insecticide applications in tropical and subtropical production regions, where these insects cause significant plant damage and yield loss. MON 87701 will be initially commercialized in South America for the control of targeted lepidopteran pests, including the velvetbean caterpillar (*Anticarsia gemmatalis*), soybean looper (*Pseudoplusia includes*), soybean axil borer (*Epinotia. aporema*), and sunflower looper (*Rachiplusia nu*). The Cry1Ac protein produced in MON 87701 is expressed at relatively high levels in leaf tissue throughout the entire growing season and provides efficacious control of these target pests.

In the U.S., the insect pressure is greatest on soybean grown in the South, where tropical and subtropical weather favors pest infestation. According to the United States Department of Agriculture (USDA), 16% of the approximately 75 million U.S. soybean acres planted in 2006, mainly those in the Southeastern and Delta states, were treated with insecticides to control the defoliating and pod-feeding insects. Given the limited number of acres that consistently have sufficient lepidopteran insect pressure to require the use of insecticides or other insect control practices, MON 87701 plantings in the U.S will be initially limited to breeding and seed multiplication activities to support the commercialization in South America.

i) Intended function of the genetic modification

MON 87701 was produced by *Agrobacterium-mediated* transformation of soybean with the plasmid PV-GMIR9 (**Figure 2**), which is a binary vector containing 2T-DNAs. The first T-DNA, designated as T-DNA I, contains the *cry1Ac* gene expression cassette. The second T-DNA, designated as T-DNA II, contains the *cp4 epsps* gene cassette. During transformation, both T-DNAs were inserted into the soybean genome. The *cp4 epsps* gene was used as the selectable marker needed for the selection of the transformed cells and plants. After the transformed cells, and subsequently the plants, were identified, the selectable marker gene no longer was needed. Therefore, a traditional breeding process was deployed to isolate plants that

[1] Hereafter referred to as Cry1Ac or Cry1Ac protein.

only contain the *cry1Ac* expression cassette (T-DNA I), thereby producing marker-free MON 87701 plants.

MON 87701 encodes for the Cry1Ac insecticidal crystal (Cry) protein (δ-endotoxin) derived from *Bacillus thuringiensis* (Bt) subsp. *kurstaki* that provides protection from feeding damage caused by targeted lepidopteran pests in soybean production regions.

ii) Mode of action of the introduced trait

The general mode of action of Cry proteins is well understood. The bacterially-produced crystal proteins are first solubilized in the insect midgut, followed by activation of the protoxins (full-length proteins) to active toxins (proteolytic-resistant cores) by midgut proteases. A similar process occurs when Cry proteins are expressed in plants. The activated proteins then bind to midgut membrane receptors in susceptible insects, insert into the apical membrane, and form pores. Formation of the pores causes loss of osmotic regulation, and eventually leads to cell lysis, which is thought to be responsible for insect death (Gill et al., 1992; Schnepf et al., 1998; Zhuang and Gill, 2003).

Cry1 protoxins (such as Cry1Ac, Cry1Ab, and Cry1F) are 130 to 140 kDa in size, and are activated by proteases to active cores of 65 to 70 kDa. The crystal solubilization is facilitated by an alkaline pH. The typical midgut pH is between 9-11 in lepidopteran larvae. During the solubilization and activation of Cry1 proteins, an N-terminal peptide of 25-30 amino acids and approximately half of the sequences from the C-terminus are cleaved (Bravo et al., 2002; Choma et al., 1990; Gill et al., 1992; Schnepf et al., 1998; Zhuang and Gill, 2003). The role of the C-terminal domain is believed to be in the formation of crystalline inclusion bodies within the Bt bacterium and is not required for insecticidal activity (De Maagd et al., 2001; Park and Federici, 2000). The 25-30 amino acid residues at the N-terminus play a role in promoting crystallization of the protoxin in the bacterium, but do not contribute to toxicity to insects (Choma et al., 1990; Gill et al., 1992; Schnepf et al., 1998). In fact, it was shown with Cry1Ac that proteolytic removal of the N-terminal peptide is essential before the protein becomes fully active (Bravo et al., 2002).

The 3-dimensional structures of three members of the Cry protein family, which may well prove to be representative of all Cry proteins, reveal the presence of three structural domains (Grochulski et al., 1995; Li et al., 1991; Morse et al., 2001). Domain I, consisting of seven α-helices, is involved in membrane insertion and pore formation. Domain II, consisting of three β-sheets in a Greek key conformation, is involved in specific receptor recognition and binding. Domain III, which consists of two β-sheets in a jellyroll conformation, has been suggested to maintain the structural integrity of the protein molecule (Li et al., 1991) and also to contribute to specificity (De Maagd et al., 2001; De Maagd et al., 2000). All three domains are included in the N-terminal portion of the protoxins during the formation of active toxins in the insect gut. Since domains II and III can both contribute to the specificity, the difference in these domains among different Cry proteins would account for the diversity of insecticidal activities. These domains may dictate whether and how binding occurs between the Cry proteins and the insect midgut. Only those insects with specific receptors are affected by Cry proteins and no toxicity is observed in species that lack these receptors (Crickmore et al., 1998; De Maagd et al., 2001).

Studies previously have been conducted to evaluate the spectrum of insecticidal activity of Cry1Ac protein produced from Bt subsp. *kurstaki* HD-73 against a variety of agronomically important insects and one non-insect arthropod taxon (Luttrell et al., 1999; MacIntosh et al., 1990). Species tested included seven species of *Lepidoptera*: beet armyworm (*Spodoptera exigua*), black cutworm (*Agrotis ipsilon*), cabbage looper (*Trichoplusia ni*), corn earworm (*Helicoverpa zea*), European corn borer (*Ostrinia nubilalis*), tobacco budworm (*Heliothis virescens*), and tobacco hornworm (*Manduca sexta*); five species of *Coleoptera*: alfalfa weevil (*Hypera postica*), cotton boll weevil (*Anthonomis grandis*), horseradish flea beetle (*Phyllotreta armoraciae*), southern corn rootworm (*Diabrotica undecimpunctata howardi*), and Japanese beetle (*Popillia japonica*); one species of *Diptera*: yellow fever mosquito (*Aedes aegypti*); one species of *Blattodea*: German cockroach (*Blatella germanica*); one species of *Hemiptera*: green peach aphid (*Myzus persicae*); one species of *Isoptera*: termite (*Reticulitermes flavipes*); and one species of mite: two-spotted spider mite (*Tetranychus urticae*). The results showed that Cry1Ac had activity against all seven of the representative lepidopteran insects. However, there was no indication of Cry1Ac activity against any of the ten non-lepidopteran species (MacIntosh et al., 1990). The results from these assays suggest that the Cry1Ac protein has insecticidal activity against lepidopteran insect pests but not against the range of non-lepidopteran pests that were tested.

b) Proposed name the product will be marketed under

Soybean containing the transformation event MON 87701 will be initially commercialized in South America for the control of targeted lepidopteran pests, and in longer term possibly any other geographies that would benefit from this trait.
There are currently no plans to produce this product in Australia and New Zealand.
A commercial trade name for the product has not been determined at the time of this submission and will be available prior to commercial launch of the product.

c) Product list

Soybean is the most commonly grown oilseed in the world. In 2007/08, approximately 218.8 MMT of harvested soybean were produced, representing 56% of the world's oilseed production (ASA, 2008; Soya and Oilseed Bluebook, 2008). Soybean is grown as a commercial crop in over 35 countries. The major producers of soybean are the U.S., Brazil, Argentina, China, and India, which accounted for approximately 91% of the global soybean production in 2007. Approximately one-third of the 2007 world soybean production occurred in the U.S. (Soya and Oilseed Bluebook, 2008). The soybean produced in China and India are primarily for domestic use, while a significant portion of that produced in U.S., Brazil, and Argentina is traded globally in the form of soybean harvested seed, soybean meal or soybean oil. Globally, the U.S. was the largest soybean export country, while Argentina led the soybean meal and soybean oil export markets in 2007 (ASA, 2008; Soya and Oilseed Bluebook, 2008).

Soybean has the remarkable ability to produce more edible protein per acre of land than any other known crop (Liu, 2004a). On average, dry soybean contains roughly 40% protein and 20% oil. It has the highest protein content among cereals and other legume species, and has the second-highest oil content among all food legumes (Liu,

2004a). Soybean is highly versatile and can be processed into a wide variety of food products. In general, soyfoods can be roughly classified into four major categories (Liu, 2004b):

1. Traditional soyfoods: As discussed above, traditional soyfoods are primarily made from whole soybean. The nonfermented traditional soyfoods include soymilk, tofu, and soybean sprouts, whereas the fermented soyfoods include soybean paste (miso), soy sauce, natto, and tempeh.

2. Soybean oil: Soybean oil constitutes approximately 71% of global consumption of edible fats and oil (ASA, 2008), and is the second largest source of vegetable oil worldwide (Soyatech, 2009). Refined, bleached, and deodorized soybean can be further processed to produce cooking oils, shortening, margarine, mayonnaise, salad dressings, and a wide variety of products that are either based entirely on fats and oils or contain fat or oil as a principal ingredient.

3. Soybean protein products: Soybean protein products are made from defatted soybean flakes, and include soybean flour, soybean protein concentrate, and soybean protein isolate. Soybean flour has a protein content of approximately 50% and is used mainly as an ingredient in the bakery industry. Soybean protein concentrate has a protein content of approximately 70% and is used widely in the meat industry as a key ingredient of meat alternative products such as soybean burgers and meatless "meatballs." Soybean protein isolate has a protein content of 90%, and possesses many functional properties such as gelation and emulsification. As a result, it can be used in a wide range of food applications, including soups, sauce bases, energy bars, nutritional beverages, infant formula, and dairy replacements.

4. Dietary supplements: Soybean is a rich source of certain phytochemicals used as dietary supplements, which include isoflavones and tocopherols. Isoflavones have been shown to inhibit the growth of cancer cells, lower cholesterol levels, and inhibit bone resorption (Messina, 1999). Tocopherols have long been recognized as a classic free radical scavenging antioxidant whose deficiency impairs mammalian fertility. In addition, new biological activities have been reported for the desmethyl tocopherols, such as γ-tocopherol, to possess anti-inflammatory, antineoplastic, and natriuretic functions (Hensley et al., 2004; IFIC, 2005; IFIC, 2006; Schafer et al., 2003). Detailed reviews of soybean as functional foods have been published (IFIC, 2005; Liu, 2004a).

2.2 History of Use

a) Donor organism

The gene encoding Cry1Ac is derived from Bt subsp. *kurstaki*, a soil microorganism that is both ubiquitous and abundant in the environment (De Boer and B. Diderichsen, 1991). Sprays of sporulated Bt have a long history of safe use for pest control in agriculture (Cannon, 1993; EPA, 1988; WHO, 1999). Microbial pesticides containing Bt Cry proteins have been used for more than 50 years and subjected to extensive toxicity testing showing no adverse effects to human health (Betz et al., 2000; EPA, 2000; James, 2003). During the last decade, a variety of crops containing Cry1 proteins from Bt have been commercialized, thus rendering these plants resistant to several insect pests (De Maagd et al., 1999). For example, corn products that produce Cry1Ab (YieldGard® Corn Borer, Bt11) and Cry1F (TC1507) proteins, and cotton products that produce the Cry1Ac protein (Bollgard® and Bollard II ®) are currently registered and sold on the market (Mendelsohn et al., 2003). Moreover, DBT418 corn that produces the tryptic core of Cry1Ac was previously deregulated by USDA (USDA, 1997). There are no known reports of allergies to Bt or its expressed proteins.

b) Host organism

i) Taxonomic classification

Cultivated soybean, *Glycine max* (L.) Merr., is a diploidized tetraploid (2n=40), which belongs to the family Leguminosae, the subfamily Papilionoideae, the tribe Phaseoleae, the genus *Glycine* Willd., and the subgenus *Soja* (Moench) F.J. Herm.

Family: Leguminosae
Subfamily: Papilionoideae
Tribe: Phaseoleae
Genus: *Glycine*
Subgenus: *Soja* (Moench) F.J. Herm.
Species: *max*

The genus *Glycine* Willd. is of Asian and Australian origin and is divided into two subgenera, *Glycine* and *Soja* (Moench) F.J. Herm. The subgenus *Glycine* consists of 22 wild perennial species, which are indigenous to Australia, west, central and south Pacific Islands, China, Russia, Japan, Indonesia, Korea, Papua New Guinea, the Philippines, and Taiwan (Hymowitz, 2004). The subgenus *Soja* includes the cultivated soybean, *G. max* (L.) Merr. and its wild annual relatives from Asia, *G. soja* Sieb. and Zucc. The list of species in the genus *Glycine* Willd. is presented in **Table 1**.

®*YieldGard, Bollgard and Bollgard II are registered trademarks of Monsanto Technology LLC.*

Table 1. List of Species in the Genus Glycine Willd., 2n Chromosome Number, Genome Symbol and Distribution

Genus		2n	Genome[1]	Distribution
Subgenus *Glycine*				
1.	*G. albicans* Tind. & Craven	40	I1	Australia
2.	*G. aphyonota* B. Pfeil	40	--[2]	Australia
3.	*G. arenaria* Tind.	40	HH	Australia
4.	*G. argyrea* Tind.	40	A2A2	Australia
5.	*G. canescens* F.J. Herm.	40	AA	Australia
6.	*G. clandestina* Wendl.	40	A1A1	Australia
7.	*G. curvata* Tind.	40	C1C1	Australia
8.	*G. cyrtoloba* Tind.	40	CC	Australia
9.	*G. dolichocarpa* Tateishi and Ohashi	80	--	(Taiwan)
10.	*G. falcate* Benth.	40	FF	Australia
11.	*G. hirticaulis* Tind. & Craven	40	H1H1	Australia
		80	--	Australia
12.	*G. lactovirens* Tind. & Craven.	40	I1I1	Australia
13.	*G. latifolia* (Benth.) Newell & Hymowitz	40	B1B1	Australia
14.	*G. latrobeana* (meissn.) Benth.	40	A3A3	Australia
15.	*G. microphylla* (Benth.) Tind.	40	BB	Australia
16.	*G. peratosa* B. Pfeil & Tind.	40	--	Australia
17.	*G. pindanica* Tind. & Craven	40	H3H2	Australia
18.	*G. pullenii* B. Pfeil, Tind. & Craven	40	--	Australia
19.	*G. rubiginosa* Tind. & B. Pfeil	40	--	Australia
20.	*G. stenophita* B. Pfeil & Tind.	40	B3B3	Australia
21.	*G. tabacina* (Labill.) Benth.	40	B2B2	Australia
		80	Complex[3]	Australia, West Central and South Pacific Islands
22.	*G. tomentella* Hayata	38	EE	Australia
		40	DD	Australia, Papua New Guinea
		78	Complex[4]	Australia, Papua New Guinea
		80	Complex[5]	Australia, Papua New Guinea, Indonesia, Philippines, Taiwan
Subgenus *Soja* (Moench) F.J. Herm.				
23.	*G. soja* Sieb. & Zucc.	40	GG	China, Russia, Taiwan, Japan, Korea (Wild Soybean)
24.	*G. max* (L.) Merr.	40	GG	Cultigen (Soybean)

[1] Genomically similar species carry the same letter symbols.
[2] Genome designation has not been assigned to the species.
[3] Allopolyploids (A and B genomes) and segmental allopolyploids (B genomes).
[4] Allopolyploids (D and E, A and E, or any other unknown combination).
[5] Allopolyploids (A and D genomes, or any other unknown combination).
Note: Table is adapted from Hymowitz (Hymowitz, 2004).

ii) Characteristics of the recipient soybean

Glycine max (L.) Merr, the cultivated soybean, is an annual crop that is planted in late spring from April to May in the north hemisphere, and from November to February in the southern hemisphere. Soybean seed germinates when the soil temperature reaches 10°C and emerges in a 5-7 day period under favourable conditions (OECD, 2000). The vegetative development phase lasts about 40 days, during which time the root nodules develop slowly, but do not become fully functional. Soybeans grow most rapidly when air temperatures are between 25 and 35°C (Beversdorf, 1993). Pods typically develop in late summer, and harvest occurs in the autumn. The life cycle of soybean is approximately 100 to 160 days, depending on the variety and the region it is cultivated. Harvesting may begin when the plants are completely dry and the seeds are liberated within the pods.

The soybean variety used as the recipient for the DNA insertion to create MON 87701 was A5547, a non-transgenic conventional variety developed by Asgrow Seed Company. A5547 is an elite maturity group V soybean variety developed and selected on the basis of its superior agronomic performance over other soybean lines (Rhodes, 1997). As a soybean variety in maturity group V, A5547 is a determinate variety adapted and most suitable for production in the Mid-South region.

In developing the data to support the safety assessment of insect-protected soybean MON 87701, the conventional soybean variety A5547 was used as the comparator. In general, the genetic background of MON 87701 was similar to that of the control, so the effect of the genetic insertion and the presence of the Cry1Ac protein could be assessed in an unbiased manner. Since MON 87701 was derived from the A5547 conventional variety, it was deemed appropriate to use the non-transformed A5547 as the control variety as its use would minimize the potential bias in subsequent comparative assessments. In addition, commercial conventional and Roundup Ready soybean (40-3-2) varieties were used as reference materials to establish ranges of responses or values representative of commercial conventional soybean varieties. The reference varieties used at each location were selected based on their availability and agronomic fit.

iii) History of soybean

Domestication of soybean is thought to have taken place in China during the Shang dynasty (approximately 1500 to 1027 B.C.) or earlier (Hymowitz, 1970). However, historical and geographical evidence could only be traced back to the Zhou dynasty (1027 to 221 B.C.) where the soybean was utilized as a domesticated crop in the northeastern part of China. By the first century A.D., the soybean probably reached central and southern China as well as peninsular Korea. The movement of soybean germplasms was probably associated with the development and consolidation of territories and the degeneration of Chinese dynasties (Ho, 1969; Hymowitz, 1970).

From the first century A.D. to approximately the 15th and 16th centuries, soybean was introduced into several countries, with land races eventually developing in Japan, Indonesia, Philippines, Vietnam, Thailand, Malaysia, Myanmar, Nepal, and Northern India. The movement of soybean throughout this period was due to the establishment

of sea and land trade routes, the migration of certain tribes from China, and the rapid acceptance of seeds as a staple food by other cultures (Hymowitz and Newell, 1981; Hymowitz et al., 1990).

Starting in the late 16th century and throughout the 17th century, soybean was used by the Europeans, and in the 17th century, soybean sauce was a common item of trade from the East to the West.

Soybean was introduced into North America in the 18th century. In 1851, soybean was introduced in Illinois and subsequently throughout the Corn Belt. In 1853, soybean seed were deposited at the New York State Agricultural Society, the Massachusetts Horticultural Society, and the Commissioner of Patents. The two societies and the Commissioner of Patents sent soybean seed to dozens of growers throughout the U.S. Soybean has been extensively cultivated and improved through conventional breeding programs following its introduction in the U.S. and has become a key source of nutrients for food and feed use in the U.S. (Hymowitz and Singh, 1987).

2.3 Nature of Genetic Modification

a) Transformation method

MON 87701 was developed through *Agrobacterium*-mediated transformation of soybean meristem tissue utilizing transformation vector, PV-GMIR9 (**Figure 2**). PV-GMIR9 is a binary vector that contains well-characterized DNA segments required for selection and replication of the plasmid vector in bacteria and transfer of the T-DNAs into plant cells. Vector PV-GMIR9 contains two separate T-DNAs (hence the descriptor "2T-DNAs") that can be effectively used to generate marker-free plants (Komari et al., 1996). The first T-DNA, designated as T-DNA I, contains the gene cassette bearing the gene of interest *cry1Ac*, and the second T-DNA, designated as T-DNA II, contains the gene cassette of selectable marker gene *cp4 epsps*. During the process of *Agrobacterium*-mediated transformation, the distinct T-DNAs containing the *cry1Ac* and *cp4 epsps* genes were integrated into the soybean genome at independent, unlinked loci, and the rest of the backbone of the vector PV-GMIR9 was not inserted into plant cells. A technique used in traditional breeding was then used to isolate plants that only contain T-DNA I (*cry1Ac* expression cassette) but do not contain T-DNA II (*cp4 epsps* expression cassette). This resulted in the production of marker-free, insect-protected soybean MON 87701.

The *Agrobacterium*-mediated soybean transformation to produce MON 87701 was based on the method described by (Martinell et al., 2002), which allows the generation of transformed plants without utilization of callus. Briefly, meristem tissues were excised from the embryos of germinated A5547 seed. After co-culturing with the *Agrobacterium* carrying the vector, the meristems were placed on selection medium containing glyphosate to inhibit the growth of untransformed plant cells, and spectinomycin and chloramphenicol to inhibit the growth of excess *Agrobacterium* so that only cells containing T-DNA II and/or T-DNA I and T-DNA II survived. The absence of the *Agrobacterium* used for transformation was confirmed by PCR targeting backbone sequence of plasmid PV-GMIR9. The meristems then were placed in media conducive to shoot and root development. Rooted plants with normal phenotypic characteristics were selected and transferred to soil for growth and further assessment.

The R_0 plants generated through this process were self-pollinated to produce the R_1 seed. During the selfing of the R_0 plants to produce the R_1 seed, the unlinked insertions of T-DNA I (*cry1Ac* gene expression cassette) and T-DNA II (*cp4 epsps* gene expression cassette) were segregated. A non-lethal dose of glyphosate herbicide was applied to R_1 plants. The resulting plants with minor injury were selected for further analyses, whereas plants showing no injury, i.e., containing T-DNA II (*cp4 epsps* gene expression cassette), were eliminated from subsequent development. Subsequently, plants containing only a single T-DNA I (*cry1Ac* gene cassette) were identified and selected by a combination of analytical techniques, including ELISA and TaqMan PCR analysis. Only R_1 plants that were homozygous for the T-DNA I cassette and not having the T-DNA II cassette were advanced for development. These R_1 plants were self-pollinated to generate a population of R_2 plants which were repeatedly self-pollinated through subsequent generations. These progeny were subjected to further molecular assessments to ensure the plants contained a single,

intact insert and phenotypic assessments to ensure the plants met commercial specifications. MON 87701 was selected as the lead event based on its superior phenotypic characteristics and molecular profile. Additional tests on MON 87701 were initiated to further characterize the genetic insertion and the expressed Cry1Ac protein, and to confirm the food, feed, and environmental safety relative to conventional soybean. The major steps involved in the development of MON 87701 are depicted in **Figure 1**.

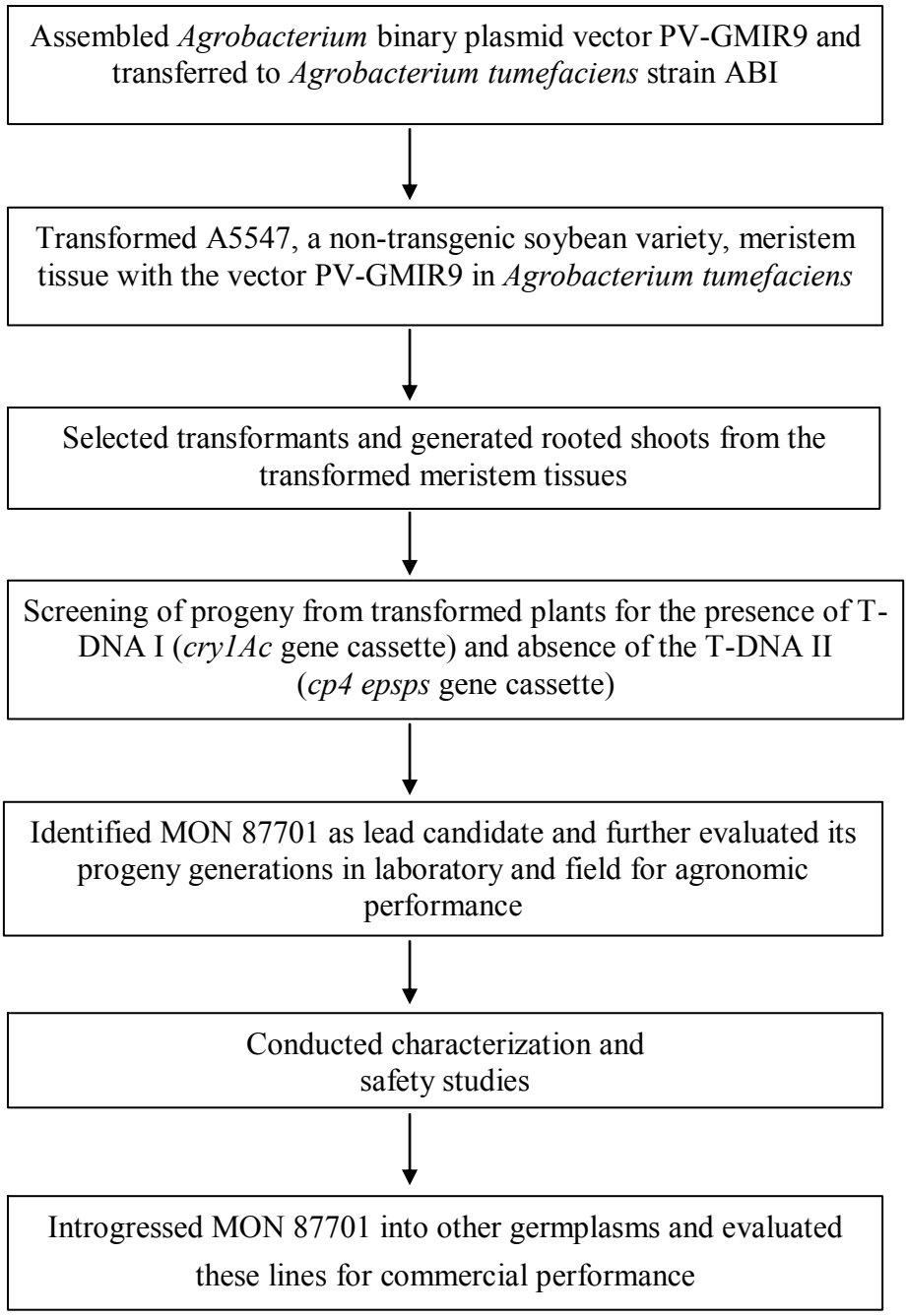


Figure 1. Schematic of the Development of MON 87701

b) Gene construct and transformation event

The vector PV-GMIR9, used for the transformation of soybean to produce MON 87701, is shown in **Figure 2** and its genetic elements are described in **Table 2**. This vector is approximately 15.5 kb and contains two T-DNAs delineated by left and right border regions. Each of the two T-DNAs contains a single expression cassette. The first T-DNA (designated as T-DNA I) contains the *cry1Ac* expression cassette, which results in the expression of the Cry1Ac protein. The *cry1Ac* expression cassette contains the *cry1Ac* coding sequence under the regulation of the *RbcS4* promoter and leader, *CTP1* chloroplast targeting sequence, and the *7S α′* 3' non-translated sequence. The second T-DNA (designated as T-DNA II) contains the *cp4 epsps* gene expression cassette. The *cp4 epsps* expression cassette contains the *cp4 epsps* coding sequence under the regulation of the *FMV* promoter, the *shkG* leader, the *CTP2* chloroplast targeting sequence and the *E9* 3' non-translated sequence. Utilizing a vector with two T-DNAs is the basis for an effective approach to generate marker-free plants. It allows for the T-DNA with the trait of interest (e.g., *cry1Ac,* T-DNA I) and the T-DNA encoding the selectable marker (e.g., *cp4 epsps*, T-DNA II) to insert into two independent loci within the genome of the plant. Following selection of the transformants, the inserted T-DNA encoding the selectable marker (e.g., T-DNA II) can be segregated from progeny through the use of traditional breeding techniques and genetic selection processes, while the inserted T-DNA containing the trait(s) of interest is maintained (e.g., T-DNA I). The result is a marker-free soybean containing only the *cry1Ac* expression cassette.

The backbone region outside of the T-DNAs contains two origins of replication for maintenance of plasmid in bacteria (OR-*ori V,* OR-*ori-pBR322*), a bacterial selectable marker gene (*aadA*), and a coding sequence for repressor of primer protein for maintenance of plasmid copy number in *E. coli* (*rop*). A description of the genetic elements and their prefixes (e.g., P-, L-, I-, TS-, OR-, B-, CS-, and T-) in PV-GMIR9 is provided in **Table 2**.

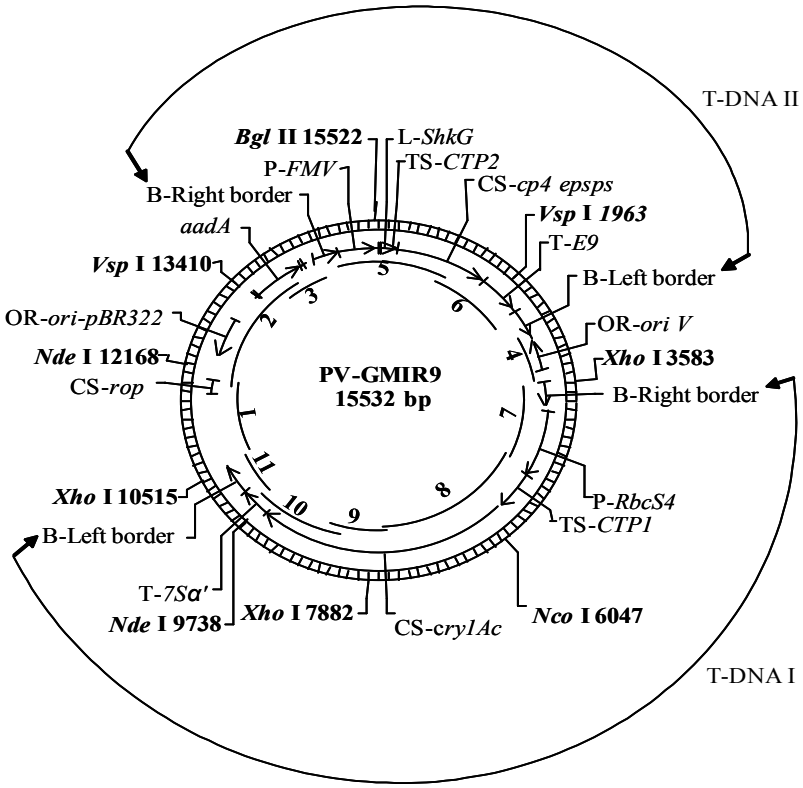


Figure 2. Genetic Elements and Restriction Sites of Vector PV-GMIR9 Used in Southern Blot Analysis (Probes 1-11)

Probe	DNA Probe	Start Position (bp)	End Position (bp)	Total Length (~kb)
1	Backbone Probe 1	10513	12013	1.5
2	Backbone Probe 2	11813	13640	1.8
3	Backbone Probe 3	13440	14549	1.1
4	Backbone Probe 4	2852	3595	0.74
5	T-DNA II Probe 5	14907	1375	2.0
6	T-DNA II Probe 6	1225	2409	1.2
7	T-DNA I Probe 7	3596	5596	2.0
8	T-DNA I Probe 8	5471	6971	1.5
9	T-DNA I Probe 9	6846	8046	1.2
10	T-DNA I Probe 10	7846	9650	1.8
11	T-DNA I Probe 11	9450	10512*	1.1

A circular map of the plasmid vector PV-GMIR9 used to develop MON 87701 is shown. Genetic elements and restriction sites used in Southern blot analyses (with positions relative to the size of the plasmid vector) are shown on the exterior of the map. The probes used in the Southern blot analyses are shown on the interior of the map. PV-GMIR9 contains two T-DNA regions designated as T-DNA I and T-DNA II. The left and right border regions of T-DNA II share 100% homology with those of T-DNA I and thus were not included in the T-DNA II analysis.

* Nucleotide 10512 is a vector backbone sequence.

Table 2. Summary of Genetic Elements in Plasmid Vector PV-GMIR9

Genetic Element	Location in Plasmid	Function (Reference)
T-DNA II (Continued from bp 15532)		
Intervening Sequence	1-14	Sequences used in DNA cloning
L[1]-*ShkG*	15-81	5' non-translated leader sequence from the *Arabidopsis ShkG* gene encoding EPSPS (Klee et al., 1987) that is involved in regulating gene expression
TS[2]-*CTP2*	82-309	Targeting sequence encoding the chloroplast transit peptide from the *ShkG* gene of *Arabidopsis thaliana* encoding EPSPS (Klee et al., 1987)
CS[3]-*cp4-epsps*	310-1677	Codon modified coding sequence of the *aroA* gene from *Agrobacterium sp.* strain CP4 encoding the CP4 EPSPS protein (Barry et al., 1997; Padgette et al., 1996)
Intervening Sequence	1678-1719	Sequences used in DNA cloning
T[4]-*E9*	1720-2362	3' non-translated sequence from *RbcS2* gene of *Pisum sativum* encoding the Rubisco small subunit, which functions to direct polyadenylation of the mRNA (Coruzzi et al., 1984)
Intervening Sequence	2363-2409	Sequences used in DNA cloning
B[5]-Left Border	2410-2851	DNA region from *Agrobacterium tumefaciens* containing the left border sequence used for transfer of the T-DNA (Barker et al., 1983a)
Vector Backbone		
Intervening Sequence	2852-2937	Sequences used in DNA cloning
OR[6]-*ori V*	2938-3334	Origin of replication from the broad host range plasmid RK2 for maintenance of plasmid in *Agrobacterium* (Stalker et al., 1981a)
Intervening Sequence	3335-3595	Sequences used in DNA cloning
T-DNA I		
B-Right Border	3596-3952	DNA region from *Agrobacterium tumefaciens* containing the right border sequence used for transfer of the T-DNA (Depicker et al., 1982a)
Intervening Sequence	3953-4061	Sequences used in DNA cloning

Table 2 (cont'd). Summary of Genetic Elements in Plasmid Vector PV-GMIR9

Genetic Element	Location in Plasmid	Function (Reference)
P^7-*RbcS4*	4062-5784	Promoter, leader, and 5' non-translated region of the *Arabidopsis thaliana RbcS4* gene encoding ribulose 1, 5-bisphosphate carboxylase small subunit 1A, (Krebbers et al., 1988). Promoter expresses in above ground tissues
TS-*CTP1*	5785-6048	Targeting sequence encoding the transit peptide of the *Arabidopsis RbcS4* encoding small subunit 1A transit peptide, from *Arabidopsis thaliana*, present to direct the *cry1Ac* protein to the chloroplast (Krebbers et al., 1988)
CS-*cry1Ac*	6049-9585	Codon-modified coding sequence of the Cry1Ac protein of *Bacillus thuringiensis* (Fischhoff and Perlak, 1995)
Intervening Sequence	9586-9594	Sequences used in DNA cloning
T-*7S α'*	9595-10033	3' region of the *Sphas1* gene of *Glycine max* encoding the 7S α' seed storage protein, β-conglycinin, including 35 nucleotides of the carboxyl terminal β - conglycinin coding region with the termination codon and the polyadenylation sequence (Schuler et al., 1982). The element functions to terminate transcription and direct polyadenylation of the mRNA.
Intervening Sequence	10034-10069	Sequences used in DNA cloning
B-Left Border	10070-10511	DNA region from *Agrobacterium tumefaciens* containing the left border sequence used for transfer of the T-DNA (Barker et al., 1983a)
Vector Backbone (Continued from bp 3595)		
Intervening Sequence	10512-11786	Sequences used in DNA cloning
CS-*rop*	11787-11978	Coding sequence for repressor of primer protein derived from ColE1 plasmid for maintenance of plasmid copy number in *E. coli* (Giza and Huang, 1989a)
Intervening Sequence	11979-12405	Sequences used in DNA cloning
OR-*ori-pBR322*	12406-12994	Origin of replication from pBR322 for maintenance of plasmid in *Escherichia coli* (Sutcliffe, 1978)
Intervening Sequence	12995-13524	Sequences used in DNA cloning

Table 2 (cont'd). Summary of Genetic Elements in Plasmid Vector PV-GMIR9

Genetic Element	Location in Plasmid	Function (Reference)
aadA	13525-14413	Bacterial promoter and coding sequence for an aminoglycoside-modifying enzyme, 3' (9)-O-nucleotidyltransferase from the transposon Tn7 (Fling et al., 1985) (GenBank accession) that confers spectinomycin and streptomycin resistance
Intervening Sequence	14414-14549	Sequences used in DNA cloning
T-DNA II (Continued from bp 2851)		
B-Right Border	14550-14906	DNA region from *Agrobacterium tumefaciens* containing the right border sequence used for transfer of the T-DNA (Depicker et al., 1982a)
Intervening Sequence	14907-14939	Sequences used in DNA cloning
P-*FMV*	14940-15503	Promoter for the 35S RNA from figwort mosaic virus (FMV) (Rogers, 2000) that directs transcription in most plant cells
Intervening Sequence	15504-15532	Sequences used in DNA cloning

L[1] -Leader; **TS**[2] - Targeting Sequence; **CS**[3] - Coding Sequence; **T**[4] - Transcription Termination Sequence; **B**[5] - Border; **OR**[6] **-** Origin of Replication; **P**[7] – Promoter.

i) T-DNA I

- **The *cry1Ac* Coding Sequence and Cry1Ac Protein**

MON 87701 expresses the insecticidal Cry1Ac protein from Bt subsp. *kurstaki*, which provides resistance to certain lepidopteran pests. Except for four additional amino acids at the N-terminus, the Cry1Ac produced in MON 87701 shares an amino acid identity of >99% with the Cry1Ac from Bt subsp. *kurstaki* and shares 100% identity with the Cry1Ac produced in Bollgard cotton. The four amino acids at the N-terminus are derived from the chloroplast targeting sequence. The deduced full-length amino acid sequence of the Cry1Ac protein is shown in **Figure 3**.

```
   1 MASSMLSSAT MVASPAQATM VAPFNGLKSS AAFPATRKAN NDITSITSNG GRVNCMQVWP
  61 PIGKKKFETL SYLPDLTDSG GRVNCMQAMD NNPNINECIP YNCLSNPEVE VLGGERIETG
 121 YTPIDISLSL TQFLLSEFVP GAGFVLGLVD IIWGIFGPSQ WDAFLVQIEQ LINQRIEEFA
 181 RNQAISRLEG LSNLYQIYAE SFREWEADPT NPALREEMRI QFNDMNSALT TAIPLFAVQN
 241 YQVPLLSVYV QAANLHLSVL RDVSVFGQRW GFDAATINSR YNDLTRLIGN YTDHAVRWYN
 301 TGLERVWGPD SRDWIRYNQF RRELTLTVLD IVSLFPNYDS RTYPIRTVSQ LTREIYTNPV
 361 LENFDGSFRG SAQGIEGSIR SPHLMDILNS ITIYTDAHRG EYYWSGHQIM ASPVGFSGPE
 421 FTFPLYGTMG NAAPQQRIVA QLGQGVYRTL SSTLYRRPFN IGINNQQLSV LDGTEFAYGT
 481 SSNLPSAVYR KSGTVDSLDE IPPQNNNVPP RQGFSHRLSH VSMFRSGFSN SSVSIIRAPM
 541 FSWIHRSAEF NNIIASDSIT QIPAVKGNFL FNGSVISGPG FTGGDLVRLN SSGNNIQNRG
 601 YIEVPIHFPS TSTRYRVRVR YASVTPIHLN VNWGNSSIFS NTVPATATSL DNLQSSDFGY
 661 FESANAFTSS LGNIVGVRNF SGTAGVIIDR FEFIPVTATL EAEYNLERAQ KAVNALFTST
 721 NQLGLKTNVT DYHIDQVSNL VTYLSDEFCL DEKRELSEKV KHAKRLSDER NLLQDSNFKD
 781 INRQPERGWG GSTGITIQGG DDVFKENYVT LSGTFDECYP TYLYQKIDES KLKAFTRYQL
 841 RGYIEDSQDL EIYSIRYNAK HETVNVPGTG SLWPLSAQSP IGKCGEPNRC APHLEWNPDL
 901 DCSCRDGEKC AHHSHHFSLD IDVGCTDLNE DLGVWVIFKI KTQDGHARLG NLEFLEEKPL
 961 VGEALARVKR AEKKWRDKRE KLEWETNIVY KEAKESVDAL FVNSQYDQLQ ADTNIAMIHA
1021 ADKRVHSIRE AYLPELSVIP GVNAAIFEEL EGRIFTAFSL YDARNVIKNG DFNNGLSCWN
1081 VKGHVDVEEQ NNQRSVLVVP EWEAEVSQEV RVCPGRGYIL RVTAYKEGYG EGCVTIHEIE
1141 NNTDELKFSN CVEEEIYPNN TVTCNDYTVN QEEYGGAYTS RNRGYNEAPS VPADYASVYE
1201 EKSYTDGRRE NPCEFNRGYR DYTPLPVGYV TKELEYFPET DKVWIEIGET EGTFIVDSVE
1261 LLLMEE.
```

Figure 3. Deduced Amino Acid Sequence of the CTP1 Targeting Sequence and the Full- Length Cry1Ac Protein Produced in MON 87701

The underlined amino acids represent the CTP1 targeting sequence (positions 1-88). The amino acid sequence of the full-length Cry1Ac protein produced in MON 87701 (positions 85-1266) consists of the deduced Cry1Ac amino acid sequence from PV-GMIR9 and the four additional amino acids in bold font (positions 85-88) derived from CTP1.

- **The *cry1Ac* regulatory sequences**

Each expression cassette contains regulatory sequences involved in the expression of the respective coding sequences. T-DNA I contains the *cry1Ac* expression cassette, which consists of the *cry1Ac* coding sequence under the regulation of the *RbcS4* promoter and leader, *CTP1* targeting sequence, and the *7S α′* 3' nontranslated sequence. The *RbcS4* promoter and leader are from the *Arabidopsis thaliana* ribulose 1,5-bisphosphate carboxylase small subunit 1A gene (Krebbers et al., 1988) and drives transcription of the *cry1Ac* gene in above-ground portions of the plant. The *CTP1* targeting sequence is the sequence encoding the transit peptide from the *Arabidopsis thaliana* small subunit 1A gene (Krebbers et al., 1988) and is present to direct the Cry1Ac protein to the chloroplast.

The *7S α′ 3′* non-translated region is from the *Glycine max* 7S seed storage protein gene (Schuler et al., 1982) and is present to terminate transcription and direct polyadenylation of the *CTP1- cry1Ac* transcript.

- **T-DNA borders**

Plasmid PV-GMIR9 contains right border and left border regions (**Figure 2** and **Table 2**) that were derived from *Agrobacterium tumefaciens* plasmids (Barker et al., 1983b; Depicker et al., 1982b). The border regions each contain a 24-25 bp sequence, called the "nick" site, which is the site of DNA exchange during transformation. The border regions delineate the T-DNA and are involved in their efficient transfer into the soybean genome. Because PV-GMIR9 is a two T-DNA vector, it contains two right border regions and two left border regions, where one set is for T-DNA I and the other set is for T-DNA II.

ii) T-DNA II

- **The *cp4 epsps* coding sequence and the CP4 EPSPS protein**

The *cp4 epsps* gene expression cassette is not present in MON 87701. The *cp4 epsps* gene expression cassette was used as a selectable marker during the transformation to produce MON 87701, but was segregated away by traditional breeding techniques at the R1 generation. The CP4 EPSPS protein confers tolerance to glyphosate and has been used safely and successfully in many Roundup Ready crops such as canola, corn, cotton, soybean, and sugar beet.

- **The *cp4 epsps* regulatory sequences**

T-DNA II contains the *cp4 epsps* expression cassette, which consists of the *cp4 epsps* coding sequence under the regulation of the Figwort Mosaic Virus (*FMV*) promoter, the *shk*G leader, the CTP2 targeting sequence and the *E9* 3' non-translated sequence. The *FMV* promoter is from the FMV 35S RNA gene (Rogers, 2000) and drives transcription of *cp4 epsps* in most plant cell types. The *shkG* leader is the 5' untranslated region (UTR) from the *Arabidopsis thaliana shkG* gene (encoding EPSPS) (Klee et al., 1987) and acts to enhance expression. The CTP2 targeting sequence is the sequence encoding the transit peptide from the *ShkG* gene of *Arabidopsis thaliana* (Klee et al., 1987) and is present to direct the CP4 EPSPS protein to the chloroplast. The *E9* non-translated region is the 3' non-translated sequence from the *RbcS2* gene of *Pisum sativum* (Coruzzi et al., 1984) and is present to direct polyadenylation of the *CTP2-cp4 epsps* transcript.

- **T-DNA borders**

The right and left borders are described under Section 2.3.b.i.

iii) Genetic elements outside of the T-DNA borders

Genetic elements that exist outside of the T-DNA borders are those that are essential for the maintenance and selection of the vector PV-GMIR9 in bacteria. The origin of replication OR-*ori V* is required for the maintenance of the plasmid in *Agrobacterium* (Stalker et al., 1981b) and is derived from the broad host plasmid RK2. The origin of replication OR-pBR322 is required for the maintenance of the plasmid in *E. coli* and is derived from the plasmid pBR322 (Sutcliffe, 1978). CS-*rop* is the coding sequence of the repressor of primer (ROP) protein and is necessary for the maintenance of plasmid copy number in *E. coli* (Giza and Huang, 1989b). The selectable marker *aadA* is a bacterial promoter and coding sequence for an enzyme from transposon Tn7 that confers spectinomycin and streptomycin resistance (Fling et al., 1985) in *E. coli* and *Agrobacterium* during molecular cloning. Because these elements are outside the border regions, they are not expected to be transferred into the soybean genome. The absence of the backbone sequence in MON 87701 has been confirmed by Southern blot analyses (see Section 2.3.c.ii.).

c) Molecular characterization

Southern blot and DNA sequence analyses were used to characterize the T-DNA insert in MON 87701. Southern analysis was used to assay the entire soybean genome for sequences derived from the transformation vector PV-GMIR9. The sequence analysis was used to determine the composition and intactness of the inserted DNA and to evaluate the region of the genomic DNA directly adjacent to the insert. The analyses were performed on the R5 generation, the same generation used to initiate the integration of MON 87701 into commercial germplasm (**Figure 12**). The Southern blot strategy was designed to ensure sufficient sensitivity while utilizing probes that span the entire transformation vector. A linear map depicting the restriction sites within the insert DNA sequence, as well as within the soybean genomic DNA immediately flanking the insert in MON 87701, is shown in **Figure 4**. A map of plasmid vector PV-GMIR9 annotated with the probes used in the Southern analysis is presented in **Figure 2**. The high level of sensitivity was demonstrated for each blot by including and detecting a $1/10^{th}$ genome equivalent of the positive control. The Southern blots were performed in a way to maximize the resolution of DNA fragments. Two restriction enzyme sets were specifically chosen to minimize the possibility that two DNA fragments could comigrate on the gel.

The DNA sequencing analyses complement the Southern blot analyses. Whereas Southern blots determined that MON 87701 contains T-DNA I-derived sequences at a single insertion site, sequencing of the insert and the flanking genomic DNA determined that T-DNA I inserted as predicted in MON 87701. Each genetic element is intact and the sequence of the insert matches the corresponding sequence in PV-GMIR9. In addition, genomic rearrangements at the insertion site were assessed by comparing the insert and flanking sequences in MON 87701 to the insertion site sequences in conventional soybean.

The results of these experiments are summarized in **Table 3**. The insert in MON 87701 matches the T-DNA sequence of PV-GMIR9 starting with the Right Border of T-DNA I and ending at Left Border of T-DNA I. The information and results derived from the molecular analyses were used to construct a linear map of the insert in MON 87701. This linear map, shown in **Figure 4** depicts restriction sites identified in the insert and the soybean genomic DNA flanking the insert, and provides information on the expected banding patterns and sizes of the DNA fragments after restriction enzyme digestions.

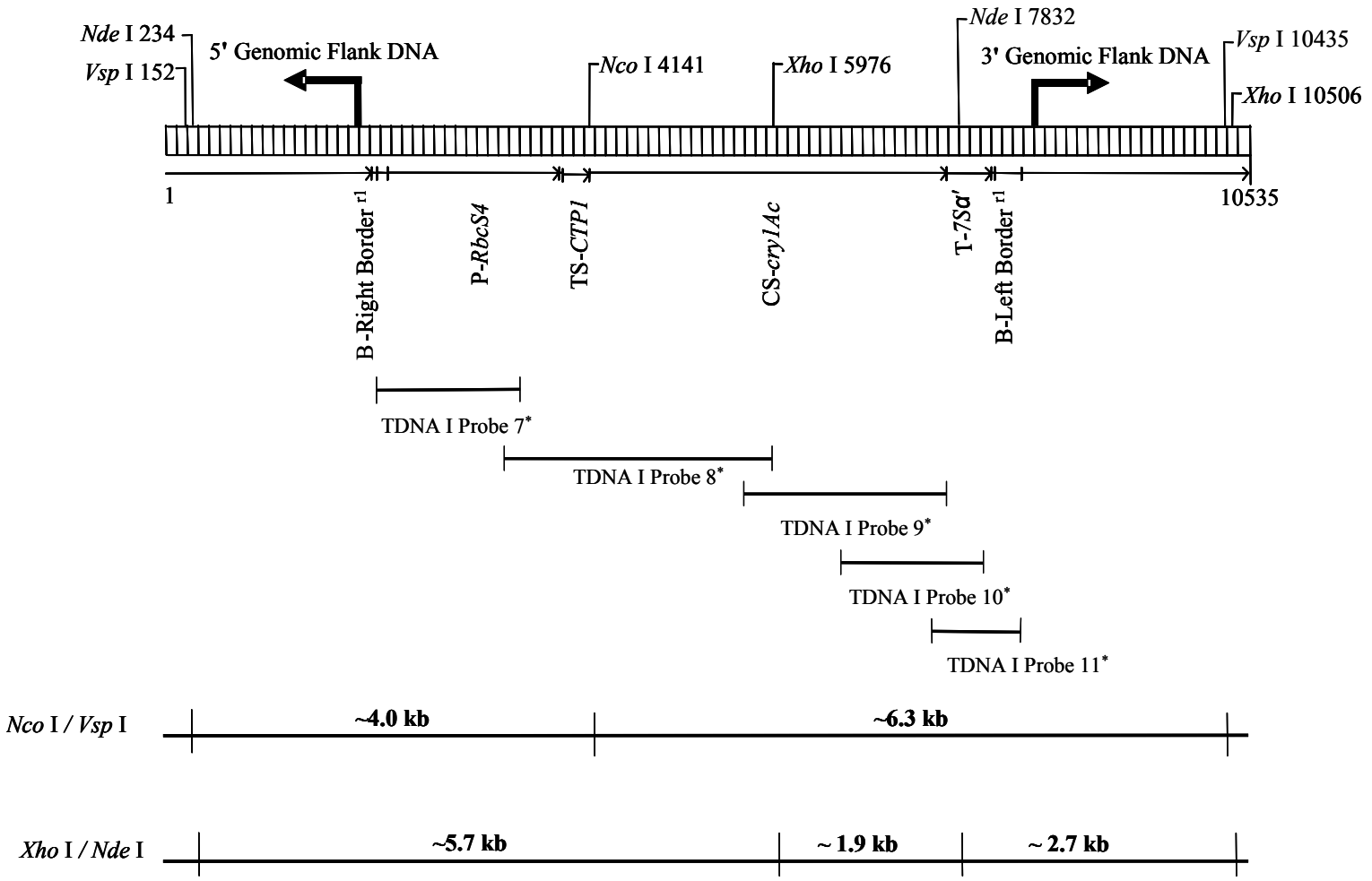


Figure 4. Schematic Representation of the Insert and Genomic Flanking Sequences in MON 87701

Linear DNA derived from T-DNA I of vector PV-GMIR9 incorporated into MON 87701. Arrows indicate the end of the insert and the beginning of soybean genomic flanking sequence. Identified on the map are genetic elements within the insert, as well as restriction sites with positions relative to the size of the linear map for enzymes used in the Southern analyses.
* These probes are not drawn to scale and are the estimated locations of the T-DNA I probes. Probes are described in **Figure 2**.

Table 3. Summary of Genetic Elements in MON 87701

Genetic Element	Location in Sequence	Function (Reference)
Sequence flanking 5' end of the insert	1-2000	Soybean genomic DNA
B[1]-Right Border	2001-2045	45 bp DNA region from the right border region remaining after integration (Depicker et al., 1982a)
Intervening Sequence	2046-2154	Sequence used in DNA cloning
P[2]- *RbcS4*	2155-3877	Promoter, leader, and 5' non-translated region of the *Arabidopsis thaliana RbcS4* gene encoding ribulose 1,5-bisphosphate carboxylase small subunit 1A (Krebbers et al., 1988)
TS[3]-*CTP1*	3878-4141	Targeting sequence encoding the transit peptide of the *Arabidopsis RbcS4* encoding small subunit 1A transit peptide, from *Arabidopsis thaliana*, present to direct the Cry1Ac protein to the chloroplast, (Krebbers et al., 1988)
CS[4]-*Cry1Ac*	4142-7678	Codon-modified coding sequence of the Cry1Ac protein of *Bacillus thuringiensis* (Fischhoff and Perlak, 1995)
Intervening Sequence	7679-7687	Sequences used in DNA cloning
T[5]-*7S α'*	7688-8126	3' region of the *Sphas1* gene of *Glycine max* encoding the 7S α' seed storage protein, β -conglycinin, including 35 nucleotides of the carboxyl terminalβ-conglycinin coding region with the termination codon and the polyadenylation sequence (Schuler et al., 1982). The element functions to terminate transcription and direct polyadenylation of the mRNA.
Intervening Sequence	8127-8162	Sequence used in DNA cloning
B[1]-Left Border	8163-8426	264 bp DNA region from the left border region remaining after integration (Barker et al., 1983a)
Sequence flanking 3' end of the insert	8427-10535	Soybean genomic DNA

[1]**B –** Border; [2]**P –** Promoter; [3]**TS –** Targeting Sequence; [4]**CS –** Coding Sequence; [5]**T –**3' non-translated transcriptional termination and polyadenylation signal sequences.

i) Insert and copy number

The number of copies and insertion sites of T-DNA I sequences in the soybean genome were evaluated by digesting the test and control genomic DNA samples with the two enzyme sets *Nco* I / *Vsp* I and *Xho* I / *Nde* I, which cleave within the insert and known flanking sequences. The enzymes used generate a restriction fragment containing T-DNA I and adjacent plant genomic DNA with a unique banding pattern. If T-DNA I sequences are present at a single integration site in MON 87701, then probing with the sequence from T-DNA I should result in the restriction fragments described in **Figure 4**. Any additional integration sites would be detected as additional bands. The blots were hybridized with overlapping T-DNA I probes spanning the entire inserted DNA sequence (Probes 7-11, **Figure 2**). Each Southern blot contained several controls. Genomic DNA isolated from the conventional soybean control, A5547, was used as a negative control to determine if the probes hybridized to any endogenous sequences. Conventional soybean spiked with either plasmid DNA or probe template was used as a positive hybridization control and to demonstrate sensitivity of the Southern blot. The results of these analyses are shown in **Figure 5** to **Figure 7**.

- **T-DNA I Probe 7**

Conventional soybean DNA digested with *Nco* I / *Vsp* I (**Figure 5**, lanes 1 and 8) or *Xho* I / *Nde* I (**Figure 5**, lanes 3 and 10) and hybridized with Probe 7 (**Figure 2**) showed no detectable hybridization bands as expected for the negative control. Conventional soybean DNA spiked with plasmid PV-GMIR9 DNA, previously digested with *Nco* I / *Bgl* II, (**Figure 2**, Probe 7) produced the expected bands at ~6.0 kb and ~9.5 kb (**Figure 5**, lane 7). In **Figure 5**, lane 6, the ~0.1 genome equivalent spike produced the expected band at ~6.0 kb, but the ~9.5 kb band is too faint to identify, since only a small portion of probe 7, which spans the Right Border region, has homology to the 9.5 kb portion of the vector. The ability to detect the spiked controls indicates that the probes are recognizing their target sequences.

MON 87701 DNA digested with *Nco* I / *Vsp* I (**Figure 5**, lanes 2 and 9) and hybridized to Probe 7 is expected to produce one band at ~ 4.0 kb. The long run (**Figure 5**, lane 2) produced a single band at ~4.1 kb (at or above the 4.1 kb marker) and the short run (**Figure 5**, lane 9) also produced a single band of the correct size. MON 87701 DNA digested with *Xho* I / *Nde* I (**Figure 5**, lanes 4 and 11) and hybridized with Probe 7 is expected to produce a single band of ~5.7 kb. The long run (**Figure 5**, lane 4) produced a single band at ~6.2 kb (at or above the 6.1 kb marker) and the short run (**Figure 5**, lane 11) produced a single band at ~5.7 kb. The apparent shift in migration of the bands in the long run versus the short run can be attributed to the method used to record the molecular weight markers on the agarose gel and on the autoradiograph and does not alter the conclusion that a single band was detected of the correct size. Thus, there is a single detectable insert containing Probe 7 sequences. The results presented in **Figure 5** indicate that the sequence covered by Probe 7 resides at a single detectable locus of integration in MON 87701.

- **T-DNA I Probes 8 and 10**

Conventional soybean DNA digested with *Nco* I / *Vsp* I (**Figure 6**, lanes 1 and 8) or *Xho* I / *Nde* I (**Figure 6**, lanes 3 and 10) and hybridized with Probes 8 and 10 (**Figure 2**) showed no detectable hybridization bands, as expected for the negative control. Pre-digested conventional soybean DNA spiked with probe template (**Figure 2**, Probes 8 and 10) generated from plasmid PV-GMIR9 produced the expected bands at ~1.5 kb, and ~1.8 kb (**Figure 6**, lanes 5 and 6). In lane 6, there is a faint band at ~3.6 kb, which likely represents a minor PCR artifact that was generated during probe template preparation. Conventional soybean DNA spiked with plasmid PV-GMIR9 DNA previously digested with *Nco* I / *Bgl* II produced the two expected bands at ~6.0 kb and ~9.5 kb (**Figure 6**, lane 7). Detection of the spiked controls indicates that the probes are recognizing their target sequences.

MON 87701 DNA digested with *Nco* I / *Vsp* I and hybridized with Probes 8 and 10 (**Figure 6**, lanes 2 and 9) produced two bands. The ~4.0 kb band is the expected size for the border fragment containing the 5' end of the inserted DNA (T-DNA I) and the adjacent genomic DNA flanking the 5' end of the insert (**Figure 4**). The ~6.3 kb band represents the 3' border fragment containing the 3' end of the inserted DNA and the adjacent genomic DNA flanking the 3' end of the insert. MON 87701 DNA digested with *Xho* I / *Nde* I (**Figure 6**, lanes 4 and 11) produced two bands. The ~5.7 kb band observed in **Figure 6** (lanes 4 and 11) is the expected size for the border fragment containing the 5' end of the inserted DNA (T-DNA I) and the adjacent genomic DNA flanking the 5' end of the insert (Figure IV-4). The ~1.9 kb band observed in **Figure 6** (lanes 4 and 11) represents the 3' border fragment containing the 3' end of the inserted DNA. The results presented in **Figure 6** indicate that sequence covered by Probes 8 and 10 resides at a single detectable locus of integration in MON 87701.

- **T-DNA I Probes 9 and 11**

Conventional soybean DNA digested with *Nco* I / *Vsp* I (**Figure 7**, lanes 2 and 11) or digested with *Xho* I / *Nde* I (**Figure 7**, lanes 4 and 13) and hybridized with Probes 9 and 11 (**Figure 2**) produced several hybridization signals. This was expected because the 7Sα' 3' non-translated region genetic element within T-DNA I originally was isolated from soybean. These hybridization signals result from the probes hybridizing to endogenous targets residing in the soybean genome and are not specific to the inserted DNA. These signals were produced in all lanes, including those containing the conventional soybean DNA material and therefore are considered to be endogenous background hybridization. Pre-digested conventional soybean DNA spiked with probe template (**Figure 2**, Probes 9 and 11) generated from plasmid PV-GMIR9, produced the expected bands at ~ 1.2 kb for probe template 9 (**Figure 7**, lanes 5 and 6) and ~1.1 kb for probe template 11 (**Figure 7**, lanes 7 and 8). Conventional soybean DNA spiked with plasmid PV-GMIR9 DNA previously digested with *Nco* I / *Bgl* II produced the two bands at ~6.0 kb and ~9.5 kb (**Figure 7**, lane 9). Detection of the spiked controls indicates that the probes are recognizing their target sequences. There is nonspecific hybridization at the bottom of the blot that spans lanes 5-13. This region of the blot corresponds to the

short run and genomic DNA was not that far in the gel. It is clear that no bands are discernable within this region of the blot.

MON 87701 DNA digested with *Nco* I / *Vsp* I and hybridized with Probes 9 and 11 (**Figure 7**, lanes 1 and 10) produced one unique band in addition to the endogenous background hybridization. The ~6.3 kb band is the expected size for the border fragment containing the 3' end of the inserted DNA (T-DNA I) and the adjacent genomic DNA flanking the 3' end of the insert (**Figure 4**). MON 87701 DNA digested with *Xho* I/ *Nde* I (**Figure 7**, lanes 3 and 12) produced three unique bands, as expected, in addition to the endogenous background hybridization. The expected band at ~5.7 kb migrated together with an endogenous hybridization signal observed in **Figure 7**, lanes 3 and 12. The ~5.7 kb band represents the 5' border fragment containing the 5' end of the inserted DNA along with the adjacent genomic DNA flanking the 5' end of the insert. The ~1.9 kb band represents a portion of the *cry1A*c expression cassette. The ~2.7 kb band represents the 3' border fragment containing the 3' end of the inserted DNA and the adjacent genomic DNA flanking the 3' end of the insert. The results presented in **Figure 7** indicate that sequences covered by Probes 9 and 11 reside at a single detectable locus of integration in MON 87701. Taken together, the data presented in **Figure 5**, **Figure** 6, and **Figure** 7 indicate that MON 87701 contains a single copy of T-DNA I at a single insertion site.

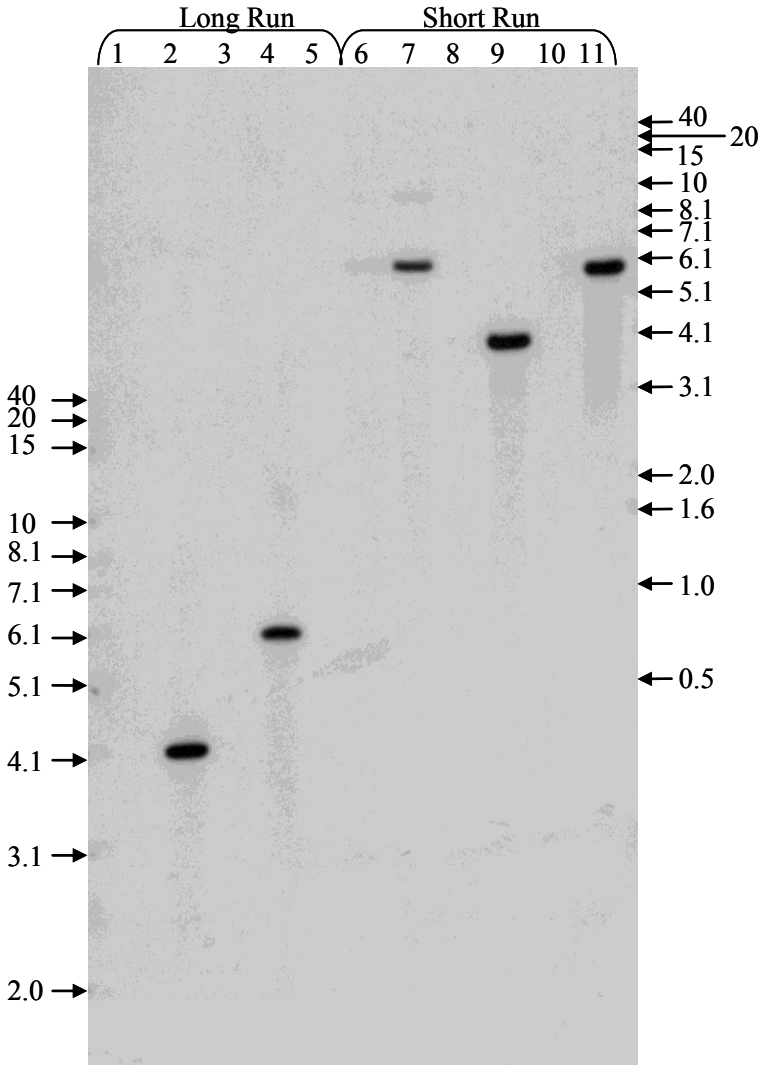


Figure 5. Southern Blot Analysis of MON 87701: T-DNA I Copy Number Analysis (Probe 7)

The blot was hybridized with one ^{32}P-labeled probe that spanned a portion of the T-DNA I sequence (**Figure 2**, Probe 7). Each lane contains ~10 µg of digested genomic DNA isolated from soybean leaf tissue. Lane designations are as follows:

Lane 1: Conventional soybean (*Nco* I / *Vsp* I)
2: MON 87701 (*Nco* I / *Vsp* I)
3: Conventional soybean (*Xho* I / *Nde* I)
4: MON 87701 (*Xho* I / *Nde* I)
5: Blank
6: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl* II) [~0.1 genome equivalent]
7: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl* II) [~1.0 genome equivalent]
8: Conventional soybean (*Nco* I / *Vsp* I)
9: MON 87701 (*Nco* I / *Vsp* I)
10: Conventional soybean (*Xho* I / *Nde* I)
11: MON 87701 (*Xho* I / *Nde* I)

→ Symbol denotes size of DNA, in kilobase pairs, obtained from MW markers on ethidium stained gel.

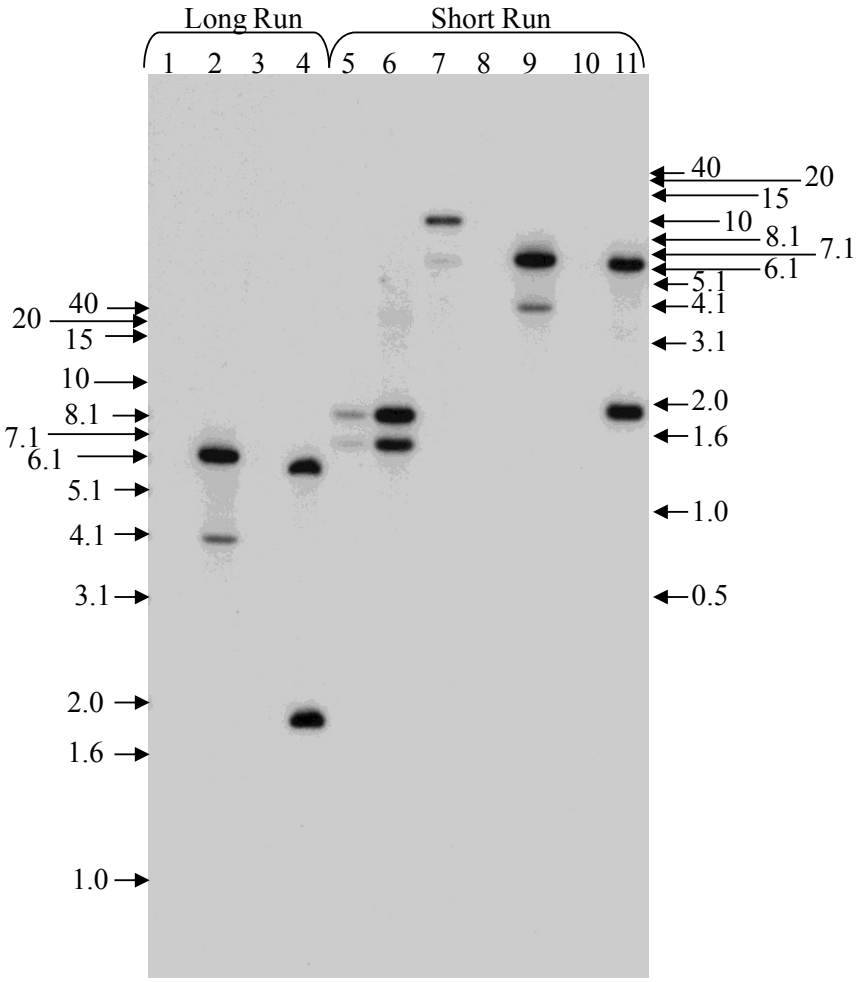


Figure 6. Southern Blot Analysis of MON 87701: T-DNA I Copy Number Analysis (Probes 8 and 10)

The blot was hybridized with two ^{32}P-labeled probes that spanned portions of the T-DNA I sequence (**Figure 2**, Probes 8 and 10). Each lane contains ~10 µg of digested genomic DNA isolated from soybean leaf tissue. Lane designations are as follows:

Lane 1: Conventional soybean (*Nco* I / *Vsp* I)
2: MON 87701 (*Nco* I / *Vsp* I)
3: Conventional soybean (*Xho* I / *Nde* I)
4: MON 87701 (*Xho* I / *Nde* I)
5: Conventional soybean spiked with probe templates (*Nco* I / *Vsp* I) [~0.1 genome equivalent]
6: Conventional soybean spiked with probe templates (*Nco* I / *Vsp* I) [~1.0 genome equivalent]
7: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl* II) [~1.0 genome equivalent]
8: Conventional soybean (*Nco* I / *Vsp* I)
9: MON 87701 (*Nco* I / *Vsp* I)
10: Conventional soybean (*Xho* I / *Nde* I)
11: MON 87701 (*Xho* I / *Nde* I)

→ Symbol denotes size of DNA, in kilobase pairs, obtained from MW markers on ethidium stained gel.

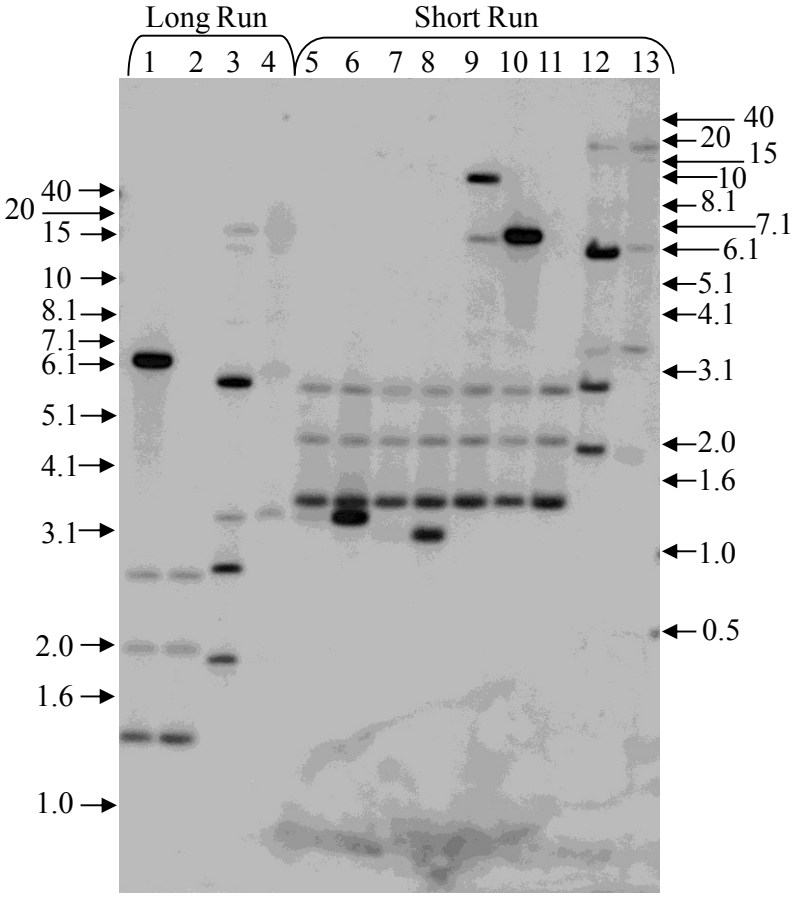


Figure 7. Southern Blot Analysis of MON 87701: T-DNA I Copy Number Analysis (Probes 9 and 11)

The blot was hybridized with two overlapping ^{32}P-labeled probes that spanned portions of the T-DNA I sequence (**Figure 2**, Probes 9 and 11). Each lane contains ~10 µg of digested genomic DNA isolated from soybean leaf tissue. Lane designations are as follows:

Lane 1: MON 87701 (*Nco* I / *Vsp* I)
2: Conventional soybean (*Nco* I / *Vsp* I)
3: MON 87701 (*Xho* I / *Nde* I)
4: Conventional soybean (*Xho* I / *Nde* I)
5: Conventional soybean spiked with probe 9 template (*Nco* I / *Vsp* I) [~0.1 genome equivalent]
6: Conventional soybean spiked with probe 9 template (*Nco* I / *Vsp* I) [~1.0 genome equivalent]
7: Conventional soybean spiked with probe 11 template (*Nco* I / *Vsp* I) [~0.1 genome equivalent]
8: Conventional soybean spiked with probe 11 template (*Nco* I / *Vsp* I) [~1.0 genome equivalent]
9: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl II*) [~1.0 genome equivalent]
10: MON 87701 (*Nco* I / *Vsp* I)
11: Conventional soybean (*Nco* I / *Vsp* I)
12: MON 87701 (*Xho* I / *Nde* I)
13: Conventional soybean (*Xho* I / *Nde* I)

→ Symbol denotes size of DNA, in kilobase pairs, obtained from MW markers on ethidium stained gel.

ii) Southern Blot Analysis to Determine the Presence or Absence of Plasmid PV-GMIR9 Backbone

The presence or absence of plasmid backbone sequences in the soybean genome was evaluated by digesting the test and control genomic DNA samples with *Nco* I / *Vsp* I or *Xho* / *Nde* I and hybridizing with backbone probes spanning the entire backbone sequence of PV-GMIR9 (**Figure 2**, Probes 1, 2, 3, and 4). If backbone sequences are present in MON 87701, then probing with backbone sequences should result in hybridizing bands. The results of this analysis are shown in **Figure 8** to **Figure 9**. Each Southern blot contains the same controls as described in Section 2.3.c.i.

- **Plasmid Backbone Probes 1, 2 and 3**

Conventional soybean control DNA was digested with *Nco* I / *Vsp* I (**Figure 8**, lanes 1 and 8) or *Xho* I / *Nde* I (**Figure 8**, lanes 3 and 10) and hybridized simultaneously with overlapping probes spanning most of the vector backbone of PV-GMIR9 (**Figure 2**, Probes 1, 2, and 3) showed no detectable hybridization bands, as expected for the negative control. Pre-digested conventional soybean DNA spiked with probe template (**Figure 2**, Probes 1, 2, and 3) generated from plasmid PV-GMIR9 produced the expected bands at ~1.5 kb, ~1.8 kb, and ~1.1 kb, respectively (**Figure 8**, lanes 5 and 6). Conventional soybean DNA spiked with plasmid PV-GMIR9 DNA previously digested with *Bgl* II / *Nco* I produced the expected size band of ~9.5 kb (**Figure 8**, lane 7). Detection of the spiked controls indicates that the probes are recognizing their target sequences.

MON 87701 DNA digested with *Nco* I / *Vsp* I (**Figure 8**, lanes 2 and 9) or *Xho* I / *Nde* I (**Figure 8**, lanes 4 and 11) and hybridized with Probes 1, 2, and 3 produced no detectable bands. There is a diffuse area of hybridization that overlaps with lane 4. Because this hybridization is not a distinct band, nor is it present in lane 11, which contains the same enzyme set, this area of hybridization is considered non-specific binding. The data indicate that MON 87701 contains no backbone elements from PV-GMIR9 that overlaps Probes 1, 2, and 3.

- **Plasmid Backbone Probe 4**

Conventional soybean control DNA digested with *Nco* I / *Vsp* I (**Figure 9**, lanes 1 and 7) or *Xho* I / *Nde* I (**Figure 9**, lanes 3 and 9) and hybridized with Probe 4 from the vector backbone of PV-GMIR9 (**Figure 2**, Probe 4) showed no detectable hybridization bands as expected for the negative control. Conventional soybean DNA spiked with plasmid PV-GMIR9 DNA previously digested with *Bgl* II / *Nco* I produced the expected band at ~6.0 kb (**Figure 9**, lanes 5 and 6). Detection of the spiked controls indicates that the probes are recognizing their target sequences.

MON 87701 DNA digested with *Nco* I / *Vsp* I (**Figure 9**, lanes 2 and 8) or *Xho* I / *Nde* I (**Figure 9**, lanes 4 and 10) and hybridized with Probe 4 produced no detectable hybridization bands, indicating that MON 87701 contains no detectable PV-GMIR9 backbone elements that are contained within Probe 4. These data, in combination with

the data presented in Section.2.3.c.ii. indicate that MON 87701 contains no detectable PV-GMIR9 backbone elements.

Figure 8. Southern Blot Analysis of MON 87701: PV-GMIR9 Backbone Sequence (Probes 1, 2, and 3)

The blot was hybridized with three overlapping ^{32}P-labeled probes that spanned a portion of the PV-GMIR9 backbone sequence (**Figure 2**, Probes 1, 2, and 3). Each lane contains ~10 µg of digested genomic DNA isolated from soybean leaf tissue. Lane designations are as follows:

Lane 1: Conventional soybean (*Nco* I / *Vsp* I)
2: MON 87701 (*Nco* I / *Vsp* I)
3: Conventional soybean (*Xho* I / *Nde* I)
4: MON 87701 (*Xho* I / *Nde* I)
5: Conventional soybean spiked with probe templates (*Nco* I / *Vsp* I) [~0.1 genome equivalent]
6: Conventional soybean spiked with probe templates (*Nco* I / *Vsp* I) [~1.0 genome equivalent]
7: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl* II) [~1.0 genome equivalent]
8: Conventional soybean (*Nco* I / *Vsp* I)
9: MON 87701 (*Nco* I / *Vsp* I)
10: Conventional soybean (*Xho* I / *Nde* I)
11: MON 87701 (*Xho* I / *Nde* I)

→ Symbol denotes size of DNA, in kilobase pairs, obtained from MW markers on ethidium stained gel.

Figure 9. Southern Blot Analysis of MON 87701: PV-GMIR9 Backbone Sequence (Probe 4)

The blot was hybridized with one ^{32}P-labeled probe that spanned a portion of the PV-GMIR9 backbone sequence (**Figure 2**, Probe 4). Each lane contains ~10 µg of digested genomic DNA isolated from soybean leaf tissue. Lane designations are as follows:

Lane 1: Conventional soybean (*Nco* I / *Vsp* I)
2: MON 87701 (*Nco* I / *Vsp* I)
3: Conventional soybean (*Xho* I / *Nde* I)
4: MON 87701 (*Xho* I / *Nde* I)
5: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl* II) [~0.1 genome equivalent]
6: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl* II) [~1.0 genome equivalent]
7: Conventional soybean (*Nco* I / *Vsp* I)
8: MON 87701 (*Nco* I / *Vsp* I)
9: Conventional soybean (*Xho* I / *Nde* I)
10: MON 87701 (*Xho* I / *Nde* I)

→ Symbol denotes size of DNA, in kilobase pairs, obtained from MW markers on ethidium stained gel.

iii) Southern Blot Analysis to Determine the Presence or Absence of T-DNA II

The presence or absence of T-DNA II sequences in the soybean genome was evaluated by digesting the test and control genomic DNA samples with the *Nco* I / *Vsp* I or *Xho* I / *Nde* I enzyme sets and hybridizing with overlapping T-DNA II probes spanning the entire T-DNA II sequence of PV-GMIR9 (**Figure 2**, Probes 5 and 6). The left and right border regions of T-DNA II share 100% homology with those of T-DNA I and thus were not included in the T-DNA II analysis. If T-DNA II sequences are present in MON 87701, then probing with the T-DNA II sequences should result in hybridizing bands. The results of this analysis are shown in **Figure 10**. The Southern blot contained the same controls as described in Section 2.3.c.i.

- **T-DNA II Probes 5 and 6**

Conventional soybean DNA digested with *Nco* I / *Vsp* I (**Figure 10**, lanes 1 and 9) or *Xho* I / *Nde* I (**Figure 10**, lanes 3 and 11) and hybridized with Probes 5 and 6 (Figure IV-2) showed no detectable hybridization bands, as expected for the negative control. Pre-digested conventional soybean DNA spiked with probe template generated from plasmid PV-GMIR9 produced the expected bands at ~2.0 kb and ~1.2 kb (**Figure 10**, lanes 6 and 7). Conventional soybean DNA spiked with plasmid PV-GMIR9 DNA previously digested with *Bgl* II/ *Nco* I produced the expected size bands of ~6.0 kb and ~9.5 kb (**Figure 10**, lane 8). Detection of the spiked controls indicates that the probes are recognizing their target sequences.

MON 87701 DNA digested with *Nco* I/ *Vsp* I (**Figure 10**, lanes 2 and 10) or *Xho* I / *Nde* I (**Figure 10**, lanes 4 and 12) produced no hybridization bands. These results indicate that MON 87701 contains no detectable T-DNA II elements.

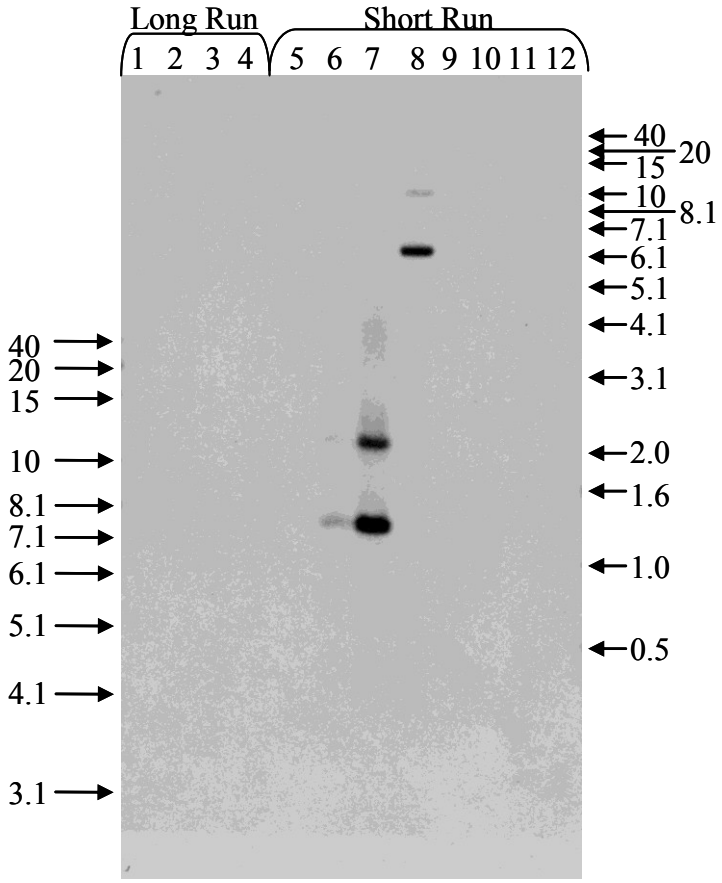


Figure 10. Southern Blot Analysis of MON 87701: T-DNA II (Probes 5 and 6)

The blot was hybridized with overlapping ^{32}P-labeled probes that spanned the T-DNA II sequence (**Figure 2**, Probes 5 and 6). Each lane contains ~10 µg of digested genomic DNA isolated from soybean leaf tissue. Lane designations are as follows:

Lane 1: Conventional soybean (*Nco* I / *Vsp* I)
2: MON 87701 (*Nco* I / *Vsp* I)
3: Conventional soybean (*Xho* I / *Nde* I)
4: MON 87701 (*Xho* I / *Nde* I)
5: Blank
6: Conventional soybean spiked with probe templates (*Nco* I / *Vsp* I) [~0.1 genome equivalent]
7: Conventional soybean spiked with probe templates (*Nco* I / *Vsp* I) [~1.0 genome equivalent]
8: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl II*) [~1.0 genome equivalent]
9: Conventional soybean (*Nco* I / *Vsp* I)
10: MON 87701 (*Nco* I / *Vsp* I)
11: Conventional soybean (*Xho* I / *Nde* I)
12: MON 87701 (*Xho* I / *Nde* I)

→ Symbol denotes size of DNA, in kilobase pairs, obtained from MW markers on ethidium stained gel.

iv) Organization and Sequence of the Insert and Adjacent Genomic DNA in MON 87701

The organization of the elements within the MON 87701 insert was confirmed by DNA sequence analyses. Several PCR primers were designed with the intent to amplify nine overlapping regions of DNA that span the entire length of the insert (see **Figure 11**). The amplified DNA fragments were subjected to DNA sequencing analyses. The DNA sequence of the MON 87701 insert is 6426 base pairs long, beginning at base 3908 of PV-GMIR9 located in the right border region and ending at base 10333 in the left border region of PV-GMIR9. A set of primers designed from the 5′ and 3′ DNA flanking sequence of MON 87701 was used to amplify and characterize the insertion site in the conventional soybean control, A5547. A sequence comparison between the PCR product generated from the conventional soybean and the sequence generated from the 5' and 3' flanking sequences of MON 87701 indicates there was a 32 bp deletion of soybean genomic DNA and a 14 bp insertion just 5' to the MON 87701 insertion site. This molecular rearrangement likely occurred in the plant during the *Agrobacterium*-mediated transformation process (Salomon and Puchta, 1998). This analysis confirms that the DNA sequences flanking the 5' and 3' ends of the insert in MON 87701 are native to the soybean genome and that no major unexpected rearrangements occurred during the transformation to produce MON 87701. Results also confirm that the arrangement of the genetic elements in MON 87701 is identical to that in plasmid PV-GMIR9 and is depicted in **Figure 4** and **Table 2**.

In **Figure 11**, the control reactions containing no template DNA (lanes 4, 8, 11, 16, 20, 25, 28, 32, and 36), and all but one of the conventional control reactions (lanes 2, 6, 9, 13, 17, 22, 26, and 34) did not generate PCR products with any of the primer sets, as expected. The conventional control reaction in lane 30 produced a product of equal size to MON 87701 (lane 31) because the primer sequences for this product are both located in the flanking genomic sequence adjacent to the 3' end of the insert in MON 87701. Plasmid PV-GMIR9 was used as a positive control template (lanes 15, 19, and 24) and produced the expected size PCR products of ~1.2 kb, 2.0 kb, and 2.0 kb, respectively, for each reaction. PCR reactions using genomic DNA from MON 87701 produced the expected size products: ~2.2 kb for Product A (lane 3); ~1.6 kb for Product B (lane 7); ~1.7 kb for Product C (lane 10); ~1.2 kb for Product D (lane 14); ~2.0 kb for Product E (lane 18); ~2.0 kb for Product F (lane 23); ~2.0 kb for Product G (lane 27); ~1.4 kb for Product H (lane 31); and ~2.9 kb for Product I (lane 35). These overlapping PCR products confirm that the organization of the insert is as expected.

To determine the sequence of the insert in MON 87701 and genomic DNA flanking the insert, the PCR Products A-I (**Figure 11**) were subjected to DNA sequencing. The consensus sequence representing the insert in MON 87701, including the genomic DNA flanking the ends of the insert, is shown in the CBI **Appendix 1** and is described in **Table 3**. This consensus sequence was generated by compiling numerous sequencing reactions using PCR products spanning the length of the insert and the 5' and 3' junctions with the flanking soybean genomic DNA. The amplification and sequencing of the insert and flanking DNA from MON 87701 establish that the arrangement and linkage of elements

in the insert are consistent with those in plasmid PV-GMIR9 and are as depicted in **Figure 4**.

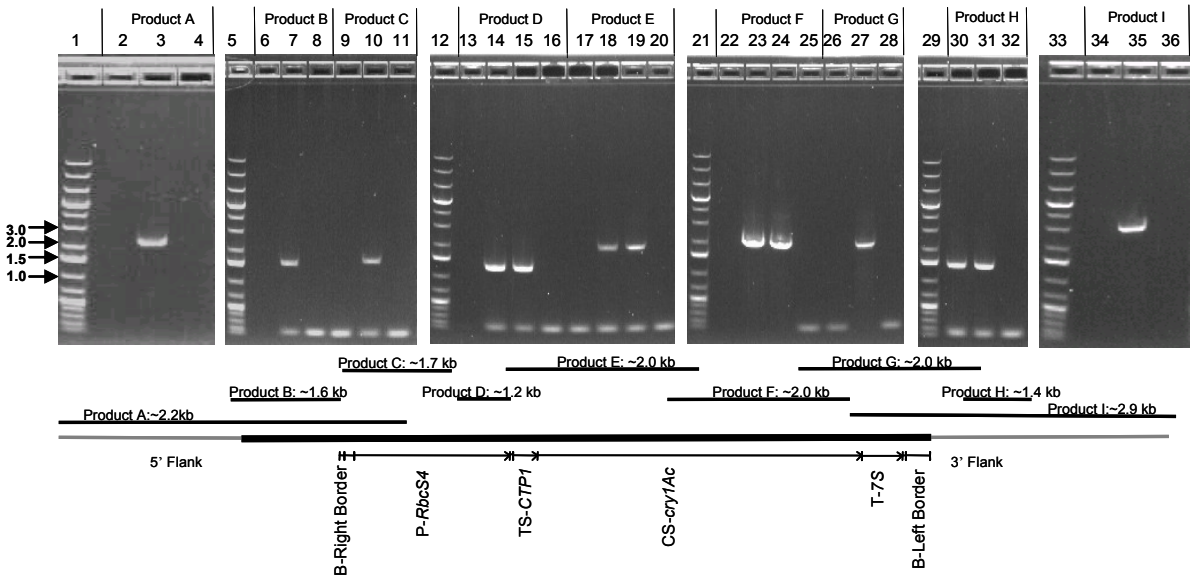


Figure 11. Overlapping PCR Analysis across the Insert in MON 87701

PCR analyses demonstrating the linkage of the individual genetic elements within the insert in MON 87701 were performed on MON 87701 genomic DNA extracted from leaf tissue (Lanes 3, 7, 10, 14, 18, 23, 27, 31 and 35). Lanes 2, 6, 9, 13, 17, 22, 26, 30, and 34 contain reactions with conventional soybean control DNA extracted from leaf tissue. Lanes 4, 8, 11, 16, 20, 25, 28, 32, and 36 are reactions containing no template DNA. Lanes 15, 19, and 24 contain reactions with PV-GMIR9 control DNA. Lanes 1, 5, 12, 21, 29, and 33 contain Fermentas GeneRuler™ 1 kb Plus DNA Ladder. Only lanes containing PCR reactions are shown in the figure. Lanes are marked to show which product has been loaded and is visualized on the agarose E-gel. The expected product size for each amplicon is provided in the illustration of the insert in MON 87701 that appears at the bottom of the figure. Three to six µl of each of the PCR products was loaded on the gel. PCR amplicons reported in this figure were not necessarily used in sequencing, but are representative of the study data.

Lanes	Lanes	Lanes	Lanes
1.) GeneRuler™ 1 kb Plus DNA Ladder	10.) MON 87701 genomic DNA	19.) PV-GMIR9 control DNA	28.) No template DNA control
2.) Conventional soybean control DNA	11.) No template DNA control	20.) No template DNA control	29.) GeneRuler™ 1 kb Plus DNA Ladder
3.) MON 87701 genomic DNA	12.) GeneRuler™ 1 kb Plus DNA Ladder	21.) GeneRuler™ 1 kb Plus DNA Ladder	30.) Conventional soybean control DNA
4.) No template DNA control	13.) Conventional soybean control DNA	22.) Conventional soybean control DNA	31.) MON 87701 genomic DNA
5.) GeneRuler™ 1 kb Plus DNA Ladder	14.) MON 87701 genomic DNA	23.) MON 87701 genomic DNA	32.) No template DNA control
6.) Conventional soybean control DNA	15.) PV-GMIR9 control DNA	24.) PV-GMIR9 control DNA	33.) GeneRuler™ 1 kb Plus DNA Ladder
7.) MON 87701 genomic DNA	16.) No template DNA control	25.) No template DNA control	34.) Conventional soybean control DNA
8.) No template DNA control	17.) Conventional soybean control DNA	26.) Conventional soybean control DNA	35.) MON 87701 genomic DNA
9.) Conventional soybean control DNA	18.) MON 87701 genomic DNA	27.) MON 87701 genomic DNA	36.) No template DNA control

→ Symbol denotes size of DNA, in kilobase pairs, obtained from MW markers on ethidium stained gel

v) Insert junction open reading frame analysis

While there was no indication of additional polypeptides produced from the MON 87701 insert other than the Cry1Ac protein, analyses of putative polypeptides encoded by DNA spanning the 5′ and 3′ junctions of the MON 87701 inserted DNA were performed using a bioinformatic comparison strategy. The purpose of the assessment is to evaluate the potential for novel open reading frames (ORFs) that may have concerns for similarity to known allergens and toxins. DNA sequence spanning the 5′ and 3′ junctions of the MON 87701 insertion site was analyzed for translational stop codons (TGA, TAG, TAA), and all ORFs originating or terminating within the MON 87701 insertion site were translated using the standard genetic code from stop codon to stop codon. Five sequences of eight amino acids or greater in length spanning the 5′ junction, and four sequences of eight amino acids or greater in length spanning the 3′ junction were considered as putative polypeptides and used as search sequences for FASTA comparisons against the AD_2009, TOX_2009 and PRT_2009 databases. In addition, the nine sequences were searched for eight amino sequences that match proteins in the AD_2009 database, using a sliding eight-amino acid window search.

Results of the FASTA sequence alignments demonstrated a lack of structurally relevant similarity between any known allergens, toxins, or bioactive proteins and the nine putative polypeptides. Results from the eight amino acid search demonstrated the lack of immunologically relevant matches between any of the putative polypeptides and the AD_2009 database. Bioinformatic analyses performed using the nine query sequences support the conclusion that even in the highly unlikely event that any of the putative junction polypeptides were translated, they would not share a sufficient degree of sequence similarity with known allergens or toxins. Therefore, there is no evidence for concern regarding health implications of the cross-junction putative polypeptides in MON 87701.

vi) Assessment of open reading frames contained in the cry1Ac coding sequence

Although DNA replication, DNA transcription and mRNA translation are of extremely high fidelity, mutation may in certain rare circumstances lead to the potential translation of mRNA on reading frames other than those defined by the intended translation start codon. In such instances, a novel protein may be produced. Due to the spontaneous nature of mutations, it is not possible to determine when or where in a coding sequence such an event may occur. In order to assess potential risks, bioinformatic analyses were performed to assess the potential of toxicity, allergenicity or biological activity of the putative peptides encoded by translation of reading frames 2 through 6 of the *cry1Ac* coding sequences. The methodology involves translation of two reading frames from the sense strand (frames 2 and 3) and three frames from the reverse complement (anti-sense) strand (frames 4, 5 and 6). Frame 1 corresponds to the *cry1Ac* coding sequence whose bioinformatics assessment will be described in Section 2.5.h.ii. Frames 2 and 3 are alternate reading frames that correspond to peptides derived from the chloroplast targeting peptide-*cry1Ac* coding sequence, beginning with nucleotide 2 or 3, respectively, through to the final nucleotide. Frames 4, 5 and 6 represent the anti-sense strand of the

chloroplast targeting peptide-*cry1Ac* coding sequence that were translated beginning with nucleotide 1, 2 or 3, respectively.

Translated sequences were compared to allergen (AD_2009), toxin (TOX_2009), and public domain (PRT_2009) sequence databases using the FASTA sequence comparison algorithm and significance criteria as described in Section 2.5.h.ii. The allergen, gliadin, and glutenin sequence database (AD_2009) was obtained from FARRP (2009), which was used as provided and contains 1,386 sequences. GenBank protein database, release 169.0 (December 16, 2008), was downloaded from NCBI and formatted for use in these bioinformatic analyses. The resulting database, referred to as the PRT_2009 database, contains 14,717,352 sequences. The toxin database is a subset of sequences derived from the PRT_2009 database that was selected using a keyword search and filtered to remove likely non-toxin proteins. The TOX_2009 database and contains 7,651 sequences.

The results of the search comparisons showed that no relevant alignments were observed against proteins in the AD_2009 database because none of the AD_2009 alignments with frames 2-6 of *cry1Ac* coding sequence met or exceeded the minimum threshold of 35% identity over 80 amino acids (Codex Alimentarius, 2003) or displayed *E*-scores less than 1e-5. Likewise, no FASTA alignments with the TOX_2009 database displayed *E*-scores less than 1e-5. When used as queries for a FASTA search the PRT_2009 database, frames 2 and 5 yielded alignments that displayed *E*-scores less than 1e-5. Inspection of the frame 2 alignments revealed that they were punctuated with numerous stop codons in the query sequence and required numerous gaps to optimize the alignment. As a result, it is unlikely these alignments reflect conserved structure. For those alignments with frame 5 yielding an *E*-score less than 1e-5, all were an identical 46 amino acid overlap displaying greater than 90% identity with patent sequences described as being "Lepidopteran-active *Bacillus thuringiensis* delta-endotoxin polynucleotides." While these alignments likely reflect conserved structure, there is no indication that they reflect the potential for adverse biological activity.

When combined, these data demonstrate the lack of relevant similarities between known allergens, toxins or other biologically active proteins for the five putative peptides derived from the *cry1Ac* coding sequence. As a result, in the event a translation product was derived from possible reading frames 2-6 of *cry1Ac*, these putative polypeptides are not expected to possess functional cross-reactivity with known allergenic proteins or be toxic or display adverse biological activity.

d) Derivation of line and generational stability

i) Southern blot analysis to examine insert stability in multiple generations of MON 87701

In order to demonstrate the stability of the T-DNA in MON 87701, Southern blot analyses were performed using DNA obtained from multiple generations of MON 87701. For reference, the breeding history of MON 87701 is presented in **Figure 12**. The specific generations tested are identified in the legend of **Figure 13**. The R_5 generation was used for the molecular characterization analyses shown in **Figure 5** through

Figure 10. To analyze stability, four additional generations were evaluated by Southern blot analysis and compared to the R_5 generation. DNA, isolated from each of the selected generations of MON 87701, were digested with the restriction enzymes *Nco* I / *Vsp* I (**Figure 4**) and hybridized with Probe 8 (**Figure 2**). Probe 8 is designed to detect both fragments generated by the *Nco* I / *Vsp* I digest. Any instability associated with the insert would be detected as novel bands within the fingerprint on the Southern blot. The results are shown in **Figure 13**. The Southern blot has the same controls as described in Section 2.3.c.i.

Conventional soybean DNA digested with *Nco* I / *Vsp* I (**Figure 13**, lane 4) and hybridized with Probe 8 showed no detectable hybridization bands, as expected for the negative control. Pre-digested conventional soybean DNA spiked with probe template (**Figure 13**, Probe 8) generated from plasmid PV-GMIR9 produced the expected band at ~1.5 kb (**Figure 13**, lanes 1 and 2). Predigested conventional soybean DNA spiked with plasmid PV-GMIR9 DNA, previously digested with *Bgl* II / *Nco* I, produced the expected size bands of ~6.0 kb and ~9.5 kb (**Figure 13**, lane 3). Detection of the spiked controls indicates that the probes are recognizing their target sequences.

DNA extracted from MON 87701 generations R_4, R_5, R_6, R_8, and R_9 digested with *Nco* I / *Vsp* I (**Figure 13**, lanes 5-9) and hybridized with Probe 8 each produced two bands of ~6.3 kb and ~4.0 kb. The ~4.0 kb band is the expected size for the 5' border fragment and the ~6.3 kb band is consistent with the expected size of the 3' border fragment. These bands are consistent with the bands detected in **Figure 6** (lanes 2 and 9) indicating that the single copy of T-DNA I in MON 87701 is stably maintained across multiple generations.

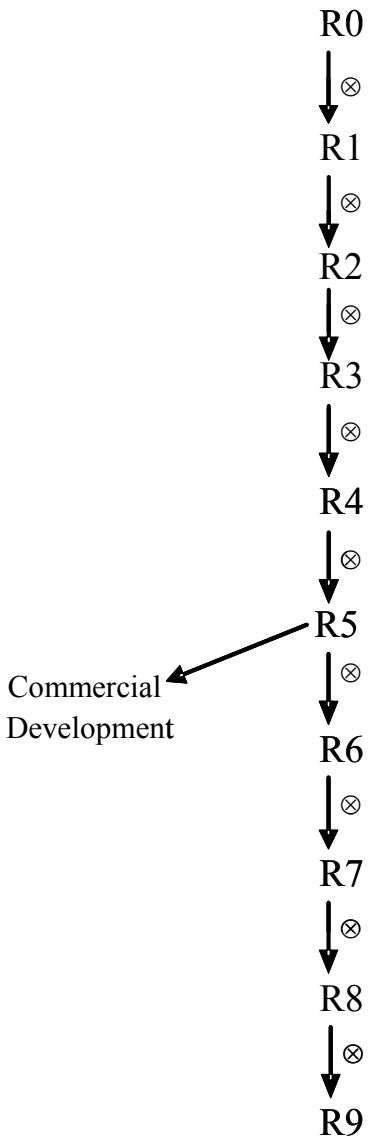


Figure 12. MON 87701 Breeding Diagram

All generations were self-pollinated (⊗). The R_5 generation seed material was used for regulatory molecular characterization and commercial development. Seed lots from the R_4, R_5, R_6, R_8, and R_9 generations were used in the molecular generational stability analysis.

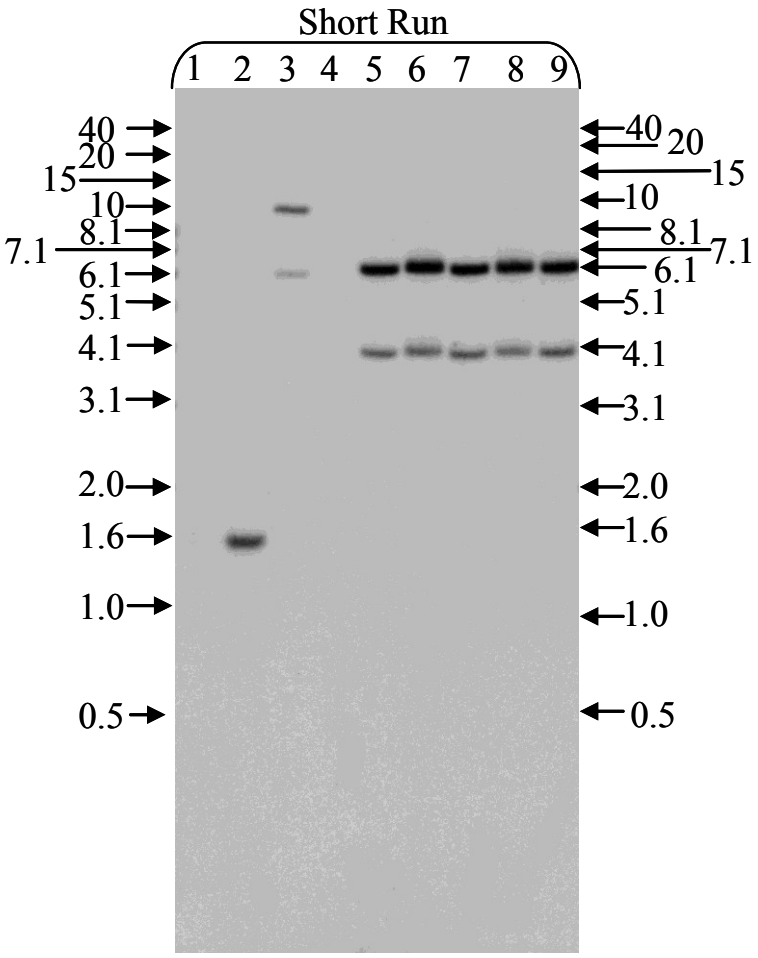


Figure 13. Insert Stability of MON 87701: T-DNA I (Probe 8)

The blot was hybridized with one ^{32}P-labeled probe that spanned a portion of the T-DNA I sequence (Figure 2, Probe 8). Each lane contains ~10 µg of digested genomic DNA isolated from soybean leaf tissue. Lane designations are as follows:

Lane 1: Conventional soybean spiked with probe templates (*Nco* I / *Vsp* I) [~0.1 genome equivalent]
2: Conventional soybean spiked with probe templates (*Nco* I / *Vsp* I) [~1.0 genome equivalent]
3: Conventional soybean (*Nco* I / *Vsp* I) spiked with PV-GMIR9 (*Nco* I / *Bgl* II) [~1.0 genome equivalent]
4: Conventional soybean (*Nco* I / *Vsp* I)
5: MON 87701 (R_4) (*Nco* I / *Vsp* I)
6: MON 87701 (R_5) (*Nco* I / *Vsp* I)
7: MON 87701 (R_6) (*Nco* I / *Vsp* I)
8: MON 87701 (R_8) (*Nco* I / *Vsp* I)
9: MON 87701 (R_9) (*Nco* I / *Vsp* I)
→ Symbol denotes size of DNA, in kilobase pairs, obtained from MW markers on ethidium stained gel.

ii) Inheritance of the genetic insert in MON87701

During development of MON 87701, segregation data were recorded to assess the heritability and stability of the *cry1Ac* gene in MON 87701. Chi-square analysis was performed over several generations to confirm the segregation and stability of the *cry1Ac* gene in MON 87701. The Chi-square analysis is based on testing the observed segregation ratio to the expected segregation ratio according to Mendelian principles.

The breeding path for generating segregation data for MON 87701 is described in **Figure 14**. The transformed R_0 plant was self-pollinated to produce R_1 seed. This seed was planted and the resulting R_1 plants were expected to segregate in a 15:1 ratio of positive to negative individual plants for the insect-protected phenotype. The 15:1 segregation ratio was expected because the *cry1Ac* gene was inserted into the soybean genome (R_0 plant) at two independently segregating loci. An individual plant (#55, designated as MON 87701), homozygous for a single copy of the *cry1Ac* gene, was identified from the R_1 segregating population by TaqMan PCR analysis.

The selected R_1 MON 87701 plant was self-pollinated to give rise to a population of R_2 plants that were repeatedly self-pollinated through the R_5 generation. At each generation, the fixed homozygous plants were tested for the expected segregation pattern of 1:0 (positive : negative) for the *cry1Ac* gene using TaqMan PCR analysis or for the presence of the Cry1Ac protein via ELISA analysis and/or protein specific lateral flow strips.

At the R_5 generation, homozygous MON 87701 soybean plants were bred via traditional breeding (bi-parental cross) with a soybean variety that did not contain the *cry1Ac* gene. The resulting F_1 plants were then self-pollinated to produce F_2 seed. The subsequent F_2 plants were tested for the presence of the MON 87701 insert by TaqMan PCR using an event-specific assay. These plants were predicted to segregate at a 1:2:1 (homozygous positive:hemizygous positive:homozygous negative) ratio according to Mendelian inheritance principles.

The heritability and stability of the *cry1Ac* gene in MON 87701 were further tested in the F_3 generation. Hemizygous positive F_2 plants were selected and self-pollinated to produce F_3 seed. The resulting F_3 plants were tested for the presence of MON 87701 by TaqMan PCR using an event-specific zygosity assay. The F_3 generation was predicted to segregate at a 1:2:1 (homozygous positive:hemizygous positive:homozygous negative) ratio.

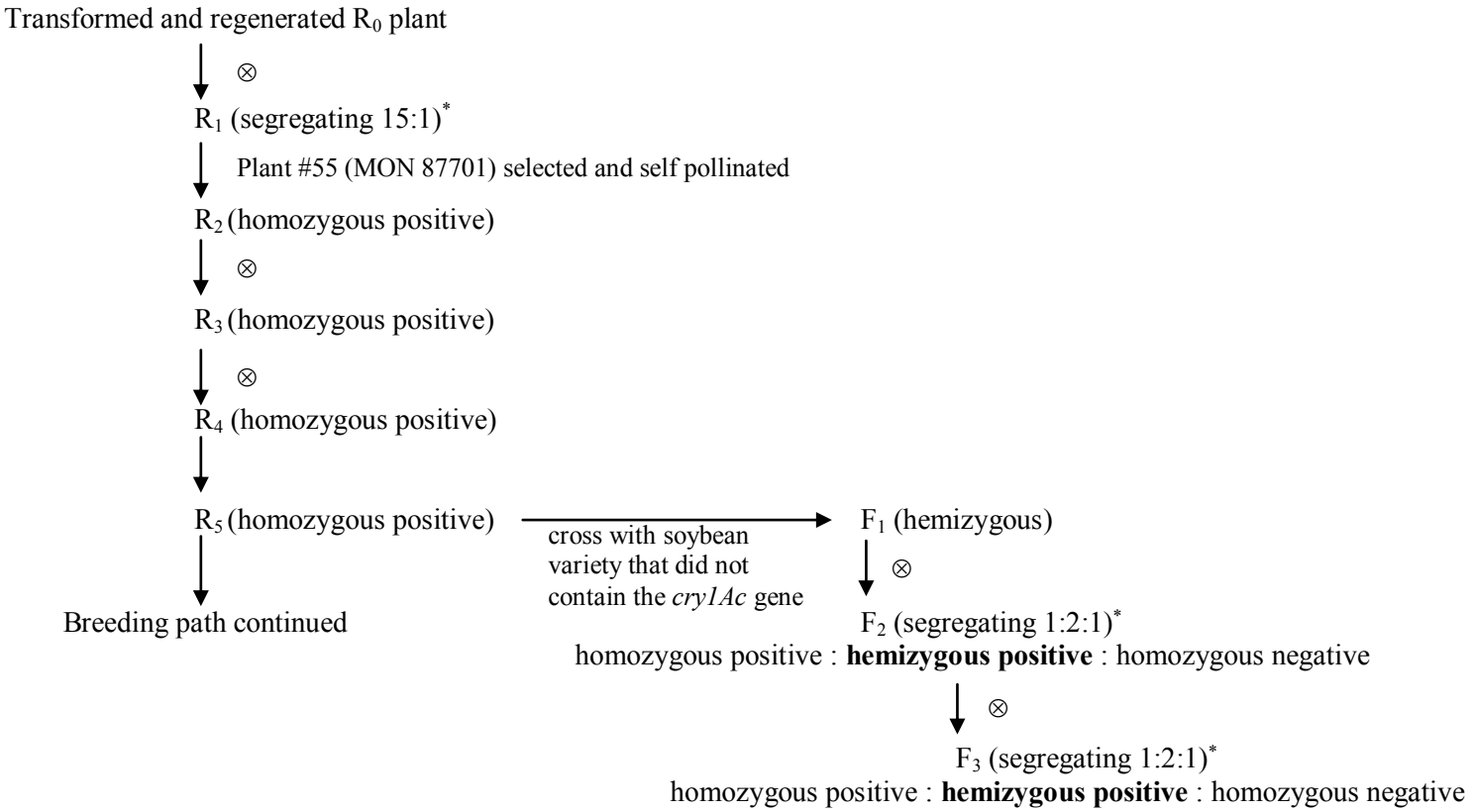
A Chi-square (χ^2) analysis was used to compare the observed segregation ratios to the expected ratios according to Mendelian principles. The Chi-square was calculated as:

$$\chi^2 = \sum [(| o - e |)^2 / e]$$

where o = observed frequency of the genotype (if PCR used) or phenotype (if ELISA used) and e = expected frequency of the genotype or phenotype. The level of statistical significance was predetermined to be 5% ($p \leq 0.05$).

The results of the χ^2 analysis of the segregating progeny of MON 87701 are presented in **Table 4** and **Table 5**. The χ^2 value in the R_1 generation indicated no statistically

significant difference between the observed and expected 15:1 segregation ratio. The insect-protected trait was subsequently fixed in the R_2, R_3, R_4, and R_5 generations, and no further segregation occurred in the generations, as expected. Following the crossing of the R_5 generation with a soybean variety that did not contain the *cry1Ac* gene, the resulting F_2 and F_3 progeny were assessed for their heritability of the *cry1Ac* gene. The χ^2 values in the F_2 and F_3 generations indicated no statistically significant difference between the observed and expected segregation ratios. These results support the conclusion that the *cry1Ac* gene in MON 87701 resides at a single locus within the soybean genome and is inherited according to Mendelian inheritance principles. These results are also consistent with the molecular characterization data that indicate MON 87701 contains a single, intact copy of the *cry1Ac* expression cassette that was inserted into the soybean genome at a single locus.



⊗ = Self pollinated

Figure 14. Breeding Path for Generating Segregation Data for MON 87701

* Chi-square analysis conducted on segregation data from the R_1, F_2, and F_3 generations.

Table 4. Segregation of the *cry1Ac* Gene during the Development of MON 87701

					Segregation			
Generation	**Expected Ratio**	**Total Plants Tested[1]**	**Observed # Plants Positive**	**Observed # Plants Negative**	**Expected # Plants Positive**	**Expected # Plants Negative**	**χ^2**	**Probability**
R_1	15:1	19	18	1	17.8	1.2	0.03	0.8590
R_2	1:0	80	80	0	80	0	Fixed	N/A
R_3	1:0	48	48	0	48	0	Fixed	N/A
R_4	1:0	598	598	0	598	0	Fixed	N/A
R_5	1:0	629	629	0	629	0	Fixed	N/A

[1] Plants were tested for the presence of the *cry1Ac* gene by protein check strips, ELISA analysis, and/or event-specific Taqman PCR. N/A = Not applicable.

Table 5. Segregation of the *cry1Ac* Gene in F_2 and F_3 Progeny from a Cross of MON 87701 with a Soybean Variety that did not Contain the *cry1Ac* Gene

					1:2:1 Segregation				
Generation[1]	**Total Plants Tested[2]**	**Observed # Plants Homozygous Positive**	**Observed # Plants Hemizygous Positive**	**Observed # Plants Homozygous Negative**	**Expected # Plants Homozygous Positive**	**Expected # Plants Hemizygous Positive**	**Expected # Plants Homozygous Negative**	**χ^2**	**Probability**
F_2	297	79	148	70	74.25	148.5	74.25	0.5	0.76
F_3	263	73	121	69	65.75	131.5	65.75	1.8	0.4069

[1] F_2 progeny were from the cross of MON 87701 homozygous positive for the *cry1Ac* gene with a soybean variety that did not contain the *cry1Ac* gene. F_3 progeny were from self-pollinated F_2 plants hemizygous positive for the *cry1Ac* gene.
[2] Plants were tested for the presence of the *cry1Ac* gene by event-specific Taqman PCR.

2.4 Antibiotic Resistance Marker Genes

No genes that encode resistance to an antibiotic marker were inserted into the soybean genome during the development of MON 87701. Molecular characterization data presented in Section 2.3.c. demonstrate the absence of the *aad*A antibiotic resistant marker in MON 87701.

a) Clinical / veterinary importance

Not applicable.

b) Viability

Not applicable.

c) Presence in food

Not applicable.

2.5 Characterisation of Novel Protein

a) Description

The Cry1Ac protein produced in MON87701 originates from *Bacillus thuringiensis*, a ubiquitous gram-positive soil bacterium that accumulates crystal proteins during sporulation. These crystal (Cry) proteins bind to the specific receptors on the midgut epithelium of targeted lepidopteran insects, and form cation selective pores, which lead to the inhibition of the digestive process and result in the insecticidal activity (Hofmann et al., 1988; Slaney et al., 1992; Van Rie et al., 1990). One valuable feature of this activity is that it is targeted to specific categories of insects, and does not impact broader insect populations or other organisms. For example, Cry1A proteins have insecticidal activity specifically against lepidopteran insects, while Cry3 proteins have insecticidal activity specifically against coleopteran insects (Höfte and Whiteley, 1989). Bt Cry1 proteins are synthesized as ~130 kDa prototoxins consisting of a three-domain toxin portion and a C-terminal extension (De Maagd et al., 2003). Domain I is involved in membrane insertion and pore formation. Domain II is involved in specific receptor recognition and binding. Domain III maintains the structural integrity of the protein molecule and also contributes to specificity (De Maagd et al., 2000 and 2001). The C-terminal portion of the Cry1A protein is thought to contribute to crystal formation via disulfide bond formation, but is not involved in determining the biological activity or specificity of the toxin toward target insects (Diaz-Mendoza et al., 2007; Miranda et al., 2001; Rukmini et al., 2000; Schnepf et al., 1998).

The Cry1Ac produced in MON 87701 is targeted to chloroplasts due to the addition of a chloroplast transit peptide (CTP) coding sequence at the 5' end of the coding sequence. Following translation and translocation into chloroplasts, the CTP is cleaved. N-terminal sequence analysis of the MON 87701-produced Cry1Ac (Section 2.5.a.iii) indicated the presence of four additional amino acids at the N-terminus compared to Cry1Ac proteins produced in Bt subsp. *kurstaki* and Bollgard cotton (**Figure 15**, **Figure 18**). The additional four amino acids are cysteine (C), methionine (M), glutamine (Q), and alanine (A). 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 CTP1 coding sequence in MON 87701. The chemistry employed in N-terminal sequencing is known to degrade cysteine (Inglis and Liu, 1970), preventing its clear identification. Except for the CTP-derived four additional amino acids at the N-terminus, the Cry1Ac that accumulates in MON 87701 shares an amino acid identity of >99% with the Cry1Ac from Bt subsp. *kurstaki* and shares 100% identity with the Cry1Ac produced in Bollgard cotton (see **Figure 15** and **Figure 18**). The presence of these four additional amino acids at the N-terminus of the MON 87701-produced Cry1Ac protein has no impact on protein activity or toxicity due to rapid proteolytic excision of the N-terminus during prototoxin activation.

The sequence encoding for the four additional amino acids derived from the CTP in MON 87701-produced Cry1Ac was included in the N-terminus of the *E. coli*-produced Cry1Ac protein that was used in the safety assessment evaluations for MON 87701. This resulted in the production of a full-length Cry1Ac protein of 1182 amino acids (1178 from Cry1Ac and four from the CTP coding region).

```
Cry1Ac          ----MDNNPNINECIPYNCLSNPEVEVLGGERIETGYTPIDISLSLTQFLLSEFVPGAGF
Bollgard        ----MDNNPNINECIPYNCLSNPEVEVLGGERIETGYTPIDISLSLTQFLLSEFVPGAGF
MON 87701       CMQAMDNNPNINECIPYNCLSNPEVEVLGGERIETGYTPIDISLSLTQFLLSEFVPGAGF

Cry1Ac          VLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIEEFARNQAISRLEGLSNLYQIYAESFRE
Bollgard        VLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIEEFARNQAISRLEGLSNLYQIYAESFRE
MON 87701       VLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIEEFARNQAISRLEGLSNLYQIYAESFRE

Cry1Ac          WEADPTNPALREEMRIQFNDMNSALTTAIPLFAVQNYQVPLLSVYVQAANLHLSVLRDVS
Bollgard        WEADPTNPALREEMRIQFNDMNSALTTAIPLFAVQNYQVPLLSVYVQAANLHLSVLRDVS
MON 87701       WEADPTNPALREEMRIQFNDMNSALTTAIPLFAVQNYQVPLLSVYVQAANLHLSVLRDVS

Cry1Ac          VFGQRWGFDAATINSRYNDLTRLIGNYTDYAVRWYNTGLERVWGPDSRDWVRYNQFRREL
Bollgard        VFGQRWGFDAATINSRYNDLTRLIGNYTDHAVRWYNTGLERVWGPDSRDWIRYNQFRREL
MON 87701       VFGQRWGFDAATINSRYNDLTRLIGNYTDHAVRWYNTGLERVWGPDSRDWIRYNQFRREL

Cry1Ac          TLTVLDIVALFPNYDSRRYPIRTVSQLTREIYTNPVLENFDGSFRGSAQGIERSIRSPHL
Bollgard        TLTVLDIVSLFPNYDSRTYPIRTVSQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHL
MON 87701       TLTVLDIVSLFPNYDSRTYPIRTVSQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHL

Cry1Ac          MDILNSITIYTDAHRGYYYWSGHQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQ
Bollgard        MDILNSITIYTDAHRGEYYWSGHQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQ
MON 87701       MDILNSITIYTDAHRGEYYWSGHQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQ

Cry1Ac          GVYRTLSSTLYRRPFNIGINNQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQ
Bollgard        GVYRTLSSTLYRRPFNIGINNQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQ
MON 87701       GVYRTLSSTLYRRPFNIGINNQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQ


Cry1Ac          NNNVPPRQGFSHRLSHVSMFRSGFSNSSVSIIRAPMFSWIHRSAEFNNIIASDSITQIPA
Bollgard        NNNVPPRQGFSHRLSHVSMFRSGFSNSSVSIIRAPMFSWIHRSAEFNNIIASDSITQIPA
MON 87701       NNNVPPRQGFSHRLSHVSMFRSGFSNSSVSIIRAPMFSWIHRSAEFNNIIASDSITQIPA

Cry1Ac          VKGNFLFNGSVISGPGFTGGDLVRLNSSGNNIQNRGYIEVPIHFPSTSTRYRVRVRYASV
Bollgard        VKGNFLFNGSVISGPGFTGGDLVRLNSSGNNIQNRGYIEVPIHFPSTSTRYRVRVRYASV
MON 87701       VKGNFLFNGSVISGPGFTGGDLVRLNSSGNNIQNRGYIEVPIHFPSTSTRYRVRVRYASV

Cry1Ac          TPIHLNVNWGNSSIFSNTVPATATSLDNLQSSDFGYFESANAFTSSLGNIVGVRNFSGTA
Bollgard        TPIHLNVNWGNSSIFSNTVPATATSLDNLQSSDFGYFESANAFTSSLGNIVGVRNFSGTA
MON 87701       TPIHLNVNWGNSSIFSNTVPATATSLDNLQSSDFGYFESANAFTSSLGNIVGVRNFSGTA

Cry1Ac          GVIIDRFEFIPVTATLEAEYNLERAQKAVNALFTSTNQLGLKTNVTDYHIDQVSNLVTYL
Bollgard        GVIIDRFEFIPVTATLEAEYNLERAQKAVNALFTSTNQLGLKTNVTDYHIDQVSNLVTYL
MON 87701       GVIIDRFEFIPVTATLEAEYNLERAQKAVNALFTSTNQLGLKTNVTDYHIDQVSNLVTYL

Cry1Ac          SDEFCLDEKRELSEKVKHAKRLSDERNLLQDSNFKDINRQPERGWGGSTGITIQGGDDVF
Bollgard        SDEFCLDEKRELSEKVKHAKRLSDERNLLQDSNFKDINRQPERGWGGSTGITIQGGDDVF
MON 87701       SDEFCLDEKRELSEKVKHAKRLSDERNLLQDSNFKDINRQPERGWGGSTGITIQGGDDVF

Cry1Ac          KENYVTLSGTFDECYPTYLYQKIDESKLKAFTRYQLRGYIEDSQDLEIYLIRYNAKHETV
Bollgard        KENYVTLSGTFDECYPTYLYQKIDESKLKAFTRYQLRGYIEDSQDLEIYSIRYNAKHETV
MON 87701       KENYVTLSGTFDECYPTYLYQKIDESKLKAFTRYQLRGYIEDSQDLEIYSIRYNAKHETV

Cry1Ac          NVPGTGSLWPLSAQSPIGKCGEPNRCAPHLEWNPDLDCSCRDGEKCAHHSHHFSLDIDVG
Bollgard        NVPGTGSLWPLSAQSPIGKCGEPNRCAPHLEWNPDLDCSCRDGEKCAHHSHHFSLDIDVG
MON 87701       NVPGTGSLWPLSAQSPIGKCGEPNRCAPHLEWNPDLDCSCRDGEKCAHHSHHFSLDIDVG

Cry1Ac          CTDLNEDLGVWVIFKIKTQDGHARLGNLEFLEEKPLVGEALARVKRAEKKWRDKREKLEW
Bollgard        CTDLNEDLGVWVIFKIKTQDGHARLGNLEFLEEKPLVGEALARVKRAEKKWRDKREKLEW
MON 87701       CTDLNEDLGVWVIFKIKTQDGHARLGNLEFLEEKPLVGEALARVKRAEKKWRDKREKLEW

Cry1Ac          ETNIVYKEAKESVDALFVNSQYDQLQADTNIAMIHAADKRVHSIREAYLPELSVIPGVNA
Bollgard        ETNIVYKEAKESVDALFVNSQYDQLQADTNIAMIHAADKRVHSIREAYLPELSVIPGVNA
MON 87701       ETNIVYKEAKESVDALFVNSQYDQLQADTNIAMIHAADKRVHSIREAYLPELSVIPGVNA

Cry1Ac          AIFEELEGRIFTAFSLYDARNVIKNGDFNNGLSCWNVKGHVDVEEQNNQRSVLVVPEWEA
```

```
Bollgard        AIFEELEGRIFTAFSLYDARNVIKNGDFNNGLSCWNVKGHVDVEEQNNQRSVLVVPEWEA
MON 87701       AIFEELEGRIFTAFSLYDARNVIKNGDFNNGLSCWNVKGHVDVEEQNNQRSVLVVPEWEA

Cry1Ac          EVSQEVRVCPGRGYILRVTAYKEGYGEGCVTIHEIENNTDELKFSNCVEEEIYPNNTVTC
Bollgard        EVSQEVRVCPGRGYILRVTAYKEGYGEGCVTIHEIENNTDELKFSNCVEEEIYPNNTVTC
MON 87701       EVSQEVRVCPGRGYILRVTAYKEGYGEGCVTIHEIENNTDELKFSNCVEEEIYPNNTVTC

Cry1Ac          NDYTVNQEEYGGAYTSRNRGYNEAPSVPADYASVYEEKSYTDGRRENPCEFNRGYRDYTP
Bollgard        NDYTVNQEEYGGAYTSRNRGYNEAPSVPADYASVYEEKSYTDGRRENPCEFNRGYRDYTP
MON 87701       NDYTVNQEEYGGAYTSRNRGYNEAPSVPADYASVYEEKSYTDGRRENPCEFNRGYRDYTP

Cry1Ac          LPVGYVTKELEYFPETDKVWIEIGETEGTFIVDSVELLLMEE
Bollgard        LPVGYVTKELEYFPETDKVWIEIGETEGTFIVDSVELLLMEE
MON 87701       LPVGYVTKELEYFPETDKVWIEIGETEGTFIVDSVELLLMEE
```

Figure 15. Amino Acid Sequence Alignment for Cry1Ac Proteins

The amino acid sequence alignment for the Cry1Ac proteins produced in *Bacillus thuringiensis* (Cry1Ac, GI-117547*), Bollgard cotton, and MON 87701. Amino acid sequence differences between Cry1Ac from Bt and the two plant-produced proteins are underlined and highlighted in gray in the Bt Cry1Ac sequence. Four amino acids originating from the CTP1 in MON 87701-produced Cry1Ac are in boldface font.

The expression level of the Cry1Ac produced in MON 87701 seed was too low and insufficient for use in the subsequent safety evaluations. Therefore, it was necessary to produce the protein in a high-expressing, recombinant microorganism in order to obtain sufficient quantities of the protein for safety evaluations. A recombinant Cry1Ac was produced in *Escherichia coli*, the sequence of which was engineered to match that of the Cry1Ac produced in MON 87701. The equivalence of the physicochemical characteristics and functional activity between the MON 87701- and *E. coli*-produced Cry1Ac proteins was confirmed by a panel of analytical techniques, including sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), western blot analysis, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), glycosylation analysis, and assay of biological activity.

The Cry1Ac protein isolated from MON 87701 harvested seed was purified and characterized, and results confirmed the equivalence of MON 87701- and *E. coli*-produced Cry1Ac proteins. SDS-PAGE demonstrated that the MON 87701-produced Cry1Ac comigrated to the same position on the gel as the *E. coli*-produced Cry1Ac protein, indicating the protein from both sources was equivalent in molecular weight. Western blot analysis conducted with a polyclonal antibody against Cry1Ac, demonstrated that the electrophoretic mobility and immunoreactivity of the MON 87701- and *E. coli*-produced Cry1Ac proteins were equivalent. The intactness of the N-terminus of the MON 87701-produced Cry1Ac protein was confirmed with an antibody which is specific to the N-terminal peptide. Tryptic peptide mapping by MALDI-TOF MS yielded peptide masses consistent with the predicted tryptic peptides generated *in silico* based on the predicted trypsin cleavage sites in the Cry1Ac sequence. In addition, the MON 87701- and *E. coli*-produced Cry1Ac proteins were found to be equivalent based on functional activities against a sensitive lepidopteran species and the lack of glycosylation. Taken together, these data provide

* GenBank gene identification number.

a detailed characterization of the Cry1Ac protein isolated from MON 87701 harvested seed and establish its equivalence to the *E. coli*-produced Cry1Ac protein. The following sections summarize the results.

i) Cry1Ac molecular weight equivalence

The equivalence in apparent molecular weights of the full-length purified MON 87701- and *E. coli*-produced Cry1Ac proteins was demonstrated using SDS-PAGE (**Figure 16**). The full-length MON 87701-produced Cry1Ac had an estimated molecular weight of 133.4 kDa (**Table 6**), and migrated to the same position on the SDS-PAGE gel as the *E. coli*-produced Cry1Ac reference standard (**Figure 16**, lane 9). The apparent molecular weight of the full-length *E. coli*-produced Cry1Ac reference protein is 131.7 kDa. The difference in the estimated molecular weights between the MON 87701-produced and *E. coli*-produced Cry1Ac full-length proteins is 1.3% (**Table 6**). Because the experimentally determined difference in apparent molecular weights met the preset acceptance criteria (≤5% difference), the MON 87701- and *E. coli*-produced Cry1Ac proteins are considered equivalent based on their molecular weights.

Table 6. Molecular Weight Difference between the Full-Length MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins Based on SDS-PAGE

Molecular Weight of Full-Length MON 87701-Produced Cry1Ac Protein	Molecular Weight of *E. coli*-Produced Cry1Ac Protein	Percent Difference from *E. coli*-Produced Cry1Ac Protein [1]
133.4 kDa	131.7 kDa	1.3 %

[1] Percent difference was calculated as follows: $\frac{133.4 - 131.7}{133.4} \times 100\% = 1.3\%$

ii) Cry1Ac immunoreactivity equivalence

A western blot analysis using goat anti-Cry1Ac serum was conducted to determine the relative immunoreactivity of the MON 87701-produced Cry1Ac protein and the *E. coli*-produced Cry1Ac reference protein. The results demonstrated that the specific anti-Cry1Ac antibody recognized the MON 87701-produced Cry1Ac protein that migrated to a similar position as the *E. coli*-produced Cry1Ac reference protein (**Figure 17).** Furthermore, the immunoreactive signal increased with increasing levels of Cry1Ac loading. The observed immunoreactivities between the full-length MON 87701- and *E. coli*-produced proteins were similar based on densitometric analysis of the western blot. Faint immunoreactive bands with molecular weights below ~133 kDa represent degradation products of Cry1Ac. Faint immunoreactive bands with molecular weights around 250 kDa also were observed, and most likely represent aggregation of the Cry1Ac protein. Both protein degradation and protein aggregation are commonly observed during protein purification of Cry proteins. Cry proteins naturally aggregate into crystal structures as has been observed for Cry1A

proteins (Güereca and Bravo, 1999), while degradation occurs primarily due to the release of endogenous proteases during the purification procedure (Gao et al., 2006).

Densitometric analysis was conducted to compare the immunoreactivity of full-length MON 87701- and *E. coli*-produced Cry1Ac proteins. The relative immunoreactivity of each protein with Cry1Ac-specific antibody was determined by averaging intensity values of six protein bands corresponding to the full-length MON 87701-produced Cry1Ac and six bands corresponding to the full-length *E. coli*-produced Cry1Ac (**Table 7**). The averaged band intensity of the signal from the MON 87701-produced Cry1Ac lanes was 33.3% less than the averaged band intensity of the signal from the *E. coli*-produced Cry1Ac lanes. The observed difference was within the preset acceptance criteria for immunoreactivity (±35% difference). Thus, the immunoblot analysis established identity of the MON 87701-produced Cry1Ac and demonstrated that the MON 87701- and *E. coli*-produced Cry1Ac proteins are equivalent based on their immunoreactivity with Cry1Ac-specific antibody.

Table 7. Comparison of Immunoreactive Signals between Full-Length MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins

Load Amount (ng)	MON 87701-produced Protein Signal Density[1]	*E. coli*-produced Protein Signal Density[1]
10	1.288	1.419
10	1.955	1.798
20	2.908	4.559
20	2.987	3.706
30	4.214	6.547
30	4.140	8.199
Sum	**17.492**	**26.228**
Average Density	**2.915**	**4.371**
Percent difference[2]	**33.3%**	

[1] The density of each band was determined by image analysis of the quantitative western blot shown in **Figure 17**. Values shown for signal density are contour quantity, i.e. average OD × contour area in mm^2.

[2] Percent difference is calculated using the equation:

$$\frac{\left| AverageDensity E.coli - AverageDensity Plant \right|}{AverageDensity E.coli} \times 100 = Percent\,Difference$$

iii) N-terminal sequence analysis

From N-terminal sequencing analysis of the first 15 amino acids of the MON 87701-produced Cry1Ac, seven definite and two tenuous amino acid assignments were made that matched the predicted N-terminal sequence for a Cry1Ac containing four amino acids derived from the CTP1 (**Figure 18**). As discussed above, the amino acid cysteine is shown in the predicted sequence at position one, based on the CTP1-Cry1Ac coding sequence in MON 87701. Cysteine is unstable during the acid hydrolysis reaction used for N-terminal sequencing, and is usually not explicitly

observed (Inglis and Liu, 1970). The clear identification of amino acids in subsequent cycles of the sequencing analysis confirmed that an unidentified amino acid was present at position one. The N-terminal sequence information, therefore, confirms the identity of the Cry1Ac isolated from MON 87701 and the intactness of its N-terminus.

iv) MALDI-TOF mass spectrum analysis

The identity of the MON 87701-produced Cry1Ac was further confirmed by tryptic peptide mass mapping analysis using MALDI-TOF MS. In general, protein identification made by peptide mapping is considered to be reliable if the measured coverage of the sequence is 15% or higher with a minimum of five matched peptides (Jensen et al., 1997). Observed tryptic peptides were considered a match to the expected tryptic mass when differences in molecular weight of less than one dalton (Da) were found between the observed and predicted fragment masses. Such matches were made without consideration for potential natural amino acid modifications, such as glycosylation. The protein sample was heat-denatured, chemically reduced, alkylated, and digested with trypsin, and the masses of the tryptic peptides were measured.

There were seventy unique peptide fragments identified that corresponded to the expected masses of the Cry1Ac trypsin-digested peptides. The identified masses were used to assemble a coverage map indicating the matched peptide sequences for the Cry1Ac protein (**Figure 19**), resulting in ~67% (787 out of 1182 total amino acids) coverage of the total protein. This analysis confirmed the identity of the MON 87701-produced Cry1Ac protein.

v) Cry1Ac functional activity equivalence

The functional activities of the MON 87701- and *E. coli*-produced Cry1Ac proteins were compared using an insect growth-inhibition bioassay. Aliquots of MON 87701- and *E. coli*-produced Cry1Ac proteins were tested in an assay using corn earworm (CEW; *Helicoverpa zea)*, an insect species known to be susceptible to Cry1Ac. Dose-response assays were performed for Cry1Ac proteins from both sources in parallel and assays were repeated on three separate days to estimate the mean EC_{50} value, the effective concentration necessary to inhibit CEW growth by 50% relative to a control population of insects not exposed to the insecticidal protein. The mean EC_{50} value determined for the MON 87701-produced Cry1Ac was 0.0039 µg Cry1Ac/ml diet. This EC_{50} value was very similar to the mean EC_{50} value of 0.0036 µg Cry1Ac/ml diet obtained for the *E. coli*-produced Cry1Ac in the same assay (**Table 8**). These results clearly demonstrate that the Cry1Ac proteins derived from MON 87701 and *E. coli* have equivalent functional activities.

Table 8. EC_{50} Values of *E. coli*-Produced and MON 87701-Produced Cry1Ac Proteins in a Corn Earworm Diet-Incorporation Bioassay

		EC_{50} (µg Cry1Ac/ml diet)[1]	
		***E. coli*-produced**	**MON 87701-produced**
Replicate[2]	1	0.0031 ± 0.00035	0.0050 ± 0.00069
	2	0.0026 ± 0.00022	0.0032 ± 0.00021
	3	0.0050 ± 0.00030	0.0034 ± 0.00035
Mean EC_{50}		0.0036	0.0039

[1] EC_{50} represents the Cry1Ac concentration needed to inhibit the growth of the target insect by 50%.
[2] Values shown for each replicate represent EC_{50} estimates ± standard error

vi) Cry1Ac glycosylation equivalence

Some eukaryotic proteins are post-translationally modified by the addition of carbohydrate moieties (Rademacher et al., 1988). These carbohydrate moieties may be complex, branched polysaccharide structures, simple oligosaccharides or monosaccharides. In contrast, the non-virulent *E. coli* strains used for cloning and expression purposes lack the necessary biochemical synthetic capacity required for protein glycosylation. Therefore, determining whether the MON 87701-produced Cry1Ac is equivalent to the *E. coli*-produced Cry1Ac requires an investigation of its glycosylation status.

To assess whether post-translational glycosylation of the MON 87701-produced Cry1Ac occurred, the purified protein sample was subjected to glycosylation analysis. The *E. coli*-produced Cry1Ac represented a negative control. The positive controls were the transferrin and horseradish peroxidase (HRP) proteins, which are known to have covalently linked carbohydrate modifications. Transferrin and HRP, as well as the purified Cry1Ac proteins isolated from MON 87701 and *E. coli,* were separated on SDS-PAGE, transferred to a PVDF membrane, and glycosylation analysis was performed to detect carbohydrate moieties on the proteins. The results of this analysis are shown in Figure VI-6. The positive controls, transferrin and HRP, were detected at the expected molecular weights of ~75 and ~45 kDa, respectively, in a concentration-dependent manner (**Figure 20**, Panel A, Lanes 2-5). No detectable signal was observed for the MON 87701- and *E. coli*-produced Cry1Ac proteins (**Figure 20**, Panel A, Lanes 6-9). To confirm that sufficient MON 87701- and *E. coli*-produced Cry1Ac proteins were present for carbohydrate detection and glycosylation analysis, the membrane was stained with Coomassie Brilliant Blue R stain to detect proteins (**Figure 20**, Panel B). With this stain, both MON 87701- and *E. coli*-produced Cry1Ac proteins were clearly detected on the membrane (**Figure 20**, Panel B, Lanes 6-9), demonstrating sufficient levels of the proteins were present for carbohydrate detection.

These results indicate that the MON 87701-produced Cry1Ac is not glycosylated and thus is equivalent to the *E. coli*-produced Cry1Ac with respect to the absence of glycosylation.

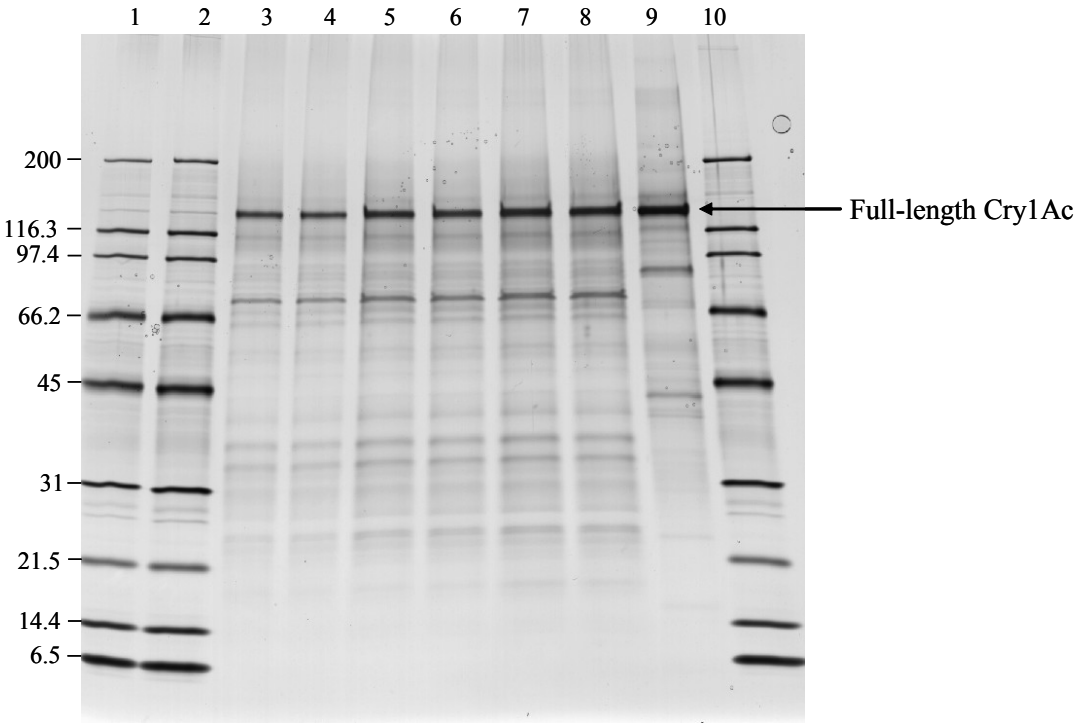


Figure 16. SDS-PAGE of MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins

Aliquots of the MON 87701-produced Cry1Ac and the *E. coli*-produced Cry1Ac protein were separated on a Tris-glycine 4-20% polyacrylamide gradient gel and stained with an Owl Silver Staining kit. Approximate molecular weights indicated (in kDa) correspond to the Broad Range Molecular Weight marker (BioRad) loaded in lanes 1, 2, and 10.

Lane	Sample	Amount loaded (ng)
1	Broad Range MW Marker	360
2	Broad Range MW Marker	360
3	MON 87701-produced Cry1Ac	95
4	MON 87701-produced Cry1Ac	95
5	MON 87701-produced Cry1Ac	189
6	MON 87701-produced Cry1Ac	189
7	MON 87701-produced Cry1Ac	284
8	MON 87701-produced Cry1Ac	284
9	*E. coli*-produced Cry1Ac	198
10	Broad Range MW Marker	360

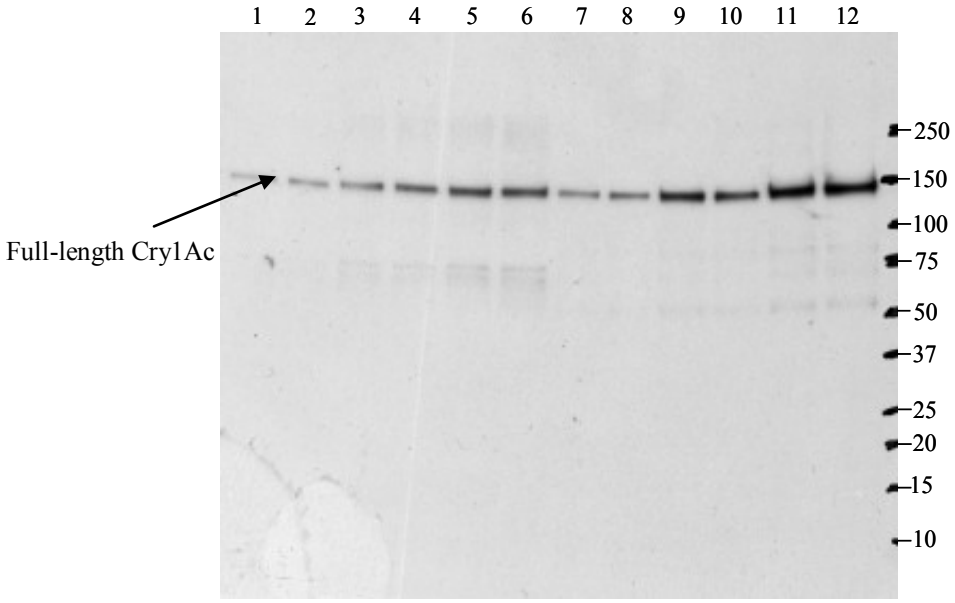


Figure 17. Western Blot Analysis of MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins

Aliquots of the purified, MON 87701- and *E. coli*-produced Cry1Ac proteins were separated by SDS-PAGE, and electrotransferred to a PVDF membrane. The membrane was probed with goat anti-Cry1Ac serum and developed using an ECL system (GE Healthcare). Approximate molecular weights (kDa) are shown on the right and correspond to the tick marks indicating the position of molecular weight markers. Amount loaded indicates amount of full-length protein.

Lane	Sample	Amount Loaded (ng)
1	MON 87701-produced Cry1Ac	10
2	MON 87701-produced Cry1Ac	10
3	MON 87701-produced Cry1Ac	20
4	MON 87701-produced Cry1Ac	20
5	MON 87701-produced Cry1Ac	30
6	MON 87701-produced Cry1Ac	30
7	*E. coli*-produced Cry1Ac	10
8	*E. coli*-produced Cry1Ac	10
9	*E. coli*-produced Cry1Ac	20
10	*E. coli*-produced Cry1Ac	20
11	*E. coli*-produced Cry1Ac	30
12	*E. coli*-produced Cry1Ac	30

Amino acid[1] residue # from the N-terminus →	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Predicted Cry1Ac Sequence[2] →	C	M	Q	A	M	D	N	N	P	N	I	N	E	C	I
Observed Sequence→	X	M	Q	A	M	D	N	(N)	P	(N)	X	X	X	X	X

1 The single letter International Union of Pure and Applied Chemistry-International Union of Biochemistry (IUPAC-IUB) amino acid code is; A, Alanine; C, Cysteine; D, Aspartic acid; E, Glutamic acid; I, Isoleucine; M, Methionine; N, Asparagine, P, Proline; and Q, Glutamine. X indicates an undesignated call in that cycle of the analysis. Parentheses indicate a tenuous designation.

2 The predicted Cry1Ac sequence was deduced from the coding region of the full length *cry1Ac* gene present in MON 87701. The observed sequence was obtained by N-terminal sequencing of the MON 87701-produced Cry1Ac.

Figure 18. N-terminal Amino Acid Sequence Analysis of the Cry1Ac Protein Purified from Harvested Seed of MON 87701

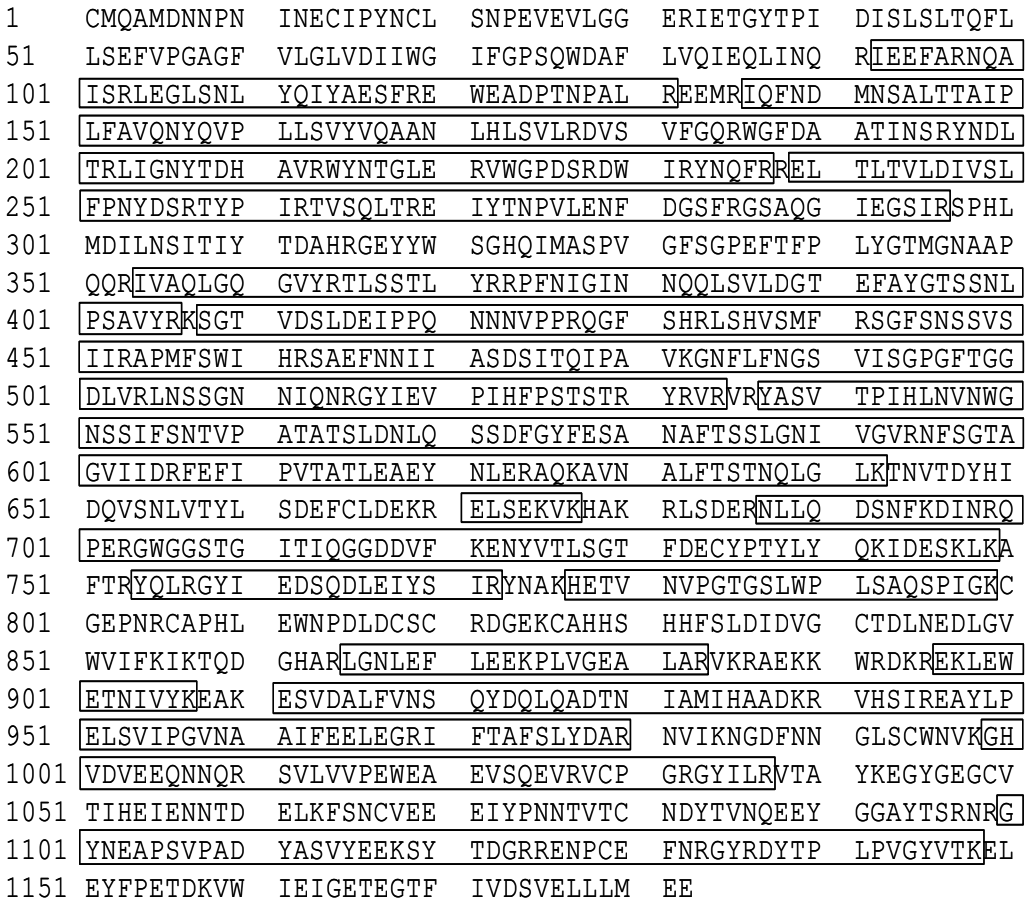


Figure 19. MALDI-TOF MS Coverage Map of the Cry1Ac Protein Isolated from MON 87701

The amino acid sequence of the plant-produced Cry1Ac protein was deduced from the coding region of the full-length *cry1Ac* gene present in MON 87701, and from N-terminal sequencing of MON 87701-produced Cry1Ac. Boxed regions indicate amino acids that were identified in tryptic peptides using MALDI-TOF MS. In total, ~67% (787 of 1182 total amino acids) of the expected protein sequence was identified.

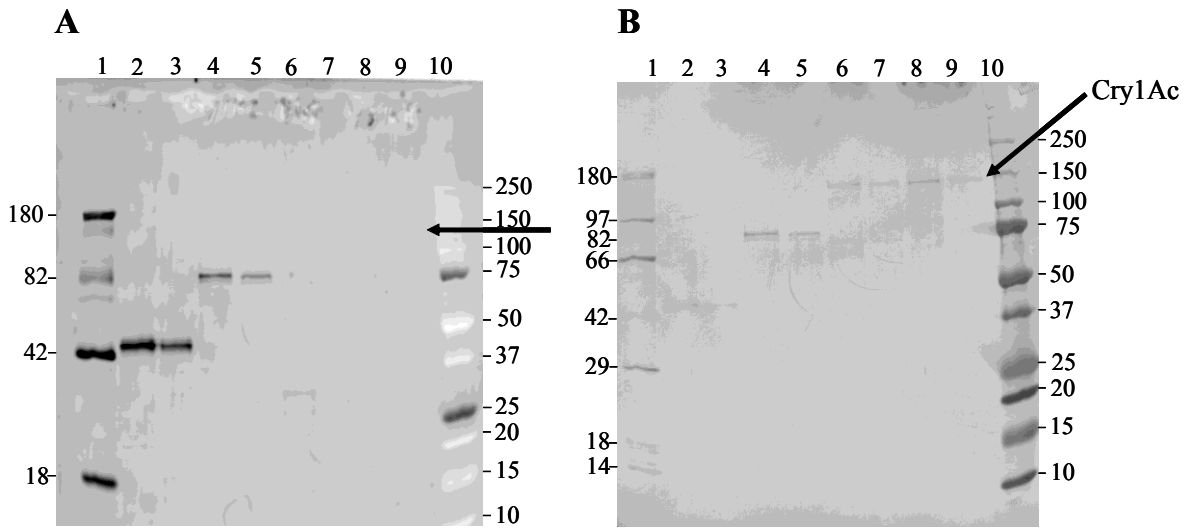


Figure 20. Glycosylation Analysis of the MON 87701-Produced and *E. coli*-Produced Cry1Ac Proteins

Aliquots of horseradish peroxidase (positive control), transferrin (positive control), the MON 87701-produced Cry1Ac protein, and *E. coli*-produced Cry1Ac protein (negative control), were separated by SDS-PAGE and electrotransferred to a PVDF membrane. For Cry1Ac samples, amount loaded indicates full-length protein amount. Approximate molecular weights indicated (in kDa) correspond to the Precision Plus pre-stained dual color molecular weight marker (BioRad) loaded in Lane 10 and the CandyCane glycosylated marker (Molecular Probes) loaded in Lane 1. (Panel A) Glycosylation Analysis: Where present, periodate-oxidized protein-bound carbohydrate moieties reacted with Pro-Q Emerald 488 glycoprotein stain and emitted a fluorescent signal at 488 nm. The arrow indicates the approximate expected position of Cry1Ac protein on the blot. (Panel B) Total Protein Staining: Following glycosylation analysis, the blot was stained for total protein using Coomassie Brilliant Blue R.

Lane	Sample	Amount Loaded (ng)
1	Candy Cane MW Marker	-
2	Horseradish Peroxidase	100
3	Horseradish Peroxidase	50
4	Transferrin	100
5	Transferrin	50
6	MON 87701-produced Cry1Ac	100
7	MON 87701-produced Cry1Ac	50
8	*E. coli*-produced Cry1Ac	100
9	*E. coli*-produced Cry1Ac	50
10	Precision Plus Dual Color MW marker	-

b) Site of expression

The levels of the Cry1Ac in various tissues of MON 87701 that are relevant to the risk assessment were assessed by a validated ELISA.[2] Tissue samples for analysis were collected from five field trials conducted in the U.S. during 2007. The trial locations were in the states of Alabama, Arkansas, Georgia, Illinois, and North Carolina, which represent relevant soybean-growing regions of the U.S. and provide a range of environmental conditions that would be typical of those encountered in the production of soybean. At each site, three replicated plots of MON 87701 and a conventional soybean control (A5547) were planted using a randomized complete block field design. Over-season leaf (OSL), forage, root, and harvested seed were collected from each replicated plot at all field sites. A description of the tissues collected is provided below.

Tissue	*Soybean development stage*	*Days after planting (DAP)*
OSL-1	V3-V4	23-34
OSL-2	V6-V8	36-45
OSL-3	V10-V12	43-57
OSL-4	V14-V16	52-70
Forage	R6	85-106
Root	R6	85-106
Mature seed	R8[1]	139-156
Pollen/Anther	R2	63

[1]Harvested at or dried to a moisture content of ~10-15%.

Pollen/anther tissue was collected at the R2 growth stage during the 2007 growing season from a field site in Jackson County, IL that was used to generate bulk quantities of MON 87701 and conventional control material. At this site, single plots were established for MON 87701 as well as the conventional soybean control. Four replicate pollen/anther samples were collected from each plot.

Cry1Ac levels were determined in all eight tissue types described above. The results obtained from ELISA analysis are summarized in **Table 9** for the various tissue types including the tissues collected throughout the growing season. The Cry1Ac levels were determined in over-season leaf (OSL1-4), forage, root, harvested seed, and pollen. The levels of Cry1Ac in tissue samples from the conventional soybean control were below the Cry1Ac assay limit of quantitation (LOQ) or limit of detection (LOD) for each tissue type.

Results showed that mean Cry1Ac levels across the five sites were highest in leaf (340 µg/g dwt in OSL-4), followed by forage (34 µg/g dwt) and mature, harvested seed (4.7 µg/g dwt). If present in root, Cry1Ac levels are less than the ELISA assay LOD of 0.347 µg/g fwt. In over-season leaf tissues harvested throughout the growing season, mean Cry1Ac levels in MON 87701 across all sites ranged from 220 – 340 µg/g dwt. In general, the mean levels of the Cry1Ac protein in leaf remained relatively constant across sampling time points, but levels of the protein were within a broader range as the growing season progressed (**Table 9**).

[2] Due to the limited quantity of material available, pollen/anther was evaluated using a non-validated, but optimized ELISA method.

Table 9. Summary of Cry1Ac Protein Levels in Tissues Collected from MON 87701 Produced across Five Sites during the U.S. 2007 Growing Season

Tissue Type	Cry1Ac µg/g fwt (SD)[1,3]	Range[4] (µg/g fwt)	Cry1Ac µg/g dwt (SD)[2]	Range (µg/g dwt)	LOQ/LOD (µg/g fwt)
OSL-1	30 (8.5)	12-40	220 (70)	110-350	2.5/0.74
OSL-2	38 (16)	18-80	260 (100)	130-500	2.5/0.74
OSL-3	34 (17)	14-77	240 (110)	94-480	2.5/0.74
OSL-4	53 (36)	15-110	340 (290)	78-960	2.5/0.74
Root	< LOD	< LOD	NA[5]	NA[5]	0.4/0.347
Forage	9.0 (8.8)	2.5-32	34 (36)	8.2-140	2.0/0.55
Harvested seed	4.2 (0.73)	3.1-5.0	4.7 (0.79)	3.4-5.7	1.0/0.47
Pollen/ anther[6]	2.3 (0.58)	1.8-3.1	NA[7]	NA[7]	ND[8]

1. Protein levels are expressed as microgram (µg) of protein per gram (g) of tissue on a fresh weight (fwt) basis.
2. Protein levels are expressed as µg/g on a dry weight (dwt) basis. The dry weight values were calculated by dividing the fwt value by the dry weight conversion factors obtained from moisture analysis data.
3. The mean and standard deviation were calculated across sites (n=15, except OSL-1 where n=13 and pollen/anther where n=4).
4. Minimum and maximum values were determined for each tissue type across sites.
5. Protein levels that were <LOD on a fwt basis were not converted to dwt values.
6. Due to limited quantity, pollen/anther material was evaluated using a non-validated, but optimized ELISA method.
7. Protein level by dry weight was not calculated due to limited quantities of pollen/anther tissue.
8. Due to limited quantities of pollen/anther tissue the LOD and LOQ were not determined.

c) Non-expression

Not applicable

d) History of human consumption

i) Safety of the donor organism : *Bacillus thuringiensis*

Bacillus thuringiensis subspecies *kurstaki* is a gram-positive bacterium that is commonly found in soil and has been used commercially in the U.S. since 1958 to produce microbial-derived products with insecticidal activity (EPA, 1988). Sprays of sporulated Bt have been used for many years in agriculture for pest control. Residual amounts of Bt spores have been detected in several fresh fruits and vegetables, grapes, milk, pasta, and bread (Frederiksen et al., 2006). Despite this exposure, no harm to mammals caused by these bacteria has been identified (Siegel, 2001). In the U.S., an exemption from the

requirement of a tolerance for the first microbial Bt product was granted in 1960 by the U. S. FDA after an extensive toxicity and infectivity evaluation program. The testing program consisted of acute, subchronic, and chronic studies, which resembled the testing required for conventional chemical pesticides. Registration was granted by the USDA later the same year. In 1971, the EPA assumed responsibility for all pesticide tolerance exemptions and product registrations. Since then, a variety of naturally occurring and genetically modified microbial Bt products have been registered and included under this tolerance exemption. Microbial pesticides containing Bt Cry1A proteins have been used for more than 50 years and subjected to extensive toxicity testing showing no adverse effects to human health (Baum et al., 1999; Betz et al., 2000; EPA, 1988; EPA, 2001; James and ISAAA, 2003; McClintock et al., 1995; Mendelsohn et al., 2003). The U.S. EPA has also evaluated the potential impacts of Cry1Ac and Cry1Ab proteins in pesticidal preparations on nontarget organisms, including mammals, birds, fish, beneficial insects, marine animals, and plants, and concluded that they do not pose a risk to these organisms (OECD, 2007).

ii) History of safe use of Cry proteins produced in crop plants

The Cry1Ac protein has a history of safe use in bacterial preparations used as biopesticides. In addition, Bollgard and Bollgard II cotton, which also produce the Cry1Ac protein, has been used commercially in the U.S. for over a decade (in the case of Bollgard) for the control of lepidopteran pests. Except for the four additional amino acids at the N-terminus, the MON 87701-produced Cry1Ac has 100% amino acid identity with the Bollgard Cry1Ac protein (**Figure 15**). A related protein, Cry1Ab, which has ~90% amino acid identity to the Cry1Ac produced in Bollgard and MON 87701, is expressed in YieldGard Corn Borer corn. The EPA has approved commercial use of the Cry1Ab and Cry1Ac proteins as expressed in corn and cotton (EPA, 2009). An exemption from the requirement for a tolerance was granted in 1996 for Cry1Ab and the genetic material necessary for its production in all plants (40 CFR § 174.511). A tolerance exemption for the Cry1Ac protein and the genetic material necessary for its production in all plants was granted on April 11, 1997 (40 CFR §174.510).

Following the deregulation of Bollgard cotton and YieldGard Corn Borer corn, both products have been grown commercially in the U.S. for more than ten years, and were both approved for importation in 2000 by FSANZ in Australia. The safety for food and feed consumption of plants expressing Cry1Ac or Cry1Ab proteins has been evaluated by several regulatory agencies around the world, including FDA, and summarized in numerous scientific publications (Betz et al., 2000; Shelton et al., 2002). Detailed human and animal safety assessments and over a decade of safe consumption of these crops confirm their safety. Collectively, this information confirms the safety of the Cry1Ac protein produced in MON 87701.

e) Oral toxicological studies

- **Acute oral toxicity studies with the Cry1Ac protein**

Most known protein toxins, including the insecticidal Cry proteins, act through acute mechanisms to exert toxicity (Hammond and Fuchs, 1998; Pariza and Johnson, 2001; Sjoblad et al., 1992). The primary exceptions to this rule consist of certain anti-nutritional proteins such as lectins and protease inhibitors, which typically require a short-

term (two to four week) feeding study to manifest toxicity (Liener, 1994). The amino acid sequence of Cry1Ac produced in MON 87701 is not similar to any of these anti-nutritional proteins or to any other known mammalian protein toxin. In addition, because Cry proteins act through acute mechanisms to control insect pests and have no activity against nontarget organisms such as mammals, the U.S. EPA has determined that a high dose acute test is sufficient to confirm their absence of toxicity towards mammals (McClintock et al., 1995). Therefore, an acute oral mouse toxicity study was considered appropriate to confirm the lack of mammalian toxicity of Cry1Ac.

An acute toxicology study was conducted with the full-length Cry1Ac that was produced in an *E. coli* expression system. This protein was shown to be physicochemically and functionally equivalent to the Cry1Ac produced in MON 87701. Cry1Ac was administered by oral gavage to 10 male and 10 female CD-1 mice at a total dose of 1290 mg/kg body wt (administered in two doses about four hours apart). Control mice were administered comparable doses of bovine serum albumin (BSA). Following dosing, all mice were subject to detailed clinical observations once daily (twice on day of dosing) for signs of mortality or toxicity. Food consumption was measured on days 0, 7, and 14. Body weights were measured prior to dosing and on study days 0, 7, and 14. All animals were sacrificed on day 14 and subjected to a gross necropsy. There were no treatment-related effects of Cry1Ac on survival, clinical observations, body weight gain, food consumption or gross pathology. A statistically significant reduction in body weight gain was observed in males but not in females dosed with 1,290 mg/kg Cry1Ac relative to BSA-treated controls; however, this result was considered equivocal because at least one male in the study experienced an interruption in water supply. In order to further investigate this possible effect on body weight, an additional group of 10 male CD-1 mice (and BSA controls) was dosed with Cry1Ac by oral gavage at a total dose of 1460 mg/kg body wt (two equal doses four hours apart). There was no effect on body weight in males dosed with 1,460 mg/kg Cry1Ac. Therefore, the effect on body weight in males at 1,290 mg/kg Cry1Ac was not reproduced in the repeat study using a higher dose and was not considered treatment-related. The NOAEL for the Cry1Ac protein was 1,460 mg/kg in males and 1,290 mg/kg in females.

f) Amino acid sequence

- **Structural similarity of Cry1Ac to known toxins or other proteins**

The assessment of the potential for protein toxicity includes bioinformatic analysis of the amino acid sequence of the introduced protein. The goal of the bioinformatic analysis is to ensure that the introduced protein does not share homology to known toxins or anti-nutritional proteins associated with adverse health effects.

Potential structural similarities shared between the Cry1Ac protein and sequences in a protein database were evaluated using the FASTA sequence alignment tool. The FASTA program directly compares amino acid sequences (i.e., primary, linear protein structure) and the alignment data may be used to infer shared higher order structural similarities between two sequences (i.e., secondary and tertiary protein structures). Proteins that share a high degree of similarity throughout the entire sequence are often homologous. Homologous proteins usually have common secondary structures, common three-dimensional configuration, and, consequently, may share similar functions.

FASTA bioinformatic alignment searches using the Cry1Ac amino acid sequence were performed with the TOXIN6 database to identify possible homology with proteins that may be harmful to human and animal health. The TOXIN6 database is a subset of 7,176 sequences derived from the protein database (PROTEIN) consisting of publicly available protein sequences from GenBank (GenBank protein database, release 163.0, December 15, 2007). Initially all header lines and the associated protein sequence in PROTEIN database were screened using all possible combinations of upper and lower case characters spelling the words "toxic" and "toxin." The resulting 9,082 header lines and associated sequences then were filtered to exclude the following terms used in combination with "toxic" or "toxin," resulting in 7,176 sequences; these terms were "synthetic," "anti," "putative," "like," "insect," "Cry," "Thuringiensis", and "toxin-reductase."

An *E-score* acceptance criteria of $<1\times10^{-5}$ for any alignment was used to identify proteins from the TOXIN6 database with potential for significant shared structural similarity and function with Cry1Ac. The *E*-score is a statistical measure of the likelihood that the observed similarity score could have occurred by chance in a search. A larger *E*-score indicates a lower degree of similarity between the query sequence and the sequence from the database. Typically, alignments between two sequences require an *E*-score of 1×10^{-5} or smaller to be considered to have sufficient sequence similarity to infer homology. The FASTA search produced 15 alignments, with the top alignment displaying 98.981% identity in a 1,178 amino acid overlap with a delta-endotoxin. Delta-endotoxin is a term that is synonymous with Bt or Cry proteins and as such, the top alignment does not indicate a potential for human or mammalian toxicity. The 14 additional alignments that were the product of the FASTA search of the TOXIN6 database were with other Cry proteins.

The results of the bioinformatic analyses demonstrated that no structurally relevant similarity exists between the Cry1Ac protein and any known toxic or other biologically active proteins that would be harmful to human or animal health.

g) Known allergenicity of source

- **Source of Cry1Ac**

The *cry1Ac* coding sequence is derived from *Bacillus thuringiensis,* an organism that is not a source of allergens and has been used commercially in the U.S. since 1958 to produce microbial-derived products with insecticidal activity (EPA, 1988).

h) Unknown allergenicity information

According to guidelines adopted by the Codex Alimentarius Commission (Codex, 2003) for the evaluation of the potential allergenicity of novel proteins, the allergenic potential of a novel protein is assessed by comparing the biochemical characteristics of the novel protein to characteristics of known allergens. A protein is not likely to be associated with allergenicity if: 1) the protein is from a non-allergenic source; 2) the protein represents only a very small portion of the total plant protein; 3) the protein does not share structural similarities to known allergens based on the amino acid sequence, and 4) the protein is rapidly digested in mammalian gastrointestinal systems. The Cry1Ac protein has been

assessed for its potential allergenicity according to the recommendations of the Codex Alimentarius Commission (Codex Alimentarius, 2003).

i) The Cry1Ac protein as a proportion of total protein

The Cry1Ac protein was detected in all plant tissues, except root, at a number of time points during the growing season. Among these tissues, harvested seed is the most relevant to the assessment of food allergenicity. The mean level of Cry1Ac in harvested seed is 4.7 µg/g dwt. The mean % dry weight of total protein in harvested seed from MON 87701 is 39.27% (or 392,700 µg/g). The percent of Cry1Ac in MON 87701 harvested seed is calculated as follows:

$$(4.7\ \mu g/g \div 392{,}700\ \mu g/g) \times 100\% \approx 0.0012\%\ \text{of total soybean protein}$$

Therefore, the Cry1Ac protein represents a very small portion of the total protein in harvested seed of MON 87701.

ii) Bioinformatics analysis of sequence similarity of the Cry1Ac protein produced in MON 87701 to allergens

In 2003, the Codex Alimentarius Commission published guidelines for the evaluation of the potential allergenicity of novel proteins based on shared amino acid sequence identity (Codex, 2003). The guideline is based on the comparison of amino acid sequences between introduced proteins and allergens, where potential allergenic cross-reactivity may exist if the introduced protein is found to have at least 35% amino acid identity with an allergen over any segment of at least 80 amino acids. The Codex guideline also recommended that a sliding window search with a scientifically justified peptide size could be used to identify immunologically relevant peptides in otherwise unrelated proteins.

The potential for allergenic cross-reactivity between the Cry1Ac produced in MON 87701 and known allergens, gliadins and glutenins was assessed using two search algorithms: 1) a FASTA sequence alignment program was used to determine if any allergens shared at least 35% amino acid identity over at least 80 amino acids with Cry1Ac, and 2) a sliding window search was used to identify any eight-amino acid matches to known allergens (2003; Thomas et al., 2005). Both bioinformatic search algorithms were used in conjunction with the Food Allergy Research and Resource Program Database allergen database (FARRP)[1] to evaluate if Cry1Ac shared any sequence similarities with known allergens. The AD8 database (release date, January 11, 2008) was assembled from sequences obtained from the FARRP allergen database. The protein sequences in the FARRP allergen database were assembled and evaluated for evidence of allergenicity by an international panel of allergy experts. For in-house searches, redundant sequences present in the FARRP database were removed and obsolete GI sequence numbers were replaced with an up-to-date GI numbers. Due to the removal of obsolete GI numbers and sequence duplicates derived from the same species, a total of 1,250 GI sequence numbers were found to be valid and used to assemble a searchable in-house database AD8.

Based on the results of the FASTA comparisons, known allergens were ranked according to their degree of similarity to Cry1Ac. Alignments for full-length search sequences may

[1] FARRP. 2009. www.allergenonline.com. University of Nebraska.

be considered relevant for similarity to known allergens if the identity is equal to or greater than 35% and equal to or greater than 80 amino acids in aligned sequence length (Codex, 2003). All alignments were inspected visually to determine if an alignment represented biologically relevant sequence similarity. None of the proteins in the AD8 database met or exceeded the threshold of 35% identity over 80 amino acids when compared to the Cry1Ac amino acid sequence. Although none of the obtained alignments satisfied minimum Codex standards, the quality of each alignment was also thoroughly evaluated for their percent identity and *E*-scores (expectation score) as produced from the FASTA bioinformatic program (Pearson, 2000; Pearson and Lipman, 1988). The analysis of the shared percent identity, length of the alignment, as well as the E-score is intended to add additional information to the search for proteins that may have potentially significant homology. The *E*-score is a statistical measure of the likelihood that the observed similarity score could have occurred by chance in a search. A larger *E*-score indicates a lower degree of similarity between the query sequence and the sequence from the database. Typically, alignments between two sequences require an *E*-score of 1×10^{-5} or smaller to be considered to have sufficient sequence similarity to infer homology. There were no alignments that had an *E*-score below 1, therefore, no alignments between the Cry1Ac sequence and sequences in the AD8 database were observed that were considered relevant from an allergenic assessment perspective.

A sliding eight-amino acid window search was performed to identify whether or not a linearly contiguous match (exact identity matches) of eight amino acids exists between the Cry1Ac amino acid sequence and any amino acid sequences contained within the AD8 allergen database . Results indicate that no alignments of eight contiguous amino acid identities were detected when the Cry1Ac amino acid sequence was compared to known allergen sequences in the AD8 database. Together, these data demonstrate that the Cry1Ac protein does not share any relevant amino acid sequence similarities with known allergens, gliadins, or glutenins.

iii) Stability of the Cry1Ac protein in simulated gastric fluid

Digestibility of the full length *E. coli*-produced Cry1Ac protein in SGF was assessed by SDS-PAGE and western blot methods. The extent of Cry1Ac digestion was evaluated by visual analysis of stained polyacrylamide gels (**Figure 21**) or by visual analysis of developed X-ray film (**Figure 22**). The LOD of Cry1Ac by Colloidal Brilliant Blue G staining was 0.0025 µg or approximately 0.3% of the protein sample loaded (0.0025 µg divided by 0.8 µg of protein loaded in each lane of the gel; **Figure 21**, panel B). The LOD of Cry1Ac by western blotting was 0.5 ng or approximately 5% of the total protein loaded (0.5 ng divided by 10 ng of the protein loaded in each lane of the gel; **Figure 22**, panel B).

Visual examination of the Colloidal Brilliant Blue G stained gel (**Figure 21**, panel A) showed that the full-length Cry1Ac was digested below the LOD within 30 s of digestion in SGF (**Figure 21**, panel A, Lane 5). Therefore, at least 99.7% (100% - 0.3% = 99.7%) of the full-length Cry1Ac was digested within 30 s of incubation in SGF. No change in the full-length Cry1Ac band intensity was observed in the absence of pepsin in the experimental controls SGF P0 and SGF P7 (**Figure 21**, panel A, lanes 3 and 12) indicating that the digestion of the Cry1Ac was due to the proteolytic activity of pepsin present in SGF and not due to instability of the protein while incubated at pH ~1.2 at ~37 °C for 60 min. A protein fragment of ~4 kDa was observed throughout the digestion

(**Figure 21**, panel A, lanes 5-11), but appears to degrade to a smaller ~3.5 kDa fragment visible at the 30 min and 60 min time points on the gel.

Western blot analysis demonstrated that the full-length Cry1Ac was digested below the LOD within 30 s of incubation in SGF (**Figure 22**, panel A, Lane 5). Based on the western blot LOD for Cry1Ac in SGF and the observation that no full-length protein or immunoreactive bands were observed on the western blot at the 30 s digestion time point, it was concluded that at least 95% (100% - 5% = 95%) of the full-length Cry1Ac protein was digested within 30 s of incubation in SGF.

Because the transiently stable Cry1Ac fragment (~4 kDa) observed in the SDS-PAGE analysis was not cross-reactive with Cry1Ac-specific antibodies, the identity of this fragment was established by N-terminal sequencing. Two sequences were identified that match internal Cry1Ac sequences starting at amino acid positions 415 and 882. This indicates that the observed transiently stable fragment was derived from Cry1Ac. As shown in **Figure 23**, the observed sequences fall in a region that is highly conserved between the Cry1Ac and Cry1Ab proteins. The long history of safe use of the Bt Cry1 proteins strongly suggests that the observed transiently stable fragments do not pose any allergenicity concerns.

iv) Stability of the Cry1Ac protein in SGF followed by SIF

To better understand the digestive fate of the full-length Cry1Ac protein and the transiently stable fragment produced in SGF, Cry1Ac was exposed to digestion with pancreatin in simulated intestinal fluid (SIF) following the digestion in SGF. After digestion of Cry1Ac in SGF for 2 min, the reaction was quenched and exposed to further digestion in SIF. The digestibility of Cry1Ac in SIF following SGF digestion was evaluated by visual analysis of stained polyacrylamide gels (**Figure 24**). The gel was loaded with ~0.8 µg total protein (based on concentration of the protein prior to the digestion in SGF) for each of the SIF digestion time points. As expected, the full-length Cry1Ac protein was undetectable at the 2 min digestion time point, while the ~ 4 kDa fragment was present (**Figure 24**, lane 3). After exposure to SIF, the ~ 4 kDa fragment was not visible at the SEQ T1 (0.5 min) digestion time point (**Figure 24**, lane 7). These data clearly indicate that the ~ 4 kDa fragment degrades rapidly (< 1 min) upon exposure to SIF.

v) Assessment of human IgE binding to MON 87701, control, and conventional soybean extracts

A study was conducted to quantitatively evaluate the binding levels of IgE antibody in sera from clinically documented, soybean-allergic patients to protein extracts prepared from MON 87701 soybean seed, a conventional soybean control (variety A5547), and 17 commercial soybean reference varieties. The reference soybean varieties were used to establish the range in soybean-specific IgE binding. The reference varieties are commercially available and included high protein, high oil, and food-grade (tofu) soybean that are already on the market and are being used for human consumption.

Sera from 13 clinically documented, soybean-allergic subjects and five non-allergic subjects were used to assess IgE binding to each soybean extract. Only soybean-allergic subjects with a documented case history of soybean allergy and a positive Double-Blind

Placebo Controlled Food Challenge (DBPCFC) were included as soybean positive subjects in this study.

Aqueous extracts were prepared from the ground soybean seeds of MON 87701, the conventional soybean control, and reference varieties. These extracts were then analyzed for soybean-specific IgE antibody binding by a validated ELISA. Each soybean extract was tested in triplicate at a concentration of 10 µg/ml of total soybean protein. Soybean-specific IgE binding was quantified by interpolation against a soybean-specific IgE standard curve and was expressed as ng of IgE/ml of serum. The standard curve was created by loading serial dilutions of human serum PEI 163[3] containing a known amount of soybean-specific IgE, into wells coated with an assay-specific soybean extract from a commercial reference variety, Hensel. The bound soybean-specific IgE was detected using biotin-conjugated antihuman IgE polyclonal antibody and a horseradish peroxidase enzyme-conjugated streptavidin.

The IgE binding values obtained for the 17 reference soybean extracts were used to calculate a 99% tolerance interval for each subject's serum. The 99% tolerance interval represents the range of IgE binding for each subject's serum to the reference soybean extracts. The tolerance interval describes the range that includes 99% of the IgE binding values and that has a statistically predicted 95% confidence level. The IgE binding values obtained for extracts prepared from MON 87701 and the conventional soybean control were compared to the tolerance interval derived for each serum. All of the IgE binding values for MON 87701 and the control, shown in **Figure 25**, are within the reference soybean tolerance limits for each subject's serum. None of the soybean varieties showed IgE binding to sera from non-allergic subjects.

The results of this assessment demonstrate that soybean-specific IgE binding to endogenous allergens in MON 87701 and the conventional soybean control are comparable with the IgE binding to commercially available soybean varieties currently on the market.

[3] PEI is the abbreviation of the Paul-Ehrlich-Institute. PEI 163 is a single soybean allergic serum that contains a known amount of soybean-specific IgE adequate for constructing the standard curve.

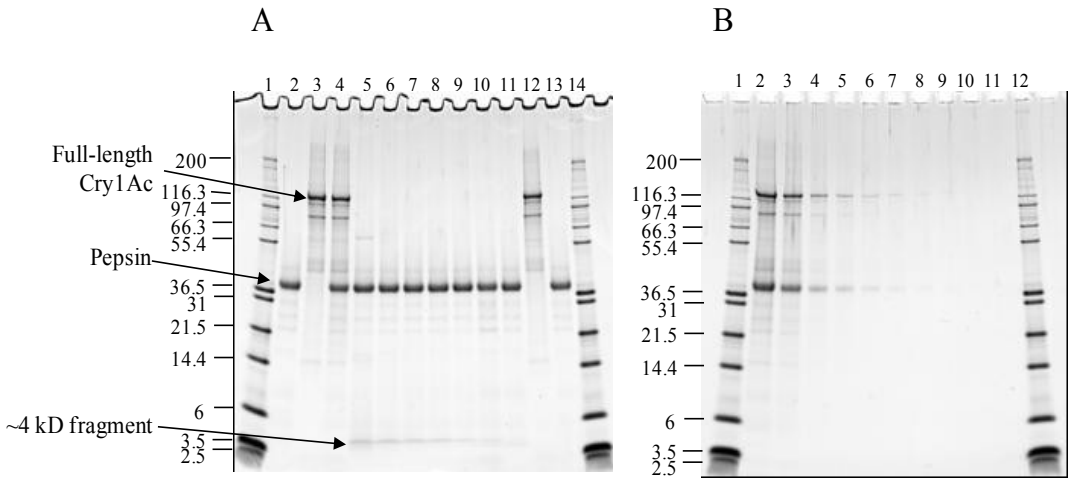


Figure 21. Colloidal Brilliant Blue G stained SDS-PAGE Gels of Cry1Ac Protein Digestion in SGF

Panel A corresponds to Cry1Ac protein digestion in SGF. Based on predigestion protein concentrations, 0.8 µg (total Cry1Ac protein) was loaded in lanes containing Cry1Ac protein. The incubation times are indicated. **Panel B** corresponds to the various amounts of SGF T0 loaded to estimate the LOD of the Cry1Ac protein. Approximate molecular weights (kDa) are shown on the left and correspond to the markers loaded in each gel.

Lane assignment for Panel A:

Lane	Sample	Incubation Time (min)
1	Molecular weight marker	—
2	SGF N0[1]	0
3	SGF P0	0
4	SGF T0	0
5	SGF T1	0.5
6	SGF T2	2
7	SGF T3	5
8	SGF T4	10
9	SGF T5	20
10	SGF T6	30
11	SGF T7	60
12	SGF P7	60
13	SGF N7	60
14	Molecular weight marker	—

Lane assignment for Panel B:

Lane	Sample	Amount (µg)
1	Molecular weight marker	—
2	SGF T0	0.8
3	SGF T0	0.4
4	SGF T0	0.1
5	SGF T0	0.05
6	SGF T0	0.02
7	SGF T0	0.01
8	SGF T0	0.005
9	SGF T0	0.0025
10	SGF T0	0.001
11	SGF T0	0.0005
12	Molecular weight marker	—

[1]A numerical code using the numbers 0 through 7 was used to distinguish incubation time points. N0, N7- negative controls (no test protein); P0, P7- protein control (no pepsin); T0-T7- incubation time point in SGF.

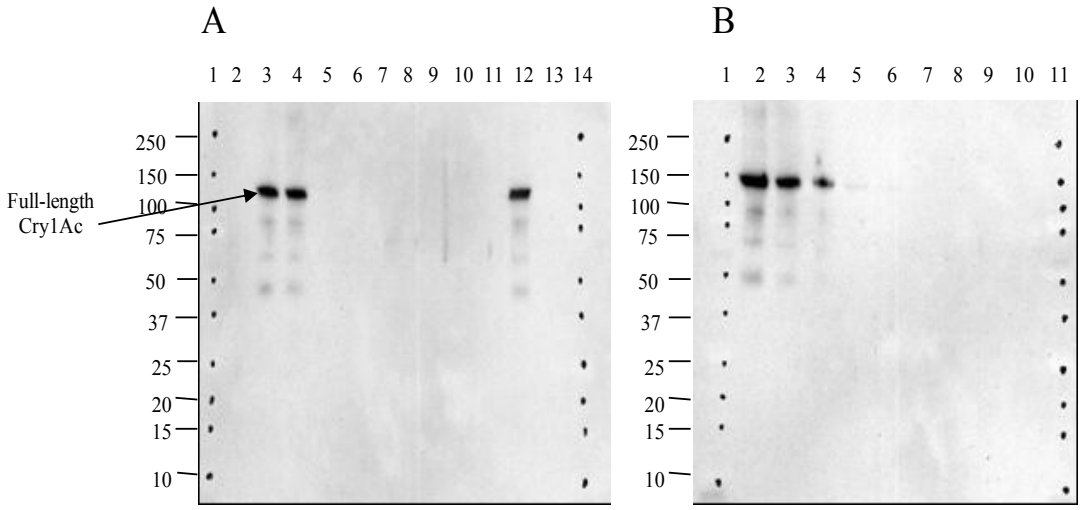


Figure 22. Western Blot Analysis of the Digestion of Cry1Ac Protein in SGF

Panel A corresponds to Cry1Ac protein digestion in SGF. Based on pre-digestion protein concentrations, 10 ng (total protein) was loaded in the lanes containing Cry1Ac protein. **Panel B** corresponds to the various amounts of SGF T0 loaded to estimate the LOD of the Cry1Ac protein. Approximate molecular weights (kDa) are shown on the left and correspond to the markers loaded in each gel. In both gels, Cry1Ac protein migrated to approximately 131.7 kDa. A 5 min exposure is shown. Blank or empty lanes are cropped and lanes renumbered relative to the raw data.

Lane assignment for Panel A:

Lane	Sample	Incubation Time (min)
1	Molecular weight marker	—
2	SGF N0[1]	0
3	SGF P0	0
4	SGF T0	0
5	SGF T1	0.5
6	SGF T2	2
7	SGF T3	5
8	SGF T4	10
9	SGF T5	20
10	SGF T6	30
11	SGF T7	60
12	SGF P7	60
13	SGF N7	60
14	Molecular weight marker	—

Lane assignment for Panel B:

Lane	Sample	Amount (ng)
1	Molecular weight marker	—
2	SGF T0	10
3	SGF T0	5
4	SGF T0	2.5
5	SGF T0	1
6	SGF T0	0.5
7	SGF T0	0.2
8	SGF T0	0.1
9	SGF T0	0.05
10	SGF T0	0.025
11	Molecular weight marker	—

[1]A numerical code using the numbers 0 through 7 was used to distinguish incubation time points. N0, N7- negative controls (no test protein), P0, P7- protein control (no pepsin), T0-T7- incubation time point in SGF.

Fragment 1, 415-424

```
cry1Ab          VDSLDEIPPQNNNVPPRQGFSHRLSH
cry1Ac          VDSLDEIPPQNNNVPPRQGFSHRLSH
observed seq        DEIPPQNNNV
```

Fragment 2, 882-894

```
cry1Ab          GEALARVKRAEKKWRDKREKLEWETNIVY
cry1Ac          GEALARVKRAEKKWRDKREKLEWETNIVY
observed seq        ARVKRAEKKWRDK
```

Figure 23. Cry1Ac and Cry1Ab Protein Sequences Identified in the ~4 kDa Stable Fragment from SGF

Alignment is shown between the N-terminal sequences from the ~4 kDa stable fragment observed in SGF analysis and the corresponding sequence regions of Cry1Ab (GenBank gene identification number 61221646) and Cry1Ac (GenBank gene identification number 117547).

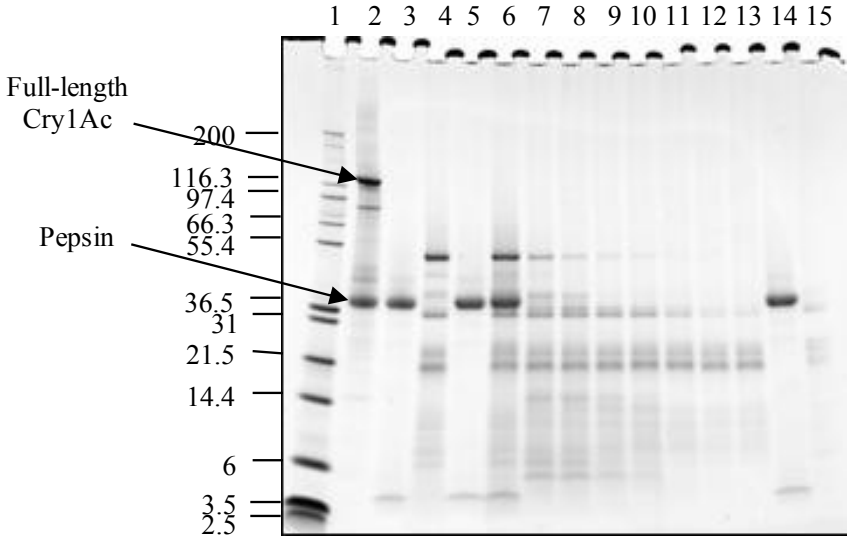


Figure 24. Colloidal Brilliant Blue G stained SDS-gel of the Cry1Ac Protein in SGF Followed by SIF

After digestion of the Cry1Ac protein in SGF for 2 min, the reaction was quenched and exposed to further digestion in SIF for the incubation times indicated.
Based on protein concentrations before digestion in SGF, 0.8 µg of total protein was loaded per lane containing Cry1Ac protein.

Lane	Sample	Incubation Time
1	Molecular weight marker	—
2	SEQ 0min	0
3	SEQ 2min	2 min[2]
4	SEQ N0[1]	0
5	SEQ P0	0
6	SEQ T0	0
7	SEQ T1	0.5 min
8	SEQ T2	2 min
9	SEQ T3	5 min
10	SEQ T4	10 min
11	SEQ T5	30 min
12	SEQ T6	1 h
13	SEQ T7	2 h
14	SEQ P7	2 h
15	SEQ N7	2 h

[1]A numerical code using the numbers 0 through 7 was used to distinguish incubation time points. N0, N7- negative controls (no test protein); P0, P7- protein control (no pancreatin); T0-T7- incubation time point in sequential digestion assay (SEQ).

[2]Indicates incubation time in SGF.

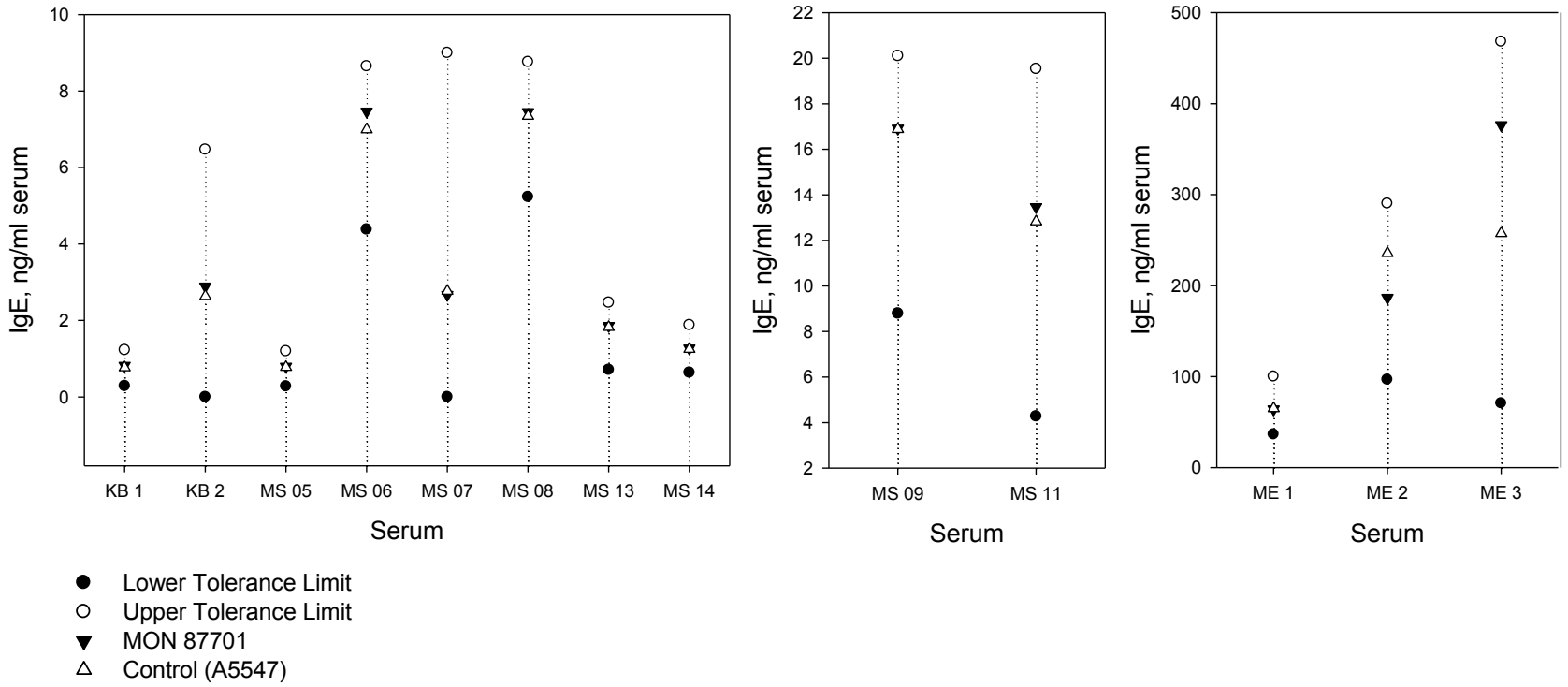


Figure 25. Serum IgE Binding Values for MON 87701, Conventional Control (A5547), and the Tolerance Limits for 17 Conventional References.

The lower and upper tolerance limits for 99% tolerance intervals with 95% confidence for each serum are the result of a tolerance interval analysis for 17 commercial soybean varieties. Lower limits of the tolerance intervals that were calculated as less than zero were reported as zero in the analysis. Data are presented in three graphs due to the difference in IgE concentration range between sera.

2.6 Characterisation of Other Novel Protein

a) Identification

Not applicable

b) Toxicity

Not applicable

2.7 Comparative Analysis

Compositional analyses were conducted on the forage and seed collected from MON 87701, the conventional soybean control (A5547), and twenty unique commercial conventional soybean varieties grown at five replicated trial sites (AL, AR, GA, IL, and NC) in a 2007 U.S. field production. Seeds were planted in a randomized complete block design with three replicates per block for MON 87701, the control, and reference soybean varieties. Samples from all three replicates of MON 87701 and control plots at each site were analyzed, whereas one replicate of the twenty unique commercial conventional soybean varieties planted across sites was analyzed. Analysis of a single replicate from each unique reference variety was sufficient to establish a 99% tolerance interval for each analyte, as described below. All samples were collected from plants grown under normal agronomic field conditions for their respective geographic regions.

A total of 64 compositional analytes (seven in forage and 57 in seed) were evaluated. Compositional analyses of the forage samples included proximates (ash, fat, moisture, and protein), carbohydrates by calculation, acid detergent fiber (ADF), and neutral detergent fiber (NDF). Seed samples were analyzed for proximates (ash, fat, moisture, and protein), carbohydrates by calculation, ADF, NDF, amino acids, fatty acids (C8-C22), trypsin inhibitors, phytic acid, lectin, isoflavones (daidzein, glycitein, and genistein), vitamin E, raffinose, and stachyose.

In a further assessment, the composition analysis data were statistically compared to that of the conventional soybean control. Of the evaluated components, nine fatty acids in harvested seed had more than 50% of the observations below the assay limit of quantitation (LOQ) and, as a result, were excluded from the statistical analysis. Therefore, the remaining 55 compositional analytes (seven in forage and 48 in seed) were statistically analyzed. A summary of the statistical results included the calculation of least square means, standard errors, and the range of observed values for MON 87701 and the conventional soybean control. A 99% tolerance interval was also calculated for each analyte that represented, with 95% confidence, 99% of the values contained in the population of commercial conventional soybean varieties. The compositional analysis data across sites (combined-site) are presented in **Table 10** for forage and in **Table 11** (nutrients) and **Table 12** (antinutrients) for harvested seed.

Each analyte for MON 87701 was statistically compared to that of the conventional soybean control in a combined-site analysis and five individual-site analyses using SAS, version 9.1, software. The combined-site analysis used a mixed model analysis

with both site and site by analyte designated as random effects. Site was treated as a random effect since it is assumed that the sites are a representative sample from a population of sites (Littell et al., 2006). The overall data set was examined for evidence of biologically relevant changes by first examining combined-site differences, followed by the individual-site assessments. Additionally, each mean test value that differed ($p<0.05$) from the control was compared to the 99% tolerance interval generated from the commercial conventional soybean varieties. A summary of the significant differences ($p<0.05$) between MON 87701 and the conventional soybean control for forage and harvested seed are presented in **Table 13**. Finally, this comparative evaluation also considered the natural ranges in soybean component levels published in the scientific literature and the International Life Sciences Institute - Crop Composition Database (ILSI-CCD)[4]. Reported literature and ILSI-CCD ranges for the analytical components present in forage and harvested seed are shown in **Table 14**.

MON 87701 is as safe and nutritious as conventional soybean based on a comprehensive compositional and nutritional assessment as described above. The combined-site analysis for seed and forage samples showed no significant difference ($p>0.05$) between MON 87701 and the conventional control for 40 of 55 comparisons. Statistically significant differences were detected for 15 analytes: alanine, 22:0 behenic acid, carbohydrates, daidzein, glycine, histidine, isoleucine, leucine, lysine, protein, serine, threonine, trypsin inhibitor, valine, and vitamin E. However, the magnitude of these differences were generally small (most <5%), were not observed consistently across all sites (individual-site analyses), and the mean values for MON 87701 were within the calculated 99% tolerance interval for the population of commercial conventional soybean varieties grown concurrently at the same field sites. Therefore, it is concluded that since the measured values fall within the range of natural variability for these soybean analytes, they were not regarded as biologically relevant. Soybean seed and forage analyte values were also comparable to values published in the scientific literature and reported in ILSI-CCD. This further supports the conclusion that the soybean seed and forage from MON 87701 are compositionally equivalent to those of conventional soybean.

The following sections provide a detailed assessment of the biological relevance of the differences observed between MON 87701 and the conventional soybean control based on the magnitude of the difference, reproducibility across sites, and comparison of mean analyte values to the 99% tolerance interval.

a) Key nutrients etc

i) Level of nutrients in soybean forage

The levels of nutrients from the combined-site analysis for forage are summarized in **Table 10**. No significant differences ($p\geq0.05$) were detected in the seven comparisons made between MON 87701 and the conventional soybean control in the combined-site analysis. Forage analyte values were within the 99% tolerance for conventional soybean varieties and comparable to values published in the scientific literature and reported in ILSI-CCD (2006).

[4] International Life Science Institute Crop Composition Database. Version 3.0. 2006. Available at http://www.cropcomposition.org/.

This supports the conclusion that the forage from MON 87701 is compositionally equivalent to that of conventional soybean.

ii) Level of nutrients in soybean seed

The levels of nutrients from the combined-site analysis for soybean seed are summarized in **Table 11**. No significant differences ($p \geq 0.05$) were detected in 27 of 40 nutrient analyte comparisons made between MON 87701 and the conventional soybean control in the combined-site analysis. Significant differences ($p < 0.05$) were detected for 13 nutrient analytes: protein, nine amino acids (alanine, glycine, histidine, isoleucine, leucine, lysine, serine, threonine, and valine), 22:0 behenic acid, carbohydrates and vitamin E (**Table 13**). The magnitude of these differences were generally small (most <5%), except in the case of vitamin E described below.

Biological relevance of these differences was assessed based on the magnitude of the difference, reproducibility across sites, and comparison of mean analyte values to the 99% tolerance interval for the population of commercial conventional soybean varieties grown concurrently at the same field sites. The protein content in MON 87701 was slightly increased, but at a small order of magnitude (3.87%) compared to the conventional soybean control. Similarly, the levels of all nine amino acids are slightly increased (2.15-4.63%) when compared to the conventional soybean control. Based on the relatively high total protein content (35-45%) in commercially available conventional soybean (**Table 14**), and the relatedness of amino acid content to total protein content, it is not unexpected that several of the amino acids in MON 87701 also had elevated levels that are significantly different from the conventional soybean control, and at a magnitude similar to that of protein. Examination of the reproducibility across sites for protein and the nine amino acids found to be different in the combined-site analysis, shows that only one analyte (histidine) was significantly different at more than one site (**Table 13**). Histidine content in MON 87701 was significantly higher than the conventional soybean control at two sites but the magnitudes of the differences were small (<6%). The differences for protein and nine amino acids are not considered to be biologically relevant changes in composition because the increase in levels were generally small, only one analyte (histidine) showed differences at more than one site, and the mean analyte levels were within the 99% tolerance interval established by commercial conventional soybean varieties grown concurrently at the same field sites. Furthermore, analyte values were comparable to values reported in ILSI-CCD.

The magnitude of differences between MON 87701 and the conventional soybean control for carbohydrates and 22:0 behenic acid were also considered to be relatively small (6.10% and 4.33%, respectively). Examination of the reproducibility across sites shows that only 22:0 behenic acid was significantly different at more than one site. Behenic acid content in MON 87701 was significantly higher than the conventional soybean control at two sites and the magnitudes of the differences were small (<8.3%). The differences for 22:0 behenic acid and carbohydrate are not considered to be biologically relevant changes in composition because the increase in levels were generally small, the differences were not reproducible at more than two sites, and the mean levels were within the 99% tolerance interval established by

commercial conventional soybean varieties grown concurrently at the same field sites. Furthermore, analyte values were comparable to those reported in ILSI-CCD.

Vitamin E levels were significantly higher (23.26%) in MON 87701 compared to the conventional soybean control. Examination of the reproducibility across sites shows that vitamin E was significantly higher than the conventional soybean control in four of five individual-site analyses, with the magnitude of differences ranging between 17-37% (**Table 13**). The differences are not biologically relevant changes in composition given that the mean levels of vitamin E in MON 87701 from the combined-site and individual-site analyses were all well within the 99% tolerance interval established by commercial conventional soybean varieties grown concurrently at the same field sites. From a nutritional perspective, vitamin E is not listed as a key nutrient in soybean by OECD (2001) for food and feed uses, though soybean oil is recognized as a source of vitamin E in the diet (Eitenmiller, 1997). No dietary impact is expected as the vitamin E levels are also comparable to the values reported in ILSI-CCD.

Based on the data and information presented above, it is concluded that the seed from MON 87701 is compositionally equivalent to conventional soybean with regard to the levels of nutrients. The differences in nutrients were limited in number, not consistently observed across all sites, and reflect the natural variation of conventional soybean, which further supports the compositional equivalence of MON 87701 to conventional soybean.

b) Other constitutes

- **Levels of antinutrients in soybean seed**

Soybean seed contains several well-described antinutritional factors (OECD, 2001) that include: trypsin inhibitors, lectins, isoflavones (daidzein, genistein and glycitein), stachyose, raffinose, and phytic acid. The levels of these components in the seed of MON 87701, the conventional soybean control (A5547), and 20 commercial conventional soybean varieties were evaluated. The levels of antinutrients from the combined-site analysis for soybean seed are summarized in **Table 12** No significant differences ($p \geq 0.05$) were detected in six of eight antinutrient analyte comparisons made between MON 87701 and the conventional soybean control in the combined-site analysis. Statistically significant differences were detected for trypsin inhibitor and daidzen (**Table 13**); however, the magnitude of these differences were small (<11%). Biological relevance of these differences was assessed based on the magnitude of the difference, reproducibility across sites, and comparison of mean analyte values to the 99% tolerance interval.

Trypsin inhibitors are heat-labile antinutrients that interfere with the digestion of proteins and result in decreased animal growth (Liener, 1994). Lectins are also heat labile, and can inhibit growth and cause death in animals if raw soybean is consumed (Liener, 1994). Both trypsin inhibitors and lectins are inactivated during processing of soybean protein products or soybean meal and, when processed appropriately, the final edible soybean fractions should contain minimal levels of these antinutrients. There were no significant differences ($p \geq 0.05$) in lectin levels between MON 87701 and conventional soybean control seed. The level of trypsin inhibitor was significantly

lower (-8.79%) in MON 87701 than the conventional control in the combined-site analysis (**Table 12**). Examination of the reproducibility across sites shows that the trypsin inhibitor levels were significantly lower (-20.48%) than the conventional soybean control in only one of five sites (**Table 13**). The difference for trypsin inhibitor is not considered to be a biologically relevant change in composition because the increase in level was small, the difference was not reproducible at more than one site, and the mean levels were within the 99% tolerance interval established by commercial conventional soybean varieties grown concurrently at the same field sites. Furthermore, the trypsin inhibitor values in MON 87701 were comparable to those reported in ILSI-CCD.

There are three principle isoflavones in soybean seed, namely, daidzein, genistein, and glycitein. Although they have been reported to possess biochemical activities including estrogenic and anti-estrogenic effects, it is not universally accepted that the isoflavones are antinutrients as they have also been reported to have beneficial antioxidant, anticarcinogenic and heart-healthy hypocholesterolemic effects (OECD, 2001). It is well documented that isoflavone levels in soybean seed are highly variable and are greatly influenced by many factors (Messina, 2001; Nelson et al., 2001; OECD, 2001).

There were no significant differences ($p \geq 0.05$) in glycitein levels between MON 87701 and conventional soybean control seed. The level of daidzein was significantly higher in MON 87701 (10.36%) than the conventional control in the combined-site analysis (**Table 12**). Examination of the reproducibility across sites shows that the daidzein levels were significantly higher in MON 87701 than the conventional soybean control in two of five individual-site analyses (<16.7%). These results are not unexpected because, as described above, soybean isoflavone levels are greatly influenced by many factors, ranging from environmental conditions, variety, and agronomic practices. The difference for daidzein is not considered to be biologically relevant because the increase in level was small, the difference was not reproducible at more than one site, and the mean levels were within the 99% tolerance interval established by commercial conventional soybean varieties grown concurrently at the same field sites. Furthermore, the daidzein values in MON 87701 were comparable to the values reported in ILSI-CCD.

Antinutrients that showed no differences in the combined-site analysis are not expected to raise biological, nutritional or toxicological concerns. However, further consideration and background is provided to complete the discussion for the group of antinutrients.

As discussed earlier, genistein is one of three principle isoflavones. A difference ($P < 0.05$) in genistein levels between the seed of MON 87701 and the conventional control was detected in one of the five sites (**Table 13**). Stachyose and raffinose are low molecular weight carbohydrates present in soybean seed that are considered to be antinutrients due to their consumption causing flatulence. Significant differences ($P < 0.05$) were detected in stachyose levels in seed from MON 87701 and the conventional soybean control at two sites (**Table 13**). The differences for genistein and stachyose were not consistently observed across all sites and no trends were observed. Furthermore, the mean levels of these components in MON 87701 seed were within the 99% tolerance interval for commercial conventional soybean

varieties. Therefore, these differences are not considered biologically relevant from a food or feed safety or nutritional perspective.

Phytic acid present in soybean seed chelates mineral nutrients, including calcium, magnesium, potassium, iron and zinc, rendering them biologically unavailable to monogastric animals consuming the seed (Liener, 2000). Unlike trypsin inhibitors, phytic acid is not heat labile, and remains stable through most soybean processing steps. No significant differences were detected in combined-site or individual-site analyses of phytic acid levels in seed from MON 87701 and its conventional soybean control (A5547).

Based on the data and information presented above, it is concluded that the seed from MON 87701 is compositionally equivalent to conventional soybean with regard to the levels of antinutrients. The differences in antinutrients were limited in number, not consistently observed across all sites, and reflect the natural variation of conventional soybean, which further supports the compositional equivalence of MON 87701 to conventional soybean.

c) Allergenic proteins

Not applicable

Table 10. Statistical Summary of Combined Site Soybean Forage Fiber and Proximate Content for MON 87701 vs. the Conventional Control (A5547)

Component (Units)[1]	MON 87701 Mean (S.E.) [Range]	A5547 Mean (S.E.) [Range]	Difference (Test minus Control) Mean (S.E.) [Range]	95% CI (Lower, Upper)	p-Value	Commercial (Range) [99% Tolerance Interval[2]]
Fiber (%DW)						
Acid Detergent Fiber	37.17 (1.72) [30.04 - 58.25]	36.53 (1.72) [27.42 - 42.06]	0.65 (2.01) [-12.02 - 10.46]	-3.42, 4.71	0.749	(27.99 - 47.33) [14.93, 56.87]
Neutral Detergent Fiber	47.16 (2.00) [37.02 - 55.99]	45.57 (2.00) [34.23 - 64.19]	1.59 (2.48) [-18.07 - 18.76]	-3.50, 6.68	0.526	(30.96 - 54.55) [21.51, 66.01]
Proximate (%DW, unless noted)						
Ash	5.84 (0.30) [5.05 - 7.46]	6.32 (0.30) [5.10 - 8.13]	-0.48 (0.33) [-1.72 - 0.92]	-1.25, 0.29	0.190	(4.77 - 8.54) [2.46, 10.14]
Carbohydrates	71.43 (1.12) [68.29 - 76.73]	70.97 (1.12) [63.68 - 74.26]	0.47 (0.61) [-2.28 - 4.62]	-0.79, 1.73	0.452	(60.61 - 77.26) [56.93, 85.88]
Moisture (% FW)	72.86 (1.19) [70.10 - 76.80]	73.41 (1.19) [69.40 - 78.10]	-0.55 (0.49) [-2.30 - 1.70]	-1.67, 0.58	0.296	(66.50 - 80.20) [57.84, 88.56]
Protein	17.39 (1.07) [13.56 - 20.03]	17.07 (1.07) [14.20 - 23.29]	0.32 (0.60) [-3.57 - 2.22]	-0.90, 1.54	0.591	(12.68 - 22.92) [7.05, 27.27]
Total Fat	5.30 (0.34) [3.60 - 6.82]	5.65 (0.34) [4.23 - 7.23]	-0.35 (0.26) [-2.76 - 0.70]	-0.89, 0.19	0.195	(3.48 - 7.88) [1.11, 9.11]

[1]DW = dry weight; FW = fresh weight; S.E. = standard error; CI = Confidence Interval.
[2]With 95% confidence, interval contains 99% of the values expressed in the population of commercial soybean varieties. Negative limits were set to zero.

Table 11. Statistical Summary of Combined-Site Soybean Seed Proximate, Amino Acid, Fatty Acid, Fiber and Vitamin Content for MON 87701 vs. the Conventional Control (A5547)

Component (Units)[1]	MON 87701 Mean (S.E.) [Range]	A5547 Mean (S.E.) [Range]	Difference (Test minus Control) Mean (S.E.) [Range]	95% CI (Lower, Upper)	p-Value	Commercial (Range) [99% Tolerance Interval[2]]
Proximate (% DW, unless noted)						
Ash	5.20 (0.18) [4.70 - 5.90]	5.14 (0.18) [4.70 - 5.88]	0.054 (0.043) [-0.14 - 0.21]	-0.046, 0.15	0.246	(4.32 - 5.62) [3.74, 6.45]
Carbohydrates	34.22 (1.50) [21.58 - 39.61]	36.44 (1.50) [29.88 - 43.48]	-2.22 (1.02) [-15.00 - 2.56]	-4.31, -0.14	0.037	(31.97 - 38.00) [28.17, 40.99]
Moisture (% FW)	7.52 (0.38) [5.86 - 10.70]	6.84 (0.38) [5.44 - 8.74]	0.68 (0.47) [-0.96 - 4.70]	-0.28, 1.64	0.159	(5.48 - 11.70) [1.45, 12.81]
Protein	39.27 (0.86) [36.49 - 42.23]	37.80 (0.86) [32.29 - 41.87]	1.46 (0.54) [-0.73 - 6.49]	0.24, 2.68	0.023	(38.14 - 42.66) [35.30, 45.38]
Total Fat	20.29 (0.78) [17.33 - 23.08]	20.12 (0.77) [17.24 - 22.55]	0.17 (0.39) [-1.82 - 1.98]	-0.71, 1.05	0.670	(17.90 - 23.56) [14.74, 25.18]
Amino Acid (% DW)						
Alanine	1.72 (0.029) [1.66 - 1.84]	1.69 (0.029) [1.59 - 1.82]	0.036 (0.016) [-0.034 - 0.099]	0.0044, 0.068	0.027	(1.66 - 1.93) [1.49, 2.02]
Arginine	2.68 (0.069) [2.36 - 3.00]	2.58 (0.069) [2.37 - 2.89]	0.096 (0.058) [-0.16 - 0.31]	-0.039, 0.23	0.138	(2.54 - 2.99) [2.22, 3.25]
Aspartic Acid	4.90 (0.10) [4.61 - 5.26]	4.85 (0.10) [4.46 - 5.34]	0.053 (0.055) [-0.23 - 0.31]	-0.059, 0.17	0.339	(4.74 - 5.50) [4.22, 5.96]
Cystine	0.62 (0.014) [0.57 - 0.67]	0.61 (0.014) [0.56 - 0.69]	0.0051 (0.014) [-0.11 - 0.066]	-0.024, 0.034	0.718	(0.53 - 0.68) [0.45, 0.77]
Glutamic Acid	7.65 (0.15) [7.25 - 8.21]	7.53 (0.15) [6.89 - 8.26]	0.12 (0.084) [-0.29 - 0.46]	-0.056, 0.29	0.177	(7.53 - 8.72) [6.60, 9.37]

Table 11. Statistical Summary of Combined-Site Soybean Seed Proximate, Amino Acid, Fatty Acid, Fiber and Vitamin Content for MON 87701 vs. the Conventional Control (A5547) (cont'd.)

Component (Units)[1]	MON 87701 Mean (S.E.) [Range]	A5547 Mean (S.E.) [Range]	Difference (Test minus Control) Mean (S.E.) [Range]	95% CI (Lower, Upper)	p-Value	Commercial (Range) [99% Tolerance Interval[2]]
Amino Acid (% DW)						
Glycine	1.75 (0.026) [1.63 - 1.89]	1.70 (0.026) [1.64 - 1.85]	0.049 (0.017) [-0.0052 - 0.12]	0.014, 0.083	0.007	(1.67 - 1.99) [1.49, 2.09]
Histidine	1.12 (0.015) [1.05 - 1.18]	1.08 (0.015) [1.03 - 1.15]	0.043 (0.011) [-0.00077 - 0.090]	0.021, 0.064	<0.001	(1.04 - 1.24) [0.94, 1.31]
Isoleucine	1.81 (0.037) [1.68 - 1.99]	1.76 (0.037) [1.64 - 1.96]	0.052 (0.020) [-0.044 - 0.12]	0.0061, 0.098	0.031	(1.73 - 2.02) [1.54, 2.14]
Leucine	3.04 (0.066) [2.82 - 3.36]	2.94 (0.066) [2.73 - 3.29]	0.095 (0.040) [-0.044 - 0.23]	0.0018, 0.19	0.046	(2.93 - 3.32) [2.64, 3.52]
Lysine	2.74 (0.060) [2.48 - 2.99]	2.62 (0.060) [2.42 - 2.91]	0.12 (0.046) [-0.12 - 0.39]	0.028, 0.21	0.012	(2.35 - 3.15) [2.05, 3.47]
Methionine	0.53 (0.012) [0.48 - 0.58]	0.53 (0.012) [0.47 - 0.59]	0.0043 (0.014) [-0.094 - 0.080]	-0.023, 0.032	0.754	(0.49 - 0.62) [0.42, 0.68]
Phenylalanine	2.15 (0.056) [1.91 - 2.48]	2.04 (0.056) [1.91 - 2.38]	0.11 (0.052) [-0.036 - 0.41]	-0.013, 0.23	0.073	(1.97 - 2.44) [1.66, 2.64]
Proline	2.01 (0.035) [1.86 - 2.16]	1.96 (0.035) [1.85 - 2.12]	0.042 (0.021) [-0.058 - 0.11]	-0.0069, 0.091	0.082	(1.92 - 2.25) [1.73, 2.35]
Serine	2.03 (0.032) [1.90 - 2.19]	1.96 (0.032) [1.87 - 2.13]	0.060 (0.019) [0.010 - 0.14]	0.020, 0.10	0.004	(1.96 - 2.30) [1.75, 2.38]
Threonine	1.60 (0.020) [1.50 - 1.72]	1.55 (0.020) [1.49 - 1.68]	0.046 (0.016) [-0.016 - 0.13]	0.0078, 0.084	0.024	(1.54 - 1.74) [1.40, 1.83]

Table 11. Statistical Summary of Combined-Site Soybean Seed Proximate, Amino Acid, Fatty Acid, Fiber and Vitamin Content for MON 87701 vs. the Conventional Control (A5547) (cont'd.)

Component (Units)[1]	MON 87701 Mean (S.E.) [Range]	A5547 Mean (S.E.) [Range]	Difference (Test minus Control) Mean (S.E.) [Range]	95% CI (Lower, Upper)	p-Value	Commercial (Range) [99% Tolerance Interval[2]]
Amino Acid (% DW)						
Tryptophan	0.51 (0.0068) [0.47 - 0.54]	0.50 (0.0068) [0.46 - 0.53]	0.011 (0.0067) [-0.039 - 0.075]	-0.0024, 0.025	0.102	(0.47 - 0.55) [0.43, 0.59]
Tyrosine	1.13 (0.034) [0.96 - 1.33]	1.10 (0.034) [0.98 - 1.22]	0.039 (0.029) [-0.11 - 0.25]	-0.028, 0.11	0.213	(1.04 - 1.31) [0.85, 1.48]
Valine	1.92 (0.032) [1.80 - 2.07]	1.86 (0.032) [1.76 - 2.04]	0.053 (0.022) [-0.033 - 0.12]	0.0029, 0.10	0.040	(1.83 - 2.13) [1.64, 2.22]
Fatty Acid (% Total FA)						
10:0 Capric Acid	0.20 (0.014) [0.14 - 0.25]	0.21 (0.014) [0.16 - 0.26]	-0.010 (0.020) [-0.11 - 0.048]	-0.053, 0.032	0.607	(0.15 - 0.27) [0.065, 0.34]
14:0 Myristic Acid	0.093 (0.0031) [0.082 - 0.10]	0.094 (0.0031) [0.083 - 0.11]	-0.00056 (0.0019) [-0.0085 - 0.0025]	-0.0048, 0.0037	0.769	(0.064 - 0.097) [0.052, 0.12]
16:0 Palmitic Acid	11.80 (0.12) [11.32 - 12.30]	11.88 (0.12) [11.50 - 12.13]	-0.079 (0.081) [-0.72 - 0.40]	-0.27, 0.11	0.359	(9.80 - 12.38) [8.88, 13.53]
16:1 Palmitoleic Acid	0.092 (0.0033) [0.073 - 0.11]	0.095 (0.0033) [0.078 - 0.11]	-0.0028 (0.0029) [-0.018 - 0.015]	-0.0097, 0.0041	0.372	(0.073 - 0.14) [0.037, 0.15]
17:0 Heptadecanoic Acid	0.094 (0.0021) [0.084 - 0.10]	0.093 (0.0021) [0.082 - 0.099]	0.0011 (0.0018) [-0.0064 - 0.0074]	-0.0030, 0.0052	0.553	(0.076 - 0.10) [0.066, 0.11]
17:1 Heptadecenoic Acid	0.041 (0.0032) [0.023 - 0.048]	0.041 (0.0032) [0.019 - 0.047]	-0.00009 (0.0040) [-0.020 - 0.022]	-0.0092, 0.0090	0.981	(0.020 - 0.064) [0.0058, 0.083]
18:0 Stearic Acid	4.59 (0.22) [3.97 - 5.36]	4.70 (0.22) [4.03 - 5.36]	-0.12 (0.11) [-0.57 - 0.29]	-0.38, 0.14	0.328	(3.21 - 5.24) [1.88, 6.25]

Table 11. Statistical Summary of Combined-Site Soybean Seed Proximate, Amino Acid, Fatty Acid, Fiber and Vitamin Content for MON 87701 vs. the Conventional Control (A5547)(cont'd.)

Component (Units)[1]	MON 87701 Mean (S.E.) [Range]	A5547 Mean (S.E.) [Range]	Difference (Test minus Control) Mean (S.E.) [Range]	95% CI (Lower, Upper)	p-Value	Commercial (Range) [99% Tolerance Interval[2]]
18:1 Oleic Acid	22.35 (1.28) [19.21 - 26.64]	22.71 (1.28) [20.34 - 28.78]	-0.36 (0.49) [-3.16 - 2.04]	-1.51, 0.79	0.486	(16.69 - 35.16) [5.01, 42.01]
18:2 Linoleic Acid	52.16 (0.95) [49.32 - 54.63]	51.76 (0.95) [47.18 - 54.07]	0.40 (0.38) [-1.35 - 2.68]	-0.48, 1.29	0.320	(44.17 - 57.72) [38.57, 66.94]
18:3 Linolenic Acid	7.24 (0.45) [5.55 - 8.41]	7.11 (0.45) [5.34 - 8.26]	0.13 (0.12) [-0.40 - 0.68]	-0.13, 0.40	0.276	(4.27 - 8.81) [2.69, 10.81]
20:0 Arachidic Acid	0.51 (0.025) [0.41 - 0.58]	0.51 (0.025) [0.41 - 0.57]	-0.0027 (0.013) [-0.044 - 0.047]	-0.032, 0.026	0.836	(0.36 - 0.55) [0.23, 0.64]
20:1 Eicosenoic Acid	0.24 (0.012) [0.19 - 0.28]	0.23 (0.012) [0.18 - 0.28]	0.0044 (0.010) [-0.065 - 0.046]	-0.020, 0.029	0.683	(0.21 - 0.30) [0.16, 0.33]
20:2 Eicosadienoic Acid	0.040 (0.0030) [0.020 - 0.054]	0.042 (0.0030) [0.020 - 0.047]	-0.0024 (0.0042) [-0.024 - 0.011]	-0.012, 0.0068	0.585	(0.016 - 0.054) [0.0029, 0.083]
22:0 Behenic Acid	0.56 (0.028) [0.46 - 0.65]	0.54 (0.028) [0.45 - 0.65]	0.023 (0.0084) [-0.00071 - 0.078]	0.0041, 0.042	0.022	(0.38 - 0.59) [0.30, 0.67]
Fiber (% DW)						
Acid Detergent Fiber	15.58 (0.49) [13.53 - 17.05]	15.62 (0.49) [14.00 - 19.02]	-0.042 (0.58) [-2.84 - 1.88]	-1.37, 1.28	0.943	(12.79 - 17.98) [11.13, 20.21]
Neutral Detergent Fiber	17.33 (0.70) [15.06 - 21.80]	17.28 (0.70) [15.02 - 22.45]	0.057 (0.74) [-6.43 - 4.47]	-1.67, 1.78	0.940	(13.32 - 23.57) [7.24, 28.70]
Vitamin (mg/100g DW)						
Vitamin E	7.69 (0.52) [6.36 - 9.62]	6.24 (0.52) [4.88 - 7.94]	1.45 (0.27) [0.57 - 2.25]	0.81, 2.09	<0.001	(1.65 - 8.08) [0, 11.09]

[1]DW = dry weight; FW = fresh weight; FA = fatty acid; S.E. = standard error; CI = Confidence Interval.
[2]With 95% confidence, interval contains 99% of the values expressed in the population of commercial soybean varieties. Negative limits were set to zero.

Table 12. Statistical Summary of Combined-Site Soybean Seed Antinutrient and Isoflavone Content for MON 87701 vs. the Conventional Control (A5547)

Component (Units)[1]	MON 87701 Mean (S.E.) [Range]	A5547 Mean (S.E.) [Range]	Difference (Test minus Control) Mean (S.E.) [Range]	95% CI (Lower, Upper)	p-Value	Commercial (Range) [99% Tolerance Interval[2]]
Antinutrient (% DW, unless noted)						
Lectin (H.U./mg FW)	0.96 (0.19) [0.062 - 2.01]	0.72 (0.19) [0.28 - 1.28]	0.24 (0.25) [-0.88 - 1.42]	-0.27, 0.74	0.354	(0.090 - 2.47) [0, 3.40]
Phytic Acid	1.85 (0.12) [1.39 - 2.29]	1.97 (0.12) [1.31 - 2.66]	-0.11 (0.097) [-0.53 - 0.31]	-0.34, 0.11	0.276	(1.10 - 2.32) [0.54, 3.05]
Raffinose	1.33 (0.19) [0.49 - 1.70]	1.34 (0.19) [0.43 - 1.85]	-0.0086 (0.074) [-0.32 - 0.19]	-0.18, 0.16	0.910	(0.52 - 1.62) [0.038, 2.24]
Stachyose	4.59 (0.63) [1.83 - 6.42]	4.93 (0.63) [2.27 - 6.65]	-0.34 (0.22) [-1.00 - 0.73]	-0.83, 0.16	0.156	(1.97 - 5.55) [0.99, 7.93]
Trypsin Inhibitor (TIU/mg DW)	26.06 (1.24) [21.65 - 32.53]	28.57 (1.24) [22.49 - 34.20]	-2.51 (0.96) [-7.75 - 6.33]	-4.48, -0.54	0.014	(20.84 - 37.24) [13.58, 46.02]
Isoflavone (mg/kg DW)						
Daidzein	667.54 (108.30) [188.96 - 983.26]	604.88 (108.30) [198.95 - 830.65]	62.65 (25.68) [-27.87 - 178.54]	3.56, 121.74	0.040	(213.98 - 1273.94) [0, 1585.14]
Genistein	655.57 (88.52) [214.73 - 863.84]	594.58 (88.52) [244.95 - 760.87]	60.99 (36.27) [-30.22 - 178.22]	-23.10, 145.09	0.132	(148.06 - 1024.50) [0, 1352.86]
Glycitein	164.87 (21.23) [61.08 - 228.79]	156.93 (21.23) [61.28 - 227.25]	7.94 (13.22) [-49.56 - 88.71]	-22.57, 38.44	0.564	(32.42 - 208.45) [0, 272.12]

[1]DW = dry weight; FW = fresh weight; FA = fatty acid; S.E. = standard error; CI = Confidence Interval.
[2]With 95% confidence, interval contains 99% of the values expressed in the population of commercial soybean varieties. Negative limits were set to zero.

Table 13. Summary of Differences (p<0.05) for the Comparison of Soybean Component Levels for MON 87701 vs. the Conventional Control (A5547)

Component (Units)[1]	MON 87701 Mean	A5547 Mean	Mean Difference (Test minus Control)		MON 87701 Range	Commercial Tolerance Interval[2]
			Mean Difference (% of A5547)	Significance (p-Value)		
Statistical Differences Observed in *Combined Site* Analysis						
Seed Proximate (% DW)						
Protein	39.27	37.80	3.87	0.023	[36.49 – 42.23]	[35.30, 45.38]
Carbohydrates	34.22	36.44	-6.10	0.037	[21.58 - 39.61]	[28.17, 40.99]
Seed Amino Acid (% DW)						
Alanine	1.72	1.69	2.15	0.027	[1.66 - 1.84]	[1.49, 2.02]
Glycine	1.75	1.70	2.88	0.007	[1.63 - 1.89]	[1.49, 2.09]
Histidine	1.12	1.08	3.94	<0.001	[1.05 - 1.18]	[0.94, 1.31]
Isoleucine	1.81	1.76	2.94	0.031	[1.68 - 1.99]	[1.54, 2.14]
Leucine	3.04	2.94	3.23	0.046	[2.82 - 3.36]	[2.64, 3.52]
Lysine	2.74	2.62	4.63	0.012	[2.48 - 2.99]	[2.05, 3.47]
Serine	2.03	1.96	3.08	0.004	[1.90 - 2.19]	[1.75, 2.38]
Threonine	1.60	1.55	2.95	0.024	[1.50 - 1.72]	[1.40, 1.83]
Valine	1.92	1.86	2.85	0.040	[1.80 - 2.07]	[1.64, 2.22]
Seed Fatty Acid (% Total FA)						
22:0 Behenic Acid	0.56	0.54	4.33	0.022	[0.46 - 0.65]	[0.30, 0.67]

Table 13. Summary of Differences (p<0.05) for the Comparison of Soybean Component Levels for MON 87701 vs. the Conventional Control (A5547) (cont'd.)

Component (Units)[1]	MON 87701 Mean	A5547 Mean	Mean Difference (Test minus Control)		Test Range	Commercial Tolerance Interval[2]
			Mean Difference (% of A5547)	Significance (p-Value)		
Statistical Differences Observed in *Combined Site* Analysis						
Seed Vitamin (mg/100g DW)						
Vitamin E	7.69	6.24	23.26	<0.001	[6.36 - 9.62]	[0, 11.09]
Seed Antinutrient (TIU/mg DW)						
Trypsin Inhibitor	26.06	28.57	-8.79	0.014	[21.65 - 32.53]	[13.58, 46.02]
Seed Isoflavone (mg/kg DW)						
Daidzein	667.54	604.88	10.36	0.040	[188.96 - 983.26]	[0, 1585.14]
Statistical Differences Observed in *More than One Individual Site*						
Seed Amino Acid (% DW)						
Arginine Site GA	2.80	2.57	8.75	0.011	[2.72 - 2.91]	[2.22, 3.25]
Arginine Site IL	2.61	2.44	6.88	0.045	[2.49 - 2.70]	
Histidine Site GA	1.15	1.09	5.17	0.019	[1.13 - 1.16]	[0.94, 1.31]
Histidine Site IL	1.11	1.05	4.90	0.036	[1.09 - 1.13]	
Tyrosine Site AL	1.32	1.20	9.95	0.034	[1.28 - 1.33]	[0.85, 1.48]
Tyrosine Site IL	1.10	1.01	9.14	0.003	[1.07 - 1.13]	

Table 13. Summary of Differences (p<0.05) for the Comparison of Soybean Component Levels for MON 87701 vs. the Conventional Control (A5547) (cont'd.)

Component (Units)[1]	MON 87701 Mean	A5547 Mean	Mean Difference (Test minus Control)		Test Range	Commercial Tolerance Interval[2]
			Mean Difference (% of A5547)	Significance (p-Value)		
Statistical Differences Observed in *More than One Individual Site*						
Seed Fatty Acid (% Total FA)						
22:0 Behenic Acid Site AR	0.47	0.46	3.08	0.037	[0.46 - 0.48]	[0.30, 0.67]
22:0 Behenic Acid Site GA	0.60	0.55	8.24	0.029	[0.58 - 0.62]	
Seed Vitamin (mg/100g DW)						
Vitamin E Site AR	6.88	5.03	36.69	<0.001	[6.77 - 7.08]	[0, 11.09]
Vitamin E Site GA	9.16	7.77	17.81	0.011	[8.51 - 9.62]	
Vitamin E Site IL	6.72	5.31	26.56	<0.001	[6.36 - 7.27]	
Vitamin E Site NC	7.83	6.14	27.55	0.017	[7.59 - 8.19]	
Seed Antinutrient (%DW)						
Stachyose Site AL	1.84	2.37	-22.36	0.024	[1.83 - 1.89]	[0.99, 7.93]
Stachyose Site NC	4.56	5.50	-17.12	0.006	[4.32 - 4.72]	
Seed Isoflavone (mg/kg DW)						
Daidzein Site AR	767.90	658.21	16.67	0.031	[747.32 – 793.95]	[0, 1585.14]
Daidzein Site IL	890.96	803.42	10.90	0.042	[834.82 – 983.26]	

Table 13. Summary of Differences (p<0.05) for the Comparison of Soybean Component Levels for MON 87701 vs. the Conventional Control (A5547) (cont'd.)

Component (Units)[1]	MON 87701 Mean	A5547 Mean	Mean Difference (Test minus Control)		Test Range	Commercial Tolerance Interval[2]
			Mean Difference (% of A5547)	Significance (p-Value)		
Statistical Differences Observed in *One Individual Site*						
Forage Fiber (% DW)						
Neutral Detergent Fiber Site AR	49.83	38.62	29.02	0.023	[46.69 - 55.99]	[21.51, 66.01]
Seed Proximate (% DW)						
Ash Site IL	5.42	5.29	2.42	0.039	[5.20 – 5.55]	[3.74, 6.45]
Carbohydrates Site IL	36.65	39.17	-6.45	0.024	[35.60 – 37.72]	[28.17, 40.99]
Seed Amino Acid (% DW)						
Isoleucine Site GA	1.81	1.74	4.23	0.029	[1.77 - 1.84]	[1.54, 2.14]
Leucine Site GA	3.04	2.91	4.59	0.014	[2.98 - 3.09]	[2.64, 3.52]
Proline Site GA	2.00	1.94	3.56	0.025	[1.99 - 2.02]	[1.73, 2.35]
Tryptophan Site NC	0.49	0.47	4.75	0.006	[0.47 - 0.51]	[0.43, 0.59]
Valine Site GA	1.91	1.84	3.96	0.035	[1.88 - 1.94]	[1.64, 2.22]
Seed Fatty Acid (% Total FA)						
16:0 Palmitic Acid Site IL	11.53	11.71	-1.48	0.025	[11.39 - 11.63]	[8.88, 13.53]
16:1 Palmitoleic Acid Site NC	0.09	0.10	-13.81	0.012	[0.084 - 0.089]	[0.04, 0.15]

Table 13. Summary of Differences (p<0.05) for the Comparison of Soybean Component Levels for MON 87701 vs. the Conventional Control (A5547) (cont'd.)

Component (Units)[1]	MON 87701 Mean	A5547 Mean	Mean Difference (Test minus Control)		Test Range	Commercial Tolerance Interval[2]
			Mean Difference (% of A5547)	Significance (p-Value)		
Statistical Differences Observed in *One Individual Site*						
Seed Fatty Acid (% Total FA)						
18:0 Stearic Acid Site NC	4.42	4.79	-7.62	0.038	[4.34 - 4.49]	[1.88, 6.25]
18:1 Oleic Acid Site NC	19.78	21.60	-8.42	0.047	[19.21 - 20.21]	[5.01, 42.01]
18:2 Linoleic Acid Site NC	54.21	52.62	3.03	0.046	[53.89 - 54.61]	[38.57, 66.94]
20:1 Eicosenoic Acid Site GA	0.24	0.22	5.27	0.035	[0.23 - 0.24]	[0.16, 0.33]
Seed Antinutrient (TIU/mg DW)						
Trypsin Inhibitor Site GA	23.28	29.27	-20.48	0.005	[21.65 - 25.24]	[13.58, 46.02]
Seed Isoflavone (mg/kg DW)						
Genistein Site AR	807.35	680.07	18.72	0.007	[771.77 - 840.99]	[0, 1352.86]

[1]DW=dry weight; FA=fatty acid; TIU= trypsin inhibitor units
[2]With 95% confidence, interval contains 99% of the values expressed in the population of commercial soybean varieties. Negative limits were set to zero.

Table 14. Literature and ILSI Ranges for Components in Soybean Forage and Seed

Forage Tissue/Component[1]	Literature Range[2]	ILSI Range[3]
Proximate (% dw)		
Ash	5.36 – 8.91	6.72 – 10.78
Carbohydrates	62.25 – 72.28	59.8 – 74.7
Moisture (% fw)	68.50 – 78.40	73.5 – 81.6
Protein	16.48 – 24.29	14.38 – 24.71
Total Fat	2.65 – 9.87	1.30 – 5.13
Fiber (% dw)		
Acid Detergent Fiber (ADF)	23.86 – 50.69	not available
Neutral Detergent Fiber (NDF)	19.61 – 43.70	not available
Seed Tissue/Component[1]	**Literature Range[2]**	**ILSI Range[3]**
Proximates (% dw)		
Ash	4.61 – 6.32	3.89 – 6.99
Carbohydrates	32.75 – 40.98	29.6 – 50.2
Moisture (% fw)	6.24 – 11.10	4.7 – 34.4
Protein	34.78 – 43.35	33.19 – 45.48
Total Fat	14.62 – 20.68	8.10 – 23.56
Fiber (% dw)		
Acid Detergent Fiber (ADF)	9.22 – 26.26	7.81 – 18.61
Neutral Detergent Fiber (NDF)	10.79 – 23.90	8.53 – 21.25
Amino Acids (% dw)		
Alanine	1.62 – 1.89	1.51-2.10
Arginine	2.57 – 3.27	2.29-3.40
Aspartic acid	4.16 – 5.02	3.81-5.12
Cystine/Cysteine	0.52 – 0.69	0.37-0.81
Glutamic acid	6.52 – 8.19	5.84-8.20
Glycine	1.59 – 1.90	1.46-2.00
Histidine	0.96 – 1.13	0.88-1.18
Isoleucine	1.59 – 2.00	1.54-2.08
Leucine	2.79 – 3.42	2.59-3.62
Lysine	2.36 – 2.77	2.29-2.84
Methionine	0.45 – 0.63	0.43-0.68
Phenylalanine	1.82 – 2.29	1.63-2.35
Proline	1.83 – 2.23	1.69-2.28
Serine	1.95 – 2.42	1.11-2.48
Threonine	1.44 – 1.73	1.14-1.86
Tryptophan	0.30 – 0.48	0.36-0.50
Tyrosine	1.27 – 1.53	1.02-1.61
Valine	1.68 – 2.09	1.60-2.20

Table 14. Literature and ILSI Ranges for Components in Soybean Forage and Seed (cont'd.)

Seed Tissue/Component[1]	Literature Range[2]	ILSI Range[3]
Fatty Acids	**(% dw)**	**(% total)**
8:0 Caprylic	not available	0.148 – 0.148
10:0 Capric	not available	not available
12:0 Lauric	not available	0.082 – 0.132
14:0 Myristic	not available	0.071 – 0.238
14:1 Myristoleic	not available	0.121 – 0.125
15:0 Pentadecanoic	not available	not available
15:1 Pentadecenoic	not available	not available
16:0 Palmitic	1.44 – 2.35	9.55 – 15.77
16:1 Palmitoleic	not available	0.086 – 0.194
17:0 Heptadecanoic	not available	0.085 – 0.146
17:1 Heptadecenoic	not available	0.073 – 0.087
18:0 Stearic	0.54 – 1.12	2.70 – 5.88
18:1 Oleic	2.87 – 8.82	14.3 – 32.2
18:2 Linoleic	6.48 – 11.6	42.3 – 58.8
18:3 Gamma Linolenic	not available	not available
18:3 Linolenic	0.72 – 2.16	3.00 – 12.52
20:0 Arachidic	0.04 – 0.7	0.163 – 0.482
20:1 Eicosenoic	0.026 – 0.057	0.140 – 0.350
20:2 Eicosadienoic	not available	0.077 – 0.245
20:3 Eicosatrienoic	not available	not available
20:4 Arachidonic	not available	not available
22:0 Behenic	0.044 – 0.073	0.277 – 0.595
22:1 Erucic	not available	not available
Vitamins (mg/100g dw)		
Vitamin E	1.29 – 4.80	0.19-6.17
Antinutrients		
Lectin (H.U./mg fw)	0.45 – 9.95	0.09 – 8.46
Trypsin Inhibitor (TIU/mg dw)	20.79 – 59.03	19.59 – 118.68
Phytic Acid (% dw)	0.41 – 1.92	0.63 – 1.96
Raffinose (% dw)	0.26 – 0.84	0.21 – 0.66
Stachyose (% dw)	1.53 – 2.98	1.21 – 3.50
Isoflavones		
Daidzein	224.03 – 1485.52	60.0 – 2453.5
Genistein	338.24 – 1488.89	144.3 – 2837.2
Glycitein	52.72 – 298.57	15.3 – 310.4

[1]fw=fresh weight; dw=dry weight
[2]Lundry et al. (2008).
[3]ILSI-CCD (2006). Available at http://www.cropcomposition.org/. Conversions: % dw x 10^4 = µg/g dw; mg/g dw x 10^3 = mg/kg dw; mg/100g dw x 10 = mg/kg dw; g/100g dw x 10 = mg/g dw.

2.8 Nutritional Impact

There is no significant change in nutritional aspects of MON 87701 soybean from the current commercial varieties of soybean. Therefore, there is no need to provide an indication of the likely dietary intake of MON 87701 according to the *Format for applying to amend the Australia New Zealand Food Standards Code – Food Produced using Gene Technology* (June 2005).

> *"If there have been any significant nutritional changes to the food, provide data on the human nutritional impact, including dietary patterns."*

2.9 Other Technical Information

a) Detection methodology

An ELISA method can be used to estimate the levels of Cry1Ac protein in tissues derived from MON 87701. The materials and methods for the ELISA analysis are given in Niemeyer and Silvanovich (2008).

b) Projected market penetration

Soybean is the most commonly grown oilseed in the world. In 2007/08, approximately 218.8 million metric tons (MMT) of harvested soybean seed were produced, representing 56% of the world's oilseed production. The impact and severity of insect pest infestations vary greatly across global soybean production regions, primarily due to the different climate and weather conditions, the distribution and environmental tolerance of insect species, and agricultural practices. MON 87701 will be initially commercialized in South America for the control of targeted lepidopteran pests. In the U.S. given the limited number of acres that consistently have sufficient lepidopteran insect pressure to require the use of insecticides or other insect control practices, MON 87701 plantings will be initially limited to breeding and seed multiplication activities to support the commercialization in South America.

Approximately 50% of the world's soybean seed supply was crushed to produce soybean meal and oil in 2007 (ASA, 2008; Soyatech, 2009), and the majority was used to supply the feed industry for livestock use or the food industry for edible vegetable oil and soybean protein isolates. Another 34% of the world soybean seed supply was traded to other geographies, with China, the EU, Japan, and Mexico being the top soybean seed import geographies (ASA, 2008). The remainder of the soybean seed produced was used as certified seed, feed, or stocks.

Global soybean plantings reached 90.5 million hectares in 2007/08, an 9.9% increase compared to 82.3 million hectares planted in 2002/03 (Soyatech, 2009). Soybean production has realized, on average, a 6.2% annual growth rate between 1995/96 to 2006/07. Increased planting flexibility, increased yield from narrow-row seeding practices, a higher rate of corn-soybean rotations, and low production costs favored expansion of soybean areas in the mid-1990s, with expansion concentrated in areas where soybean yields were highest.

There are no current plans to produce MON 87701 commercially in Australia or New Zealand. However, soybean grains and products containing processed soybean may potentially be imported into Australia and New Zealand from countries where MON 87701 is intended for commercialisation.

PART 3 REGULATORY / LEGISLATIVE IMPLICATIONS

3.1 Other approvals

a) Relevant overseas approvals

Monsanto has submitted a food and feed safety and nutritional assessment summary for MON 87701 to the United States Food and Drug Administration (FDA) and has also requested a Determination of Nonregulated Status for MON 87701 from the United States Department of Agriculture-Animal and Plant Health and Inspection Service in 2009. Similarly, food, feed and environmental submissions were made to CFIA, Heath Canada in June 2009 and submission for cultivation was made to CTNBio, Brazil in June 2009. Like wise, an import submission for food and feed use of MON 87701 was made to the European Food Safety Authority (EFSA) in August 2009.

Regulatory submission have been or will be made to countries that import significant quantities of soybean or its processed fractions from South America and have established regulatory approval processes in place. These include submissions to a number of foreign governmental regulatory authorities including, but not limited to, Ministry of Agriculture (MOA) of China, Ministry of Health, Labor and Welfare (MHLW) and Ministry of Agriculture, Forestry and Fisheries (MAFF) of Japan, and the Korean Food and Drug Administration (KFDA) and the Rural Development Administration (RDA).

b) Approval refusal

No application has been rejected or withdrawn by any regulatory body.

3.2 Regulatory Impact Statement

As described previously, soybean is one of the most agriculturally important crops in the world. This application – if approved – will ensure imports of food derived from MON 87701 soybeans comply with the *Australian New Zealand Food Standards Code*. This will ensure that there is no potential for trade disruption on regulatory grounds.

PART 4 STATUTORY DECLARATION – AUSTRALIA

I, Amanda Forster, declare that the information provided in this application fully sets out the matters required and that the same are true to the best of my knowledge and belief, and that no information has been withheld that might prejudice this application.

Signature: _____

Declared before me ……

This …….. day of …….. 2009.

PART 5 REFERENCES

ASA. 2008. Soy Stats 2008. American Soybean Association, St. Louis, Missouri.

Barker, K., I. Chibata, K. Nakayama, K. Takinami, and H. Yamada. 1983a. Nucleotide sequence of the T-DNA Region from the *Agrobacterium tumefaciens* Octopine Ti Plasmid pTi15955. Plant Mol. Biol. **2**:335-350.

Barker, R.F., K.B. Idler, D.V. Thompson, and J.D. Kemp. 1983b. Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octapine Ti Plasmid pTi15955. Plant. Mol. Biol. 2:335-350.

Barker, T., H. Campos, M. Cooper, D. Dolan, G. Edmeades, J. Habben, J. Schlusser, D. Wright, and C. Zinselmeier. 2005. Improving Drought Tolerance in Maize. Pages 173-253 in Plant Breeding Reviews. Vol 25, J. Janick, (ed.) John Wiley and Sons, Inc., Hoboken, New Jersey.

Barry, G.F., G.M. Kishore, S.R. Padgette, and W.C. Stallings. 1997. Glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthases. U.S.A. Patent 5,633,435. http://www.patentstorm.us/patents/7214535.html.

Baum, J.A., T.B. Johnson, and B.C. Carlton. 1999. *Bacillus thuringiensis.* Natural and recombinant bioinsecticide products. Pages Pp 189-209 in Methods in Biotechnology. Pesticides: Use and Delivery. Vol 5, F.R. Hall and J.J. Menn, (eds.) Humana Press, Inc., Totowa, New Jersey.

Betz, F.S., B.G. Hammond, and R.L. Fuchs. 2000. Safety and advantages of *Bacillus thuringiensis*-protected plants to control insect pests. Reg. Toxicol. and Pharmacol. 32:156-173.

Boyer, J.S. 1982. Plant Productivity and Environment. Science 218:443-448.

Bravo, A., J. Sanchez, T. Kouskoura, and N. Crickmore. 2002. N-terminal activation is an essential early step in the mechanism of action of the Bacillus thuringiensis Cry1Ac insecticidal toxin. J Biol Chem 277:23985-23987.

Cannon, R.J.C. 1993. Prospects and progress for *Bacillus thuringiensis*-based pesticides. Pesticide Science 37:331-335.

Choma, C.T., W.K. Surewicz, P.R. Carey, M. Pozsgay, T. Raynor, and H. Kaplan. 1990. Unusual proteolysis of the protoxin and toxin from Bacillus thuringiensis. Structural implications. Eur J Biochem 189:523-527.

Codex Alimentarius. 2003. Guideline For The Conduct Of Food Safety Assessment Of Foods Derived From Recombinant-DNA Plants. Pages 5-20. ftp://ftp.fao.org/es/esn/food/guide_plants_en.pdf.

Coruzzi, G., R. Broglie, C. Edwards, and N. Chua. 1984. Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. EMBO 3:1671-1679.

Crickmore, N., D.R. Zeigler, J. Feitelson, E. Schnepf, J. Van Rie, D. Lereclus, J. Baum, and D.H. Dean. 1998. Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. Microbiology and Molecular Biology Reviews 62:807-813.

De Boer, A.S., and B. B. Diderichsen. 1991. On the safety of Bacillus subtilis and B. amyloliquefaciens: a review. Appl Microbiol Biotechnol 36:1-4.

De Maagd, R.A., A. Bravo, and N. Crickmore. 2001. How Bacillus thuringiensis has evolved specific toxins to colonize the insect world. Trends Genet. 17:193-199.

De Maagd, R.A., D. Bosch, and W. Stiekema. 1999. toxin-mediated insect resistance in plants. Trends Plant Sci 4:9-13.

De Maagd, R.A., A. Bravo, C. Berry, N. Crickmore, and H.E. Schnepf. 2003. Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu. Rev. Genet. 37:409-433.

De Maagd, R.A., M. Weemen-Hendriks, W. Stiekema, and D. Bosch. 2000. *Bacillus thuringiensis* delta-endotoxin Cry1C domain III can function as a specificity determinant for Spodoptera exigua in different, but not all, Cry1-Cry1C hybrids. Appl Environ Microbiol 66:1559-1563.

Depicker, A., S. Stachel, P. Dhaese, P. Zambryski, and H.M. Goodman. 1982a. Nopaline synthase: transcript mapping and DNA sequence. J. Molecular Applied Genetics 1:561-573.

Depicker, A., S. Stachel, P. Dhaese, P. Zambryski, and J.M. Goodman. 1982b. Nopaline Synthase: Transcript Mapping and DNA Sequence. J. Molec. Appl. Genet. 1:561-573.

Diaz-Mendoza, M., G.P. Farinos, P. Castanera, P. Hernandez-Crespo, and F. Ortego. 2007. Proteolytic processing of native Cry1Ab toxin by midgut extracts and purified trypsins from the Mediterranean corn borer Sesamia nonagrioides. J Insect Physiol 53:428-435.

Eitenmiller, R.E. 1997. Vitamin E Content of Fats and Oils - Nutritional Implications. Food Technology 51(5):78-81.

EPA. 2008. Administrative Revisions to Plant-Incorporated Protectant Tolerance Exemptions. Direct Final Rule. 72 FR 20431.

EPA. 2009. Current & Previously Registered Section 3 PIP Registrations. U.S. Environmental Protection Agency Washington, D.C.

EPA, U.S. 1988. Guidance for the reregistration of pesticide products containing *Bacillus thuringiensis* as the active ingredient. U.S. Environmental Protection Agency. NTIS PB 89-164198.

EPA, U.S. 2000. *Bacillus thuringiensis* Cry1Ab delta-endotoxin and the genetic material necessary for its production in corn (MON 810) Fact Sheet. Biopesticide Fact Sheet 006430.

EPA, U.S. 2001. Biopesticides Registration Action Document: Bacillus thuringiensis (Bt) Plant-incorporated Protectants (October 15, 2001). U.S. Environmental Protection Agency http://www.epa.gov/pesticides/biopesticides/pips/bt_brad.htm.

Fischhoff, D.A., and F.J. Perlak. 1995. Synthetic plant genes.

Fling, M.E., J. Kopf, and C. Richards. 1985. Nucleotide sequence of the transposon Tn*7* gene encoding an aminoglycoside-modifying enzyme, 3"(9)-O-nucleotidyltransferase. Nucleic Acids Res. 13:7095-7106.

Frederiksen, K., H. Rosenquist, K. Jorgensen, and A. Wilcks. 2006. Occurrence of natural Bacillus thuringiensis contaminants and residues of Bacillus thuringiensis-based insecticides on fresh fruits and vegetables. Appl Environ Microbiol 72:3435-3440.

Gao, Y., K.J. Fencil, X. Xu, D.A. Schwedler, J.R. Gilbert, and R.A. Herman. 2006. Purification and characterization of a chimeric Cry1F delta-endotoxin expressed in transgenic cotton plants. J Agric Food Chem 54:829-835.

Gill, S.S., E.A. Cowles, and P.V. Pietrantonio. 1992. The mode of action of *Bacillus thuringiensis* endotoxins. Ann. Rev. Entomol. 37:615-636.

Giza, P.E., and R.C. Huang. 1989a. A self-inducing runaway-replication plasmid expression system utilizing the Rop protein. Gene 78:73-84.

Giza, P.E., and R.C. Huang. 1989b. A self-inducing runaway-replication plasmid expression system utilizing the Rop protein. Gene 78:73-84.

Grochulski, P., L. Masson, S. Borisova, M. Pusztai-Carey, J.L. Schwartz, R. Brousseau, and M. Cygler. 1995. *Bacillus thuringiensis* Cry1A(a) insecticidal toxin: crystal structure and channel formation. J Mol Biol 254:447-464.

Güereca, L., and A. Bravo. 1999. The oligomeric state of *Bacillus thuringiensis* Cry toxins in solution. Biochim Biophys Acta 1429:342-350.

Hammond, B., and R.L. Fuchs. 1998. Safety evaluation for new varieties of food crops developed through biotechnology. Pages Pp 61-79 in Biotechnology and Safety Assessment, 2nd ed, J.A. Thomas, (ed.) Taylor & Francis, Philadelphia, Pennsylvania.

Hensley, K., E.J. Benaksas, R. Bolli, P. Comp, P. Grammas, L. Hamdheydari, S. Mou, Q.N. Pye, M.F. Stoddard, G. Wallis, K.S. Williamson, M. West, W.J. Wechter, and R.A. Floyd. 2004. New perspectives on Vitamin E: γ–tocopherol and carboxyethylhydroxy chroman metabolites in biology and medicine. Free Radical Biol & Med 36:1-15.

Ho, T. 1969. The loess and the origin of Chinese agriculture. Am Hist Rev 75:1-36.

Hofmann, C., H. Vanderbruggen, H. Hoefte, J.V. Rie, S. Jansens, and H.V. Mellaert. 1988. Specificity of *Bacillus thuringiensis* delta-endotoxins is correlated with the presence of high-affinity binding sites in the brush border membrane of target insect midguts. Proceedings of the National Academy of Sciences of the United States of America 85:7844-7848.

Höfte, H., and H.R. Whiteley. 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol Rev 53:242-255.

Hymowitz, T. 1970. On domestication of soybean. Economic Botany 24:408-421.

Hymowitz, T. 2004. Speciation and cytogenetics. Soybeans: improvement, production, and uses.

Hymowitz, T., and C.A. Newell. 1981. Taxonomy of the genus glycine, domestication and uses of soybeans. Econ. Bot. 35:272-288.

Hymowitz, T., and R.J. Singh. 1987. Taxonomy and Speciation. Soybean Monograph, Soybeans: Improvement, Production and Uses:23-48.

Hymowitz, T., R.J. Singh, and R.P. Larkin. 1990. Long distance dispersal: The case for the allopolyploid *glycine tabacina* Benth. and *G. tomentella* Hayata in the west-central pacific. Micronesia 23:5-13

IFIC. 2005. Functional foods fact sheet : Soy.

IFIC. 2006. Functional food fact sheet: antioxidants.

Inglis, A.S., and T. Liu. 1970. The stability of cysteine and cystine during acid hydrolysis of proteins and peptides. J. Biol. Chem. 245:112-116.

James, C. 2003. Global review of commercialized transgenic crops: 2002. in ISSA Briefs No. 24. ISAAA Briefs no. 29. Feature: *Bt* maize, Ithaca, New York.

James, C., and ISAAA. 2003. Global status of commercialized transgenic crops: 2003. Monsanto Technical. Ithaca, New York.

Jensen, O.L., A.v. Podtelejnikov, and m. Mann. 1997. Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. Anal. Chem. 69:4741-4750.

Klee, H.J., Y.M. Muskopf, and C.S. Gasser. 1987. Cloning of an *Arabidopsis thaliana* gene encoding 5-enolpyruvylshikimate- 3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. Mol. Gen. Genet. 210:437-442.

Komari, T., Y. Hiei, Y. Saito, N. Murai, and T. Kumashiro. 1996. Vectors carrying two separate T-DNAs for co-transformation of higher plants mediated by *Agrobacterium tumefaciens* and segregation of transformants free from selection markers. The Plant J. 10:165-174.

Krebbers, E., J. Seurinck, L. Herdies, A.R. Cashmore, and M.P. Timko. 1988. Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of Arabidopsis thaliana. Plant Mol Biol 11:745-759.

Li, J.D., J. Carroll, and D.J. Ellar. 1991. Crystal structure of insecticidal delta-endotoxin from Bacillus thuringiensis at 2.5 A resolution. Nature 353:815-821.

Liener, I.E. 1994. Implications of antinutritional components in soybean foods. Critical Reviews in Food Science and Nutrition 34:31-67.

Littell, R., G. Milliken, W. Stroup, r. Wolfinger, and O. Schabenberger. 2006. SAS for Mixed Models. SAS Institute Inc, Cary, NC.

Liu, K.S. 2004a. Soybeans as a powerhouse of nutrients and phyhtochemicals. American Oil Chemists' Society, Champaign, Illinois.

Liu, K.S. 2004b. Soybeans as functional foods and ingredients, Champaign, IL.

Luttrell, R., L. Wan, and K. Knighten. 1999. Variation in susceptibility of noctuid (Lepidoptera) larvae attacking cotton and soybean to purified endotoxin proteins and commercial formulations of *Bacillus thuringiensis*. Journal of Economic Entomology 92:21-32.

MacIntosh, S.C., T.B. Stone, S.R. Sims, P.L. Hunst, J.T. Greenplate, P.G. Marrone, F.J. Perlak, D.A. Fischhoff, and R.L. Fuchs. 1990. Specificity and efficacy of purified *Bacillus thuringiensis* proteins against agronomically important insects. J. Invertebr. Pathol. 56:258-266.

Martinell, B.J., L.S. Julson, C.A. Emler, Y. Huang, D.E. McCabe, and E.J. Williams. 2002. Soybean agrobacterium transformation method. United States Patent 6,384,301.

McClintock, J.T., C.R. Schaffer, and R.D. Sjoblad. 1995. A Comparative Review of the Mammalian Toxicity of Bacillus Thuringiensis-Based Pesticides. Pesticide Science 45:95-105.

Mendelsohn, M., J. Kough, Z. Vaituzis, and K. Matthews. 2003. Are *Bt* crops safe? Nature Biotechnology 21:1003-1009.

Messina, M. 1999. Legumes and soybeans: overview of their nutritional profiles and health effects. Am J Clin Nutri 70:439S-450S.

Messina, M. 2001. Soy and isoflavones.

Miranda, R., F.Z. Zamudio, and A. Bravo. 2001. Processing of Cry1Ab delta-endotoxin from Bacillus thuringiensis by Manduca sexta and Spodoptera frugiperda midgut proteases: role in protoxin activation and toxin inactivation. Insect Biochem Mol Biol 31:1155-1163.

Morse, R.J., T. Yamamoto, and R.M. Stroud. 2001. Structure of Cry2Aa suggests an unexpected receptor binding epitope. Structure 9:409-417.

Nelson, R., V. Lozovaya, A. Lygin, and J. Widholm. 2001. Variations in isoflavones in seeds of domestic and exotic soybean germplasm. Crop Sci Soc Am Annual Meeting Abstracts.

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

OECD. 2007. Consensus Document on Safety Information on Transgenic Plants Expressing *Bacillus thuringiensis*-Derived Insect Control Proteins. Organization of Economic Cooperation and Development No. 42, ENV/JM/MONO(2007)14.

Padgette, S.R., D. Re, G. Barry, D. Eichholtz, X. Delannay, R.L. Fuchs, G. Kishore, and R.T. Fraley. 1996. New weed control opportunities: development of soybeans with a Roundup Ready gene. CRC Handbook 4:53-84.

Pariza, M.W., and E.A. Johnson. 2001. Evaluating the safety of microbial enzyme preparations used in food processing: update for a new century. Regul Toxicol Pharmacol 33:173-186.

Park, H.W., and B.A. Federici. 2000. Domain I plays an important role in the crystallization of Cry3A in Bacillus thuringiensis. Mol Biotechnol 16:97-107.

Pearson, W.R. 2000. Flexible sequence similarity searching with the FASTA3 program package. Methods Mol Biol 132:185-219.

Pearson, W.R., and D.J. Lipman. 1988. Improved tools for biological sequence comparison. Pages 2444-2448 in Proceedings of the National Academy of Science, USA.

Rademacher, T.W., R.B. Parekh, and R.A. Dwek. 1988. Glycobiology. Annu Rev Biochem 57:785-838.

Rhodes, W.K. 1997. Soybean cultivar A5547. U.S. Patent 5,659,113. http://www.freepatentsonline.com/5659113.html
http://www.patentstorm.us/patents/5659113/claims.html
http://xrint.com/patents/us/5659113 [Accessed 6-Jun-2008].

Rogers, S.G. 2000. Promoter for transgenic plants. United States Patent Number 6.018.100.

Rukmini, V., C.Y. Reddy, and G. Venkateswerlu. 2000. *Bacillus thuringiensis* crystal delta-endotoxin: role of proteases in the conversion of protoxin to toxin. Biochimie 82:109-116.

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

Schafer, F.Q., E.E. Kelley, and G.R. Buettner. 2003. Oxidative stress and antioxidant intervention. World Scientific, New Jersey.

Schnepf, E., N. Crickmore, J. Van Rie, D. Lereclus, J. Baum, J. Feitelson, D.R. Zeigler, and D.H. Dean. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 62:775-806.

Schuler, M.A., E.S. Schmitt, and R.N. Beachy. 1982. Closely related families of genes code for the alpha and alpha' subunits of the soybean 7S storage protein complex. Nuc. Acid. Res. 10:8225-8244.

Shelton, A.M., J.Z. Zhao, and R.T. Roush. 2002. Economic, ecological, food safety, and social consequences of the deployment of bt transgenic plants. Annu Rev Entomol 47:845-881.

Siegel, J.P. 2001. The mammalian safety of *Bacillus thuringiensis*-based insecticides. Journal of invertebrate pathology 77:13-21.

Sjoblad, R.D., J.T. McClintock, and R. Engler. 1992. Toxicological considerations for protein components of biological pesticide products. Regul. Toxicol. Pharmacol. 15:3-9.

Slaney, A.C., H.L. Robbins, and L. English. 1992. Mode of action of Bacillus thuringiensis toxin CryIIIA: An analysis of toxicity in Leptinotarsa decemlineata (Say) and Diabrotica undecimpunctata howardi Barber. Insect Biochemistry and Molecular Biology 22:9-18.

Soya and Oilseed Bluebook. 2008. Statistics. Soyatech, Manitoba, Canada.

Soyatech. 2009. Statistics. Manitoba, Canada.

Stalker, D.M., C.M. Thomas, and D.R. Helinsk. 1981a. Nucleotide sequence of the region of the origin of replication of the broad host range plasmid RK2. Mol. Gen. Genet. 181:8-12.

Stalker, D.M., C.M. Thomas, and D.R. Helinski. 1981b. Nucleotide sequence of the region of the origin of replication of the broad host range plasmid RK2. Mol Gen Genet 181:8-12.

Sutcliffe, J.G. 1978. Nucleotide sequence of the ampicillin resistance gene of *Escherichia coli* plasmid pBR322. Proc. Natl. Acad. Sci. USA 75:3737-3741.

Thomas, K., G. Bannon, S. Hefle, C. Herouet, M. Holsapple, G. Ladics, S. MacIntosh, and L. Privalle. 2005. In silico methods for evaluating human allergenicity to novel proteins: international bioinformatics workshop meeting report, 23-24 February. Toxicological Sciences 88:307-310.

USDA-NASS. 2007. Acreage 2007 (June Report). United States Department of Agriculture National Agricultural Statistics Service, Washington, D.C.

USDA. 1997. Dekalb Genetics Corp.: Availability of determination of nonregulated status for genetically engineered corn.62 FR Pages 17143-17144. http://www.aphis.usda.gov/brs/aphisdocs2/96_29101p_com.pdf [Accessed May 22, 2009].

Van Rie, J., S. Jansens, H. Hofte, D. Degheele, and H. Van Mellaert. 1990. Receptors on the brush border membrane of the insect midgut as determinants of the specificity of *Bacillus thuringiensis* Delta-endotoxins. Applied and environmental microbiology:1378-1385.

WHO, I. 1999. Environmental Health Criteria 217: *Bacillus thuringiensis*. Pages 1-81. World Health Organization, International Program on Chemical SafetyGeneva, Switzerland http://www.inchem.org/documents/ehc/ehc/ehc217.htm [Accessed February 5, 2007].

Zhuang, M., and S.S. Gill. 2003. Mode of action of *Bacillus thuringiensis* toxins. Pages Pp 213-236 in Chemistry of Crop Protection, Progress and Prospects in Science and Regulation, G. Voss and G. Ramos, (eds.) Wiley-VCH, Weinheim, Germany.