Early Food Safety Evaluation for a Glyphosate N-Acetytransferase Protein: GAT4601

Submitting Company:

Pioneer Hi-Bred International, Inc.
7100 NW 62nd Avenue
PO Box 1014
Johnston, IA 50131-1014

Submitted by:

Tracy Rood, US Registration Manager
Pioneer Hi-Bred International, Inc.
7250 NW 62nd Ave.
PO Box 552
Johnston, IA 50131
Telephone: 515-270-4036
Fax: 515-334-4478

June 16, 2006

No CBI

Table of Contents

Key to Abbreviations

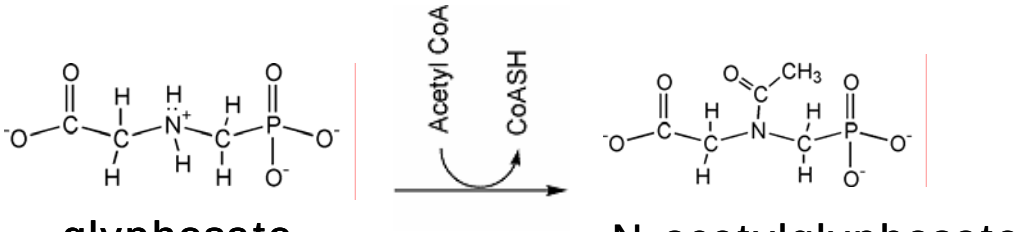
~	approximately
ALS	acetolactate synthase
ATCC	American Type Culture Collection
B. licheniformis	*Bacillus licheniformis*
BAR	phosphinothricin acetyltransferase from *Streptomyces hygroscopicus*
CAMV	Cauliflower Mosaic Virus
DNA	deoxyribonucleic acid
E score	expectation score
EPSPS	enolpyruvulshikimate-3-phosphate synthase
E. coli	*Escherichia coli*
GAT	glyphosate N-acetyltransferase
GAT4601	specific GAT protein
gat4601	specific *gat* gene
gm-als	native soybean acetolactate synthase gene
gm-hra	modified version of soybean acetolactate synthase gene
GNAT family	GCN5-related family of N-acetyltransferases
ILSI	International Life Sciences Institute
kDa	kilodalton
MALDI-MS	matrix assisted laser desorption ionization mass spectrometry
mg/kg	milligrams per kilogram
mM	millimolar
NCBI	National Center for Biotechnology Information
OECD	Organisation for Economic Cooperation and Development
PAT	phosphinothricin acetyltransferase from *Streptomyces viridochomogenes*
RNA	ribonucleic acid
SAMS	S-adenosyl-L-methionine synthase
SCP1	synthetic core promoter
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGF	simulated gastric fluid
SIF	simulated intestinal fluid
TMV	Tobacco Mosaic Virus
µL	microliter
µg	microgram
USDA	United States Department of Agriculture
UTR	untranslated region

1. Name, Description and Function of GAT Proteins

Glyphosate acetyltransferase (GAT) proteins are N-acetyltransferase proteins that were optimized by gene shuffling to more efficiently acetylate the broad-spectrum herbicide glyphosate. Transgenic plants expressing GAT proteins acetylate glyphosate to non-phytotoxic N-acetylglyphosate and are thereby rendered tolerant to glyphosate herbicides.

GAT proteins acetylate the secondary amine of glyphosate using acetyl coenzyme A as an acetyl donor (Figure 1).

Figure 1. Enzymatic Activity of GAT Proteins



glyphosate N-acetylglyphosate

GAT

Phytotoxic **Non-phytotoxic**

GAT proteins are members of the GCN5-related family of N-acetyltransferases (also known as the GNAT family). The GNAT superfamily is one of the largest enzyme superfamilies recognized to date with over 10,000 representatives from all kingdoms of life. Members of the GNAT family all contain a highly conserved GNAT motif but have diverse functions and high sequence diversity (Vetting *et al.,* 2005). GNAT proteins are known to have a number of metabolic functions including detoxification (Dyda *et al*., 2000).

2. Description of the Intended Effect of GAT Proteins

Expression of GAT proteins in transgenic crops provides tolerance to the broad spectrum herbicide glyphosate. Post emergence application of glyphosate very effectively controls grasses and annual broadleaf weeds. Glyphosate has proven to be a very popular herbicide because of its effectiveness and safety. According to USDA, in 2005 glyphosate tolerant soybeans accounted for 87% of all soybeans planted in the United States (http://www.ers.usda.gov/Data/BiotechCrops/ExtentofAdoptionTable3.htm).

Glyphosate inhibits the enzyme enolpyruvulshikimate-3-phosphate synthase (EPSPS) in the plant chloroplast-localized pathway that leads to the biosynthesis of aromatic amino acids. Some microbial EPSPS enzyme variants are insensitive to glyphosate inhibition. When expressed in chloroplasts of transgenic plants, the insensitive enzymes confer tolerance to glyphosate.

GAT provides an alternative mechanism of resistance to glyphosate. GAT proteins detoxify glyphosate to the non-herbicidal form N-acetylglyphosate. This detoxification mechanism is similar to that of the phosphinothricin acetyltransferase (PAT or BAR) enzymes from *Streptomyces*, which detoxify phosphinothricin- or bialaphos-based herbicides by adding an acetyl group (De Block *et al.,* 1987).

3. Identity and Source of Introduced Genetic Material

GAT4601 is a synthetic protein based on N-acetyltransferase protein sequences from *Bacillus licheniformis*, a gram positive saprophytic bacterium that is widespread in nature and thought to contribute substantially to nutrient cycling due to the diversity of enzymes produced by members of its species.

B. licheniformis is used for the production of a number of enzymes such as proteases and amylases that have wide application in the detergent industry. Proteases are also used for dehairing and batting in the leather industry, and amylases are used for desizing of textiles and sizing of paper (EPA, 1996). *B. licheniformis* has been used in the United States, Canada, and Europe in the fermentation industry for production of food enzymes (*e.g.,* alpha-amylase, cyclodextrin glycosyltransferase, hemicellulase, protease, and pullulanase; Rey *et al.,* 2004).

All *B. licheniformis* cultures available from the American Type Culture Collection (ATCC) are classified as Biosafety Level 1. Items that the ATCC classifies at Biosafety Level 1 have no known history of causing disease in humans or animals based on their assessment of potential risk using U.S. Public Health Service guidelines, with assistance provided by ATCC scientific advisory committees. *B. licheniformis* is considered a Class 1 microorganism under the National Institutes of Health (NIH) guidelines for research involving recombinant DNA molecules and also falls under Class 1 containment under the European Federation of Biotechnology guidelines. Class 1 organisms have the lowest level of risk and are defined as those that are unlikely to cause human disease.

To synthesize the *gat4601* gene, three distinct alleles of the *gat* gene were isolated from different strains of *B. licheniformis*. *B. licheniformis* strain 401 was purchased from ATCC (catalog number 14580); two other *B. licheniformis* strains (B6 and DS3) were isolated from soil samples and genotyped by 16S RNA sequencing. Native GAT enzymes were capable of acetylating glyphosate, but at a very slow rate.

To develop an enzyme useful for engineering glyphosate tolerant plants, DNA shuffling was used to improve glyphosate N-acetyltransferase activity of GAT. DNA shuffling is a process that recombines genetic diversity from parental genes to create libraries of gene variants that are screened to identify those progeny with improved properties. This recombination and selection process can be repeated using improved progeny as parents for the next iteration of shuffling. The gene shuffling process and GAT protein optimization are described fully in Castle *et al.,* 2004 and at http://www.isb.vt.edu/articles/sep0403.htm. After seven rounds of shuffling, the GAT4601 protein was identified. Transformation of plants with the *gat4601* gene enabled regeneration of highly tolerant transgenic soybean plants.

The sequence below (Figure 2) represents the translated protein sequence of the *gat4601* gene. GAT4601 is 146 amino acids in length and has an approximate molecular weight of 17 kDa. The asterisk (*) indicates the translational stop codon. There are 23-24 amino acid changes (15-16 of which are conservative) between the shuffled GAT4601 protein and the original three native GAT proteins.

Figure 2. Translated Protein Sequence of *gat4601* Gene

```
  1    MIEVKPINAE DTYELRHRIL RPNQPIEACM FESDLLRGAF HLGGFYRGKL
 51    ISIASFHQAE HSELQGQKQY QLRGMATLEG YREQKAGSTL VKHAEEILRK
101    RGADMLWCNA RTSASGYYKK LGFSEQGEIF DTPPVGPHIL MYKRIT*
```

The *gat4601* coding sequence is under control of a constitutive synthetic SCP1 promoter. The GAT4601 protein is typically co-expressed with a modified soybean acetolactate synthase (*gm-hra*) gene that confers tolerance to the acetolactate synthase (ALS)-inhibiting class of herbicides. The *gm-hra* gene is expressed constitutively using an endogenous soybean promoter from the S-adenosyl-L-methionine synthase (SAMS) gene.

The following table (Table 1) outlines the identities and sources of the introduced genetic material in soybean plants expressing the GAT4601 protein.

Table 1. Identity and Sources of Introduced Genetic Material

Genetic Element	Size (base pairs)	Description
SCP1 promoter	486	Constitutive synthetic core promoter consisting of a portion of the CaMV 35S promoter (Odell *et al.,* 1985) and the Rsyn7-Syn II Core consensus promoter (Bowen *et al*., 2000, 2003).
TMV omega 5'-UTR	67	An enhancer element derived from the Tobacco Mosaic Virus omega 5' untranslated leader (Gallie and Walbot, 1992).
Gat4601 gene	441	Synthetic glyphosate N-acetyltransferase (*gat*) gene (Castle *et al.,* 2004) based on the *Bacillus licheniformis gat* gene.
*pin*II terminator	316	Terminator region from *Solanum tuberosum* proteinase inhibitor II (*pin*II) gene (Keil *et al.,* 1986; An *et al.,* 1989).
SAMS promoter	645	SAMS promoter portion of the regulatory region of the S-adenosyl-L-methionine synthase (SAMS) gene from soybean (Falco and Li, 2003).
SAMS 5'-UTR	59	5' untranslated region of the SAMS gene from soybean (Falco and Li, 2003).
SAMS intron	591	Intron within the 5'-untranslated region of the SAMS gene from soybean (Falco and Li, 2003).
SAMS 5'-UTR	16	5' untranslated region (UTR) of the SAMS gene from soybean (Falco and Li, 2003).
gm-hra gene	1971	Modified version of the acetolactate synthase gene from soybean, with 5 additional nucleotides on the 5' end derived from the S-adenosyl-L-methionine synthase gene and two nucleotide changes within the coding sequence.
gm-als terminator	651	Native terminator from the soybean acetolactate synthase gene.

4. Assessment of Allergenicity Potential of GAT4601

No single factor has been recognized as the primary indicator for allergenic potential, and no validated animal model that is predictive of allergenic potential is available. Therefore, a weight-of-evidence approach, which takes into account a variety of relevant factors and experimental observations used to derive an overall assessment of the allergenic potential of a novel protein, was applied to evaluating the allergenic potential of the GAT4601 protein (Codex, 2003). The allergenicity potential assessments are typically based on what is known about food allergens, including the history of exposure and safety of the gene(s) source; molecular structure of the proteins (*e.g.,* amino acid sequence identity to known human allergens); physicochemical

properties such as stability to pepsin digestion *in vitro* (Thomas *et al.,* 2004) or other enzymes such as pancreatin; glycosylation status, and an estimate of the exposure of the novel protein(s) to the gastrointestinal tract where absorption occurs (*e.g.,* digestibility, protein abundance in the crop, and food/feed processing effects).

The allergenic potential of GAT4601 was assessed by: 1) bioinformatic comparison of the amino acid sequence of the GAT4601 protein with known protein allergen sequences; 2) evaluation of the stability of the microbially produced and purified GAT4601 protein from *E. coli* using *in vitro* gastric and intestinal digestion models; 3) determination of the protein glycosylation status; and 4) assessment of the *gat4601* gene source and history of use or exposure.

4.1. Amino Acid Sequence Homology of GAT4601 to Known Protein Allergens

Bioinformatic analyses were conducted to evaluate the potential allergenicity of the GAT4601 protein. The amino acid sequence of GAT4601 was compared to the FARRP6 database (*i.e.,* version 6 (January 2006) of the University of Nebraska Allergen Database; http://www.allergenonline.com/about.asp) containing the amino acid sequences of known allergenic proteins. Potential identities between GAT4601 and proteins in the allergen database were evaluated using the FASTA34 sequence alignment program (Pearson and Lipman, 1988) set to the default parameters (word size = 2, scoring matrix = BLOSUM50, gap creation penalty = -10, gap extension penalty = -2, E score cutoff = 10). The top scoring 20 alignments were returned and reviewed for identities greater than or equal to 35% over 80 or greater residues.

None of the alignments met or exceeded the 35% threshold. In addition, GAT4601 was evaluated for any eight or greater contiguous identical amino acid matches to the allergens contained in the database noted above. The use of a match of eight contiguous, identical amino acids appears to have some relevance based upon the minimum peptide length for a T cell-binding epitope (Metcalfe *et al*., 1996). Results of the evaluation showed there were no eight or greater contiguous identical amino acid matches observed with GAT4601.

4.2. Lability of GAT4601 to Pepsin in Simulated Gastric Fluid (SGF)

A factor that may increase the likelihood of allergic oral sensitization to proteins is the stability of the protein to gastro-intestinal digestion. Proteins that are highly digestible could be expected to have less opportunity to exert adverse health effects when consumed. The ability of food allergens to remain stable long enough to cross the mucosal membrane of the intestinal tract where absorption occurs is important in the context of a weight-of-evidence approach to understanding a protein's potential allergenic risk (Metcalfe *et al.,* 1996; FAO/WHO, 2001; CODEX, 2003; Thomas *et al*., 2004).

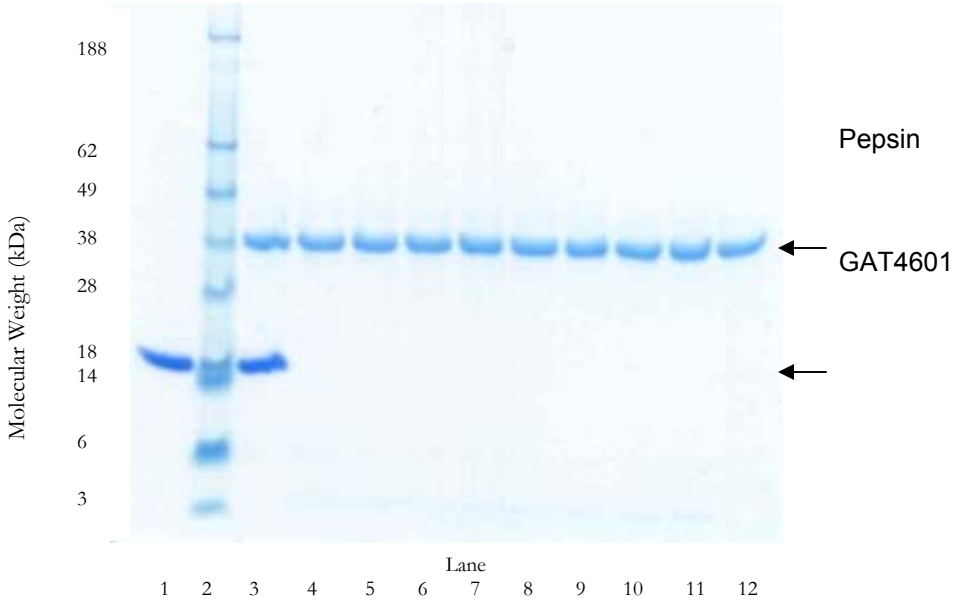
Simulated mammalian gastric fluid (SGF) was used to assess the susceptibility of microbially expressed and purified GAT4601 protein to proteolytic digestion by pepsin *in vitro*. The International Life Sciences Institute (ILSI) has standardized the pepsin digestibility assay protocol in a multi-laboratory evaluation (Thomas *et al.*, 2004). The SGF formulation, time course, and experimental parameters followed in the evaluation of GAT4601 were similar to conditions used in the ILSI multi-laboratory evaluation.

Bovine serum albumin (BSA) was used as a positive control for this study as it is known to hydrolyze readily in pepsin, and β-lactoglobulin was used as a negative control since it is known to persist in pepsin (data not shown). The GAT4601, BSA and β-lactoglobulin were incubated in SGF containing pepsin at pH 1.2 for specific time intervals and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The molar ratio of enzyme to protein in the study was ~0.02 mM pepsin to ~0.015 mM GAT4601, or ~1.3:1. This is equivalent to ~ 3:1 pepsin to GAT4601 ratio on a weight basis (Thomas *et al*., 2004).

A very faint low molecular weight band was visible in lanes 4-11 near the dye front at or below the 3 kDa marker (Figure 3). Additional work is being conducted to characterize the band and determine if it is a product of the GAT4601 protein.

Results of the SGF study, shown in Figure 3, demonstrated that the GAT4601 protein is rapidly (< 30 seconds) hydrolyzed in SGF containing pepsin at pH 1.2 as demonstrated by SDS-PAGE analysis.

Figure 3. Lability of GAT4601 to Pepsin in SGF: Scanned Image of SDS-PAGE Gel



Lane	Sample Identification
1	GAT4601 in water ~60 minutes (2.3 µg)
2	SeeBlue molecular weight marker
3	GAT4601 “Time 0” (2.3 µg)
4	GAT4601 0.5 minutes in SGF
5	GAT4601 1 minute in SGF
6	GAT4601 2 minutes in SGF
7	GAT4601 5 minutes in SGF
8	GAT4601 10 minutes in SGF
9	GAT4601 20 minutes in SGF
10	GAT4601 30 minutes in SGF
11	GAT4601 60 minutes in SGF
12	SGF control (-GAT4601) 60 minutes

4.3. Lability of GAT4601 to Pancreatin in Simulated Intestinal Fluid (SIF)

The ability of food allergens to remain stable long enough to cross the mucosal membrane of the intestinal tract where absorption can occur is important in the context of a weight-of-evidence approach to understanding a protein’s potential allergenic risk. In order to assess lability of the GAT4601 protein in the intestinal tract, microbially expressed and purified GAT4601 protein was

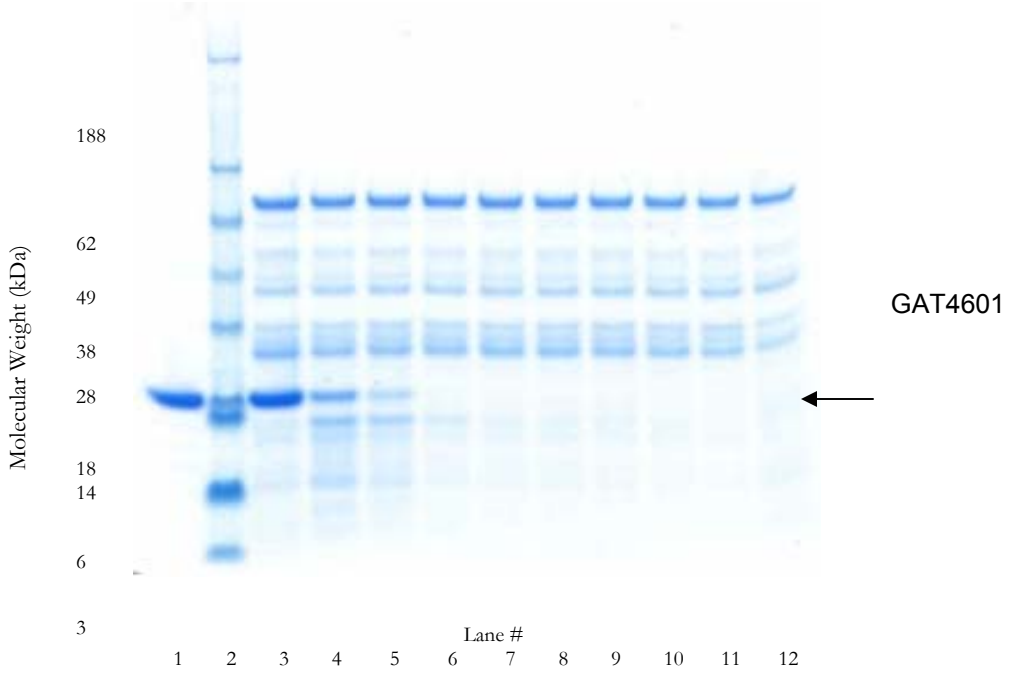
incubated in simulated intestinal fluid (SIF) containing pancreatin prepared as described in the United States Pharmacopoeia (Anonymous, 1995) for specific time intervals and analyzed by SDS-PAGE. SIF contained 0.25 mg/ml of GAT4601 protein in 50 mM KH_2PO_4, 1% w/v pancreatin, pH 7.5. This is equivalent to ~ 40:1 pancreatin:GAT4601 ratio on a weight basis. ß-lactoglobulin and BSA were used as controls (data not shown).

Porcine pancreatin (from Sigma; meets United States Pharmacopoeia standards) contains many enzymes, including amylase, lipase and protease. In Figure 4, the various proteins in pancreatin can be seen in lane 12.

The faint low molecular weight band that was seen in SGF was not seen in SIF, indicating complete lability of the GAT4601 protein.

The GAT4601 protein is rapidly (< 2 minutes) hydrolyzed in SIF containing pancreatin at pH 7.5 as demonstrated by SDS-PAGE analysis (Figure 4).

Figure 4. Lability of GAT4601 to Pancreatin in SIF: Scanned Image of SDS-PAGE Gel



Lane	Sample Identification
1	GAT4601 in water ~60 minutes (3.3 µg)
2	SeeBlue molecular weight marker
3	GAT4601 “Time 0” (3.3 µg)
4	GAT4601 0.5 minutes in SIF
5	GAT4601 1 minute in SIF
6	GAT4601 2 minutes in SIF
7	GAT4601 5 minutes in SIF
8	GAT4601 10 minutes in SIF
9	GAT4601 20 minutes in SIF
10	GAT4601 30 minutes in SIF
11	GAT4601 60 minutes in SIF
12	SIF control (-GAT4601) 60 minutes

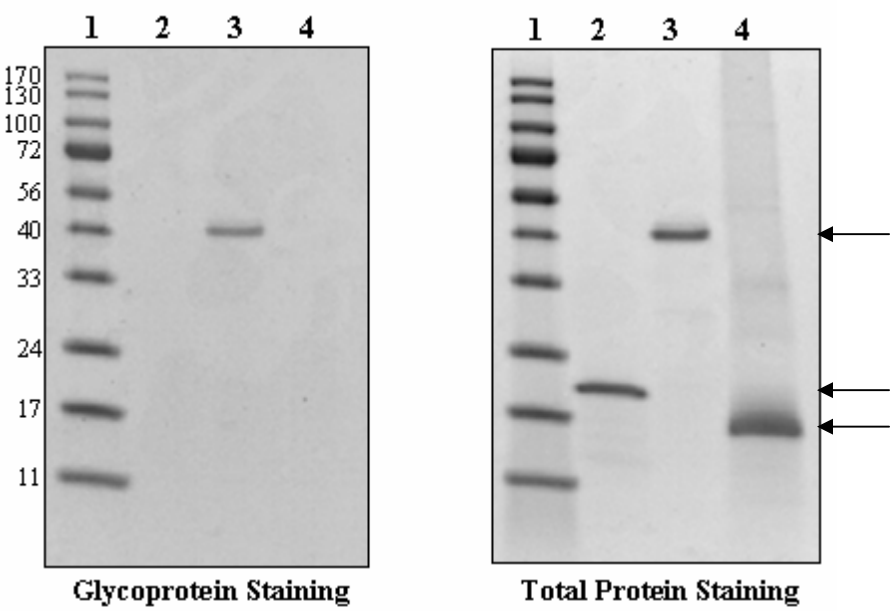
4.4 Glycosylation Analysis of GAT4601

Allergenic proteins are often glycosylated, but this relationship is not absolute, as numerous proteins that are not considered allergens are glycosylated, while some allergenic proteins are not. However, the absence of glycosylation, in the context of other weight-of-evidence data, provides additional support for the conclusion that a protein is non allergenic.

To examine whether GAT4601 protein isolated from soybean tissue was modified by N-linked glycosylation, the protein was run on an SDS-PAGE along with a glycoprotein-positive control (horseradish peroxidase) and glycoprotein-negative control (soybean trypsin inhibitor). Glycoproteins were detected with the GelCode Glycoprotein Staining Kit from Pierce according to the manufacturer's instructions. Briefly, the gel was first fixed with 50% methanol for 30 minutes and washed with 3% acetic acid. Then the gel was incubated with oxidizing solution for 15 minutes and washed with 3% acetic acid. Glycoproteins on the gel were stained with the staining reagent. Finally the gel was treated with a reducing reagent and extensively washed with 3% acetic acid and deionized water. Following glycoprotein detection, the same gel was stained with Coomassie blue to visualize total proteins.

Results showed that GAT4601 protein, isolated and purified from leaf tissue from transgenic soybean plants transformed with the *gat4601* gene, was not glycosylated (Figure 5). Lack of glycosylation was also demonstrated for microbially expressed and purified GAT4601 protein from *E. coli* (data not shown).

Figure 5. Lack of Glycosylation of GAT4601: Scanned Image of SDS-PAGE Gel Stained for Glycoproteins and Total Protein



1- Prestained protein markers (Fermentas Lot 2801)
2- Soybean trypsin inhibitor (Pierce, 1 μg)
3- Horseradish peroxidase (Pierce, 1 μg)
4- GAT4601 purified from soybean tissues (~1.5 μg)

Horseradish
Peroxidase

Soybean Trypsin Inhibitor

GAT4601

4.5. GAT4601 Gene Source and History of Exposure

The *gat4601* gene, which codes for the GAT4601 protein, is derived from *Bacillus licheniformis*, a ubiquitous gram-positive soil bacteria that has a history of safe use for the production of food enzymes in the United States, Canada, and Europe (e.g., alpha-amylase, cyclodextrin glycosyltransferase, hemicellulase, protease, pullulanase), biocontrol agents (EU Commission; SCF/CS/ADD/AMI52; July 2000; USFDA; CFSAN; Office of Food Additive Safety; July 2001) and as a probiotic (Kritas *et al*., 2006; Alexopoulos *et al.,* 2004a and b). (Also see Section 6 below. *B. licheniformis* does not have a history of causing clinical allergy.

4.6. Conclusions on Allergenicity Potential of GAT4601

Bioinformatic analyses revealed no similarities to known protein allergens for the GAT4601 protein sequence. None of the proteins identified met or exceeded the threshold of greater than or equal to 35% identity over 80 or greater residues. Furthermore, no contiguous stretches of eight or greater amino acids were shared between the GAT4601 protein and proteins in the allergen database. These data indicate the lack of both amino acid identity and immunologically relevant similarities between the GAT4601 protein and known protein allergens. The GAT4601 protein was rapidly hydrolyzed in both simulated gastric and intestinal fluids. Further, the GAT4601 protein is not glycosylated and its source organism, *B. licheniformis*, has a history of safe use in the food industry. Taken together, these data support the conclusion that the GAT4601 protein is not a potential allergen.

5. Assessment of Toxicity Potential of GAT4601

The potential toxicity of GAT4601 was assessed by bioinformatic comparison of the amino acid sequence of the GAT4601 protein with publicly available protein sequences. Proteins most similar to GAT4601 were manually inspected to identify any that could be potentially toxic to humans or animals. A close match could be an indicator of toxicological potential of GAT4601. In addition, the acute oral toxicity of GAT4601 in mice (dosing via gavage) was evaluated.

5.1. Assessment of Amino Acid Homology of GAT4601 to Known Protein Toxins

A global sequence similarity search of the GAT4601 protein sequence against the NCBI Protein dataset was conducted using the BLASTp algorithm. A sequence file comprising the translation of the *gat4601* gene was queried using the BLASTP 2.2.12 algorithm against Release 153.0 (4/15/06) of the Genpept “nr” dataset, which incorporates non-redundant entries from all GenBank nucleotide translations along with protein sequences from SWISS-PROT (http://www.expasy.org/sprot/), PIR (http://pir.georgetown.edu/), PRF (http://www4.prf.or.jp/en/), and PDB (http://www.wwpdb.org/).

One of the most important parameters to monitor when performing similarity searches is the expectation, or E score. This E score represents the probability that a particular alignment is due to random chance and can be used to evaluate the significance of an alignment. The calculated E score depends on the overall length of the aligned sequences (including inserted gaps), the number of identical and conserved residues within the alignment, and the size of the database (Pearson and Lipman, 1988; Baxevanis and Ouellette, 1998). When examining an alignment between two protein sequences, a very low E score is more likely to reflect a true similarity while a high E score is more likely to be produced by chance and therefore less biologically relevant.

A cutoff expectation (E) score of 1.0 was used to generate biologically meaningful similarity between the GAT4601 protein and proteins in the NCBI GenPept database. Although a statistically significant sequence similarity generally requires a match with an E score of less than

0.01 (Pearson, 2000), a cutoff of E < 1.0 insures that proteins with even limited similarity will not be overlooked in the search. Low complexity filtering was turned off and the maximum number of alignments returned was set at 2000.

The GAT4601 similarity search identified 192 proteins that were within these criteria. The top three accessions represent GAT gene variants (Castle *et al*., 2004), while the fourth accession is a closely related acetyltransferase protein from *B. licheniformis*. Thirty-five other accessions represent putative or predicted acetyltransferases from other *Bacillus* species, such as *B. subtilis, B. cereus,* and *B. thuringiensis*. The remaining 153 matching accessions represent both known and putative acetyltransferase proteins from various bacterial, archaebacterial, and eukaryotic species.

None of the similar proteins returned by the search were identified as toxins, demonstrating that GAT4601 protein is unlikely to share relevant sequence similarities with known protein toxins and is therefore unlikely to be a toxin itself.

5.2. Results of Mouse Acute Oral Study Using GAT4601

Exposure to most proteins does not cause adverse effects, however, those that do cause toxicity are believed to act through acute mechanisms of action (Sjoblad *et al.,* 1992; Pariza and Johnson, 2001; Hammond and Fuchs, 1998). Exceptions have been identified including certain antinutrient proteins (*e.g.,* lectins and protease inhibitors). However, GAT4601 does not belong to any of these families and, as presented in Sections 4.2 and 4.3, is readily digested unlike most antinutritional proteins. Therefore, an acute oral mouse toxicity assessment was considered appropriate to assess the safety of the GAT4601 protein. The oral route of exposure was selected because it is the most likely route of exposure for humans.

A single dose of GAT4601 protein (containing at least 95% microbially expressed and purified GAT protein) was administered by oral gavage to groups of five fasted male and five fasted female Crl:CD®-1(ICR)BR mice at a dose of 2000 mg/kg (OECD, 2001). Control groups of five fasted male and five fasted female mice were administered bovine serum albumin at a dose of 2000 mg/kg, or vehicle alone, once by oral gavage.

The GAT4601 protein used for this study was produced in *E. coli* and purified using a two-step purification method. The GAT4601 protein was characterized by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to determine molecular weight, western blotting to determine immunoreactivity, N-terminal amino acid sequencing and matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) to determine protein sequence, and N-linked glycoprotein staining to demonstrate lack of glycosylation. (Data demonstrating equivalency between plant-produced and microbially expressed GAT4601 are not shown, but will be submitted in a future Premarket Biotechnology Notice).

The mice were observed for mortality, body weight gain, and clinical signs for 14 days post dosing, after which they were sacrificed and subjected to gross necropsy to detect observable evidence of organ or tissue damage or dysfunction.

All mice survived until the scheduled sacrifice on Day 14. No clinical signs of systemic toxicity or test substance-related body weight losses were observed in any mice. No gross lesions were observed in the mice at necropsy.

Under the conditions of this study, administration of GAT4601 protein to male and female mice at a dose of 2000 mg/kg produced no test substance-related clinical signs of toxicity, body weight losses, gross lesions, or mortality. From this study, it was concluded that the GAT4601 protein is not acutely toxic.

5.3. Conclusions on Toxicity Potential of GAT4601

Bioinformatic analyses revealed GAT4601 to be similar to other N-acetyltransferase proteins. No biologically relevant sequence similarities were observed between known protein toxins and the GAT4601 protein sequence. Along with the lack of acute toxicity in mice, these data support the conclusion that the GAT4601 protein is not acutely toxic.

6. Information on History of Safe Consumption of GAT Proteins in Food

The GAT4601 protein retains the characteristics found in other N-acetyltransferases that are ubiquitous in plants and microorganisms (Neuwald and Landsman, 1997). GAT4601 contains the definitive motif for the GNAT family of N-acetyltransferases (Marchler-Bauer *et al.,* 2005). This superfamily of enzymes is present in all organisms, including plants, mammals, fungi, algae, and bacteria (Dyda *et al*., 2000).

Although GAT4601 is a synthetic protein, it is 84% identical and 94-95% similar at the amino acid level to the translated protein sequences of each of the three original *gat* alleles from *B. licheniformis* from which *gat4601* was derived. Due to its ubiquitous presence as spores in soil and dust, *B. licheniformis* is widely known as a contaminant of food but is not associated with any adverse effects. *B. licheniformis* was determined by EPA to present low risk of adverse effects to human health and the environment and was subsequently granted a TSCA section 5(h)(4) exemption (EPA, 1996).

7. Overall Conclusions

Glyphosate acetyltransferase (GAT) proteins are N-acetyltransferase proteins that were optimized by gene shuffling to more efficiently acetylate the broad-spectrum herbicide glyphosate. Transgenic plants expressing GAT proteins acetylate glyphosate to non-phytotoxic N-acetylglyphosate and are thereby rendered tolerant to glyphosate herbicides.

GAT4601 is a synthetic protein based on N-acetyltransferase protein sequences from *Bacillus licheniformis*, a gram positive saprophytic bacterium that is widespread in nature and thought to contribute substantially to nutrient cycling due to the diversity of enzymes produced by members of its species.

To synthesize the *gat4601* gene, three distinct alleles of the *gat* gene were isolated from different strains of *B. licheniformis*. Native GAT enzymes were capable of acetylating glyphosate, but at a very slow rate. To develop an enzyme useful for engineering glyphosate tolerant plants, DNA shuffling was used to improve glyphosate N-acetyltransferase activity of GAT. DNA shuffling is a process that recombines genetic diversity from parental genes to create libraries of gene variants that are screened to identify those progeny with improved properties. This recombination and selection process can be repeated using improved progeny as parents for the next iteration of shuffling. After seven rounds of shuffling, the GAT4601 protein was identified. Transformation of plants with the *gat4601* gene enabled regeneration of highly tolerant transgenic soybean plants.

Using FDA's guidance for the early food safety evaluation of new proteins in new plant varieties that are under development, the GAT4601 was evaluated for its allergenicity and toxicity potential.

The allergenic potential of GAT4601 was assessed by: 1) bioinformatic comparison of the amino acid sequence of the GAT4601 protein with known protein allergen sequences; 2) evaluation of the stability of the microbially produced and purified GAT4601 protein from *E. coli* using *in vitro*

gastric and intestinal digestion models; 3) determination of the protein glycosylation status; and 4) assessment of the *gat4601* gene source and history of use or exposure.

Bioinformatic analyses revealed no similarities between known protein allergens and the GAT4601 protein sequence. Furthermore, no short (≥eight amino acids) polypeptide matches were shared between the GAT4601 protein and proteins in the allergen database. These data indicate the lack of both amino acid identity and immunologically relevant similarities between the GAT4601 and known allergens. The GAT4601 protein was rapidly hydrolyzed in both simulated gastric and intestinal fluids. Further, the GAT4601 protein is not glycosylated and its source organism, *B. licheniformis*, has a history of safe use in the food industry. Taken together, these data support the conclusion that the GAT4601 protein is not a potential allergen.

Bioinformatic analyses revealed GAT4601 to be similar to other N-acetyltransferase proteins. No biologically relevant sequence similarities were seen between known protein toxins and the GAT4601 protein sequence. There was no evidence of acute toxicity in mice. These data support the conclusion that the GAT4601 protein is not acutely toxic.

The GAT4601 protein retains the characteristics found in other N-acetyltransferases that are ubiquitous in plants and microorganisms. GAT4601 contains the definitive motif for the GNAT family of N-acetyltransferases. This superfamily of enzymes is present in all organisms, including plants, mammals, fungi, algae, and bacteria (Dyda *et al*., 2000). Although GAT4601 is a synthetic protein, it is 84% identical and 94-95% similar at the amino acid level to the translated protein sequences of each of the three original *gat* alleles from *B. licheniformis* from which *gat4601* was derived. Due to its ubiquitous presence as spores in soil and dust, *B. licheniformis* is widely known as a contaminant of food but is not associated with any adverse effects. *B. licheniformis* was determined by EPA to present low risk of adverse effects to human health and the environment and was subsequently granted a TSCA section 5(h)(4) exemption.

Based on the data and information provided in this submission, we have determined that the GAT4601 protein is unlikely to cause an allergic reaction in humans or be a toxin in humans or animals.

8. References

Alexopoulos, C., Georgoulakis, I..E, Tzivara, A., Kritas, S.K., Siochu, A., and Kyriakis, S.C. 2004a. Field evaluation of the efficacy of a probiotic containing *Bacillus licheniformis* and *Bacillus subtilis* spores, on the health status and performance of sows and their litters. J. Anim. Physiol. Anim. Nutr. 88: 381-392.

Alexopoulos, C., Georgoulakis, I.E., Tzivara, A., Kyriakis, C.S., Govaris, A., and Kyriakis, S.C. 2004b. Field evaluation of the efficacy of a probiotic containing *Bacillus licheniformis* and *Bacillus subtilis* spores, on the health status and performance of grower and finisher pigs. J. Vet. Med. A Physiol. Clin. Med. 51: 306-312.

An, G., Mitra, A., Choi, H.K., Costa, M.A., An, K., Thornburg, R.W., and Ryan, C.A. 1989. Functional analysis of the 3' control region of the potato wound-inducible proteinase inhibitor II gene. Plant Cell 1: 115-122.

Anonymous. 1995. Simulated Gastric Fluid and Simulated Intestinal Fluid. In The United States Pharmacopeia 23. The National Formulary 18; The United States Pharmacopeil Convention, Inc.: Rockville, MD, p. 2053.

Baxevanis, A.D. 2005. Assessing pairwise sequence similarity: BLAST and FASTA. In: Bioinformatics: A practical guide to the analysis of genes and proteins, Third Edition, John Wiley & Sons, Inc., New York. Chapter 11:296-324

Bowen, B.A., Bruce, W.B., Lu, G., Sims, L.E., and Tagliani, L.A. 2000. Synthetic promoters. United States Patent No. 6072050.

Bowen, B.A., Bruce, W.B., Lu, G., Sims, L.E., and Tagliani, L.A. 2003. Synthetic promoters. United States Patent No. 6555673 B1.

Castle, L.A., Siehl, D.L., Gorton, R., Patten, P.A., Chen, Y.H., Bertain, S., Cho, H.-J., Duck, N.B., Wong, J., Liu, D., and Lassner, M.W. 2004. Discovery and directed evolution of a glyphosate tolerance gene. Science 304: 1151-1154.

Codex Alimentarius Commission, Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Twenty-Fifth Session, Rome, Italy, June 30 – July 5, 2003. Appendix III, Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants, and Appendix IV, Annex of the assessment of possible allergenicity, 47-60.

De Block, M., Botterman, J., Vandewiele, M., Dockx, J., Thoen, C., Gosselé, V., Rao Movva, N., Thompson, C., Van Montagu, M. and Leemans, J. 1987. Engineering herbicide resistance in plants by expression of a detoxifying enzyme. EMBO J. 6: 2513-2518.

Dyda, F., Klein, D.C., and Hickman, A.B. 2000. GCN5-related N-acetyltransferases: a structural overview. Annu. Rev. Biophys. Biomol. Struct. 29: 81-103.

EPA. 1996. *Bacillus licheniformis* TSCA Section 5(h)(4) Exemption: Final Decision Document. http://earth1.epa.gov/biotech_rule/pubs/fra/fd005.htm

Falco, C.S. and Li, Z. 2003. S-Adenosyl-L-Methionine Synthetase Promoter and its Use in Expression of Transgenic Genes in Plants. United States Patent Application No. 2003/0226166.

FAO/WHO. Evaluation of allergenicity of genetically modified foods. Report of a Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology, January 22-25, 2001. Rome, Italy, 2001.

Gallie, D.R. and Walbot, V. 1992. Identification of the motifs within the tobacco mosaic virus 5'-leader responsible for enhancing translation. Nucleic Acids Res. 20: 4631-4638.

Hammond, B. and Fuchs, R.L. 1998. Safety evaluation of food crops developed through biotechnology. In: J.A. Thomas, ed, Biotechnology and Safety Assessment, 2nd edition, Taylor & Francis, pp 61-79.

Keil, M., Sanches-Serrano, J., Schell, J., and Willmitzer, L. 1986. Primary structure of a proteinase inhibitor II gene from potato. Nucleic Acids Res. 14: 5641-5650.

Liener, I.E. 1983. Removal by processing of naturally occurring toxicants and antinutrients. In: L.W. Shemit, ed, Chemistry and World Food Supplies: The New Frontiers. Pergamon Press, NY, NY, pp 453-463.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., Bryant, S.H. 2005. CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res. 33: D192-6.

Metcalfe, D.D., Astwood, J.D., Townsend, R., Sampson, H.A., Taylor, S.L., and Fuchs, R.L. 1996. Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Crit. Rev. Food Sci. Nutr. 36: S165-S186.

Neuwald, A.F. and Landsman, D. 1997. GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. Trends in Biochem. Sci. 22: 154-155

OECD [Organisation for Economic Co-operation and Development]. 2001. Guideline 420: Acute oral toxicity—Fixed dose procedure. Paris: OECD.

Odell, J.T., Nagy, F., and Chua, N.-H. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature 313: 810-812.

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

Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological comparison. Proc. Natl. Acad. Sci. USA 85: 2440-2448.

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

Rey, M., Ramaiya, P., Nelson, B.A., Brody-Karpin, S.D., Zaretsky, E.J., Tang, M., de Leon, A.L, Xiang. H., Gusti, V., Clausen, I.G., Olsen, P.B., Rasmussen, M.D., Andersen, J.T., Jørgensen. P., Larsen, T.S., Sorokin, A., Bolotin. A., Lapidus, A., Galleron, N., Ehrlich, S.D., and Berka, R.M. 2004. Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. Genome Biology 5: R77.

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

Thomas, K., Aalbers, M., Bannon, G.A., Bartels, M., Dearman, R.J., Esdaile, D.J., Fu, T.J., Glatt, C.M., Hadfield, N., Hatzos, C., Hefle, S.L., Heylings, J.R., Goodman, R.E., Henry, B., Herouet, C., Holsapple, M., Ladics, G.S., Landry, T.D., MacIntosh, S.C., Rice, E.A., Privalle, L.S., Steiner, H.Y., Teshima, R., van Ree, R., Woolhiser, M., and Zawodny, J. 2004. A multi-laboratory evaluation of a common *in vitro* pepsin digestion assay protocol used in assessing the safety of novel proteins. Reg. Toxicol. Pharmacol. 39: 87-98.

Vetting, M., de Carvalho, L.P.S., Yu, M., Hegde, S.S., Magnet, S, Roderick, S., and Blanchard, J. 2005. Structure and functions of the GNAT superfamily of acetyltransferases. Archives of Biochemistry and Biophysics 433: 212-226.

PIONEER®
A DUPONT COMPANY

DuPont Agriculture & Nutrition
Pioneer Emerson Building
7250 N.W. 62nd Ave.
P.O. Box 552
Johnston, IA 50131-0552
(515) 270 3202 Tel

June 29, 2006

Mary Ditto, Ph.D. (HFS-255)
Office of Food Additive Safety
5100 Paint Branch Parkway
College Park, MD 20740-3835

Dear Dr. Ditto,

Enclosed please find a CD containing the electronic version of an Early Food Safety Evaluation (EFSE) submission for GAT4601, a glyphosate N-acetyltransferase protein. This EFSE was submitted for FDA review on June 16, 2006.

The CD contains a Word document of the EFSE itself, as well as PDF files of all the literature cited.

Please let me know if I can provide you with any additional information.

Sincerely,

Tracy A. Rood
US Registration Manager
Pioneer Hi-Bred International, Inc., a DuPont company
7250 NW 62nd Ave.
Johnston, IA 50131
Phone: 515-270-4036
Fax: 515-334-4478
Email: tracy.rood@pioneer.com

June 16, 2006

Mary Ditto, Ph.D. (HFS-255)
Office of Food Additive Safety
5100 Paint Branch Parkway
College Park, MD 20740-3835

Dear Dr. Ditto,

Enclosed please find an Early Food Safety Evaluation (EFSE) submission for GAT4601, a glyphosate N-acetyltransferase protein. Expression of GAT4601 in transgenic plants confers tolerance to broad spectrum herbicide glyphosate. Pioneer Hi-Bred International, Inc. (Pioneer) is using the *gat4601* gene in the product development of herbicide-tolerant soybeans.

The submission covers background information about the GAT4601 protein and information pertinent to the assessment of allergenicity and toxicity potential. Pioneer is requesting that FDA evaluate this EFSE submission and provide us with a response.

Under separate cover I am sending you a CD with an electronic version of this submission, along with PDF files of all the literature cited.

Please let me know if I can provide you with any additional information.

Sincerely,

Tracy A. Rood
US Registration Manager
Pioneer Hi-Bred International, Inc., a DuPont company
7250 NW 62nd Ave.
Johnston, IA 50131
Phone: 515-270-4036
Fax: 515-334-4478
Email: tracy.rood@pioneer.com