Early Food Safety Evaluation for a Glyphosate N-Acetytransferase Protein: GAT4621

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

January 31, 2007

No CBI

Table of Contents

Key to Abbreviations

~	approximately
ATCC	American Type Culture Collection
BAR	phosphinothricin acetyltransferase from *Streptomyces hygroscopicus*
B. licheniformis	*Bacillus licheniformis*
DNA	deoxyribonucleic acid
E score	expectation score
E. coli	*Escherichia coli*
GAT	glyphosate N-acetyltransferase
GAT4601	specific GAT protein from 7th round of gene shuffling
gat4601	specific *gat* gene from 7th round of gene shuffling
GAT4621	specific GAT protein from 11th round of gene shuffling
gat4621	specific *gat* gene from 11th round of gene shuffling
GNAT family	GCN5-related family of N-acetyltransferases
ILSI	International Life Sciences Institute
MALDI-MS	matrix assisted laser desorption ionization mass spectrometry
NCBI	National Center for Biotechnology Information
OECD	Organisation for Economic Cooperation and Development
PAT	phosphinothricin acetyltransferase from *Streptomyces viridochomogenes*
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGF	simulated gastric fluid
SIF	simulated intestinal fluid

Abbreviations of units of measurement and of physical and chemical quantities are presented according to the standard format described in Instructions to Authors in the Journal of Biological Chemistry (http://www.jbc.org/).

1. Name, Description and Function of GAT4621

GAT4621 is a glyphosate acetyltransferase protein based on N-acetyltransferase protein sequences from *Bacillus licheniformis*, a gram positive saprophytic bacterium that is widespread in nature. The GAT4621 protein is 147 amino acids in length and has an approximate molecular weight of 17 kDa (Figure 1).

Figure 1. Amino Acid Sequence of the GAT4621 Protein

```
  1    MAIEVKPINA EDTYDLRHRV LRPNQPIEAC MFESDLTRSA FHLGGFYGGK
 51    LISVASFHQA EHSELQGKKQ YQLRGVATLE GYREQKAGSS LVKHAEEILR
101    KRGADMIWCN ARTSASGYYR KLGFSEQGEV FDTPPVGPHI LMYKRIT
```

The GAT4621 protein is very similar to the GAT4601 protein, which was the subject of New Protein Consultation 003 submitted to FDA on June 16, 2006. The GAT4601 protein is present in Pioneer's 356043 soybean, which was the subject of BNF0108 submitted to FDA on November 16, 2006.

GAT4601 and GAT4621 are 91% identical and 96% similar (Figure 2). Shaded amino acids represent conservative amino acid differences between GAT4601 and GAT4621, and boxed amino acids represent non-conservative differences. The GAT4621 protein has higher catalytic efficiency for glyphosate than GAT4601, as discussed below on pp. 6-7.

Figure 2. Sequence Comparison of GAT4621 and GAT4601

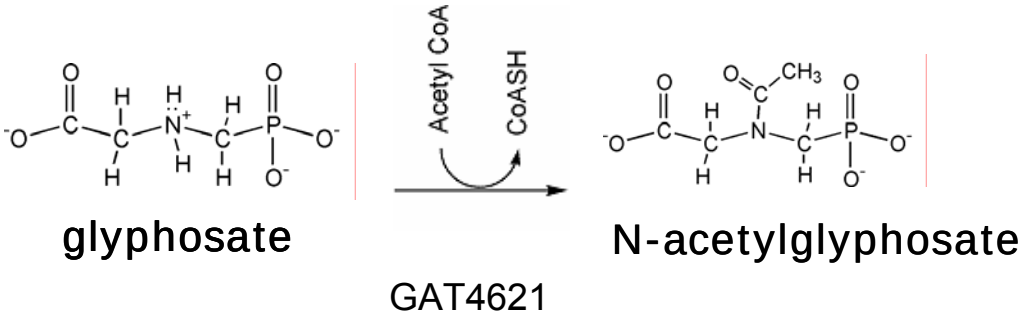
```
                1                                                  50
GAT4621    (1) MAIEVKPINAEDTYDLRHRVLRPNQPIEACMFESDLTRSAFHLGGFYGGK
GAT4601    (1) -MIEVKPINAEDTYELRHRILRPNQPIEACMFESDLLRGAFHLGGFYRGK

               51                                                 100
GAT4621   (51) LISVASFHQAEHSELQGKKQYQLRGVATLEGYREQKAGSSLVKHAEEILR
GAT4601   (50) LISIASFHQAEHSELQGQKQYQLRGMATLEGYREQKAGSTLVKHAEEILR

              101                                            147
GAT4621  (101) KRGADMIWCNARTSASGYYRKLGFSEQGEVFDTPPVGPHILMYKRIT
GAT4601  (100) KRGADMLWCNARTSASGYYKKLGFSEQGEIFDTPPVGPHILMYKRIT
```

Glyphosate acetyltransferase (GAT) proteins are members of the GCN 5-related family of N-acetyltransferases (also known as the GNAT superfamily). The GNAT superfamily is one of the largest enzyme superfamilies recognized to date with over 10,000 representatives from plants, animals and microbes. Members of the GNAT superfamily all contain a highly conserved GNAT motif but have 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 the GAT4621 Protein

Expression of the GAT4621 protein in transgenic crops provides tolerance to the broad spectrum herbicide glyphosate. The GAT4621 protein detoxifies glyphosate to the non-herbicidal form N-acetylglyphosate (Figure 3). 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).

Figure 3. Enzymatic Activity of GAT4621



3. Identity and Source of Introduced Genetic Material

In order to develop a GAT protein that would confer commercial levels of tolerance to the herbicide glyphosate when expressed in plants, an in-house collection of several hundred *Bacillus* isolates was screened. A mass spectrometry method was developed to detect low levels of N-acetylglyphosate, the non-phytotoxic end product of glyphosate acetylation. The *Bacillus* isolates were grown to stationary phase, permeabilized, and incubated with glyphosate and acetyl coenzyme A, and the supernatants were screened for the presence of N-acetylglyphosate by mass spectrometry. Several strains of *Bacillus licheniformis* exhibited GAT activity and had the greatest reproducible accumulation of N-acetylglyphosate (http://www.isb.vt.edu/articles/sep0403.htm and Castle *et al*., 2004).

To isolate the gene encoding GAT, recombinant *E. coli* expressing genomic DNA fragments from *B. licheniformis* were assayed by the mass spectrometry method. DNA sequences of multiple genomic fragments specifying GAT activity from *B. licheniformis* strain B6 and *B. licheniformis* strain 401 (purchased from ATCC—catalog number 14580) were determined, and the corresponding genes were cloned. A polymerase chain reaction (PCR) survey of *B. licheniformis* strains revealed a third gene variant in isolate DS3. The B6, 401 and DS3 *gat* genes were used as parents for fragmentation-based multigene shuffling to create enzymes with improved activity on the substrate glyphosate.

DNA shuffling is a process that recombines genetic diversity from parental genes to create libraries of gene variants that are then screened to identify those progeny with improved properties (Stemmer, 1994; Crameri *et al*., 1998). This process of fragmentation and recombination followed by selection can be repeated using those progeny with improved properties as parents for the next round of shuffling. In the case of the *gat4621* gene, this process was repeated eleven times using a combination of multi-gene shuffling and the introduction of genetic diversity via PCR.

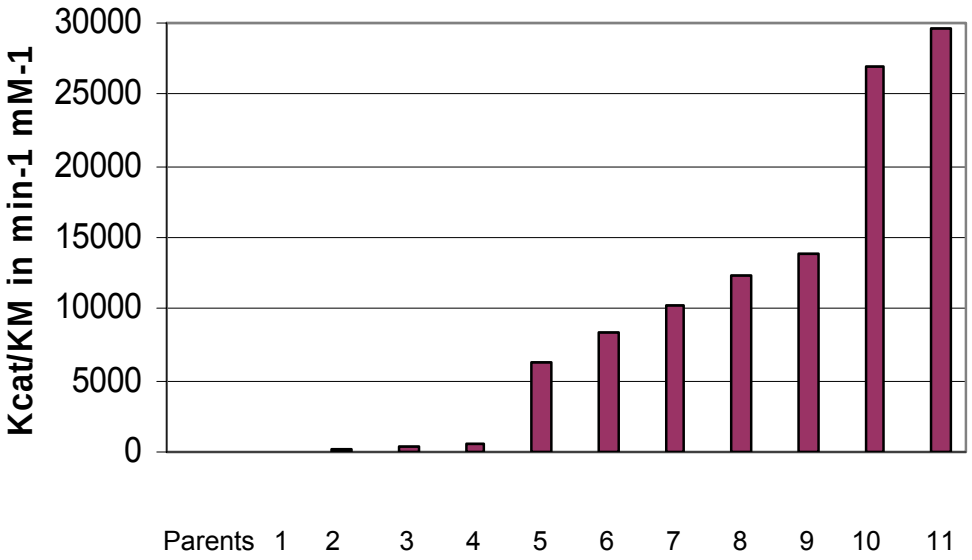
To initiate the first round of gene shuffling, the three native *B. licheniformis gat* genes were used as parental templates. The initial diversity represented among the three native GAT protein sequences occurred at 12 of the 146 total amino acid positions. Libraries of shuffled gene variants were created, expressed in *E. coli*, and screened. Shuffled variants that resulted in the accumulation of more N-acetylglyphosate than the parental controls were selected for further rounds of shuffling. In each round of DNA shuffling, approximately 5,000 gene variants were screened and 24-48 purified enzymes were analyzed to determine their kinetic properties. Typically, three to twelve improved variants exhibiting a high k_{cat}, a low K_M, or a high k_{cat}/K_M ratio were chosen to be the parents for the next round.

In enzyme kinetics, k_{cat} is a measure of the turnover rate or speed of the reaction. The higher the k_{cat}, the faster the enzyme reaction. K_M is the affinity of the enzyme for a substrate or tightness of binding of the substrate to the enzyme. The lower the K_M, the greater the affinity of the enzyme for the substrate. A k_{cat}/K_M ratio is the common way to express the catalytic efficiency of the enzyme.

Several GAT enzymes from the third round of gene shuffling had rate constants (k_{cat}/K_M ratios) about 100-fold improved over the original enzymes, but the gene variants were unable to confer glyphosate tolerance to transgenic plants. At the fifth round of shuffling, two advances were made: 1) the additional introduction of diversity by PCR incorporation of oligonucleotides based on related DNA sequences from *Bacillus cereus* and *Bacillus subtilis* during the fragment reassembly step that allowed for substitutions at 27 amino acid residues; and 2) a functional pre-screen based on resistance of GAT-expressing *E. coli* to glyphosate (Castle *et al*., 2004). At the sixth round of shuffling, a spectrophotometric assay was used that allowed the variants to be screened for relative k_{cat} and K_M values without protein purification. At the end of the seventh round of gene shuffling, the *gat4601* gene encoding the GAT4601 protein was identified. An additional oligonucleotide incorporation of sequence diversity based on acetyltransferase proteins with 30-60% identity to GAT was introduced in the eighth round of shuffling.

At the end of the eleventh round of gene shuffling, the rate constant of the best round 11 variant analyzed was approximately 7000-fold improved over the native enzymes (Castle *et al*., 2004; Siehl *et al*., 2005). The average k_{cat}/K_M of the parental enzymes was 4.2 min^{-1} mM^{-1}, and the average of the best round 11 variant was approximately 29,600 min^{-1} mM^{-1} (Figure 4).

Figure 4. Kinetic Improvement of GAT Enzymes Through Shuffling



Taken from Siehl *et al*., 2005.

In order to optimize the gene from the best round 11 variant for plant expression, codon changes in the gene sequence were made to eliminate rare plant codons. These changes did not alter the encoded protein sequence. Additionally, a GCT codon for alanine was inserted at amino acid position 2. This, along with a plant promoter, resulted in a consensus translation initiation site thought to be best for protein production in plants (Joshi *et al.,* 1997). The plant-optimized gene was given the designation *gat4621*.

The GAT4621 protein, encoded by the *gat4621* gene, is 75-78% identical and 90-91% similar at the amino acid level to each of the three native GAT enzymes from which it was derived, compared to 94% identity of each of the native enzymes to each other (Table 1). There are 32-36 amino acid changes (22-23 of which are conservative) between the shuffled GAT4621 protein and any one of the original three native GAT proteins.

Table 1. Comparison of Sequence Identity Between Parental GAT and GAT4621 Proteins

	GAT from strain 401	**GAT from strain B6**	**GAT from strain DS3**	**GAT4621**
GAT from strain 401	100%	94% identical	94% identical	78% identical 91% similar
GAT from strain B6		100%	94% identical	76% identical 91% similar
GAT from strain DS3			100%	75% identical 90% similar
GAT4621				100%

4. Assessment of Allergenicity Potential of GAT4621

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 GAT4621 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 GAT4621 was assessed by: 1) bioinformatic comparison of the amino acid sequence of the GAT4621 protein with known protein allergen sequences; 2) evaluation of the stability of the microbially produced and purified GAT4621 protein from *E. coli* using *in vitro* gastric and intestinal digestion models; 3) glycosylation analysis of the GAT4621 sequence; and 4) assessment of the *gat4621* gene source and history of use or exposure.

4.1. Amino Acid Sequence Homology of GAT4621 to Known Protein Allergens

Bioinformatic analyses were conducted to evaluate the potential allergenicity of the GAT4621 protein. The amino acid sequence of the GAT4621 protein was compared to a database of allergens from the Food Allergy Research and Resource Program (FARRP), University of Nebraska, Allergen Database (Version 7.0, January 2007), which contains the amino acid sequences of known and putative allergenic proteins. Potential identities between the GAT4621 protein 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 resulting 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.

The GAT4621 amino acid sequence was also evaluated for any eight or greater contiguous identical amino acid matches to the same database of allergens 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 B 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 the GAT4621 amino acid sequence.

4.2. Lability of GAT4621 to Pepsin in Simulated Gastric Fluid (SGF)

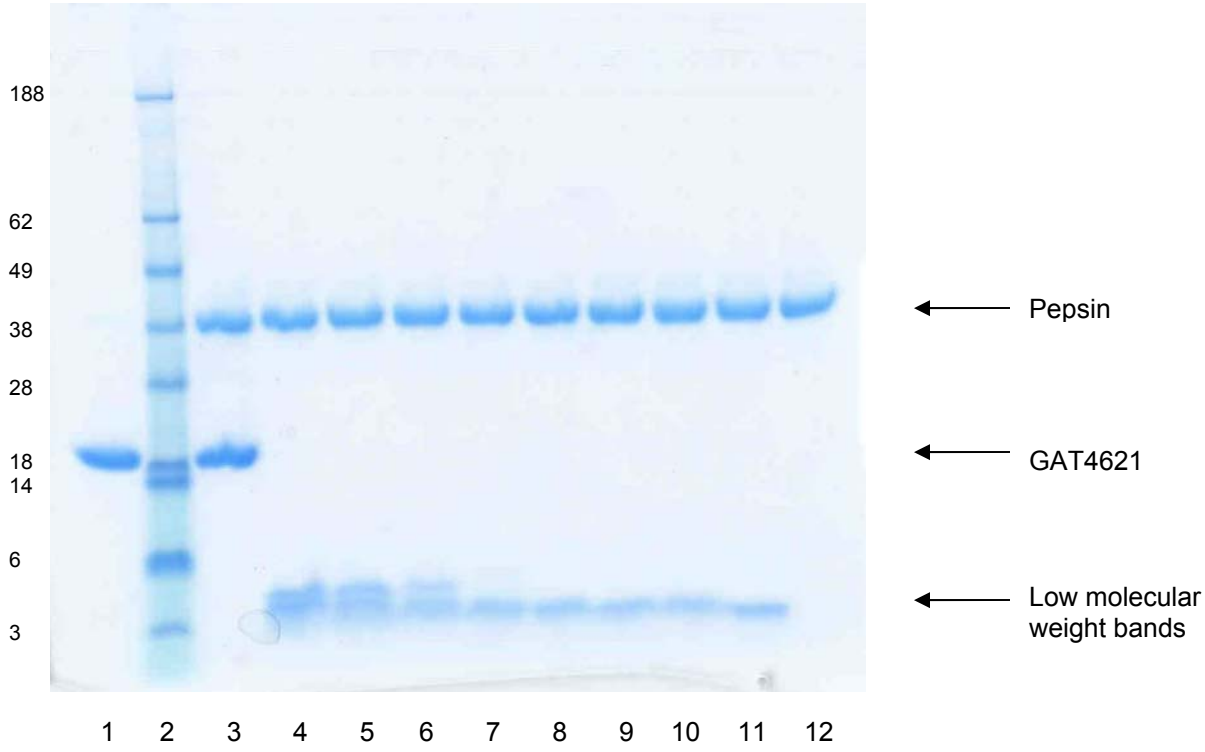
A factor that may increase the likelihood of allergic oral sensitization to proteins is the stability of the protein to gastro-intestinal digestion. Proteins that are highly digestible could be expected to have less opportunity to exert adverse health effects when consumed. The ability of food allergens 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).

SGF was used to assess the susceptibility of microbially expressed and purified GAT4621 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 GAT4621 were similar to conditions used in the ILSI multi-laboratory evaluation.

Bovine serum albumin (BSA) and β-lactoglobulin were used as positive and negative controls, respectively (data not shown). The GAT4621, 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 pepsin to GAT4621 protein in the study was ~0.02 mM pepsin to ~0.015 mM GAT4621, or ~1.3:1. This is equivalent to ~ 3:1 pepsin to GAT4621 ratio on a weight basis (Thomas *et al*., 2004).

Results of the SGF study are shown in Figure 5. The GAT4621 parent protein was not detectable at 30 seconds in SGF (Figure 5, lane 4). Two faint low molecular weight bands were visible in lanes 4-6 near the dye front, and the lower of the two bands persisted through 60 minutes (lane 11). These bands are likely a mix of breakdown products from the GAT4621 protein. Results of the SGF study demonstrate that the GAT4621 protein is rapidly (< 30 seconds) hydrolyzed in SGF containing pepsin at pH 1.2, as shown by SDS-PAGE analysis.

Figure 5. Lability of GAT4621 to Pepsin in SGF: Scanned Image of SDS-PAGE Gel



Lane	Load Volume (µl)	Sample Identification
1	20	GAT4621 (~2.3 µg) water ~60 minutes
2	13	SeeBlue molecular weight marker
3	20	GAT4621 (~2.3 µg) "Time 0"
4	20	GAT4621 30 seconds
5	20	GAT4621 1 minute
6	20	GAT4621 2 minutes
7	20	GAT4621 5 minutes
8	20	GAT4621 10 minutes
9	20	GAT4621 20 minutes
10	20	GAT4621 30 minutes
11	20	GAT4621 60 minutes
12	20	SGF control (pepsin) ~60 minutes

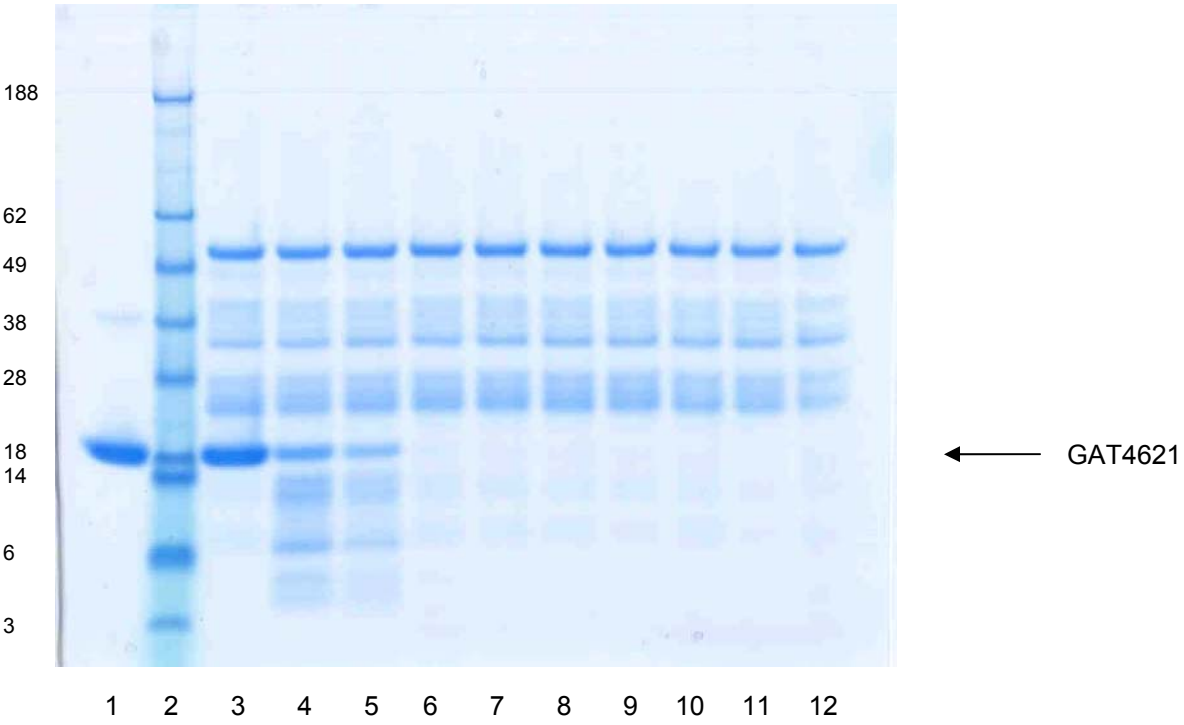
4.3. Lability of GAT4621 to Pancreatin in Simulated Intestinal Fluid (SIF)

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 GAT4621 protein in the intestinal tract, microbially expressed and purified GAT4621 protein was incubated in 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 GAT4621 protein in 50 mM KH_2PO_4, 1% w/v pancreatin, pH 7.5. This is equivalent to an ~ 40:1 pancreatin to GAT4621 protein ratio on a weight basis. ß-lactoglobulin and BSA were used as controls (data not shown).

Porcine pancreatin contains many enzymes, including amylase, lipase and protease. In Figure 6, the mixture of proteins in pancreatin can be seen as multiple stained protein bands in lane 12.

Results of the SIF study are shown in Figure 6. The GAT4621 protein was not detectable at 5 minutes in SIF (Figure 6, lane 7). The low molecular weight bands that were seen in SGF were not seen in SIF, indicating complete lability of the GAT4621 protein. Results of the SIF study demonstrate that the GAT4621 protein is rapidly (< 5 minutes) hydrolyzed in SIF containing pancreatin at pH 7.5, as shown by SDS-PAGE analysis.

Figure 6. Lability of GAT4621 to Pancreatin in SIF: Scanned Image of SDS-PAGE Gel



Lane	Load Volume (µl)	Sample Identification
1	20	GAT4621 (~2.3 µg) water ~60 minutes
2	13	SeeBlue molecular weight marker
3	20	GAT4621 (~2.3 µg) “Time 0”
4	20	GAT4621 30 seconds
5	20	GAT4621 1 minute
6	20	GAT4621 2 minutes
7	20	GAT4621 5 minutes
8	20	GAT4621 10 minutes
9	20	GAT4621 20 minutes
10	20	GAT4621 30 minutes
11	20	GAT4621 60 minutes
12	20	SIF control (pancreatin) ~60 minutes

4.4 Glycosylation Analysis of the GAT4621 Sequence

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.

Microbially produced GAT4621 protein is not glycosylated (data not shown). Data are not yet available for plant-produced protein. However, the GAT4621 protein encoded by the plant transcription unit does not have the amino acid consensus sequence for N-linked glycosylation (asparagine, followed by any amino acid except proline, followed by serine or threonine). Therefore, the plant-expressed GAT4621 is not expected to be post-translationally modified with N-linked glycosylation.

4.5. GAT4621 Gene Source and History of Exposure

The *gat4621* gene, which codes for the GAT4621 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). *B. licheniformis* does not have a history of causing adverse effects (also see Section 6 below).

4.6. Conclusions on the Allergenicity Potential of GAT4621

Bioinformatic analyses revealed no similarities to known protein allergens for the GAT4621 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 GAT4621 protein and proteins in the allergen database. These data indicate the lack of both amino acid identity and immunologically relevant similarities between the GAT4621 protein and known protein allergens. The GAT4621 protein was rapidly hydrolyzed in both simulated gastric and intestinal fluids. Further, the GAT4621 protein is not expected to be N-link glycosylated in plants 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 GAT4621 protein is not a potential allergen.

5. Assessment of Toxicity Potential of GAT4621

The potential toxicity of GAT4621 was assessed by bioinformatic comparison of the amino acid sequence of the GAT4621 protein with publicly available protein sequences. Proteins most similar to GAT4621 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 GAT4621. In addition, the acute oral toxicity of GAT4621 in mice (dosing via gavage) was evaluated.

5.1. Assessment of Amino Acid Homology of GAT4621 to Known Protein Toxins

A global sequence similarity search of the GAT4621 protein sequence against the NCBI Protein dataset was conducted using the BLASTp algorithm. A sequence file comprising the translation of the *gat4621* gene was queried using the BLASTP 2.2.13 algorithm against Release 157.0 (12/18/06) 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, 2005). 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 GAT4621 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 to 2000.

The GAT4621 similarity search identified 225 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-seven other accessions represent putative or predicted acetyltransferases from other *Bacillus* species, such as *B. subtilis, B. cereus,* and *B. thuringiensis*. The remaining 184 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 GAT4621 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 GAT4621

Oral exposure to most proteins does not cause adverse effects. Those that do cause toxicity are believed to act through acute mechanisms of action (Sjoblad *et al.,* 1992; Hammond and Fuchs, 1998; Pariza and Johnson, 2001). An acute oral mouse toxicity study of the GAT4621 protein was conducted. The oral route of exposure was selected because it is the most likely route of exposure for humans.

A single dose of GAT4621 protein preparation (containing approximately 82% microbially expressed, purified GAT4621 protein) was administered by oral gavage to groups of five fasted male and five fasted female Crl:CD®-1(ICR)BR mice at a target dose of 2000 mg/kg body weight (OECD, 2001). The actual dose of purified GAT4621 protein was 1640 mg/kg. A control group of five fasted male and five fasted female mice was administered bovine serum albumin at a dose of 2000 mg/kg or water alone at an equivalent dose volume to the GAT4621 treated mice.

The GAT4621 protein used for this study was produced in and purified from *E. coli* BL21 (DE3) by cation exchange chromatography, anion exchange chromatography, hydrophobic interaction chromatography and diafiltration. The GAT4621 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 peptide mass and, indirectly, protein sequence, and glycoprotein staining to demonstrate lack of glycosylation.

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

All mice survived until the scheduled euthanization on Day 14. No clinical signs of 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 GAT4621 protein to male and female mice at a dose of 1640 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 GAT4621 protein is not acutely toxic.

5.3. Conclusions on Toxicity Potential of GAT4621

Bioinformatic analyses revealed GAT4621 to be similar to other N-acetyltransferase proteins. No biologically relevant sequence similarities were observed between known protein toxins and the GAT4621 protein sequence. Along with the lack of acute toxicity in mice, these data support the conclusion that the GAT4621 protein is not acutely toxic.

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

The GAT4621 protein retains the characteristics found in other N-acetyltransferases that are ubiquitous in plants and microorganisms (Neuwald and Landsman, 1997). GAT4621 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 GAT4621 is a optimized protein, it is 75-78% identical and 90-91% similar at the amino acid level to the translated protein sequences of each of the three original *gat* alleles from *B. licheniformis* from which *gat4621* 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

Using FDA's guidance for the early food safety evaluation of new proteins in new plant varieties that are under development, the GAT4621 protein was evaluated for its allergenicity and toxicity potential.

The allergenic potential of GAT4621 was assessed by: 1) bioinformatic comparison of the amino acid sequence of the GAT4621 protein with known protein allergen sequences; 2) evaluation of the stability of the microbially produced and purified GAT4621 protein from *E. coli* using *in vitro* gastric and intestinal digestion models; 3) N-linked glycosylation analysis of the GAT4621 sequence; and 4) assessment of the *gat4621* gene source and history of use or exposure.

Bioinformatic analyses revealed no similarities between known protein allergens and the GAT4621 protein sequence. Furthermore, no short (≥ eight amino acids) polypeptide matches were shared between the GAT4621 protein and proteins in the allergen database. These data indicate the lack of both amino acid identity and immunologically relevant similarities between the GAT4621 and known allergens. The GAT4621 protein was rapidly hydrolyzed in both simulated gastric and intestinal fluids. Further, the GAT4621 protein is not expected to be N-link glycosylated in plants 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 GAT4621 protein is not a potential allergen.

Bioinformatic analyses revealed GAT4621 to be similar to other N-acetyltransferase proteins. No biologically relevant sequence similarities were seen between known protein toxins and the GAT4621 protein sequence. There was no evidence of acute toxicity in mice. These data support the conclusion that the GAT4621 protein is not acutely toxic.

The GAT4621 protein retains the characteristics found in other N-acetyltransferases that are ubiquitous in plants and microorganisms. GAT4621 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 GAT4621 is a optimized protein, it is 75-78% identical and 90-91% similar at the amino acid level to the translated protein sequences of each of the three original *gat* alleles from *B. licheniformis* from which *gat4621* 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 GAT4621 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.

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: Baxevanis, A.D. and Ouellette, B.F.F., eds, Bioinformatics: A practical guide to the analysis of genes and proteins. 3rd Edition, John Wiley & Sons, Inc., New York, pp 296-324.

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

FAO/WHO. 2001. 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.

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.

Joshi, C.P., Zhou H., Huang X. and Chiang V.L. 1997. Context sequences of translation initiation codon in plants. Plant Mol. Biol. 35: 993-1001.

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.

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.

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.