SUMMARY

(In accordance with 40 CFR part 152, this summary is available
for public release after registration)

STUDY TITLE

Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Maize Event
DAS-40278-9 for Potential Protein Allergenicity and Toxicity

DATA REQUIREMENTS

N/A

AUTHOR(S)

P. Song

STUDY COMPLETED ON

28 – June – 2010

PERFORMING LABORATORY

Regulatory Sciences and Government Affairs—Indianapolis Lab
Dow AgroSciences LLC
9330 Zionsville Road
Indianapolis, Indiana 46268-1054

LABORATORY STUDY ID

101709

Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Maize Event DAS-40278-9 for Potential Protein Allergenicity and Toxicity

SUMMARY

A plant-optimized aryloxyalkanoate dioxygenase-1 (*aad*-1) gene, originally from common soil bacterium *Sphingobium herbicidovorans,* was integrated into maize (*Zea mays*) to produce event DAS-40278-9 by Whiskers®-mediated transformation using a DNA fragment released by *Fsp* I from plasmid pDAS1740. The AAD-1 protein, encoded by the *aad*-1 gene, provides tolerance to 2,4-dicholorophenoxyacetic acid (2,4-D) and aryloxyphenoxypropionate (*e.g*. haloxyfop, cyhalofop, quizalofop, etc.) based herbicides. DNA sequences flanking the *aad-1* insert in event DAS-40278-9 maize have been cloned and characterized. The DNA sequence of the insert is identical to the corresponding portion in the *Fsp* I fragment of pDAS1740 except for truncations at the 5' end of the 2 MAR elements flanking the AAD-1 expression cassette in an opposite orientation. The junctions of the insert and flanking genomic borders were identified and screened for "novel" reading frames spanning the junction sites. A total of 10 "novel" reading frames were identified and evaluated for potential allergenicity and toxicity using bioinformatics tools. Searches of the putative "novel" reading frames against a peer reviewed allergen database (FARRP Allergen Database Version 10, Released in January, 2010) did not generate any significant amino acid sequence similarities with known allergens. Similarly, searches against the GenBank non-redundant protein sequences "nr" did not detect any protein sequence similarities with toxic proteins harmful to humans or animals.

STUDY TITLE

Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Maize Event
DAS-40278-9 for Potential Protein Allergenicity and Toxicity

DATA REQUIREMENTS

N/A

AUTHOR(S)

P. Song 317-337-3434
[psong@dow.com]

STUDY COMPLETED ON

28 – June – 2010

PERFORMING LABORATORY

Regulatory Sciences and Government Affairs—Indianapolis Lab
Dow AgroSciences LLC
9330 Zionsville Road
Indianapolis, Indiana 46268-1054

LABORATORY STUDY ID

101709

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

Compound: Maize AAD-1 event DAS-40278-9

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Maize Event DAS-40278-9 for Potential Protein Allergenicity and Toxicity

- **STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS:**

  No claim of confidentiality, on any basis whatsoever, is made for any information contained in this document. I acknowledge that information not designated as within the scope of FIFRA sec. 10(d)(1)(A), (B), or (C) and which pertains to a registered or previously registered pesticide is not entitled to confidential treatment and may be released to the public, subject to the provisions regarding disclosure to multinational entities under FIFRA sec. 10(g).

  Company: Dow AgroSciences LLC

  Company Agent: L.A. Tagliani

  Title: Regulatory Manager

  Signature: _____

  Date: 23-June-2010

THIS DATA MAY BE CONSIDERED CONFIDENTIAL IN COUNTRIES OUTSIDE THE UNITED STATES.

STATEMENT OF COMPLIANCE WITH GOOD LABORATORY PRACTICE STANDARDS

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Maize Event DAS-40278-9 for Potential Protein Allergenicity and Toxicity

Study Initiation Date: 26/04/2010

This report represents data generated after the effective date of the EPA FIFRA Good Laboratory Practice Standards.

United States Environmental Protection Agency
Title 40 Code of Federal Regulations Part 160
FEDERAL REGISTER, August 17, 1989

Organisation for Economic Co-Operation and Development
ENV/MC/CHEM(98)17, Paris January 26, 1998

At the time this study was conducted, it was not subject to the Good Laboratory Practice Standards and was, therefore, not monitored by the quality assurance unit.
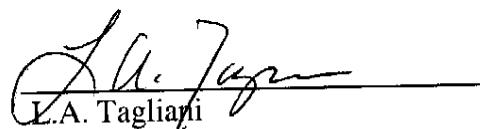
L.A. Tagliani   23 - June - 2010
Sponsor   Date
Dow AgroSciences LLC
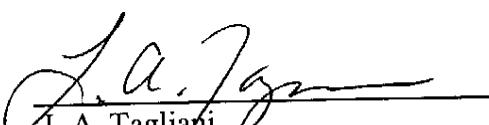
L.A. Tagliani   23 - June - 2010
Submitter   Date
Dow AgroSciences LLC
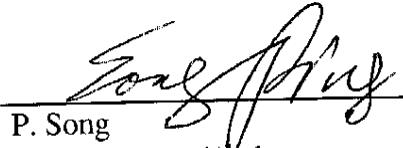
P. Song   28 - June - 2010
Study Director/Author   Study Completion Date
Dow AgroSciences LLC

QUALITY ASSURANCE STATEMENT

Compound: Maize AAD-1 event DAS-40278-9

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Maize Event DAS-40278-9 for Potential Protein Allergenicity and Toxicity

Study Initiation Date: 26/04/2010   Study Completion Date: 28/04/2010

NON-GLP STUDY

SIGNATURE PAGE

P. Song
P. Song
Author
Dow AgroSciences LLC

28-June-2010
Date

M. Zhuang
M. Zhuang
Peer Reviewer
Dow AgroSciences LLC

June, 23, 2010
Date

R. Herman
R. Herman
Peer Reviewer
Dow AgroSciences LLC
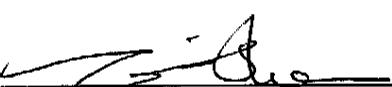
23-June-2010
Date

G. Shan
G. Shan
Science Leader
Dow AgroSciences LLC
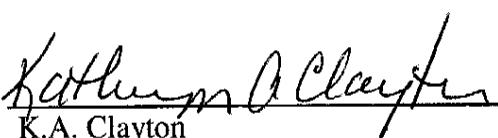
23 June 2010
Date

K.A. Clayton
K.A. Clayton
Global Leader, Biotechnology Regulatory
Sciences
Dow AgroSciences LLC

28 JUNE 2010
Date

STUDY PERSONNEL

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Maize Event DAS-40278-9 for Potential Protein Allergenicity and Toxicity

Principal Analyst: N/A
(Principle Investigator)

Analysts: N/A

TABLE OF CONTENTS

Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Maize Event DAS-40278-9 for Potential Protein Allergenicity and Toxicity

ABSTRACT

A plant-optimized aryloxyalkanoate dioxygenase-1 (*aad*-1) gene, originally from common soil bacterium *Sphingobium herbicidovorans,* was integrated into maize (*Zea mays*) to produce event DAS-40278-9 by Whiskers®-mediated transformation using a DNA fragment released by *Fsp* I from plasmid pDAS1740. The AAD-1 protein, encoded by the *aad*-1 gene, provides tolerance to 2,4-dicholorophenoxyacetic acid (2,4-D) and aryloxyphenoxypropionate (*e.g.* haloxyfop, cyhalofop, quizalofop, etc.) based herbicides. DNA sequences flanking the *aad-1* insert in event DAS-40278-9 maize have been cloned and characterized. The DNA sequence of the insert is identical to the corresponding portion in the *Fsp* I fragment of pDAS1740 except for truncations at the 5' end of the 2 MAR elements flanking the AAD-1 expression cassette in an opposite orientation. The junctions of the insert and flanking genomic borders were identified and screened for "novel" reading frames spanning the junction sites. A total of 10 "novel" reading frames were identified and evaluated for potential allergenicity and toxicity using bioinformatics tools. Searches of the putative "novel" reading frames against a peer reviewed allergen database (FARRP Allergen Database Version 10, Released in January, 2010) did not generate any significant amino acid sequence similarities with known allergens. Similarly, searches against the GenBank non-redundant protein sequences "nr" did not detect any protein sequence similarities with toxic proteins harmful to humans or animals.

INTRODUCTION

A plant-optimized aryloxyalkanoate dioxygenase-1 (*aad*-1) gene, originally from common soil bacterium *Sphingobium herbicidovorans,* was integrated into maize (*Zea mays*) to produce event DAS-40278-9 by Whiskers®-mediated transformation using a DNA fragment released by *Fsp* I from plasmid pDAS1740. The AAD-1 protein, encoded by the *aad*-1 gene, provides tolerance to 2,4-dicholorophenoxyacetic acid (2,4-D) and aryloxyphenoxypropionate (*e.g.* haloxyfop, cyhalofop, quizalofop, etc.) based herbicides. DNA sequences flanking the insert in event DAS-40278-9 maize have been cloned and characterized (1). The DNA sequence of the insert is identical to the corresponding portion in the *Fsp* I fragment of pDAS1740 except for truncations at the 5' end of the 2 MAR elements flanking the AAD-1 expression cassette in an opposite orientation.

Theoretically, the DNA sequences surrounding the junctions of the insert and its flanking borders could create potential "novel" reading frames spanning those junction sites. In the safety assessment of transgenic crops, one of the concerns is that "novel" proteins could be expressed that might have a potential to elicit allergic or toxic reactions in humans. Therefore, the potential novel reading frames spanning the junctions of the insert and border sequences can be analyzed for sequence similarity to known allergens or toxins as an indication of a safety concern should those reading frames actually be expressed. For this study, "novel" reading frames are defined very conservatively as any reading frame spanning the junctions regardless of the presence of a start codon and the number of amino acid residues.

To assess potential allergenicity using bioinformatics tools, two criteria for evaluating structural similarities between query proteins and known allergens are currently used based on amino acid sequence alignments (2, 3, 4). The first criterion is a search over 80-amino-acid stretches (sliding window search) to detect >35% identity between a query protein and known allergens. The window size of 80 amino acids was selected to correspond with a typical domain size in a protein, and recognizes that single protein domains may contain epitopes that mediate antibody

binding. The second criterion involves evaluating short amino-acid stretches for identity between the query protein and known allergens. As stated in the report of Codex Ad Hoc Working Group on Allergenicity (2), "the size of the contiguous amino acid search should be based on a scientifically justified rationale in order to minimize the potential for false negative or false positive results". Window sizes of 6 to 8 amino acids have been suggested based on hypothetical epitope sizes, however, use of window sizes of less than 8 amino acids have been largely abandoned based on the high probability of random alignments that are of no predictive value (5, 6). The use of any short-alignment criteria for predicting the allergenic potential of proteins has also been recently criticized (7, 8, 9, 10). For evaluation of potential protein toxicity, structure similarity between a query protein and known protein toxins are identified using local sequence alignment search tools such as BLAST and FASTA algorithms against a database of all available protein sequences.

The purpose of this study is to identify the potential "novel" reading frames at the junction of the border and insert sequences of event DAS-40278-9 and evaluate them for potential allergenicity and toxicity using bioinformatics tools along with updated allergen and non-redundant protein databases.

METHODS

Search for "Novel" Reading Frames

DNA sequences including the whole insert and its border regions of maize event DAS-40278-9 were analyzed with an in-house Perl script to search six-frame translations from stop codon to stop codon across all the identified junctions (Figure 1). For each reading frame (RF), the exact locations of 5' and 3' stop codons were identified. Those RFs spanning the junctions were considered as "novel" and evaluated for any potential protein allergenicity or toxicity using bioinformatics tools.

Query Sequence Preparation

Each putative reading frame sequence was prepared in FASTA format for the use of FASTA and BLASTp search programs.

Allergenicity Assessment

For the allergenicity assessment, the amino acid sequence of each RF was compared with a peer-reviewed database containing 1471 known and putative allergens as well as celiac-induction protein sequences residing in the FARRP dataset (Version 10, Released in January 2010, University of Nebraska, http://www.allergenonline.org/). Potential identities between the RF peptide sequences and proteins in the allergen database were evaluated with the FASTA program (v34) using the default algorithm parameters (Matrix = BLOSUM50; Expectation = 10; Gap Penalties = -12/-2; *ktup*=2). The FASTA search was run by an in-house Perl script in a UNIX computer with Linux operation system. If a query sequence is longer than 80 amino acids, the script parses the query sequence into a complete (overlapping) set of 80 amino acid long fragments and each fragment is subjected to a FASTA search. A greater than 35% identity threshold over any 80 or more amino acid sequences between a query sequence and an allergen was used to indicate the potential for cross-reactivity. To ensure that high identity over a short stretch (for example, 80% over 60 amino acids) will not be overlooked, a calculation, (Identity% × number of overlapped amino acids)/80, was implemented as a conversion to check the criteria of >35% over 80 amino acids when the FASTA alignment (overlapped amino acids) is less than 80 amino acids. Reading frames shorter than 29 amino acids were not evaluated using FASTA search since >35% identity requires at least a match of 29 amino acids over 80 amino acids. RF peptide sequences were also screened for any matches of 8 contiguous amino acids to the allergens contained in the database noted above as long as an RF is equal to or longer than 8 amino acids. This was done using an in-house Perl script that generates all sequentially possible (overlapping) 8-residue peptides from a query protein, followed by Fuzzpro program (Emboss Package v2.10.0) search that compares each query "word" with all allergen sequences in the database for perfect matches.

Toxicity Assessment

To assess potential toxicity of the *in silico* translated peptides of the RFs, a similarity search was conducted using the BLASTp algorithm. Reading frames were queried using the BLASTp 2.2.21 algorithm against non-redundant protein dataset (update to April 23, 2010), which incorporates non-redundant entries from all GenBank and RefSeq nucleotide translations, including non-redundant GenBank CDS translation along with protein sequences from SWISS-PROT (http://www.expasy.org/sprot/), PIR (http://pir.georgetown.edu/), PRF (http://www.prf.or.jp/aboutdb-e.html), and PDB (http://www.wwpdb.org/). BLASTp searches were done in the NCBI (National Center of Biotechnology Information) BLAST website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). BLASTp searches were performed in an internal UNIX computer using the default setting of algorithm parameters (Matrix = BLOSUM 62, Gap Costs: Existence: 11, Extension: 1, Word size=3) except that the low complexity filter was off and Expectation =1. Although a statistically significant sequence similarity generally requires an alignment with an expectation value less than 0.01, a threshold of E-value < 1.0 ensures that proteins with even limited similarity will not be overlooked in the search (11).

RESULTS AND CONCLUSIONS

A total of 10 reading frames spanning the junctions across the DAS-40278-9 insert and its border regions were identified (Table 1). When the amino acid sequences of the 10 reading frames were compared with the FARRP allergen dataset (Version 10, January 2010), no matches of eight or greater contiguous amino acids were observed in any of the translated sequences. Since the RF 1_-2, 1_-3, and 2_+2 are less than 29 amino acids, only 7 RFs were subject to search against the allergen database using the FASTA program. No over threshold identities (greater than 35% identity over greater than or equal to 80 amino acid residues) were detected in the FASTA search outputs when using the peptide sequences from the 7 RFs as query (Table 2, APPENDIX 1).

When the 10 reading frames were subjected to BLASTp analysis against the GenBank non-redundant protein dataset, no alignments with E-value less than 1 were returned (Table 2, APPENDIX 2).

In conclusion, bioinformatics evaluation of the 10 putative "novel" reading frames did not generate any significant amino acid sequence similarities with known allergens or toxic proteins that are harmful to humans or animals.

REFERENCES

1. Mo, J., Zhou, N., Poorbaugh, J. 2009. Cloning and Characterization of DNA Sequence in the Insert and the Flanking Border Regions of AAD-1 Corn Event DAS-40278-9. Dow AgroSciences Study Report 091023.

2. Codex Alimentarius Special Publications, FOODS DERIVED FROM MODERN BIOTECHNOLOGY (Second Edition), 2009; GUIDELINE FOR THE CONDUCT OF FOOD SAFETY ASSESSMENT OF FOODS DERIVED FROM RECOMBINANT-DNA PLANTS; ANNEX 1: ASSESSMENT OF POSSIBLE ALLERGENICITY, 22-27.

3. Ladics G. S. 2008. Current Codex guidelines for assessment of potential protein allergenicity. Food Chem Toxicol 2008, 46:S20-S23.

4. FAO/WHO (World Health Organization): Evaluation of Allergenicity of Genetically Modified Foods. Report of Joint FAO/WHO Expert Consultation. Rome: Food and Agriculture Organization of the United Nations. 2001

5. Silvanovich A, Nemeth M.A., Song P, Herman R, Tagliani L, Bannon, G. A. 2006. The value of short amino acid sequence matches for prediction of protein allergenicity. Tox Sci., 90:252-258.

6. Stadler M. B., Stadler, B. M. 2003. Allergenicity prediction by protein sequence. FASEB J., 17:1141-1143.

7. Goodman R. E., Vieths S., Sampson H. A., Hill D., Ebisawa M., Taylor S. L., van Ree R. 2008. Allergenicity assessment of genetically modified crops – what makes sense? Nat Biotech, 26:73-81.

8. Thomas K., Herouet-Guicheney C., Ladics G., McClain S., MacIntosh S., Privalle L., Woolhiser M., 2008. Current and future methods for evaluating the allergenic potential of proteins: International workshop report 23-25 October 2007. Food Chem Tox, 46:3219-3225.

9. Cressman R. F., Ladics G., 2009. Further evaluation of the utility of "sliding window" FASTA in predicting cross-reactivity with allergenic proteins. Regul Toxicol Pharmacol, 54:S20-S25.

10. Herman R., Song P, ThirumalaiswamySekhar, A., 2009. Value of eight-amino-acid matches in predicting the allergenicity status of proteins: an empirical bioinformatic investigation. Clinical and Molecular Allergy, 7:9.

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

Table 1. Deduced Amino Acid Sequences of Reading Frames across the Junctions in Maize
Event DAS-40278-9

Reading Frame Name (Junction_Frame)	Nucleotide location	Number of amino acids	Deduced amino acid sequence
1_+1	1588−1878	97	ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLH SGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSL VRPARTIELQSTAPYLEAEYNEG_
1_+2	1772−1960	63	AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMK VSYDLQQSQNTMNHKVIEARNIRRNKYF_
1_-1	1888−1682	69	IVANLHCIPLQGTVLLIVTQWFVLDARERDRRFSAS LGRLERRNTLLLGMDLYVVGYKRGLLNGEKLMR_
1_-2	1881−1843	13	LTFIVFRFKVRCC_
1_-3	1877−1800	26	PSLYSASRYGAVDCNSMVRAGRTREG_
2_+1	6670−6783	38	KLYIYIFYPKAPQGVALGVRTDSKHRQLARQVGGV SLI_
2_+2	6656−6712	19	RHIYKNYIFIFFTQKHRKG_
2_+3	6621−6737	39	IFFKENSIKSLKGIYIKTIYLYFLPKSTARGSPGCADG L_
2_-1	6736−6620	39	SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLK NI_
2_-2	6792−6676	39	ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVK NINI_

Table 2. Summary of Results from BLASTp Search for Sequence Similarities of Putative
Reading Frames across Junctions in Maize Event DAS-40278-9

| Reading Frame (Junction_Frame)	Length (aa)	Match of 8 or more residues with known allergens	FASTA Search (>35% identity over ≥ 80 residues)	Number of BLASTp hits (E()<1)
1_+1	97	No	No	0
1_+2	63	No	No	0
1_-1	69	No	No	0
1_-2	13	No	N/A	0
1_-3	26	No	N/A	0
2_+1	38	No	No	0
2_+2	19	No	N/A	0
2_+3	39	No	No	0
2_-1	39	No	No	0
2_-2	39	No	No	0

N/A = Not applicable
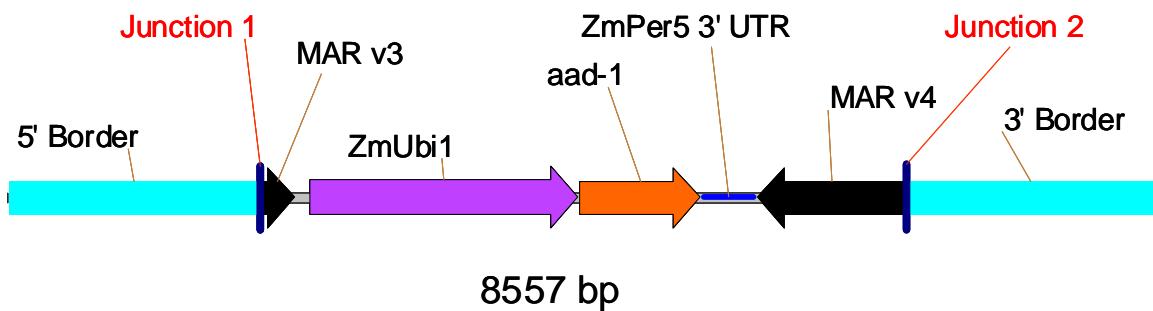
Figure 1. Diagram of the Insert, its Flanking Borders, and Junction Sites in Maize Event DAS-40278-9

APPENDIX

1. FASTA Search Outputs of the Putative Reading Frames (>28 aa) in Maize Event DAS-40278-9 against the Allergen Database V10

    There were 7 reading frames (>28 aa). FASTA search output files are electronically stored in a secured computer in Dow AgroSciences and are available for view in PDF format.

2. BLASTp Search Outputs Using Putative Reading Frames as Queries in Maize Event DAS-40278-9

    BLASTp search output files of the 10 putative reading frames are electronically stored in a secured computer in Dow AgroSciences and are available for view in PDF format.