



HPPD W336 PROTEIN
AMINO ACID SEQUENCE HOMOLOGY
SEARCH WITH KNOWN ALLERGENS

DATA REQUIREMENT
No applicable guidelines

PSI NUMBER: TX99L065

AUTHOR: A. CAPT

TESTING FACILITY:

Bayer CropScience
355, rue Dostoïevski
BP 153
06903 Sophia Antipolis Cedex
France

SPONSOR:

Bayer AG
Bayer CropScience
Alfred Nobel Str. 50
40789 Monheim
Germany

STUDY COMPLETED ON: SEPTEMBER 11, 2009

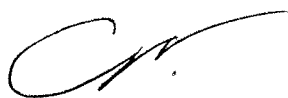
PAGE 1 OF 8



M-355637-01-1

APPROVALS PAGE

Author:



September 11, 2009

Annabelle Capt, Ph.D.
Regulatory Toxicology
Bioscience
Bayer CropScience
355 Rue Dostoïevski
BP 153
06903 Sophia-Antipolis Cedex
France

Head of Unit:



11 SEPTEMBER 2009

Corinne Herouet-Guicheney, Ph.D.
Regulatory Toxicology
Bioscience
Bayer CropScience
355 Rue Dostoïevski
BP 153
06903 Sophia-Antipolis Cedex
France

TABLE OF CONTENTS

APPROVALS PAGE.....	2
TABLE OF CONTENTS.....	3
SUMMARY	4
1. MATERIAL AND METHODS	5
1.1. Amino acid query sequence	5
1.2. Allergen database.....	5
1.3. Epitope homology search	5
1.4. Overall homology search	5
1.5. Potential N-Glycosylation site search.....	6
2. RESULTS.....	6
2.1. Epitope homology search	6
2.2. Overall homology search	6
2.3. Potential N-Glycosylation site search.....	6
3. CONCLUSION	7
4. REFERENCES	7
5. ACRONYMS AND ABBREVIATIONS	8

SUMMARY

This study evaluated the potential amino acid sequence similarity of the single mutated para-hydroxyphenylpyruvate dioxygenase (HPPD W336) protein with known allergens by using several *in silico* approaches.

- An epitope search was carried out to identify any short sequences of amino acids that might represent an isolated shared allergenic epitope. This epitope search compared the amino acid sequence of the HPPD W336 protein subdivided into 8 amino acid blocks, with all known allergens present in the public allergen database AllergenOnline (www.allergenonline.com; release 9.2, 1386 sequences). The algorithm used was FindPatterns (GCG package) and the criterion indicating potential allergenicity was a 100 % identity on a window of 8 amino acids with an allergenic protein.
- An overall identity search was carried out by using FASTA algorithm, which compares the complete amino acid sequence of the HPPD W336 protein with all protein sequences present in the AllergenOnline database. The criterion indicating potential allergenicity was a 35% identity over at least 80 consecutive amino acids with an allergenic protein.
- An 80-mer allergenic identity search was performed to compare the query sequence subdivided into 80 amino acid blocks, with all known allergens present in the AllergenOnline database. The criterion indicating potential allergenicity was a 35 % identity with an allergenic protein.

Furthermore, this study considered the potential N-glycosylation sites by searching their known consensus sequence, potentially found in allergenic proteins.

The results of the epitope homology search showed no identity with epitopes from known allergenic proteins. In addition, the overall and 80-mer identity searches showed no relevant sequence similarity between HPPD W336 and any known allergens from the AllergenOnline database.

No potential N-glycosylation site was identified on the amino acid sequence of the HPPD W336 protein.

In conclusion, it is unlikely that the HPPD W336 protein possesses allergenic properties.

1. Material and methods

1.1. AMINO ACID QUERY SEQUENCE

The 358 amino acid query sequence was coded using the one-letter code adopted by the Commission on Biochemical Nomenclature of the [IUPAC-IUB \(1984\)](#).

As described in the document number [M-228149-01-1 \(De Beuckeleer, 2004\)](#), the query sequence corresponding to the HPPD W336 protein is as follows:

```
MADLYENPMGLMGFEFIEFASPTPGTLEPIFEIMGFTKVATHRSKNVHLYRQGEINLILN  
NEPNSIASYFAAEHGPSVCGMAFRVKDSQKAYNRALELGAQPIHIDTGPMELNLPKIKGI  
GGAPLYLIDRFGEFGSSIIDYDFVYLEGVERNVPVGAGLKVIDHLTHNVYRGRMVYWANFYE  
KLFNFREARYFDIKGEYTGLTSKAMSAPDGMIRIPLNEESSKGAGQIEEFMLMQFNNGEIQ  
HVAFLTDDLVTWDALKKIGMRMTAPPDTYYEMLEGRLPDHGEPVDQLQARGILLDGSS  
VEGDKRLLLQIFSETLMGPVFFEFIQRKGDDGFGEWNFKALFESIERDQVRRGVLTAD
```

Date of search: August 19, 2009.

1.2. ALLERGEN DATABASE

The allergen database used was AllergenOnline, version 9.2, 2009. AllergenOnline allergen database (www.allergenonline.com) is a free, publically available, archived resource list of known and putative allergens and sequences. The database is updated annually by searching NCBI and IUIS annotated sequences and by evaluating the candidate entries for evidence of protein allergenicity (i.e. IgE binding test) and food allergy (e.g. clinical test). A peer review panel of food allergy experts from academia is in charge of this curation. The exact list of experts is reported on the website. They identify whether proteins are allergens, putative allergens or unlikely to be allergenic based on predefined criteria, which are described on the website. Version 9.2 of the database includes 1 386 unique sequences that are clustered into 502 allergen groups based on species (n=236) and sequence identities ([Thomas et al., 2008](#)).

1.3. EPITOPE HOMOLOGY SEARCH

The epitope homology study was carried out by comparing the query sequence, subdivided into 8 amino acid blocks, with all known allergens present in the allergen database.

The algorithm used was FindPatterns (release 10.3, November 19, 2002) from Genetic Computer Group (GCG), which enabled the search of the potential common epitopes between the translated sequences and allergenic sequences present in a large reference allergen database.

The criterion indicating potential allergenicity was a 100 % identity on a window of at least 8 amino acids with an allergenic protein.

1.4. OVERALL HOMOLOGY SEARCH

The rationale is based on the FAO/WHO expert panel ([FAO/WHO, 2001](#)) recommending a criteria of >35% identity over any segment of 80 or more amino acids as an indication of possible cross-reactivity for allergens. These criteria were adopted by the *Codex Alimentarius* Commission

(CAC) as the primary sequence search criteria for use in flagging proteins that might be of some concern of cross-reactivity for genetically modified plants (CAC, 2003).

Two kinds of comparison were carried out:

- **Overall identity search:** comparison of the complete query sequence with all the sequences available in the allergen database. The overall homology search used the FASTA program (version 3.47, 2004; Pearson and Lipman, 1988) and the BLOSUM62 scoring matrix. Only the matches of at least 35% identity over 80 amino acids or more were considered relevant.
- **Eighty-mer allergenic identity search:** comparison of the query sequence subdivided in 80-mers blocks (segments of aa 1-80, then 2-81, 3-82, etc) with all known allergens present in the allergen database. The search is performed with every possible 80 amino acid segment of the query protein. This 80-mer homology search used the FASTA program and BLOSUM50 scoring matrix. Only the matches of at least 35% identity are considered were relevant.

1.5. POTENTIAL N-GLYCOSYLATION SITE SEARCH

The best-studied mode of glycosylation is the formation of an N-glycosidic linkage to Asparagin in the polypeptide chain. The necessary (but not sufficient) criterion for protein N-glycosylation is the presence of the sequence N-X~(P)-S/T, where N = Asparagin, X~(P) = any amino acid except Proline (P), S = Serin and T = Threonin, in the query sequence. Although rare, the sequence motif N-X-C can also be an acceptor site (where N = Asparagin, X = any amino acid and C = Cystein).

Therefore, the consensus sequences searched were of the following type:
N - X~(P) - [S,T] or N - X - C.

2. Results

2.1. EPITOPE HOMOLOGY SEARCH

No identities were found between the 8 linearly contiguous amino acid blocks, which compose the HPPD W336 protein, and known allergens from the AllergenOnline database.

2.2. OVERALL HOMOLOGY SEARCH

No similarities were found between the HPPD W336 protein and known allergens from the AllergenOnline database, based on a '35% identity over an 80 amino acid segment' matching criteria and an 80-mer search.

2.3. POTENTIAL N-GLYCOSYLATION SITE SEARCH

No potential N-glycosylation site was identified on the HPPD W336 protein amino acid sequence by using the N - X~(P) - [S,T] or N-X-C consensus sequences.

3. Conclusion

The lack of any significant amino acid sequence homology with known allergens supports that it is unlikely that the HPPD W336 protein possesses allergenic properties.

4. References

DART #	References
M-276947-01-1	<i>Codex Alimentarius</i> commission (CAC). 2003. <i>Codex</i> principles and guidelines on food derived from biotechnology. <i>Codex Alimentarius</i> Commission. CAC/GL44-2003 and CAC/GL45-2003.
M-232685-01-1	De Beuckeleer, M. 2004. Description of the amino acid sequence of the HPPDase W336 protein. Unpublished. Bayer Cropscience. 5 pages.
M-215805-01-1	FAO/WHO. 2001. Report of a joint Food and Agriculture Organization of the United Nations (FAO)/World Health Organization (WHO) expert consultation on foods derived from biotechnology. Rome, Italy, 22-25 January, 2001.
M-264607-01-1	IUPAC-IUB. 1984. Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature and symbolism for amino acids and peptides. Recommendations 1983. <i>Eur. J. Biochem.</i> 138: 9-37. http://www.chem.qmw.ac.uk/iupac/AminoAcid/
M-303903-01-1	Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. <i>Proc. Natl. Acad. Sci. USA.</i> 85:2444-2448.
M-307889-01-1	Thomas, K., Herouet-Guicheney, C., Ladics, G., McClain, S., MacIntosh, S., Privalle, L. and Woolhiser, M. 2008. Current and future methods for evaluating the allergenic potential of proteins: international workshop report 23-25 October 2007. <i>Food Chem Toxicol.</i> 46:3219-3225.

5. Acronyms and Abbreviations

%	Percent(age)
aa	Amino acid(s)
C	Cystein
CAC	<i>Codex Alimentarius</i> commission
E-Value	Expect(ed) value
GCG	Genetic Computer Group
FAO	Food and Agriculture Organization
HPPD or HPPDase	para-hydroxyphenyl piruvate dioxygenase
IgE	Immunoglobulin E
IUPAC	International Union of Pure and Applied Chemistry
IUB	International Union of Biochemistry
IUIS	International Union of Immunological Societies
JCBN	Joint Commission on Biochemical Nomenclature
N	Asparagin
NCBI	National Center for Biotechnology Information
S	Serin
T	Threonin
X	Any amino acid
X~(P)	Any amino acid except Proline
WHO	World Health Organization