



PAT/*pat* PROTEIN
AMINO ACID SEQUENCE HOMOLOGY
SEARCH WITH KNOWN ALLERGENS

DATA REQUIREMENT
No applicable guidelines

STUDY IDENTIFICATION NUMBER: TXSRT010

AUTHOR: [REDACTED]

TESTING FACILITY:

Bayer S.A.S.
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: MARCH 05, 2012

PAGE 1 OF 11



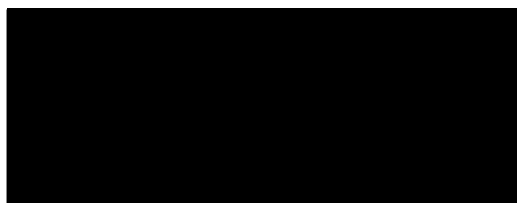
M-217013-05-1

PAT/*pat* Protein
Amino acid sequence homology search with known allergens

This page is deliberately left blank for local regulatory submission purposes.

APPROVALS PAGE

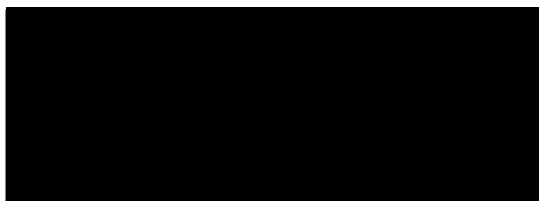
Author:



Yardi 05, 2012

Human and Animal Safety
Regulatory Science - BioScience
Bayer SAS
Bayer CropScience
355 Rue Dostoïevski
BP 153
06903 Sophia Antipolis Cedex
France

Head of Unit:



Nad 05, 2012

Human and Animal Safety
Regulatory Science - BioScience
Bayer SAS
Bayer CropScience
355 Rue Dostoïevski
BP 153
06903 Sophia Antipolis Cedex
France

TABLE OF CONTENTS

APPROVALS PAGE	3
TABLE OF CONTENTS	4
SUMMARY	5
1. MATERIAL AND METHODS	6
1.1. Amino acid query sequence.....	6
1.2. Allergen database	6
1.3. Epitope homology search	6
1.4. Overall homology search	7
1.5. Potential N-Glycosylation site search.....	7
2. RESULTS.....	8
2.1. Epitope homology search	8
2.2. Overall homology search	8
2.3. Potential N-Glycosylation site search.....	8
3. CONCLUSION	8
4. REFERENCES.....	9
5. ACRONYMS AND ABBREVIATIONS	10
FINAL REPORT AMENDMENT	11
END OF REPORT	11

SUMMARY

This study evaluated the potential amino acid sequence similarity of the PAT/*pat* 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 potential allergenic epitope. This epitope search compared the amino acid sequence of the PAT/*pat* protein subdivided into 8 amino acid blocks, with all known allergens present in the public allergen database AllergenOnline (www.allergenonline.org). The program used was SeqMatchAll 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 PAT/*pat* 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.

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 identity search showed no relevant sequence similarity between PAT/*pat* and any known allergens from the AllergenOnline database.

No potential N-glycosylation sites were identified on the amino acid sequence of the PAT/*pat* protein.

In conclusion, the lack of any significant amino acid sequence homology with known allergens supports that it is unlikely that the PAT/*pat* protein possesses allergenic properties.

1. Material and methods

1.1. AMINO ACID QUERY SEQUENCE

The 183 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-084188-01-1 ([De Beuckeleer, 2004](#)), the query sequence corresponding to the PAT/*pat* protein is as follows:

```
1  MSPERRPVEI RPATAADMAA VCDIVNHYIE TSTVNFRTPE QTPQEWIDDL ERLQDRYPWL
61  VAEVEGVVAG IAYAGPWKAR NAYDWTVEST VYVSHRHQRL GLGSTLYTHL LKSMEAQGFK
121 SVVAVIGLPN DPSVRLHEAL GYTARGTLRA AGYKHGGWHD VGFWQRDFEL PAPRPVVRPV
181  TQI
```

Date of search: February 22, 2012

1.2. ALLERGEN DATABASE

AllergenOnline allergen database (www.allergenonline.org) 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 12 of the database includes 1603 unique sequences that are clustered into 603 allergen groups based on species (n=273) and sequence identities.

Name	Database type	Number of sequences	Version	Date of release
AllergenOnline	Allergen database	1603	12	February, 2012

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 program used was the SeqMatchAll tool from the European Molecular Biology Open Software Suite (EMBOSS) ([Rice et al., 2000](#)) (version 6.3.1, July 2010), 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](#)).

The complete query sequence was compared with all the sequences available in the allergen database. The overall homology search used the FASTA program (version 35.04, 2009; [Pearson and Lipman, 1988](#)) and the BLOSUM50 scoring matrix. An E-value threshold of 100 was used. Only the matches of at least 35% identity over 80 amino acids or more were considered 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 PAT/*pat* protein, and known allergens from the AllergenOnline database.

2.2. OVERALL HOMOLOGY SEARCH

No similarities were found between the PAT/*pat* protein and known allergens from the AllergenOnline database, based on a '35% identity over an 80 amino acid segment' matching criteria.

2.3. POTENTIAL N-GLYCOSYLATION SITE SEARCH

No potential N-glycosylation sites were identified on the PAT/*pat* protein amino acid sequence by using the N - X~(P) - [S,T] and 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 PAT/*pat* protein possesses allergenic properties.

4. References

DART #	References
M-276947-01-1	<i>Codex Alimentarius</i> commission (CAC). 2003. <i>Codex principles and guidelines on food derived from biotechnology. Codex Alimentarius Commission. CAC/GL44-2003 and CAC/GL45-2003.</i>
M-135134-03-1	De Beuckeleer, M. 2004. Description of the amino acid sequence of the PAT protein encoded from the <i>pat</i> gene. Study Report <i>pat/pat aas/01</i> . Unpublished. Bayer CropScience. Jan 19, 2004. 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-410793-01-1	Rice, P., Longden, I. and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. <i>Trends in Genetics.</i> 16:276-277.

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
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
PAT	Phosphinothricin acetyltransferase
S	Serin
T	Threonin
X	Any amino acid
X~(P)	Any amino acid except Proline
WHO	World Health Organization

FINAL REPORT AMENDMENT

There is no final report amendment at this time.