12

In silico prediction of potential allergenicity of proteins according to the FAO/WHO guidelines with the help of Allermatch™

Mark W.E.J. Fiers[#], Gijs A. Kleter[##], Ad A.C.M. Peijnenburg[##], Herman Nijland[#], Jan Peter Nap[#] and Roeland C.H.J. van Ham[#,]*

Abstract

Allermatch™ (http://allermatch.org) is a novel web tool for the efficient and standardized prediction of potential allergenicity of proteins according to the current recommendations of the FAO/WHO Expert Consultation, as outlined in the Codex alimentarius. A query amino-acid sequence is compared against the Allermatch™ Allergen database based on current SwissProt and WHO-IUIS allergen lists. The web tool uses a sliding window to identify stretches of 80 amino acids with more than 35% similarity, or identical small stretches of at least six amino acids. The outcome of the analyses is presented in a concise format. Allermatch™ is likely to contribute to improved, transparent and more consistent analyses of potential allergenicity of genetically modified food prior to market release. In the future, the FAO/WHO guidelines may be improved upon. Different methods that could enhance the predictive value of allergen prediction are discussed.
Keywords: allergenicity prediction; FAO/WHO; web tool

Introduction

The safety of genetically modified foods must be assessed before authorities in most nations will consider granting market approval. An important issue in the food-safety assessment is the evaluation of the potential allergenicity of food derived from biotechnology. Food allergy is an immunoglobulin-E-mediated response to food components and is part of a wider group of adverse reactions to food termed ‘food sensitivity’. Food allergy may have symptoms that vary from itching, vomiting and diarrhoea to life-threatening anaphylaxis. As all known food allergens are proteins, the introduction of a new (‘foreign’) protein in food by genetic engineering can cause allergic reactions in a ‘worst case’ scenario. Potential allergenicity of a protein is a complex issue and various tests are used to predict potential allergenicity, including bioinformatics, in-vitro digestibility of the protein, and binding to antisera of allergic patients (Stiekema and Nap 2004; FAO and WHO 2003).

[#] Applied Bioinformatics, Plant Research International, Plant Sciences Group, Wageningen University and Research Centre, P.O. Box 16, 6700 AA Wageningen, The Netherlands
[##] RIKILT Institute of Food Safety, Wageningen University and Research Centre, P.O. Box 230, 6700 AE Wageningen, The Netherlands
[*] Corresponding author: E-mail: roeland.vanham@wur.nl

The FAO/WHO's Codex alimentarius and an Expert Consultation Group have established guidelines to assess potential allergenicity of proteins with bioinformatics in a step-by-step procedure (FAO and WHO 2001; FAO and WHO 2003). Eventually, these guidelines will have to be incorporated into law by all FAO/WHO member states. The guidelines aim to assess whether a given primary protein sequence is sufficiently similar to sequences of known allergenic proteins to give reason for concern. The recommended procedure to establish the potential for allergenicity is as follows (FAO and WHO 2001):

1) Obtain the amino-acid sequences of known allergens in public protein databases in FASTA format (using the amino acids from the mature proteins only, disregarding the leader sequences, if any are annotated)
2) Prepare a complete set of 80-amino-acid length sequences derived from the expressed protein (again disregarding the leader sequence, if any)
3) Compare each of the sequences of (2) with all sequences of (1), using the program FASTA (Pearson and Lipman 1988) with default settings for gap penalty and width.

According to the Codex alimentarius potential allergenicity should be considered, (FAO and WHO 2003) when there is either:

a) more than 35 % similarity over a window of 80 amino acids in the amino-acid sequence of the query protein (without the leader sequence, if any) with an entry known as allergen;

or

b) a stretch of identity of 6 to 8 contiguous amino acids.

If either analysis points to possible allergenicity, the allergenicity of the protein should be verified using serum-binding tests and/or in-vivo methods such as patient panels, skin prick tests or animal exposure tests (Stiekema and Nap 2004).

Features of the Allermatch™ web tool

The Allermatch™ web tool complies with the FAO/WHO criteria given above. The first step was to create databases for the analysis. These databases were established in three steps. First, a Swissprot allergen database was created by extracting all 334 proteins from SwissProt annotated as an allergen (Boeckmann et al. 2003. SwissProt version 44.1, July 5 2004, http://www.expasy.org/cgi-bin/lists?allergen.txt). Leader sequences were, if annotated, trimmed and the mature protein sequences were stored in the Allermatch™ Swissprot allergen database. Second, all 632 entries (excluding some duplicates) from the WHO-IUIS allergen list (King et al. 1994) were extracted from the public databases SwissProt (Boeckmann et al. 2003. Version 44.1, July 5 2004), PIR (Wu et al. 2003) and GenPept (http://www.ncbi.nlm.nih.gov). It should be noted that the WHO-IUIS list contains SwissProt sequences which are not on the SwissProt allergen list and that the SwissProt allergen list contains sequences which are not on the WHO-IUIS list. Annotated leader sequences were trimmed and the sequences were stored in the Allermatch™ WHO/IUIS allergen database. Joining the above databases and removing redundancies created the Allermatch™ combined allergen database. The combined SwissProt and WHO-IUIS allergen databases contained 236 duplicate sequences (Figure 1). The resulting non-redundant Allermatch™ allergen database

contains 730 allergen sequences. The current version of the Allermatch™ webtool allows analysis of a given query protein with any of the three databases created but uses the combined database per default. In the future, it will be possible to upload local sequences to be used as database.

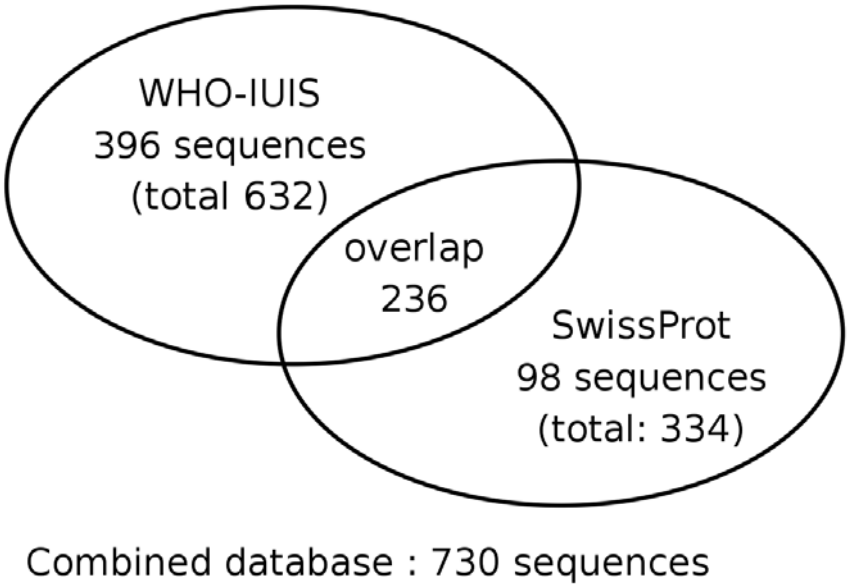

Figure 1. Venn diagram showing the relationship between the two databases used by Allermatch™

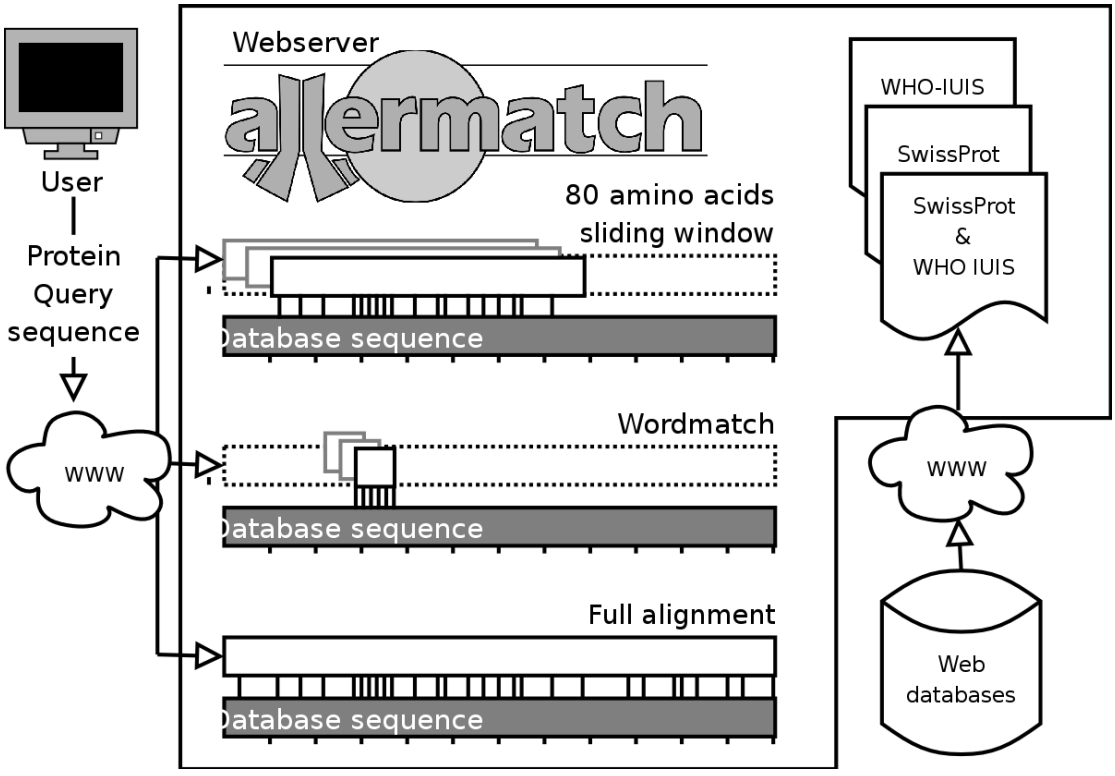

Figure 2. Schematic representation of the Allermatch™ webtool

For protein-sequence alignment, Allermatch™ uses the FASTA program (Pearson and Lipman 1988) version 3.4t2 with default settings (ktup = 2, matrix = Blosum50,

Gap open = -10, Gap extend = -2). All other software is written in Python and runs on a Suse Linux Enterprise Server 8 using mod_python and an Apache webserver (version 1.3.26). Allermatch™ provides three different search modes to assess and visualize the potential allergenicity of proteins (Figure 2).

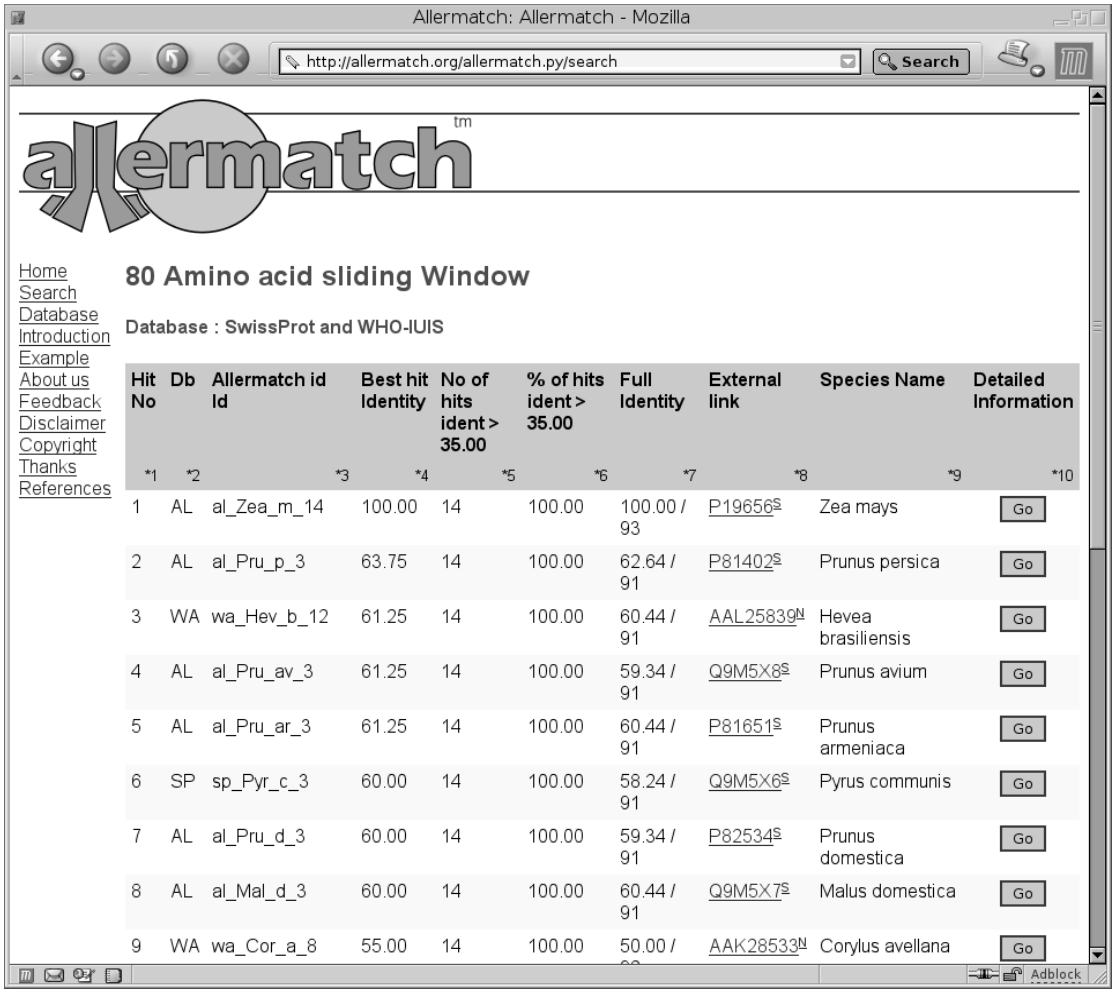

Figure 3. Screenshot of a results screen of an 80-amino-acid alignment. This figure shows an overview of all matches found with the 80-amino-acid sliding-window method on an pollen allergen sequence from *Zea mays* (Zea m 14). The columns represent: 1) the number of the hit, sorted on column 4; 2) the database from which the sequence was derived; 3) the allergen identifier; 4) the best 80-amino-acid similarity of all matched windows; 5) and 6) the number and percentage of windows with a similarity above 35%; 7) the percentage similarity and the number of similar amino acids in a full alignment of the query sequence with this database allergen; 8) the SwissProt identifier and a link to the SwissProt web site; 9) the species name from which the allergen sequence derives; and 10) a link to a page with more details on this specific hit

These three search modes are:

1) Mode 1: 80-amino-acid sliding window
   The query protein sequence is divided into windows of 80 amino acids using a sliding window with steps of a single amino acid. Each of these windows is compared with all sequences in the Allermatch™ allergen database. All database entries showing a similarity higher than a given threshold percentage (default is 35%; a user can adjust the threshold percentage if so desired) to any of all 80-amino-acid query-sequence windows, are identified. Upon completion of the analysis, a table is generated that shows all database entries identified (Figure 3).
   For each database entry, the highest similarity score is given, as well as the number of 80-amino-acid windows having a similarity above the cut-off percentage. For each entry identified, more detailed information can be retrieved on the similarity between the allergen and query sequence, for example, the areas of both proteins within all 80-amino-acid windows that score above the threshold percentage.
   If the similarity score calculated by FASTA applies to stretches smaller than 80 amino acids, Allermatch™ converts such a similarity score to an 80-amino-acid window in a linear fashion. For example, a 40% similarity on a stretch of 40 amino acids converts to 20% similarity in an 80-amino-acid window. This criterion implies that sequences shorter than 80 amino acids need to have higher similarity in order to be identified as a potential allergen.
2) Mode 2: Wordmatch
   The second method looks for short sub-sequences (words) that have a perfect match with a database entry. The word size is configurable (default is 6 amino acids). The resulting output is similar to the output given by Mode 1. All database entries with at least one hit are listed and for each entry more detailed information can be retrieved upon request (Figure 4).
3) Mode 3: Full alignment
   The Allermatch™ webtool offers the full alignment of the query sequence with the Allermatch™ allergen database entries. A FASTA alignment of the entire input protein allows one to obtain a global view of the query's protein similarities with a known allergen and may help to position regions of potential allergenicity in the primary structure of the protein. Upon parsing of the FASTA output, information from the Allermatch™ Allergen database is added and presented. This full alignment is not part of the recommendations of the FAO/WHO guidelines. It is added as an additional useful tool for further research.

Validation of the Allermatch™ webtool

A major issue in the prediction of potential allergenicity of a protein from a biosafety point of view is the likelihood of error. The algorithms used for prediction should be as accurate as possible and have as low an error rate as possible. One can identify two types of errors: a query protein that is identified as a potential allergen, while in fact it is not (i.e. a false positive), and a query protein that is excluded from the possibility of being an allergen, while in fact it is (i.e. a false negative).

Both error types are estimated for each of the three databases evaluated. For the sliding-window approach an 80-amino-acid window with a 35% similarity cut-off is used and for the wordmatch approach 6, 7 and 8 amino-acid word sizes are tested.

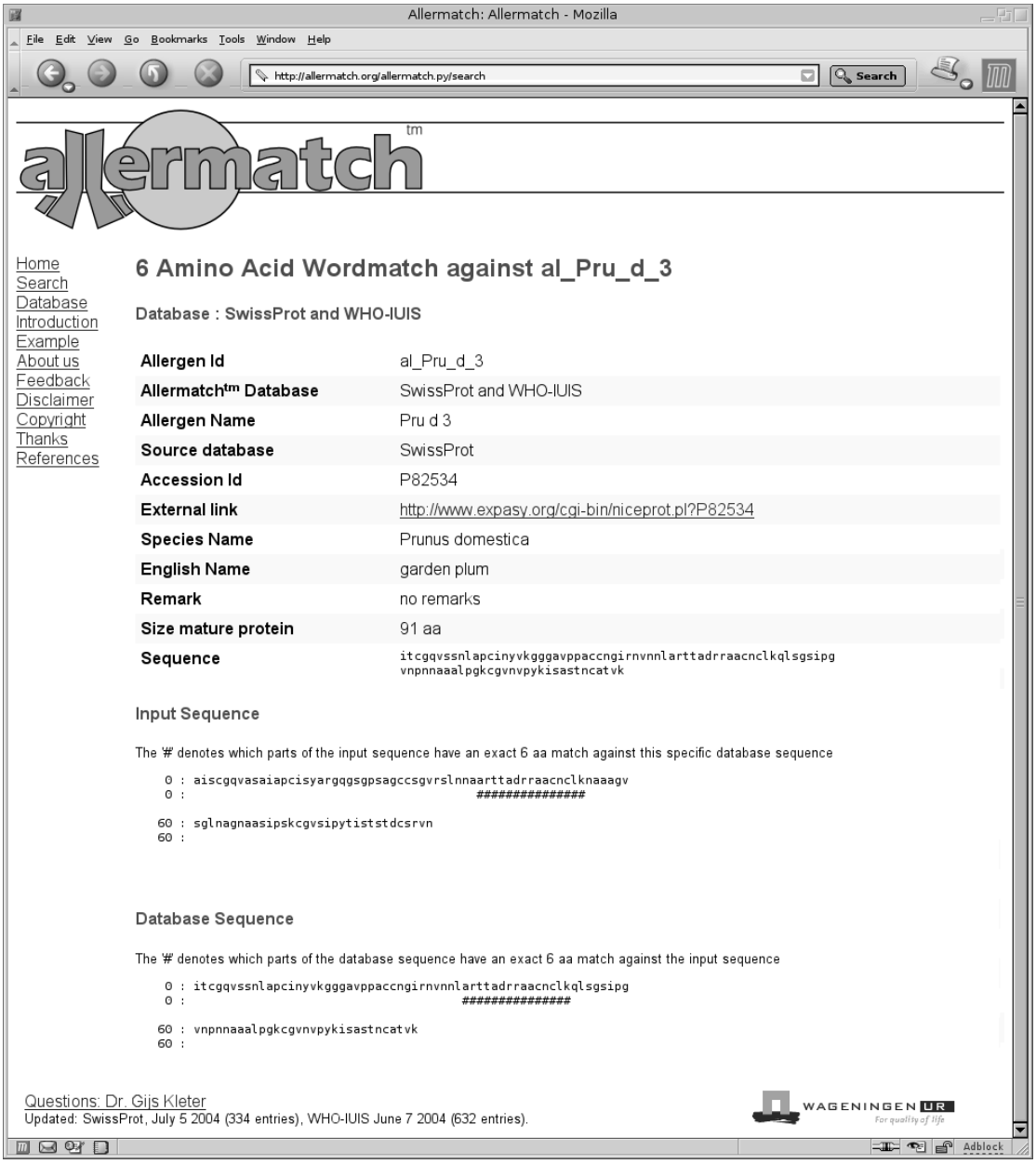

Figure 4. Screenshot of a detailed view of a single wordmatch analysis on the same sequence as used for Figure 3. The image shows detailed information on the sequence from the Allermatch™ allergen database matched. Below two alignments can be seen, the first alignment shows which parts of the input sequence have a 6-amino-acid exact match with the database sequence (marked with #). The second alignment displays the same for the allergen database sequence

Estimation of the error rate of false negatives

It is not easy to investigate the detection rate of false negatives by the algorithms employed in the Allermatch webtool, as there are no proteins known as allergen while they are not represented in the databases used (as there should be none). As an approximation we have determined the number of 'orphan' entries in the Allermatch allergen database. An orphan entry is an entry that, according to the Allermatch analysis, has no relationship to any other entry in the database except itself. Such an orphan entry would represent a false negative relative to that database if this sequence were not present in this database. This approach is also called 'jack-knifing'. The

number of orphan entries in a database is an approximation of the false-negative rate. The results of performing this analysis on all three databases and using the first two analysis methods are summarized in Table 1.

Table 1. Prediction quality of the FAO/WHO methods
The number and percentage of false negative and false positive hits is shown here for all FAO/WHO-recommended method/database combinations. Result set 1 describes the number of false negative hits observed in a leave-one-out method. The next result set (2) shows the same results but corrected for those sequences that were not able to generate a hit against themselves due to the short length of the sequences. The last (3) result set shows the observed number of false positives when testing 12 non-allergenic sequences (see table 2) against the Allermatch™ web tool. Each of the result sets consists of two columns; the first column shows the number of false hits and the total number of sequences in this set. The second column shows the percentage of false hits

Database	Method	Word size	Result set 1 False negatives		Result set 2 False negatives (corrected)		Result set 3 False positives	
			Number	%	Number	%	Number	%
SwissProt	Window	n.a.	71 / 334	21.3	57 / 320	17.8	3 / 12	25.0
	Wordmatch	6	54 / 334	16.2	n.a.	n.a.	7 / 12	58.3
		7	69 / 334	20.7	n.a.	n.a.	6 / 12	50.0
		8	78 / 334	23.4	n.a.	n.a.	3 / 12	25.0
WHO-IUIS	Window	n.a.	99 / 632	15.7	78 / 611	12.8	4 / 12	33.3
	Wordmatch	6	58 / 632	9.2	n.a.	n.a.	9 / 12	75.0
		7	98 / 632	15.5	n.a.	n.a.	8 / 12	66.7
		8	117 / 632	18.5	n.a.	n.a.	3 / 12	25.0
SwissProt & WHO-IUIS	Window	n.a.	101 / 730	13.8	77 / 706	10.9	5 / 12	41.7
	Wordmatch	6	55 / 730	7.5	n.a.	n.a.	9 / 12	75.0
		7	95 / 730	13.0	n.a.	n.a.	8 / 12	66.7
		8	115 / 730	15.8	n.a.	n.a.	3 / 12	25.0

In examining the false negative results, various sequences were observed that did not produce a hit against themselves (data not shown). On closer inspection, this was found to be due to the short length of these protein sequences. If a sequence is shorter than 28 amino acids, even 100% similarity will convert the similarity to less than 35% after conversion to an 80-amino-acid window. This may overestimate the error rate. Therefore, we also determined the false-negative rates with those sequences not able to generate a hit against themselves excluded. Even after this correction the wordmatch method, with a 6-amino-acid word length, gives a lower percentage of false negatives than the sliding-window approach.

Estimation of the error rate of false positives

The second control examines 12 proven non-allergenic sequences against the Allermatch™ databases. Non-allergenicity can, for example, be based on non-reactivity of these proteins towards IgE sera of allergy patients or the inability to cause IgE responses in experimental animals (see Table 2). It should be noted that such data exist only for a limited number of proteins, which also accounts for the size of this dataset. A non-allergenic sequence is not supposed to generate a hit; therefore we consider each hit a false positive. Results are summarized in Table 1, result set 3.

Table 2. Sequences used for the negative control

Protein	Host organism	Evidence for non-allergenicity	Accession	Reference
Amaranth seed albumin	*Amaranthus hypochondriacus*	IgG response, but no raised IgE levels, after administration (intranasal and intraperitoneal) of amaranth seed albumin to mice	GenPept CAA77664	(Chakraborty, Chakraborty and Datta 2000)
T1	*Catharanthus roseus*	No reaction of recombinant T1 in IgE sera binding, basophile histamine release, and skin prick testing using patients allergic to the related birch pollen allergen Bet v 1	Not applicable	(Laffer et al. 2003)
Mite ferritin heavy chain	*Dermatophagoides pteronyssinus*	Reaction of mite ferritin with IgG, but not with IgE, of sera from patients allergic to house-dust mite	GenPept AAG02250	(Epton et al. 2002)
Maltose-binding protein	*Escherichia coli*	No reaction with IgE sera from patients allergic to natural rubber latex (maltose binding protein used as part of fusion proteins with latex allergens)	SwissProt P02928	(Rihs et al. 2003)
Human serum albumin	*Homo sapiens*	No reaction of human serum albumin with IgE sera of patients allergic to cat and porcine serum albumin	SwissProt P02768	(Hilger et al. 1997)
Human heat-shock protein 70	*Homo sapiens*	No reaction of human heat-shock protein 70 with IgE-sera of patients allergic to heat-shock protein 70 from *Echinococcus granulosus*	SwissProt P08107	(Ortona et al. 2003)
Human beta-2-glycoprotein I	*Homo sapiens*	Presence of IgM antibodies, but not of IgE antibodies, directed against human beta-2-glycoprotein I in sera from atopic eczema/dermatitis patients	SwissProt P02749	(Szakos et al. 2004)
Guayule rubber-particle protein	*Parthenium argentatum*	No cross-reactivity between proteins from guayule and latex using IgE-sera from patients allergic to latex	Swissprot Q40778	(Siler, Cornish and Hamilton 1996)
Purle acid phosphatase 1	*Solanum tuberosum*	Stimulation of IgG-, but no or only low stimulation of IgE-antibodies following administration of potato acid phosphatase to mice (oral and intraperitoneal)	TrEMBL Q6J5M7	(Dearman and Kimber 2001)
Purle acid phosphatase 2	*Solanum tuberosum*	See above	TrEMBL Q6J5M9	(Dearman and Kimber 2001)
Purle acid phosphatase 3	*Solanum tuberosum*	See above	TrEMBL Q6J5M8	(Dearman and Kimber 2001)
Potato lectin	*Solanum tuberosum*	Stimulation of IgG-, but no or only low stimulation of IgE-antibodies following administration of potato lectin to mice (intraperitoneal)	TrEMBL Q9S8M0	(Dearman et al. 2003)

Discussion

Prediction of allergenicity can broadly be done in two ways: one can look for linear or conformational epitopes (Bredehorst and David 2001). The first method tries to assess whether two proteins share similarities in the primary sequence, whereas the second method looks at similarities in 3D structure. The Codex guidelines recommend a combination of both approaches. Short exact word matches and positive hits in the sliding 80-amino-acid window may indicate potential linear epitopes and similar 3D structures, respectively.

Examination of the false-negative rate (see Table 1, result set 1) clearly shows a link between the database size and the false-negative hit rate. As the database grows in size, the false-negative hit rate reduces. This is to be expected since the probability that an allergen with sufficient similarity is present in the database is bigger when the database is larger. However, another (part of an) explanation could be that the larger database has more isoallergen families (a group of allergens with minor sequence differences) present, which will diminish the chance of false negatives since fewer sequences will be 'orphans'. Another factor possibly influencing the false-negative hit rate are signal peptides that could still be present in the database due to poor annotation of these proteins. Positive hits might be generated by these signal peptides for proteins that should have been orphans.

When evaluating the false positive hits we see a similar trend; the number of false positives grows with the database size, as is to also be expected since the chance of a random hit increases with a larger database. In contrast to the false-negative hit rates however, the sliding-window method gives a lower percentage of erroneous hits here. The results of this test might overestimate the number of false positives, since a number of these non-allergens are related to and display similarities with their allergenic counterparts, *i.e.* T1 is related to Bet v 1 (Laffer et al. 2003), human serum albumin is related to animal serum albumins (Hilger et al. 1997) and human heat-shock protein 70 is similar to heat-shock proteins from fungi and other allergens (Ortona et al. 2003) (Table 2). A true selection of unrelated, non-allergenic proteins is therefore likely to give a lower false-positive rate.

These results show that by choosing a database and algorithm one can influence the error rates towards either a higher rate of false positives or towards more false negatives. A too high detection rate of false positives would generate an unnecessary and undesirable burden of additional testing of proteins used in genetic engineering. On the other hand, a too high detection rate of false negatives would generate undesirable potential health risks for consumers. Either error is undesirable, but because this bioinformatics analysis identifies proteins for further testing of true allergenicity, a 'better safe than sorry' strategy could be opted for. Such a strategy would obviously strive to minimize the detection rate of false negatives by using the results of both the sliding window and the six-amino-acid word match against the largest Allermatch™ combined allergen database. Positive results from these analyses should first be analysed in depth by checking medical literature on these proteins. Ultimately all valid predictions will, as suggested by FAO/WHO, have to be tested further with methods as skin prick tests or animal models. Even after these tests there is no absolute certainty that the protein in question will never elicit an allergenic reaction. In time people might still become sensitized to the protein as a novel allergen, or only a very small part of the population is sensitive to cross-reacting allergens, too small to have been noticed in the tests.

In general, one should keep in mind that performance of these algorithms is far from perfect. This is in agreement with other literature where similar results for the FAO/WHO methods are shown and other algorithms proven to give better results (Soeria-Atmadja et al. 2004; Zorzet, Gustafsson and Hammerling 2002; Kleter and Peijnenburg 2002). These supplementary methods include, for example, advanced motif discovery methods where a complete allergen database is scanned for highly represented motifs. These motifs are then used to identify possible allergenicity (Stadler and Stadler 2003). In addition, a machine-learning approach was described using FASTA and a neural network to compare query proteins with allergens (Zorzet, Gustafsson and Hammerling 2002).

In the public domain, three other websites exist that assess potential allergenicity of proteins based on their primary sequence:

- SDAP http://fermi.utmb.edu/SDAP/
- FARRP http://www.allergenonline.com
- AllerPredict http://research.i2r.a-star.edu.sg/Templar/DB/Allergen/

These websites are also able to perform complete FASTA alignments (SDAP, Farrp), 80-amino-acid sliding window (SDAP, Allerpredict) and 6 to 8 amino acids exact matches (SDAP, Allerpredict).

Allermatch™ will greatly enhance and improve the prediction of allergenicity according to current guidelines in the Codex by combining all recommended algorithms in a single website. In the future, the Allermatch™ web tool will stay updated with the public allergen databases on a regular basis and the requirements by law on assessing allergenicity. To increase the predictive power, supplementary bioinformatics facilities will be added. Such additional facilities may include, among other things, the possibility to do batch analyses, to upload users' own databases, and to use supplementary tools such as the examples described above. Feedback from users will help us to identify particular issues that address their needs.

References

Boeckmann, B., Bairoch, A., Apweiler, R., et al., 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, 31 (1), 365-370.

Bredehorst, R. and David, K., 2001. What establishes a protein as an allergen? *Journal of Chromatogry. B, Biomedical Sciences and Applications*, 756 (1/2), 33-40.

Chakraborty, S., Chakraborty, N. and Datta, A., 2000. Increased nutritive value of transgenic potato by expressing a nonallergenic seed albumin gene from *Amaranthus hypochondriacus*. *Proceedings of the National Academy of Sciences of the United States of America*, 97 (7), 3724-3729.

Dearman, R.J. and Kimber, I., 2001. Determination of protein allergenicity: studies in mice. *Toxicology Letters*, 120 (1/3), 181-186.

Dearman, R.J., Stone, S., Caddick, H.T., et al., 2003. Evaluation of protein allergenic potential in mice: dose-response analyses. *Clinical and Experimental Allergy*, 33 (11), 1586-1594.

Epton, M.J., Smith, W., Hales, B.J., et al., 2002. Non-allergenic antigen in allergic sensitization: responses to the mite ferritin heavy chain antigen by allergic and non-allergic subjects. *Clinical and Experimental Allergy*, 32 (9), 1341-1347.

FAO and WHO, 2001. *Evaluation of allergenicity of genetically modified foods: report of a Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology, 22 – 25 January 2001*. FAO, Rome. [http://www.who.int/foodsafety/publications/biotech/en/ec_jan2001.pdf]

FAO and WHO, 2003. *Codex principles and guidelines on foods derived from biotechnology*. FAO, Rome. [ftp://ftp.fao.org/codex/standard/en/CodexTextsBiotechFoods.pdf]

Fiers, M.W., Kleter, G.A., Nijland, H., et al., 2004. Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics*, 5 (1), 133.

Hilger, C., Kohnen, M., Grigioni, F., et al., 1997. Allergic cross-reactions between cat and pig serum albumin. Study at the protein and DNA levels. *Allergy*, 52 (2), 179-187.

King, T.P., Hoffman, D., Lowenstein, H., et al., 1994. Allergen nomenclature. *International Archives of Allergy and Immunology*, 105 (3), 224-233.

Kleter, G.A. and Peijnenburg, A.A., 2002. Screening of transgenic proteins expressed in transgenic food crops for the presence of short amino acid sequences identical to potential, IgE - binding linear epitopes of allergens. *BMC Structural Biology*, 2 (1), 8. [http://www.biomedcentral.com/1472-6807/2/8]

Laffer, S., Hamdi, S., Lupinek, C., et al., 2003. Molecular characterization of recombinant T1, a non-allergenic periwinkle (*Catharanthus roseus*) protein, with sequence similarity to the Bet v 1 plant allergen family. *Biochemical Journal*, 373 (Pt 1), 261-269.

Ortona, E., Margutti, P., Delunardo, F., et al., 2003. Molecular and immunological characterization of the C-terminal region of a new *Echinococcus granulosus* Heat Shock Protein 70. *Parasite Immunology*, 25 (3), 119-126.

Pearson, W.R. and Lipman, D.J., 1988. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, 85 (8), 2444-2448.

Rihs, H.P., Dumont, B., Rozynek, P., et al., 2003. Molecular cloning, purification, and IgE-binding of a recombinant class I chitinase from *Hevea brasiliensis* leaves (rHev b 11.0102). *Allergy*, 58 (3), 246-251.

Siler, D.J., Cornish, K. and Hamilton, R.G., 1996. Absence of cross-reactivity of IgE antibodies from subjects allergic to *Hevea brasiliensis* latex with a new source of natural rubber latex from guayule (*Parthenium argentatum*). *Journal of Allergy and Clinical Immunology*, 98 (5 Pt 1), 895-902.

Soeria-Atmadja, D., Zorzet, A., Gustafsson, M.G., et al., 2004. Statistical evaluation of local alignment features predicting allergenicity using supervised classification algorithms. *International Archives of Allergy and Immunology*, 133 (2), 101-112.

Stadler, M.B. and Stadler, B.M., 2003. Allergenicity prediction by protein sequence. *Faseb Journal*, 17 (9), 1141-1143.

Stiekema, W.J. and Nap, J.P., 2004. Bioinformatics for biosafety: predicting the potential allergenicity of GM Food. *In:* Nap, J.P. and Atanassov, A. eds. *Genomics for biosafety in plant biotechnology*. IOS Press, Amsterdam, 98-114. NATO science series vol. 359.

Szakos, E., Lakos, G., Aleksza, M., et al., 2004. Association between the occurrence of the anticardiolipin IgM and mite allergen-specific IgE antibodies in children with extrinsic type of atopic eczema/dermatitis syndrome. *Allergy*, 59 (2), 164-167.

Wu, C.H., Yeh, L.S., Huang, H., et al., 2003. The protein information resource. *Nucleic Acids Research*, 31 (1), 345-347.

Zorzet, A., Gustafsson, M. and Hammerling, U., 2002. Prediction of food protein allergenicity: a bioinformatic learning systems approach. *In Silico Biology*, 2 (4), 525-534.

Endnote

This article is an extended version of an article published in BMC Bioinformatics by the same authors (Fiers et al. 2004).