

Presence of potential allergy-related linear epitopes in novel proteins from conventional crops and the implication for the safety assessment of these crops with respect to the current testing of genetically modified crops

Gijs A. Kleter* and Ad A. C. M. Peijnenburg

RIKILT Institute of Food Safety, PO Box 230, NL 6700 AE Wageningen, the Netherlands

Received 16 May 2003;

revised 1 July 2003;

accepted 1 July 2003.

*Correspondence (fax +31 317 417 717;

e-mail gijs.kleter@wur.nl)

Abstract

Mitochondria of cytoplasmic male sterile crop plants contain novel, chimeric open reading frames. In addition, a number of crops carry endogenous double-stranded ribonucleic acid (dsRNA). In this study, the novel proteins encoded by these genetic components were screened for the presence of potential binding sites (epitopes) of allergy-associated IgE antibodies, as was previously done with transgenic proteins from genetically modified crops. The procedure entails the identification of stretches of at least six contiguous amino acids that are shared by novel proteins and known allergenic proteins. These stretches are further checked for potential linear IgE-binding epitopes. Of the 16 novel protein sequences screened in this study, nine contained stretches of six or seven amino acids that were also present in allergenic proteins. Four cases of similarity are of special interest, given the predicted antigenicity of the identical stretch within the allergenic and novel protein, the IgE-binding by a peptide containing an identical stretch reported in literature, or the multiple incidence of identical stretches of the same allergen within a novel protein. These selected stretches are present in novel proteins derived from oilseed rape and radish (ORF138), rice (dsRNA), and fava bean (dsRNA), and warrant further clinical testing. The frequency of positive outcomes and the sizes of the identical stretches were comparable to those previously found for transgenic proteins in genetically modified crops. It is discussed whether novel proteins from conventional crops should be subject to an assessment of potential allergenicity, a procedure which is currently mandatory for transgenic proteins from genetically modified crops.

Keywords: allergy, cytoplasmic male sterility, double-stranded RNA, food crops, genetic modification, novel proteins.

Introduction

As a natural form of 'genetic engineering', conventional food crops may contain novel open reading frames (*orfs*) that could potentially encode novel proteins. Two types of novel *orfs* have been well characterized. First of these are chimeric *orfs* in the mitochondrial DNA that have been formed by recombination events in cytoplasmic male sterile plants. Cytoplasmic male sterile lines with these novel *orfs* are widely

exploited in crop breeding because they eliminate the need for manual emasculation of female plants. The changes in mitochondrial DNA associated with cytoplasmic male sterility in a number of crops were recently reviewed (Schnable and Wise, 1998). Studies on the expression of these *orfs* focused primarily on RNA-transcripts, while in some cases, the protein that is also encoded by these *orfs* has been identified, e.g. in sunflower and wheat (Moneger *et al.*, 1994; Song and Hedgcoth, 1994). Sterile male plants can be reverted to

fertility if functional nuclear restorer genes are introduced through the male parent.

Second, *orfs* of endogenous extrachromosomal, self-replicating double-stranded RNA (dsRNA) are present in a number of food crops. Given the similarities between dsRNA from unrelated food crops (beans and rice) and other viruses, it has been postulated that these dsRNA have a viral ancestor (Gibbs *et al.*, 2000). In contrast to dsRNA of some plant viruses, these dsRNA molecules are capable of replicating themselves, but do not give rise to viruses. The molecular characterization of dsRNA in rice, fava bean and pear has been documented, including the amino acid sequences encoded by the *orfs* (Fukuhara, 1999; Osaki *et al.*, 1998; Pfeiffer, 1998). RNA-dependent RNA-polymerase encoded by the rice dsRNA has been detected in, and purified from, cultured rice cells (Horiuchi *et al.*, 2001).

The novel proteins encoded by these *orfs* in conventional crops are not subject to safety evaluations in current regulatory practice, contrary to 'foreign' transgenic proteins in genetically modified crops. During the regulatory safety assessment of genetically modified crops, the potential allergenicity of transgenic proteins is one of the main issues to be addressed. An internationally acknowledged decision tree is used by safety evaluators world-wide for the assessment of potential allergenicity (Metcalf *et al.*, 1996). Various methods are devised by this decision tree, depending upon the origin of the foreign gene and intermediate outcomes. These methods include:

- Comparison of the primary structure (amino acid sequence) of the novel protein with the structures of known allergenic proteins. Both overall structural similarities and the occurrence of short identical stretches between the novel and allergenic proteins are screened for. Short identical stretches may constitute, among others, linear epitopes for IgE-antibodies, which are involved in the recognition of allergens. By contrast, non-linear, three-dimensional IgE epitopes containing amino acid residues that are spatially grouped together but separated within the primary structure (Bredehorst and David, 2001), will not be identified by screening for linear stretches.
- Digestibility of the novel protein. This is assessed *in vitro* in model solutions containing gastro-intestinal protein-degrading enzymes. Degradation in the stomach, for example, is assessed by the incubation of a protein with simulated gastric fluid, which is an acidic hydrochloric solution containing the acid protease pepsin. Protein integrity can be determined by gel electrophoresis of the incubation mixture sampled at various time points,

followed by the staining of separated protein bands in the gel. If a protein is found to be stable in gastro-intestinal digestion, it is more likely to interact with the gut-associated immune system and thereby become an allergen.

- Binding of antisera from allergic patients. If the source of the foreign gene is an allergen, then the antisera of patients allergic to that allergen should be used for screening. Similarly, if the novel protein exhibits relevant similarity to an allergenic protein, antisera from patients that are allergic to the latter protein should be used.
- Further clinical testing, including skin prick tests and food challenges with allergy patients, may be considered necessary if the previous antisera binding tests yield negative results.

Against this background, international guidance is currently being developed by the Task Force on Food Derived from Biotechnology of the Codex alimentarius Committee of the joint Food and Agriculture Organization and World Health Organization of the United Nations (FAO/WHO, 2002). The standards of this committee practically have legal status and will be implemented by its Member States. The guidance developed by this committee will therefore set minimal requirements for the safety evaluation of genetically modified foods. The Task Force is currently in the final stage of completing guidance for the safety assessment of genetically modified foods.

In its preparations of the guidance for safety assessment, the Task Force has gathered scientific advice from recent FAO/WHO Expert Consultations. One of these consultations addressed the potential allergenicity of genetically modified foods (FAO/WHO, 2001). During this consultation, the experts advised an updating of the decision tree for the assessment of potential allergenicity of genetically modified foods. Among others, it was advised to include targeted antisera screens for novel proteins from gene sources without a history of allergy. The antisera used in such screens should be taken from patients that are allergic to organisms that are broadly related to the gene source. In addition, the use of animal models was recommended. For a comparison of the amino acid sequences of novel proteins to those of allergenic proteins, it was devised to lower the threshold for short identical stretches from eight to six contiguous amino acids, based on new insights on the minimal size of IgE epitopes. In this revised approach, the outcomes of the various studies should indicate the probability that the novel protein would be an allergen. In case of identical sequences of six or more contiguous amino acids, for example, the novel protein should be classified as 'likely allergenic'.

Table 1 Investigated novel mitochondrial proteins

Source		Amino acid sequence		
<i>orf</i>	Plant	Size (aa)	Accession (gi)	Reference
<i>orfA501</i>	<i>Allium schoenoprasum</i> (chives), CMS1 cytoplasm	167	–	Engelke and Tatioglu (2002)*
<i>orf138</i>	<i>Brassica napus</i> (oilseed rape) and <i>Raphanus sativus</i> (radish), ogu cytoplasm	138	732177	Bonhomme <i>et al.</i> (1992); Krishnasamy and Makaroff (1993)†
<i>orf222</i>	<i>Brassica napus</i> (oilseed rape), nap cytoplasm	222	1800191	L'Homme <i>et al.</i> (1997)
<i>orf224</i>	<i>Brassica napus</i> (oilseed rape), pol cytoplasm	224	257560	Singh and Brown (1991)‡
<i>orfB</i>	<i>Daucus carota</i> (carrot)	216	10801544	Nakajima <i>et al.</i> (2001)§
<i>orfH522</i>	<i>Helianthus annuus</i> (sunflower), PET1 cytoplasm	174	12987	Koehler <i>et al.</i> (1991)
<i>orf79</i>	<i>Oryza sativa</i> (rice), Bo cytoplasm	79	440860	Akagi <i>et al.</i> (1994)¶
<i>orf239</i>	<i>Phaseolus vulgaris</i> (common bean)	239	169321	Johns <i>et al.</i> (1992)
<i>orf256</i>	<i>Triticum aestivum</i> (wheat)	256	903885	Rathburn and Hedgcock (1991)**
<i>orf77</i>	<i>Zea mays</i> (maize), S cytoplasm	40	–	Zabala <i>et al.</i> (1997)
<i>orf355</i>	<i>Zea mays</i> (maize), S cytoplasm	355	2266936	Zabala <i>et al.</i> (1997)
<i>orf13</i>	<i>Zea mays</i> (maize), T cytoplasm	115	82756	Dewey <i>et al.</i> (1986)
<i>orf107</i>	<i>Sorghum bicolor</i> (sorghum), A3 cytoplasm	107	7489656	Tang <i>et al.</i> (1996)††

aa = amino acid; gi = general identifier for protein accession file; *orf* = open reading frame.

*Protein sequence translated from DNA sequence in reference (ExPASy SequenceTool).

†Edited sequences B, E, F and H present in cultivated radish also included (Yamagashi and Terachi, 2001).

‡Edited sequence also included (Handa *et al.*, 1995; Stahl *et al.*, 1994).

§Database accession sequence differs from literature on positions 28 and 122. Both the sequences from literature and the accession file have been included, each in edited and nonedited form (Nakajima *et al.*, 2001).

¶DNA sequence of *orf79* provided in literature reference.

**Literature sequence is edited form of accession file sequence; both sequences have been included.

††Edited sequence also included (Tang *et al.*, 1996).

As discussed above, the information from amino acid sequence comparisons has a role in the safety assessment, especially for proteins derived from sources without a history of allergenicity. We have previously assessed the consequences of the devised lower threshold for the identical stretches of the primary structures that are shared by transgenic and allergenic proteins. Based upon the positive outcomes for a majority of these transgenic proteins, we have proposed and applied further refinements by: (i) a comparison of the identical stretches with linear IgE-epitopes described in literature, and (ii) a prediction of the antigenicity of the stretch, i.e. the likelihood that it will be bound by antibodies. Positive outcomes of this new strategy warrant further clinical testing in order to verify true allergenicity (Kleter and Peijnenburg, 2002).

The same approach has been applied in this study to 16 novel (putative) proteins in food crops, i.e. 13 proteins encoded by *orfs* in the mitochondria of cytoplasmic male sterile crops and three proteins encoded by endogenous dsRNA. The data from this work provide a reference for current studies focusing on the potential allergenicity of genetically modified foods.

Results

Sixteen novel proteins have been studied, including 13 proteins encoded by mitochondrial *orfs* of cytoplasmic male sterile food crops and three proteins encoded by *orfs* of endogenous dsRNA in food crops (Tables 1 and 2).

Number of identical stretches

In nine of the 16 novel proteins, identical stretches have been identified with a length of at least six amino acids that are also present in known allergenic proteins (Table 3). Thirty-nine

Table 2 Investigated dsRNA-derived novel proteins

Plant	Amino acid sequence		
	Size (aa)	Accession (gi)	Reference
<i>Oryza sativa</i> (rice)	4572	1030716	Moriyama <i>et al.</i> (1995)
<i>Vicia faba</i> (fava bean)	5825	7460238	Pfeiffer (1998)
<i>Pyrus pyrifolia</i> (sand pear)	477	3970837	Osaki <i>et al.</i> (1998)

Table 3 Summary of sizes of identical stretches

Novel protein		Number of identical stretches*	
Identifier	Source	6 aa	7 aa
ORFA501	<i>Allium schoenoprasum</i>	–	–
ORF138	<i>Brassica napus</i> and <i>Raphanus sativus</i>	1 (1)†	–
ORF222	<i>Brassica napus</i>	1 (1)	–
ORF224	<i>Brassica napus</i>	–	–
ORFB	<i>Daucus carota</i>	–	–
ORFH522	<i>Helianthus annuus</i>	3 (2)	–
ORF79	<i>Oryza sativa</i>	–	–
ORF239	<i>Phaseolus vulgaris</i>	1 (1)	–
ORF256	<i>Triticum aestivum</i>	1 (1)	–
ORF77	<i>Zea mays</i>	–	–
ORF355	<i>Zea mays</i>	2 (2)	–
ORF13	<i>Zea mays</i>	–	–
ORF107	<i>Sorghum bicolor</i>	–	–
dsRNA	<i>Oryza sativa</i>	16 (13)	2 (2)
dsRNA	<i>Vicia faba</i>	17 (13)	3 (2)
dsRNA	<i>Pyrus pyrifolia</i>	1 (1)	–
Total		43 (35)	5 (4)

aa = amino acids.

*Values between brackets are identical stretches shared with allergens listed by IUIS/WHO.

†The sequence KEEKKE is present threefold in ORF138, and twofold as partially overlapping sequences in Phl p 13.

of the 48 identical stretches involve allergenic proteins that are officially listed by the International Union of Immunological Societies and by the World Health Organization (IUIS/WHO; <http://www.allergen.org>). The size of 43 identical stretches is six amino acids, while that of the five remaining stretches is seven amino acids. In two instances, multiple identical stretches of the same allergen are present in a novel

protein. Three copies of the KEEKKE sequence of grass pollen allergen Phl p 13, for example, are present in ORF138 associated with the 'ogu' cytoplasmic male sterility of oilseed rape and radish (Table 4).

Literature search for linear IgE epitopes

The identical stretches have been checked for their possible coincidence with documented linear IgE binding epitopes. This is based in particular on data on peptides that would (or would not) bind IgE from allergic patients' antisera. Other relevant data include, for example, the effect of amino acid substitutions on IgE binding.

For one identical stretch, ESLLKAN, literature data indicate that it is part of an IgE-binding peptide of an allergenic protein. This stretch is equivalent between the novel protein derived from the dsRNA of the fava bean (*Vicia faba*) and two tropomyosin allergens from the lobster (Pan s 1) and shrimp (Met e 1). Another peptide containing the ESLLKAN subsequence, however, was not bound in the same study (Table 4; Ayuso *et al.*, 2002).

Antigenicity prediction

Two identical stretches have been predicted to be the most probable antigenic site, i.e. they contained the highest point of the antigenicity prediction plot, made according to the method of Hopp and Woods (1981) (window size 6 amino acids). One stretch, KEEKKE, is equivalent between ORF138 of oilseed rape/radish and the grass pollen allergen Phl p 13 (Table 4; Figure 1). The other stretch, EIEEFR, is equivalent between dsRNA of rice (*Oryza sativa*) and grass pollen allergens Phl p 11 and Lol p 11.

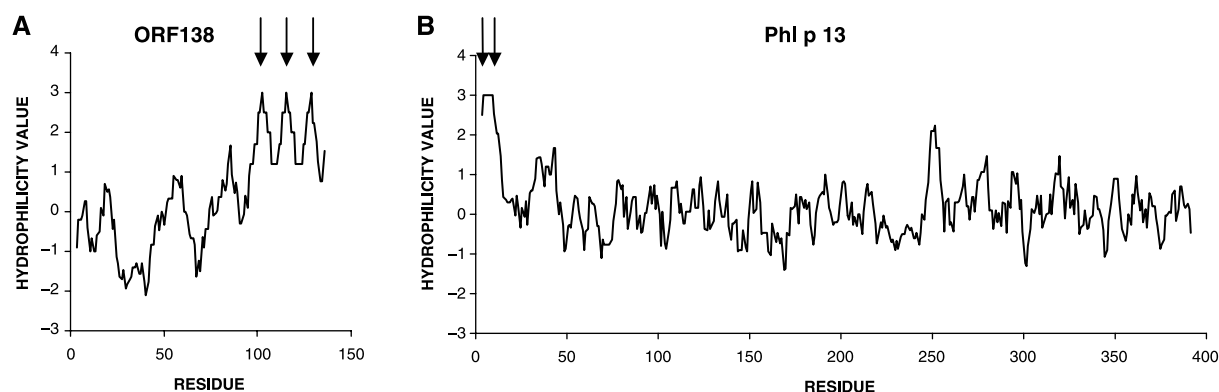


Figure 1 Antigenicity prediction plots created with the Hopp and Woods method, window size six amino acids, of the ORF138 protein of cytoplasmic male sterile brassicas (A), and the grass pollen allergen Phl p 13 (B). The sequence of the non-RNA edited form of ORF138 has been used for plotting. In the plots of the edited sequences of ORF138, the KEEKKE repeats also display the highest scores. Location of the KEEKKE identical stretch shared by both proteins is indicated with an arrow.

Table 4 Selected identical stretches of special interest

Novel protein*	Identical peptide	Allergen†	Remark
ORF138 <i>Brassica napus</i> and <i>Raphanus sativus</i> (oilseed rape and radish)	KEEKKE	Grass pollen allergen Phl p 13	Identical stretch: (1) predicted to be antigenic determinant of allergenic and novel proteins (2) three repeats with identical stretch are present in novel protein, potential crosslinking of IgE if bound by these stretches‡ (3) KEEKKE corresponds to the KXEE/KEXE (X = empty, T or A) IgE-binding motif sequence of latex allergen Hev b 5 in immunized mice§
dsRNA <i>Oryza sativa</i> (rice)	EIEEFR	Grass pollen allergens Phl p 11 and Lol p 11	identical stretch predicted to be antigenic determinant of allergenic protein
dsRNA <i>Vicia faba</i> (fava bean)	KYESNK LSLVTK	Mite allergens Eur m 14 and Der p 14	Double stretch shared with same allergenic proteins, potential crosslinking of IgE if bound by both stretches
dsRNA <i>Vicia faba</i> (fava bean)	ESLLKAN	Shrimp allergen Met e 1 and lobster allergen Pan s 1	identical stretch reported to be part of 15-mer peptide that is bound by IgE from shrimp-allergic patients' antisera¶

*Accessions: ORF138 (*Brassica napus* and *Raphanus sativus*): gi 732177; dsRNA (*Oryza sativa*): gi 1030716; ds RNA (*Vicia faba*): gi 7460238.

†Accessions: Phl p 13: gi 4826572; Phl p 11: gi 23452313; Lol p 11: gi 626029, gi 1582249; Eur m 14: gi 6492307; Der p 14: gi 20385544; Met e 1: gi 607633, gi 6094504; Pan s 1: gi 3080761, gi 14285797.

‡Besides the nonedited sequence, the RNA-edited sequences B, E, and H of ORF138 from cultivated radish also contain three KEEKKE repeats, whereas the edited sequence F with a 13-amino acid deletion of the C-terminal sequence contains two KEEKKE repeats (Yamagashi and Terachi, 2001).

§Reference on IgE-binding sequence motif of Hev b 5: Slater *et al.* (1999).

¶15-mer peptide from shrimp allergen Pen a 1 sequence containing ESLLKAN was bound by one of 18 shrimp-allergic patients' antisera, while another 15-mer peptide containing ESLLKAN was not recognized by any antisera (Ayuso *et al.*, 2002).

Discussion

In the present study, we have sought to identify short identical stretches (peptides) occurring in novel proteins from conventional crops and allergenic proteins. For this purpose, Internet-accessible tools were used. Initially, the amino acid sequence of the novel protein was screened against the sequences of known allergenic proteins stored in a database. Short matches were identified using the BLAST algorithm.

This exercise corresponds with the common practice of 'sequence homology' that is part of the assessment of potential allergenicity of transgenic proteins expressed in genetically modified crops. Besides bioinformatic analysis of sequence homology, other factors that have been recommended for the assessment of potential allergenicity include degradation of the novel protein under conditions simulating processing and digestion, reactions with antisera from allergic individuals, animal testing, and, eventually, clinical testing (FAO/WHO, 2001). Another point of consideration is the amount of novel protein that consumers will be exposed to through the consumption of products containing this protein. There is still uncertainty over the dose that is required to

sensitize individuals to an allergen (FIFRA SAP, 2000), while minimum thresholds have been described for the provocation of allergic (anaphylactic) reactions after oral ingestion of allergenic proteins by individuals who have already been sensitized (Bindslev-Jensen *et al.*, 2002).

One of the rationales for screening for sequence homology is that short identical stretches from an allergenic protein may form 'linear' epitopes, which bind allergy-related IgE antibodies. Linear epitopes have been well documented in biomedical literature for a number of allergens. It should be noted here that non-linear epitopes would be missed in this procedure, given that their constituent amino acids are separated along the linear amino acid sequence, but close to each other in the three-dimensional structure of the protein. The identification of non-linear epitopes is still in progress. For both linear and non-linear epitopes, the accessibility of the constituent amino acid residues is important for IgE binding. In addition to epitopes containing amino acid residues, protein-bound carbohydrates can also form (part of) IgE-binding epitopes in allergenic proteins (Garcia Casado *et al.*, 1996). These carbohydrate epitopes are not retrieved by amino acid sequence homology searches.

In a previous study we applied the practice of sequence homology to identify identical stretches in transgenic proteins and allergenic proteins at least six amino acids long. This was done in order to assess the effect of lowering the currently recommended threshold of eight amino acids to six, as was recently discussed within the FAO/WHO. As a result, two-thirds (22) of the 33 tested transgenic proteins showed a similarity of six or seven contiguous amino acids to allergenic proteins. Because many of the positive results probably represent 'false positives', we recommended a new strategy which included additional steps to check if the positive results represent true linear IgE-binding epitopes. This involves a comparison of the identical stretch with linear epitopes reported in literature for the specific allergenic protein that contains this stretch. For those allergenic proteins that are less well documented, antigenicity prediction has been recommended to predict the most probable site of the protein where the antibodies would bind. Both for epitopes predicted by literature surveys and antigenicity prediction plots, further clinical testing has been recommended before reaching a final conclusion. Of the 22 proteins that were positive in the sequence comparison, two contained identical stretches that are part of linear epitopes reported in literature. Another protein contains an identical stretch that coincided with the highest point of the antigenicity prediction plot in the allergenic counterpart and that is therefore the most probable antibody-binding site.

In the current study, nine of the 16 novel proteins showed identical stretches of at least six amino acids length that also reside in allergenic proteins. For novel proteins encoded by dsRNA from rice and fava bean, the number of overall identical stretches appears to be comparatively high, i.e. 18 and 20 identities, respectively, against nil to three stretches in the other proteins (Table 3). This appears to correlate with the large size of the two dsRNA-encoded proteins, i.e. 4572 and 5825 amino acids, compared to that of the other proteins, i.e. 40–477 amino acids (Tables 1 and 2). The two large dsRNA-encoded sequences are in fact polyproteins harbouring various subsequences, which accounts for their large size.

Further refinement has been applied by a literature screen for true linear IgE-binding epitopes and antigenicity prediction according to the Hopp and Woods method. The results that are considered to be of interest involve four occurrences of identical stretches in three novel proteins, i.e. ORF138 of brassicas (oilseed rape, radish), and the dsRNA-encoded novel proteins of rice and fava bean (Table 4).

The KEEKKE stretch that ORF138 shares with the grass pollen allergen Phl p 13 is predicted to be antigenic in both the

allergenic and novel protein (Figure 1; Table 4). No literature data on linear IgE-binding epitopes of Phl p 13 have been found. The KEEKKE repeat, however, corresponds to KXEE/KEXE (X = empty, T or A), which has previously been identified as an IgE-binding motif sequence of the latex allergen Hev b 5 in immunized mice (Slater *et al.*, 1999). It is of particular interest that this stretch occurs three times in the ORF138 molecule, at positions 100–105, 113–118 and 126–131. If KEEKKE were a true epitope recognized by the antisera of allergic patients, the ORF138 protein would be able to crosslink IgE-antibodies on the surface of mast cells, which would subsequently trigger the release of mediators of the allergic response. For IgE-crosslinking, a minimal distance between the epitopes would be required. Compared to the size of peptides that have previously been reported to elicit mast cell (basophil) reactions, e.g. 20-mer peptides from ovalbumin (Honma *et al.*, 1996), the subsequence 100–131 should be large enough for IgE-crosslinking. The KEEKKE stretch is part of a polygalacturonic acid-binding N-terminal motif of the Phl p 13 protein, which is a polygalacturonidase (Petersen *et al.*, 2001). ORF138 is an oligomeric mitochondrial membrane protein with one putative hydrophobic membrane-spanning domain at positions 24–42 (Grelon *et al.*, 1994; Krishnasamy and Makaroff, 1994). The mitochondria of male-sterile radish and brassica 'cybrids' contains ORF138 protein, while the restoration of fertility of these male sterile plants is associated with decreased levels of ORF138 in reproductive tissues (Bellaoui *et al.*, 1999; Krishnasamy and Makaroff, 1994).

Moreover, the KYESNK and LSLVTK stretches that dsRNA from fava bean shared with mite allergens Eur m 14 and Der p 14 would in theory be able to crosslink IgE-antibodies on mast cells (Table 4). The antigenicity prediction for these sequences displays positive values (lower peaks) for KYESNK and slightly positive and negative values for LSLVTK in Eur m 14, Der p 14 and the dsRNA-encoded protein from fava bean. In addition, these sequences resided on remote positions in the linear amino acid sequence of the novel protein, i.e. 415–420 (KYESNK) and 2098–2103 (LSLVTK). The data therefore do not provide further evidence in favour or against the possibility of IgE-crosslinking.

Another identical stretch, EIEEFR, which is shared by the protein encoded by dsRNA from rice and the grass pollen allergens Phl p 11 and Lol p 11, forms a predicted antigenic site within the allergenic proteins (Table 4). The same stretch coincides with a lower peak of the antigenicity prediction plot of the novel protein. This difference in peak heights relates to contributions from flanking amino acid residues with high hydrophilicity values in the allergenic proteins,

i.e. (T/S)CSEIEEFRDRA (highest-scoring window underlined), compared to less hydrophilic flanks in the novel protein, i.e. LNLIEEFRQAN (window yielding local peak underlined). Phl p 11 and Lol p 11 show similarities to soybean trypsin inhibitor and the olive pollen allergen Ole e 1 (Marknell DeWitt *et al.*, 2002; Van Ree *et al.*, 1995). In addition, IgE-epitopes in Lol p 11 consist of both carbohydrate and peptide groups (Van Ree *et al.*, 1995).

The stretch ESLLKAN from both the novel protein encoded by dsRNA of fava bean and the allergenic tropomyosins Met e 1 from shrimp and Pan s 1 from lobster is part of an IgE-reactive 15-mer peptide reported in the literature (Table 4; Ayuso *et al.*, 2002). This peptide corresponds to a part of the sequence of the shrimp allergen Pen a 1 which has not been retrieved from the protein database. This indicates that not all known sequences of allergens have been filed in such databases. Another 15-mer peptide containing ESLLKAN and partially overlapping with the first peptide, however, did not react with IgE-antisera in the same report (Ayuso *et al.*, 2002). This indicates that the flanking residues of ESLLKAN in the first peptide that are not present in the latter peptide are at least required for IgE-binding.

In all of the four selected cases, further clinical testing should be carried out to reach a definitive conclusion. Out of these four cases, only ORF138 corresponds with multiple criteria for selection, i.e. predicted antigenicity and multiple copies in the novel protein. In addition, the identical stretch coincides with a three and four amino acid motif epitope of latex allergens reported in the literature.

Compared to the previously reported results for transgenic proteins, there appear to be no gross differences overall, i.e. the maximum size of identical stretches is seven amino acids in both studies and also for only a limited number of the identical stretches, indications of true epitopes exist. Therefore, neither of these identical stretches would have been identified if a threshold of eight amino acids were employed, which is currently common practice for transgenic proteins. Based upon these results, it appears that novel proteins in conventional food crops and transgenic proteins in genetically modified crops do not show any particular difference in the outcome of the screening for linear IgE-binding epitopes. From this perspective, it may be concluded that both groups should be subject to the same safety assessment procedure.

If the same procedure were to be applied to conventional crops as to genetically modified crops, this would in practice be hampered by several constraints. For example, in the case of genetic modification, the characteristics of the transgenic protein will be well-known on beforehand. For some conven-

tional crops, such as the cytoplasmic male sterile and dsRNA containing crops chosen for this study, novel open reading frames are well-characterized. However, in the majority of crops generated by conventional and mutation breeding, only the effects on genes of interest may be known, whereas other mutations may have remained unnoticed. Besides a precise characterization of known novel proteins expressed in conventional crops, a more holistic approach should therefore be applied to the identification of potential novel allergens. This may entail a combination of proteomic techniques for protein-profiling combined with the detection of cross-reactive proteins in a crop with antibodies from allergic patients. This approach has led to the recent identification of allergenic proteins in extracts of allergens, such as hazelnut (Beyer *et al.*, 2002a), sesame seed (Beyer *et al.*, 2002b) and shrimp (Yu *et al.*, 2003). In addition, proteins from the Nangai nut, a novel food, were found to cross-react with antisera from patients allergic to pollen. These patients have been selected based upon common occurrences of allergies to plant foods among them (Sten *et al.*, 2002). Interestingly, the feasibility of holistic techniques, including proteomics coupled to immunochemical detection, for the identification of potential novel allergens in genetically modified crops, is currently being studied within the EU-sponsored research project GMOCARE (GMOCARE, 2003).

Another category of novel proteins of interest, but not examined in this study, are those encoded by plant viruses. Coat and movement proteins, for example, may be expressed at substantial levels in virus-infected crops. It remains to be examined whether these viral proteins contain IgE-binding epitopes. In fact, the levels of viral gene products to which consumers may be exposed through the consumption of conventional crops have been considered as a reference for safe use in regulatory pre-market safety assessments of genetically modified crops carrying viral transgenes (e.g. FSANZ, 2000a,b).

In conclusion, no particular overall difference in the prediction of linear IgE-binding epitopes has been observed between novel proteins from conventional crops and transgenic proteins from genetically modified crops, as was previously reported. A limited number of identical stretches would be eligible for further clinical testing. It is recognized that the currently tested approach is targeted at genetic alterations that have been accurately characterized at the molecular level. Because alterations in conventional crops are generally less well characterized, modern holistic techniques for allergen identification can supplement the current approach with the aim of a comprehensive screening methodology for crops in general.

Experimental procedures

The methodology is the same as that previously used for screening transgenic proteins from genetically modified crops for potential linear, IgE-binding epitopes (Kleter and Peijnenburg, 2002).

Primary protein structures (amino acid sequences)

Protein accession files containing the amino acid sequences of specific novel proteins expressed in food crops have been obtained from the Entrez Protein database <<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein>>. Some sequences have been retrieved from the literature, e.g. *orf77* of maize (*Zea mays*, S-cytoplasm; Zabala *et al.*, 1997) and *orfA501* of chives (*Allium schoenoprasum*, translated from a DNA sequence described by Engelke and Tatlioglu, 2002). An overview of the sequences of novel proteins employed in this study is provided in Tables 1 and 2.

Comparison of amino acid sequences of novel proteins and allergenic proteins

The amino acid sequences of novel proteins have been screened for short, identical stretches that are also present in allergenic proteins. For this purpose, BLAST searches for 'short nearly exact matches' have been performed on-line on the Entrez website <<http://www.ncbi.nlm.nih.gov/BLAST/>>. The novel protein's sequence has been entered and, among others, the query limit 'allergen' has been selected to ensure that the searched Entrez Protein's database accessions contain most of the known allergen sequences. Some proteins that are retrieved by the query limit 'allergen' are not allergens themselves, but, for example, allergen-binding antibodies, sequences resembling allergen sequences and proteins purified together with allergens.

Antigenicity prediction

Plots have been created according to the method described by Hopp and Woods, which assigns fixed values to each amino acid residue, indicative of their hydrophilicity and tendency to be part of antibody-binding epitopes. An averaging window of six amino acids has been used, as was originally devised (Hopp and Woods, 1981). For this purpose, the Colorado State University's online facility to create such plots has been used <<http://arbl.cvmbs.colostate.edu/molkit/hydropathy/index.html>>. It has been checked whether the highest peak in each plot, i.e. the most probable antigenic site of the

plotted protein sequence, coincides with an identical stretch previously identified in novel and allergenic proteins.

Literature search

Data have been collected on linear, IgE-binding epitopes of the allergens that contained identical stretches with the aid of the PubMed online bibliography <<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>>. It was also verified whether the reported epitopes coincided with the identical stretch. Additional information on epitopes can be obtained from a number of online databases on allergens (Kleter and Peijnenburg, 2002).

Acknowledgements

The authors gratefully acknowledge financial support from the Ministry of Agriculture, Nature and Food Quality, scientific programmes 378 and 390. In addition, the authors wish to thank Dr M. Bremer for her comments on the manuscript.

Supplementary material

An Appendix has been provided, which is available at <<http://www.blackwellpublishing.com/products/journals/suppmat/PBI/PBI035/PBI035sm.htm>>, with a detailed account of the method and the results in table format. The identical stretches of six or more contiguous amino acids that are present in both novel and allergenic proteins are listed, as well as the outcomes of the Hopp and Woods antigenicity prediction method and literature data on the coincidence of the identical stretch with linear IgE-epitopes of allergenic proteins.

References

- Akagi, H., Sakamoto, M., Chou, S., Shimada, H. and Fujimura, T. (1994) A unique sequence located downstream from the rice mitochondrial *atp6* may cause male sterility. *Curr. Genet.* **25**, 52–58.
- Ayuso, R., Lehrer, S.B. and Reese, G. (2002) Identification of continuous, allergenic regions of the major shrimp allergen Pen a 1 (tropomyosin). *Int. Arch. Allergy Immunol.* **127**, 27–37.
- Bellaoui, M., Grelon, M., Pelletier, G. and Budar, F. (1999) The restorer *Rfo* gene acts post-translationally on the stability of the ORF138 Ogura CMA-associated protein in reproductive tissues of rapeseed brassicas. *Plant Mol. Biol.* **40**, 893–902.
- Beyer, K., Bardina, L., Grishina, G. and Sampson, H.A. (2002b) Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: seed storage proteins as common food allergens. *J. Allergy Clin. Immunol.* **110**, 154–159.
- Beyer, K., Grishina, G., Bardina, L., Grishin, A. and Sampson, H.A. (2002a) Identification of an 11S globulin as a major hazelnut food

- allergen in hazelnut-induced systemic reactions. *J. Allergy Clin. Immunol.* **110**, 517–523.
- Bindslev-Jensen, C., Briggs, D. and Osterballe, M. (2002) Can we determine a threshold level for allergenic foods by statistical analysis of published data in the literature? *Allergy*, **57**, 741–746.
- Bonhomme, S., Budar, F., Lancelin, D., Small, I., Defrance, M.C. and Pelletier, G. (1992) Sequence and transcript analysis of the Nco2.5 Ogura-specific fragment correlated with cytoplasmic male sterility in *Brassica* cybrids. *Mol. Gen. Genet.* **235**, 340–348.
- Bredehorst, R. and David, K. (2001) What establishes a protein as an allergen? *J. Chromatogr. B, Biomed. Sci. Appl.* **756**, 33–40.
- Dewey, R.E., Levings, C.S. and Timothy, D.H. (1986) Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male-sterile cytoplasm. *Cell*, **44**, 439–449.
- Engelke, T. and Tatlioglu, T. (2002) A PCR-marker for the CMS, inducing cytoplasm in chives derived from recombination events affecting the mitochondrial gene *atp9*. *Theor. Appl. Genet.* **104**, 698–702.
- FAO/WHO (2001) *Joint FAO/WHO expert consultation on foods derived from biotechnology – Allergenicity of genetically modified foods*. Rome, 22–25 January 2001. Rome: Food and Agriculture Organisation of the United Nations. <http://www.who.int/fsf/Documents/Biotech_Consult_Jan2001/report20.pdf>.
- FAO/WHO (2002) *Draft guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants (ALINORM 01/34A)*. Rome: Codex Alimentarius Committee, Food and Agriculture Organisation of the United Nations. <<ftp://ftp.fao.org/codex/alinorm01/al0134ae.pdf>>.
- FIFRA SAP (2000) *A set of scientific issues being considered by the Environmental Protection Agency regarding: assessment of scientific information concerning StarLink corn*. Arlington, VA: Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel, US Environmental Protection Agency. <<http://www.epa.gov/scipoly/sap/2000/november/one.pdf>>.
- FSANZ (2000a) *Draft risk analysis report, application A383, food derived from insect and potato leafroll virus-protected potato lines RBMT21-129, RBMT21-350, and RBMT22-82*. Canberra and Wellington: Food Standards Australia New Zealand. <http://www.foodstandards.gov.au/_srcfiles/A383%20FA.pdf>.
- FSANZ (2000b) *Draft risk analysis report, application A384, food derived from insect and potato virus Y-protected potato lines RBMT15-101, SEMT15-02 and SEMT15-15*. Canberra and Wellington: Food Standards Australia New Zealand. <http://www.foodstandards.gov.au/_srcfiles/A384%20FA.pdf>.
- Fukuhara, T. (1999) Double-stranded RNA in rice. *J. Plant Res.* **112**, 131–138.
- Garcia Casado, G., Sanchez Monge, R., Chrispeels, M.J., Armentia, A., Salcedo, G. and Gomez, L. (1996) Role of complex asparagine-linked glycans in the allergenicity of plant glycoproteins. *Glycobiology*, **6**, 471–477.
- Gibbs, M.J., Koga, R., Moriyama, H., Pfeiffer, P. and Fukuhara, T. (2000) Phylogenetic analysis of some large double-stranded RNA replicons from plants suggests they evolved from a defective single-stranded RNA virus. *J. Gen. Virol.* **81**, 227–233.
- GMOCARE (2003) *GMOCARE, new methodologies for assessing the potential of unintended effects in genetically modified food crops, QLK1-1999-00765*. Wageningen, Netherlands: RIKILT – Institute of Food Safety. <<http://www.entransfood.com/RTD-projects/GMOCARE/default.htm>>.
- Grelon, M., Budar, F., Bonhomme, S. and Pelletier, G. (1994) Ogura cytoplasmic male-sterility (CMS)-associated *orf138* is translated into a mitochondrial membrane polypeptide in male-sterile *Brassica* cybrids. *Mol. Gen. Genet.* **243**, 540–547.
- Handa, H., Gualberto, J.M. and Grienemberger, J.M. (1995) Characterization of the mitochondrial *orfB* gene and its derivative, *orf224*, a chimeric open reading frame specific to one mitochondrial genome of the 'Polima' male-sterile cytoplasm in rapeseed (*Brassica napus* L.). *Curr. Genet.* **28**, 546–552.
- Honma, K., Kohno, Y., Saito, K., Shimoto, N., Horiuchi, T., Hayashi, H., Suzuki, N., Hosoya, T., Tsunoo, H. and Niimi, H. (1996) Allergenic epitopes of ovalbumin (OVA) in patients with hen's egg allergy: inhibition of basophil histamine release by haptenic ovalbumin peptide. *Clin. Exp. Immunol.* **103**, 446–453.
- Hopp, T.P. and Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA*, **78**, 3824–3828.
- Horiuchi, H., Udagawa, T., Koga, R., Moriyama, H. and Fukuhara, T. (2001) RNA-dependent RNA polymerase activity associated with endogenous double-stranded RNA in rice. *Plant Cell Physiol.* **42**, 197–203.
- Johns, C., Lu, M., Lyznik, A. and MacKenzie, S. (1992) A mitochondrial DNA sequence is associated with abnormal pollen development in cytoplasmic male sterile bean plants. *Plant Cell*, **4**, 435–449.
- Kleter, G.A. and Peijnenburg, A.A.C.M. (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 Struct. Biol.* **2**, 8. <<http://www.biomedcentral.com/1472-6807/2/8>>.
- Koehler, R.H., Horn, R., Loessl, A. and Zetsche, K. (1991) Cytoplasmic male sterility in sunflower is correlated with the co-transcription of a new open reading frame with the *atpA* gene. *Mol. Gen. Genet.* **227**, 369–376.
- Krishnasamy, S. and Makaroff, C.A. (1993) Characterization of the radish mitochondrial *orfB* locus: possible relationship with male sterility in Ogura radish. *Curr. Genet.* **24**, 156–163.
- Krishnasamy, S. and Makaroff, C.A. (1994) Organ-specific reduction in the abundance of a mitochondrial protein accompanies fertility restoration in cytoplasmic male-sterile radish. *Plant Mol. Biol.* **26**, 935–946.
- L'Homme, Y., Stahl, R.J., Li, X.-Q., Hameed, A. and Brown, G.G. (1997) *Brassica nap* cytoplasmic male sterility is associated with expression of a mtDNA region containing a chimeric gene similar to the *pol* CMS-associated *orf224* gene. *Curr. Genet.* **31**, 325–335.
- Marknell DeWitt, Å., Niederberger, V., Lehtonen, P., Spitzauer, S., Sperr, W.R., Valent, P., Valenta, R. and Lidholm, J. (2002) Molecular and immunological characterization of a novel timothy grass (*Phleum pratense*) pollen allergen, Phl p 11. *Clin. Exp. Allergy*, **32**, 1329–1340.
- 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**(Suppl.), S165–S186.
- Moneger, F., Smart, C.J. and Leaver, C.J. (1994) Nuclear restoration of cytoplasmic male sterility in sunflower is associated with the tissue-specific regulation of a novel mitochondrial gene. *EMBO J.* **13**, 8–17.
- Moriyama, H., Nitta, T. and Fukuhara, T. (1995) Double-stranded RNA in rice: a novel RNA replicon in plants. *Mol. Gen. Genet.* **248**, 364–369.

- Nakajima, Y., Yamamoto, T., Muranaka, T. and Oeda, K. (2001) A novel *orfB*-related gene of carrot mitochondrial genomes that is associated with homeotic cytoplasmic male sterility (CMS). *Plant Mol. Biol.* **46**, 99–107.
- Osaki, H., Kudo, A. and Ohtsu, Y. (1998) Nucleotide sequence of seed- and pollen-transmitted double-stranded RNA, which encodes a putative RNA-dependent RNA polymerase, detected from Japanese pear. *Biosci. Biotechnol. Biochem.* **62**, 2101–2106.
- Petersen, A., Suck, R., Hagen, S., Cromwell, O., Fiebig, H. and Becker, W.M. (2001) Group 13 grass allergens: structural variability between different grass species and analysis of proteolytic stability. *J. Allergy Clin. Immunol.* **107**, 856–862.
- Pfeiffer, P. (1998) Nucleotide sequence, genetic organization and expression strategy of the double-stranded RNA associated with the '447' cytoplasmic male sterility trait in *Vicia faba*. *J. Gen. Virol.* **79**, 2349–2358.
- Rathburn, H.B. and Hedgcoth, C. (1991) A chimeric open reading frame in the 5' flanking region of *coxI* mitochondrial DNA from cytoplasmic male-sterile wheat. *Plant Mol. Biol.* **16**, 909–912.
- Schnable, P.S. and Wise, R.P. (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci.* **3**, 175–180.
- Singh, M. and Brown, G.G. (1991) Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. *Plant Cell*, **3**, 1349–1362.
- Slater, J.E., Paupore, E.J. and O'Hehir, R.E. (1999) Murine B-cell and T-cell epitopes of the allergen Hev b 5 from natural rubber latex. *Mol. Immunol.* **36**, 135–143.
- Song, J. and Hedgcoth, C. (1994) A chimeric gene (*orf256*) is expressed as protein only in cytoplasmic male-sterile lines of wheat. *Plant Mol. Biol.* **26**, 535–539.
- Stahl, R., Sun, S., L'Homme, Y., Ketela, T. and Brown, G.G. (1994) RNA editing of transcripts of a chimeric mitochondrial gene associated with cytoplasmic male-sterility in Brassica. *Nucl. Acids Res.* **22**, 2109–2113.
- Sten, E., Stahl Skov, P., Andersen, S.B., Torp, A.M., Olesen, A., Bindslev-Jensen, U., Poulsen, L.K. and Bindslev-Jensen, C. (2002) Allergenic components of a novel food, Micronesian nut Nangai (*Canarium indicum*), shows IgE cross-reactivity in pollen allergic patients. *Allergy*, **57**, 398–404.
- Tang, H.V., Pring, D.R., Shaw, L.C., Salazar, R.A., Muza, F.R., Yan, B. and Schertz, K.F. (1996) Transcript processing internal to a mitochondrial open reading frame is correlated with fertility restoration in male-sterile sorghum. *Plant J.* **10**, 123–133.
- Van Ree, R., Hoffman, D.R., Van Dijk, W., Brodard, V., Mahieu, K., Koeleman, C.A.M., Grande, M., Van Leeuwen, W.A. and Aalberse, R.C. (1995) *Lol p XI*, a new major grass pollen allergen, is a member of a family of soybean trypsin inhibitor-related proteins. *J. Allergy Clin. Immunol.* **95**, 970–978.
- Yamagashi, H. and Terachi, T. (2001) Intra- and inter-specific variations in the mitochondrial gene *orf138* of Ogura-type male-sterile cytoplasm from *Raphanus sativus* and *Raphanus raphanistrum*. *Theor. Appl. Genet.* **103**, 725–731.
- Yu, C.J., Lin, Y.F., Chiang, B.L. and Chow, L.P. (2003) Proteomics and immunological analysis of a novel shrimp allergen, Pen m 2. *J. Immunol.* **170**, 445–453.
- Zabala, G., Gabay Laughnan, S. and Laughnan, J.R. (1997) The nuclear gene *Rf3* affects the expression of the mitochondrial chimeric sequence R. implicated in S-type male sterility in maize. *Genetics*, **147**, 847–860.