Transcriptional frameshifts contribute to protein allergenicity

Benoit Thouvenot, … , Claude Favrot, Bernard E. Bihain


Graphical abstract

Find the latest version:

https://jci.me/126275/pdf
Transcriptional frameshifts contribute to protein allergenicity

Benoit Thouvenot,1 Olivier Roitel,1 Julie Tomasina,1 Benoît Hilselberger,1 Christelle Richard,1 Sandrine Jacqueten,1 Françoise Codreanu-Morel,2 Martine Morisset,3 Gisèle Kanny,4 Etienne Beaudouin,5 Christine Delebarre-Sauvage,4 Thierry Olivry,7 Claude Favrot,8 and Bernard E. Bihain1

1Genclis SA, Vandoeuvre-lès-Nancy, France. 2Unité Nationale d’Immunologie et d’Allergologie, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg. 3Unité d’Allergologie, Centre Hospitalier Universitaire Angers, Angers, France. 4Service de Médecine Interne, Immunologie Clinique et Allergologie, Hôpitaux de Brabois, Centre Hospitalier Universitaire de Nancy, Vandoeuvre-lès-Nancy, France. 5Unité d’Allergologie, Centre Hospitalier Régional de Metz, Metz, France. 6Allergology Center Saint-Vincent de Paul Hospital, Groupe Hospitalier de l’Institut Catholique de Lille, Lille, France. 7Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina, USA. 8Clinic for Small Animal Internal Medicine, University of Zurich, Zurich, Switzerland.

Transcription infidelity (TI) is a mechanism that increases RNA and protein diversity. We found that single-base omissions (i.e., gaps) occurred at significantly higher rates in the RNA of highly allergenic legumes. Transcripts from peanut, soybean, sesame, and mite allergens contained a higher density of gaps than those of nonallergens. Allergen transcripts translate into proteins with a cationic carboxy terminus depleted in hydrophobic residues. In mice, recombinant TI variants of the peanut allergen Ara h 2, but not the canonical allergen itself, induced, without adjuvant, the production of anaphylactogenic specific IgE (sIgE), binding to linear epitopes on both canonical and TI segments of the TI variants. The removal of cationic proteins from bovine lactoserum markedly reduced its capacity to induce sIgE. In peanut-allergic children, the sIgE reactivity was directed toward both canonical and TI segments of Ara h 2 variants. We discovered 2 peanut allergens, which we believe to be previously unreported, because of their RNA-DNA divergence gap patterns and TI peptide amino acid composition. Finally, we showed that the sIgE of children with IgE-negative milk allergy targeted cationic proteins in lactoserum. We propose that it is not the canonical allergens, but their TI variants, that initiate sIgE isotype switching, while both canonical and TI variants elicit clinical allergic reactions.

Introduction

Clinical allergy resulting from the inappropriate response of mammalian immune systems to otherwise harmless exogenous substances has become an important public health concern in Western countries (1, 2). The prevalence of food allergy in the US is estimated to be 10.8% (3); that of peanut allergy doubled in the last decade, so that it now affects 1.2% to 2% of children and 0.6% to 0.8% of adults (4). Our current understanding of the molecular cascade of events leading to the development of allergic diseases remains limited (2). IgE plays a central role in allergy (5), as this isotype is produced in response to a cytokine/allergic diseases remain unknown (2).

To date, the intrinsic structures or patterns common to all allergenic proteins that are responsible for inducing a specific IgE (sIgE) production remain unknown (2).

We and others have shown that RNA sequences occasionally differ from their canonical DNA templates (8–16). These RNA-DNA divergences (RDD) occur during, or immediately after, transcription and are collectively referred to as transcription infidelity (TI) (15) or RNA editing (ref. 16; for a review, see ref. 17). TI results from RNA Pol II errors, causing base substitutions within and across base families. Despite several regulatory mechanisms preventing the translation of erroneous messages (i.e., a nonsense-mediated decay; ref. 18), RNAs carrying TI events are often translated (8, 13).

Herein, we report what we believe is a novel RDD event: single-base omissions that create virtual RNA gaps causing translational frameshifts. As a result of such omissions, proteins translated from gapped RNAs contain 2 parts: (a) an aminoterminal segment whose sequence is the same as the canonical one and (b) a different carboxy terminal TI segment that, because of the degeneracy of the genetic code favoring basic aa, is cationic. The causal role of TI variants in eliciting IgE production was established in mice using both recombinant TI variants of the major peanut allergen Ara h 2 and the purified, low-abundance, cationic fraction of bovine lactoserum. We then translated these findings from mice to humans in 3 different contexts. We identified 2 peanut allergens, which we believe to be novel, based on their RDD
gap patterns and TI peptide aa composition. We then showed that peanut-allergic patients have sIgE directed toward both canonical and TI segments of Ara h 2. Finally, we found that the serum IgE from IgE-negative milk-allergic children reacted with the low-abundance, cationic (that is TI variant containing) rather than the anionic (i.e., TI variant depleted) fraction of bovine lactose- rum. Together, these findings indicate that single-gap TI contributes to protein allergenicity.

Results

RDD gap rates correlate with the differential allergenicity of 3 legumes. We began by analyzing mismatches between the genomes and transcriptomes of 3 legumes: the highly allergenic peanut (Arachis hypogea), the moderately allergenic soybean (Glycine max), and the rarely allergenic green bean (Phaseolus vulgaris); see Supplemental Methods for the bioinformatics method (ref. 19; supplemental material available online with this article; https://doi.org/10.1172/JCI126275DS1). We first calculated the RDD gap rates (i.e., the overall TI gap burden) following the method depicted in Figure 1A. With this calculation, we showed that A and T omissions were significantly more frequent in the peanut than in the soybean and rarest in the green bean, thus mirroring their allergenicity (Figure 1B).

In the peanut, gaps more commonly affected A and T repeats, and the RDD gap rate increased with the length of the repeats despite their decreasing frequency (Figure 1C). In peanuts, when comparing the composition of bases flanking A and T gaps with those around As and Ts unaffected by transcription errors, we observed that the 5 to 6 bases surrounding gapped As and Ts appeared to influence the presence of the TI events (Figure 1D). The most influential bases were those downstream of As and upstream of Ts (see Supplemental Figure 1 for the Multiple EM for Motif Elicitation [MEME] analysis of DNA motifs observed at each specific gap position; http://meme-suite.org/doc/overview.html).

Frameshifted RNAs translate into cationic peptides. The in silico translation of single-base omissions from peanut mRNA sequences showed a modest, but highly significant, increase in the percentage of basic aa and a profound decrease in that of acidic aa (Figure 2A); these variations are consistent with the genetic code containing relatively more codons for basic than acidic aa. When combined, these changes cause a 2- to 3-log increase in the isoelectric point of the TI peptides translated from the sequence downstream of a gap (Figure 2B). We then evaluated the impact of deleting each codon as a function of the first base of the next codon, assuming an equal probability for the 4 bases. Taking into account the peanut codon usage bias and the aa composition of the canonical proteins, we calculated the probabilities of gain, loss, or conservation of each aa (Table 1). Subsequently, we estimated the theoretical composition of peanut proteins translated from frameshifted RNA and found that more than 90% of both observed basic aa gains and acidic aa losses were explained by this simple calculation (Table 2). Therefore, because of the biased degeneracy of the genetic code that favors basic (i.e., positively charged aa), the consequence of frameshifts occurring at any given position and affecting any base is the potential translation of a cationic segment. The small difference in the predicted versus the observed upward shift in the isoelectric point is likely due to all bases not being equally affected by TI gaps, as shown above (Figure 1A).

The RDD gap patterns and predicted aa composition of TI peptides are different in allergen versus nonallergen coding transcripts. To better understand the RDD gaps at the individual transcript level, we compared the RDD profiles in 11 of the 17 peanut allergens recognized by the International Union of Immunological Societies (IUIS) whose genomic sequences are sufficiently annotated with those of 10,224 transcripts encoding peanut proteins not currently recognized as allergens (Table 3). We did not include 6 allergens in this analysis because of the high stringency constraints of our bioinformatics protocols: a perfect match to the respective DNA reference sequence was unavailable for Ara h 11, 12, 16, and 17; the position coverage was insufficient for Ara h 5; and Ara h 4 is, in fact, an isoform of Ara h 3 (20).

To summarize, we detected several essential differences between mRNAs coding for allergens and those translated into nonallergen proteins. The expression levels of peanut allergen-coding transcripts (ACTs) were significantly higher than those of nonallergen-coding transcripts (NACTs) (Table 3). ACTs encoded significantly smaller proteins and had a significantly higher gap rate compared with NACTs (Table 3). Finally, the aa composition of TI variants translated from gaps occurring in ACTs was significantly different from that of NACTs. Indeed, TI variants from peanut ACTs contained a significantly lower number of hydrophobic aa and had a higher proportion of hydrophilic, basic, and acidic aa compared with those encoded by peanut NACTs.

To verify that these observations were also valid for other allergens, we repeated the same calculations for 2 other familiar sources of food allergens (soybeans and sesame seeds) and one major environmental allergen, the Dermatophagoides pteronyssinus house dust mite (Tables 4, 5, and 6 and Supplemental Methods). In these 3 allergen sources, we found the same high-density patterns of RDD gaps spanning the entire ACT as in peanut allergens. Similarly, their ACTs also had a significantly higher expression level and translated into significantly smaller proteins compared with NACTs. Finally, we observed the same differences in TI peptide aa composition between ACTs and NACTs, with only some exceptions: in the sesame, for example, the percentage of hydrophilic aa was not significantly different between ACTs and NACTs. Using in silico translation of TI gaps from both ACTs and NACTs, we verified that the 2-log increase in the TI peptide isoelectric point was, indeed, the consequence of gaps in all 4 allergen sources (data not shown).

Together, our observations indicate that, in all these 4 distinct allergen sources, transcripts encoding allergens translate into a higher, albeit variable, number of TI variants containing a cationic TI peptide at its carboxy terminal end.

The TI variants of Ara h 2, but not the canonical allergen itself, induce the production of specific IgE in mice. To experimentally verify the hypothesis that TI variants translated from gapped RNA contribute to protein allergenicity, we first focused on the peanut because of the well-established central role played by a single allergen (Ara h 2; UniProt Q6PSU2; LOC112707245 on the A. hypogea genome and 100% identical to our genome assembly) in about 90% of peanut-allergic human patients (21). The transcripts encoding Ara h 2 contain 13 RDD gaps (Figure 3A). We produced recombinant Ara h 2 (rAra h 2) in its canonical sequence and 3 of its TI variants (v36, v38, and v40) translated from frameshifted cDNA.
Figure 1. RDD gap rates in legumes. (A) Method of calculation of the RDD gap rates. (B) RDD gap rates related to the numbers (indicated as n) of A and T repeats for peanut, soybean, and green bean. (C) Peanut RDD gap rates related to the numbers of A and T repeats. (D) Base composition surrounding (-10 and +10) RDD gap positions (b0). t test, Welch’s t test and Mann-Whitney U tests with Holm-Bonferroni correction: *P < 0.05; ***P < 0.001. (B and C) Box-and-whisker plots: box, interquartile range; bar, median; whiskers, 95% CI.
All 3 TI variants of Ara h 2 induced IgE reactivity (Figure 3D). Seventy-six days after the injections of the canonical rAra h 2 caused no detectable increase in circulating sIgE, a challenge with this protein led to a clinical response that was significantly more noticeable than that of nonsensitized controls (Figure 3F). This observation was not surprising, as canonical Ara h 2 injections induced the production of specific IgG and adoptive transfer experiments have shown previously that clinical signs of allergy can also be IgG mediated in mice (23–25).

We then used direct and inhibition Western immunoblotting with canonical Ara h 2 as an inhibitor to verify the slgE elicited by TI variants recognized the TI or the canonical segments of the Ara h 2 variants. We found that the canonical rAra h 2 nearly extinguished slgE reactivity toward the canonical rAra h 2, while it only reduced — but did not suppress — that present against TI variant linear epitopes (Figure 3G). With this set of experiments, we established that it is the TI variants of an allergen, and not its canonical version, that induce the slgE production and clinical signs of allergy in sensitized mice; furthermore, we confirmed that the slgE from Ara h 2–sensitized mice reacted with linear epitopes located on both canonical and TI segments of the TI variants.

Adding a cationic TI peptide to the carboxy terminal end of a nonallergenic protein is sufficient for eliciting the production of slgE in mice. We next tested to determine whether the induction of slgE could be experimentally induced against a nonallergenic protein by adding a TI peptide to its carboxy terminal end. We selected a peanut protein (ARR17L encoded by a gene in our A. hypogea genome assembly 100% identical to LOC107470167 in the Arachis duranensis genome encoding a 2-component response regulator ARR17 like; see Supplemental Methods); it is of a size similar to that of Ara h 2 (20 kDa), has no aa sequence identity with any known allergen, and is encoded by a transcript without any identifiable RDD gaps. We thus produced 2 recombinant ARR17L proteins: one with the added carboxy terminal cationic TI peptide (ARR17L-TIP) and one with the noncationic peptide resulting from the corresponding canonical translation (ARR17L-CAN) (Figure 4). Three injections, at a 1-week interval, of 400 μg of either ARR17L chimera elicited the production of slgM and slgG against the respective chimera, but slgE rose significantly only against the ARR17L-TIP (Figure 4). Therefore, the mere addition of a cationic TI peptide to the carboxy terminus of a nonallergenic protein appears sufficient to induce the production of slgE after injection into mice without any adjuvant.

The concomitant, but not the sequential, injections of TI variants of Ara h 2 induce the production of slgE. A high diversity of TI variants appears to be a general characteristic of ACTs compared with NACTs (Tables 3–6). This finding implies that, in any given allergen preparation, several TI variants of the same canonical allergen will coexist. We thus questioned whether the different TI variants might “cooperate” to induce the production of slgE.

In Figure 5, we show that injections of a cocktail of 3 TI variants and the canonical protein, each at a dose 4-fold lower (100 μg) than that used in the experiment reported in Figure 3 and Figure 4 (400 μg per injection), significantly increased the production of Ara h 2–specific IgM, IgG, and IgE. In contrast, the sequential injections, at the same 1-week interval, of each TI variant to achieve the same cumulative dose only caused a substantial (P

---

**Table 1. Changes in aa composition in proteins translated from frameshift are explained by probabilities of gain, loss, and conservation of residues**

<table>
<thead>
<tr>
<th>aa</th>
<th>Probability of gain</th>
<th>Probability of loss</th>
<th>Probability of conservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acidic</td>
<td>0.036</td>
<td>0.120</td>
<td>0</td>
</tr>
<tr>
<td>Basic</td>
<td>0.116</td>
<td>0.111</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Acidic aa are D and E and Basic aa are R, K, and H.
< 0.05) increase in Ara h 2–specific IgM and IgG without detectable sIgE. These data suggest that the TI variants present in ACTs have a cooperative effect on the production of sIgE. Therefore, the injection of either a single TI variant at a high dose (400 μg) or the combination of 3 TI variants and the canonical Ara h 2 at low doses (100 μg each) is sufficient to induce the production of IgE without the need for any adjuvant.

The removal of low-abundance cationic proteins from bovine whey markedly reduces its ability to cause IgE production in mice. Bioinformatic data indicated that TI events in the peanut only affected a minute percentage (less than 0.5%) of transcribed mRNAs; assuming that all TI-containing transcripts would be translated, we estimated that TI variants represent less than 4% of the total peanut protein content. As these TI variants are predicted to be cationic, their separation and removal based on a higher isoelectric point are therefore possible.

To test such a method of TI variant depletion, we selected whey (i.e., lactoserum), a milk derivative devoid of caseins. Bovine whey was separated using ion-exchange chromatography into 2 fractions with an isoelectric cutoff set at 7.4. We first calculated that the fraction with an isoelectric point higher than this cutoff represented less than 5% of whey proteins, a finding consistent with our previous calculation of the proportion of TI variants among peanut proteins (see above). This cationic fraction with a high isoelectric point had a profile in SDS-PAGE that was different from those of the total whey and the low isoelectric point noncationic fraction (Figure 6A).

In a first experiment, a low dose (60 μg) of either total, cationic, or noncationic whey fractions was administered to mice intragastrically to closely mimic the physiological conditions of milk exposure (Figure 6B). Remarkably, at that dose, only the cationic whey fraction — but not the total or noncationic fractions — induced a statistically significant increase in sIgE relative to that of control mice on day 55 (Figure 6C).

As there were notable interindividual variations in the IgE response to this cationic whey fraction, we conducted a second experiment in which a higher number of mice were given either a high dose (3.5 mg) of total whey or a 25-fold lower dose (150 μg) of either cationic or noncationic whey fractions. We showed that the low dose of cationic whey induced the production of sIgE at the same level as that induced by the 25-fold higher amount of total whey (Figure 6D). In contrast, the ingestion of the low dose of noncationic whey led to significantly lower sIgE production, with a titer plateauing between 35 and 42 days, while that observed for total or cationic whey continued to increase over time significantly (Figure 6D). A limitation of our fractionation method, however, is that we cannot exclude that the noncationic whey fraction might still contain some TI variants, as the addition of a cationic TI peptide might not always be sufficient to shift an entire variant’s isoelectric point above that of our 7.4 cutoff point. Together, our results suggest that proteins that trigger sIgE production in response to the ingestion of whey are present in this food’s cationic fraction, which should include the TI variants of whey allergens. Nevertheless, we cannot eliminate the possibility that a difference in the identity of the proteins present in the cationic and anionic fractions also might have contributed to the notable difference seen in immunogenicity.

The analysis of RDD gap patterns in the peanut transcriptome leads to the discovery of 2 prevalent allergens for peanut-allergic humans. To identify allergens that we believe to be previously unreported, we screened the available peanut transcriptome for mRNAs encoding proteins not known currently as allergens, but that had an expression level, an RDD gap profile, and a TI segment aa composition suggesting that the translated canonical protein could be an allergen in peanut-allergic human patients. We discovered a transcript with 12 RDD gaps and a gap density compa-
rable to that of Ara h 2 (13 gaps spanning 312 nucleotides coding for a 19 kDa protein). The canonical sequence of this believed to be novel peanut allergen (NPA) encodes a predicted 23.4 kDa protein with a suspected S26 type I signal peptidase domain (NCBI CD06530; Figure 7A). In the peanut wild ancestor \textit{A. duranensis}, the LOC107473664 gene encodes this protein, which we designated as NPA23. This NPA23 has no sequence homology with any of the known peanut allergens, but it contains very short aa sequences shared with 4 allergens in the Allergome database (http://www.allergome.org/) (Figure 7A). Two of these allergens are uncommon, but there could be a potential crossreactivity of NPA23 with the cat allergen Fel d 2.0101 and the dog allergen Can f 3.0101. The canonical NPA23 protein was produced in \textit{E. coli} and purified as a single product forming several oligomers, as shown by SDS-PAGE in Figure 7B. We found the sera of all 10 children with clinical peanut allergy (among our patients; Table 7) to have detectable sIgE against this NPA23 on immunoblots, while only 1 in 5 atopic patients without a peanut allergy and none of 5 healthy controls had such detectable reactivity (Figure 7B). Of potential clinical relevance was the observation that the atopic child with detectable NPA23-sIgE was, in fact, clinically allergic to cats.

Using the same approach, we found another peanut transcript with 47 RDD gaps, which encoded a putative 27.2 kDa canonical protein (NPA27) with no sequence identity with any of the known allergens. This NPA27 contained a thaumatin-like domain (NCBI CD09218, TLP-PA) that explains its low-sequence homology with 4 other allergens (Supplemental Figure 3). In the peanut wild ancestor \textit{Arachis ipaensis}, the LOC107629116 gene encodes the NPA27. We produced recombinant NPA27 in \textit{E. coli}, and this allergen was recognized by serum sIgE from peanut-allergic patients, albeit at an intensity and frequency lower than that of NPA23 (Supplemental Figure 3).

To more precisely estimate the prevalence of IgE seropositivity to the NPA23 and NPA27, we tested a larger group (including previous patients tested above by immunoblotting) of peanut-allergic and tolerant children by ELISA (Table 7). We detected sIgE against NPA23 and NPA27, respectively, in the serum of 24 of 52 (46%) and 25 of 52 (48%) peanut-allergic children and 10 of 38 and 8 of 38 tolerant children.

**Table 4. Comparison of allergen and nonallergen transcripts for soybean**

<table>
<thead>
<tr>
<th></th>
<th>Allergen genes (n = 18)</th>
<th>Nonallergen genes (n = 24,238)</th>
<th>P value (allergens vs. nonallergens)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Protein size</td>
<td>334</td>
<td>178</td>
<td>500</td>
</tr>
<tr>
<td>Transcript expression levels</td>
<td>2.31</td>
<td>2.31</td>
<td>0.06</td>
</tr>
<tr>
<td>Gap density</td>
<td>0.06</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Earliest occurrence of T1 event (0 NH2 -&gt; 1 COOH)</td>
<td>0.07</td>
<td>0.08</td>
<td>0.26</td>
</tr>
<tr>
<td>Latest occurrence of T1 event (0 NH2 -&gt; 1 COOH)</td>
<td>0.87</td>
<td>0.28</td>
<td>0.74</td>
</tr>
<tr>
<td>Acidic aa (%)</td>
<td>5.72</td>
<td>6.23</td>
<td>4.58</td>
</tr>
<tr>
<td>Basic aa (%)</td>
<td>16.20</td>
<td>10.69</td>
<td>13.98</td>
</tr>
<tr>
<td>Hydrophobic aa (%)</td>
<td>43.07</td>
<td>15.5</td>
<td>43.04</td>
</tr>
<tr>
<td>Hydrophilic aa (%)</td>
<td>35.01</td>
<td>14.39</td>
<td>32.4</td>
</tr>
</tbody>
</table>

Number of aa indicates protein size; transcript expression is expressed in percentages of RPKM, with the highest expression serving as a reference; gap diversity corresponds to the number of gaps, taking into account the protein size. The range of occurrence of T1 events has a value between 0 and 1, with 0 indicating that it occurs at the N-terminus and 1 at the C-terminus. The statistical test used for comparisons was Welch’s t test with Holm–Bonferroni correction.

**Table 5. Comparison of allergen and nonallergen transcripts for sesame**

<table>
<thead>
<tr>
<th></th>
<th>Allergen genes (n = 9)</th>
<th>Nonallergen genes (n = 17,285)</th>
<th>P value (allergens vs. non allergens)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Protein size</td>
<td>316</td>
<td>179</td>
<td>494</td>
</tr>
<tr>
<td>Transcript expression levels</td>
<td>70.86</td>
<td>29.57</td>
<td>0.68</td>
</tr>
<tr>
<td>Gap density</td>
<td>0.04</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Earliest occurrence of T1 event (0 NH2 -&gt; 1 COOH)</td>
<td>0.18</td>
<td>0.16</td>
<td>0.19</td>
</tr>
<tr>
<td>Latest occurrence of T1 event (0 NH2 -&gt; 1 COOH)</td>
<td>0.95</td>
<td>0.02</td>
<td>0.80</td>
</tr>
<tr>
<td>Acidic aa (%)</td>
<td>5.72</td>
<td>5.17</td>
<td>4.64</td>
</tr>
<tr>
<td>Basic aa (%)</td>
<td>17.14</td>
<td>10.97</td>
<td>14.84</td>
</tr>
<tr>
<td>Hydrophobic aa (%)</td>
<td>44.98</td>
<td>12.47</td>
<td>48.51</td>
</tr>
<tr>
<td>Hydrophilic aa (%)</td>
<td>31.80</td>
<td>12.04</td>
<td>32.00</td>
</tr>
</tbody>
</table>

Number of aa indicates protein size; transcript expression is expressed in percentages of RPKM, with the highest expression serving as a reference; gap diversity corresponds to the number of gaps, taking into account the protein size. The range of occurrence of T1 events has a value between 0 and 1, with 0 indicating that it occurs at the N-terminus and 1 at the C-terminus. The statistical test used for comparisons was Welch’s t test with Holm–Bonferroni correction.
The Journal of Clinical Investigation

RESEARCH ARTICLE

Table 6. Comparison of allergen and nonallergen transcripts for D. pteronyssinus house dust mite

<table>
<thead>
<tr>
<th></th>
<th>Allergen genes (n = 17)</th>
<th>Nonallergen genes (n = 5,587)</th>
<th>P value (allergens vs. nonallergens)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Protein size</td>
<td>420</td>
<td>406</td>
<td>650</td>
</tr>
<tr>
<td>Transcript expression levels</td>
<td>16.49</td>
<td>13.22</td>
<td>1.97</td>
</tr>
<tr>
<td>Gap density</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Earliest occurrence of TI event (0 NH2 → 1 COOH)</td>
<td>0.16</td>
<td>0.22</td>
<td>0.34</td>
</tr>
<tr>
<td>Latest occurrence of TI event (0 NH2 → 1 COOH)</td>
<td>0.87</td>
<td>0.10</td>
<td>0.71</td>
</tr>
<tr>
<td>Acidic aa (%)</td>
<td>3.54</td>
<td>5.31</td>
<td>4.51</td>
</tr>
<tr>
<td>Basic aa (%)</td>
<td>13.95</td>
<td>9.20</td>
<td>12.43</td>
</tr>
<tr>
<td>Hydrophilic aa (%)</td>
<td>51.78</td>
<td>13.44</td>
<td>54.3</td>
</tr>
<tr>
<td>Hydrophilic aa (%)</td>
<td>30.73</td>
<td>12.09</td>
<td>28.77</td>
</tr>
</tbody>
</table>

The number of aa indicates protein size; transcript expression is expressed in percentages of RPKM, with the highest expression serving as a reference; gap diversity corresponds to the number of gaps, taking into account the protein size. The range of occurrence of TI events has a value between 0 and 1, with 0 indicating that it occurs at the N-terminus and 1 at the C-terminus. The statistical test used for comparisons was Welch’s t test with Holm-Bonferroni correction.

(26% and 21%) peanut-tolerant atopic children, but in none of the 6 healthy children (Figure 7C); these proportions were significantly higher in peanut-allergic than in peanut-tolerant atopic and control children (P < 0.01 and P < 0.03 for NPA23 and NPA 27, respectively). If we were to use a cutoff of a mean ± 2 SD of the IgE concentration obtained in peanut-tolerant atopic children, the proportion of children with positive sIgE reactivity against NPA23 would be 18 of 52 (35%) peanut-allergic but only 2 of 38 (5%) peanut-tolerant atopic children. Similarly, that toward NPA27 would be above this threshold in 12 of 52 (23%) peanut-allergic but only 3 of 38 (8%) peanut-tolerant atopic children. Therefore, both immunoblotting and ELISA results concur in demonstrating that both NPA23 and NPA27 are 2 prevalent, yet minor, peanut allergens.

The sIgE of peanut-allergic children recognizes both canonical and TI segments of Ara h 2. We hypothesized that if TI proteins were responsible for the development of peanut allergy in human patients, some of their sIgE should be directed against the cationic carboxy terminal TI segment of the Ara h 2 TI protein variants.

We first tested sIgE reactivity against both the canonical Ara h 2 and 3 of its TI variants in a cohort of 51 children with clinically proven peanut allergy as well as 37 peanut-tolerant atopic children and 6 healthy controls (among our patients; Table 7). In peanut-tolerant atopic or healthy controls, we found sIgE above the cutoff (mean ± 2 SD of peanut-tolerant children) against rAra h 2 and its v36, v38, and v40 variants in 1 of 37 (3%), 2 of 43 (5%), 1 of 39 (3%), and 2 of 38 (3%) children, respectively (data not shown). In contrast, we found sIgE against rAra h 2 and its v36, v38, and v40 variants in 30 of 37 (81%), 37 of 51 (73%), 33 of 46 (72%), and 34 of 42 (81%) of peanut-allergic children, respectively (Figure 8A).

The difference among the sIgE levels against each of the 4 proteins was highly significant (P < 10⁻¹⁰) between peanut-tolerant and peanut-allergic children. In the 7 peanut-allergic children without detectable rAra h 2 sIgE by ELISA, we found sIgE to the v40 variant in only 1 patient. Finally, the sIgE binding to the v40 variant was significantly higher than that measured against the canonical Ara h 2, despite this particular TI variant missing 1 of the 8 Ara h 2 immunodominant epitopes (26).

We then conducted ELISA inhibition experiments using the sera of 10 peanut-allergic children. These inhibition assays showed that an excess of canonical Ara h 2 inhibited significantly more of the reactivity to the identical canonical protein than that against its 3 TI variants (Figure 8B). These observations were confirmed using a 2-fold higher or lower concentration of inhibiting canonical Ara h 2 (data not shown). Of note is that the proportions of IgE specific for the TI segment of each TI variant varied between patients (Table 8). These inhibition experiments indicated that, on average, 20% of IgE bound the TI segment of Ara h 2.

Together, these observations suggested that, as in mice, some of the sIgE of peanut-allergic humans recognize both the canonical and TI Ara h 2 variants. Furthermore, the sIgE reactivity against the TI segment of an Ara h 2 variant was always associated with one against the canonical segment.

Children with IgE-negative milk allergy have detectable IgE against the cationic TI protein-enriched fraction of lactoserum. Children with milk allergy are diagnosed after the development of clinical signs upon the ingestion of milk and their resolution with the avoidance of dairy products (Table 9). While most of the children with clinical milk allergy have detectable sIgE (IgE-positive milk allergy), some do not (the so-called IgE-negative milk allergy or non-IgE-mediated milk allergy). We hypothesized that the lack of detection of sIgE in children with IgE-negative milk allergy could be due to an insufficient amount of TI proteins in the standard milk extract used for sIgE detection by ELISA. As shown in the preceding sections, TI variants of allergens are cationic, so we processed bovine whey (lactoserum) based on an isoelectric point of 7.4 to separate its anionic from its cationic fraction. We tested by ELISA the sera of 7 children with IgE-negative milk allergy, 7 with IgE-positive milk allergy, and 10 controls for IgE reactivity against lactoserum and its anionic fraction. As shown in Figure 9, the ratio of cationic-to-anionic IgE reactivity was significantly higher in IgE-negative than in IgE-positive milk-allergic children (P < 0.005) and other allergic children (P < 0.05). This observation suggests that in the IgE-negative milk allergic children, the previously undetectable IgE are directed more often against the cationic fraction of lactoserum enriched in low-abundance TI allergens than to its anionic fraction that contains proteins translated from canonical sequences.
In this paper, we provide 8 independent lines of evidence supporting the idea that RDD gap rates are both associated with allergenic proteins translated from their canonical RNA, but that it is induced by the presence of low-abundance TI variants translated from gapped transcripts. Herein, we showed the following: (a) the RDD gap rates are significantly higher in an allergenic legume (the peanut) compared with 2 other legumes of lower allergenicity; (b) RDD gap rates are significantly higher in an allergenic legume (the peanut) compared with 2 other legumes of lower allergenicity; (c) 3 recombinant nonglycosylated TI variants of the major peanut allergen Ara h 2, but not the protein resulting from canonical transcripts, induce IgE isotype switching in mice at low doses when added concomitantly, but also when added singly at a higher dose; in both cases, adjuvants were not needed; (d) adding a TI peptide to a nonallergenic peanut protein is sufficient to cause IgE production — again without the need for an adjuvant; (e) the removal of the cationic peptide–enriched fraction from whey, which is expected to contain its putative TI variants, markedly reduces its capacity to induce sIgE production in mice; the whey fraction enriched in cationic peptides is sufficient to induce sIgE at a dose 25-fold lower than that of total whey; (f) the analysis of the RDD gap pattern of peanut transcripts led to the identification of 2 previously unknown prevalent peanut allergens; (g) the sIgE of children with a clinically proven peanut allergy are reactive, not only with the canonical Ara h 2, but also with the TI segments present on at least 3 Ara h 2 TI variants; and (h) finally, we established that all 7 tested children with IgE-negative milk allergy had, in fact, detectable sIgE against the cationic fraction of whey, which is enriched in TI proteins that are typically in low abundance in standard milk extracts used for ELISA.

Discussion
In this paper, we provide 8 independent lines of evidence supporting the concept that IgE production is not due to abundant allergenic proteins translated from their canonical RNA, but that it is induced by the presence of low-abundance TI variants translated from gapped transcripts. Herein, we showed the following: (a) the RDD gap rates are significantly higher in an allergenic legume (the peanut) compared with 2 other legumes of lower allergenicity; (b) at the single-transcript level, the RDD gap density is significantly higher in allergen-coding than in nonallergen-coding mRNAs, and this observation was verified in the peanut, the soybean, and the sesame seed, as well as in a classic environmental allergen, the D. pteronyssinus house dust mite; (c) 3 recombinant nonglycosylated TI variants of the major peanut allergen Ara h 2, but not the protein resulting from canonical transcripts, induce IgE isotype switching in mice at low doses when added concomitantly, but also when added singly at a higher dose; in both cases, adjuvants were not needed; (d) adding a TI peptide to a nonallergenic peanut protein is sufficient to cause IgE production — again without the need for an adjuvant; (e) the removal of the cationic peptide–enriched fraction from whey, which is expected to contain its putative TI variants, markedly reduces its capacity to induce sIgE production in mice; the whey fraction enriched in cationic peptides is sufficient to induce sIgE at a dose 25-fold lower than that of total whey; (f) the analysis of the RDD gap pattern of peanut transcripts led to the identification of 2 previously unknown prevalent peanut allergens; (g) the sIgE of children with a clinically proven peanut allergy are reactive, not only with the canonical Ara h 2, but also with the TI segments present on at least 3 Ara h 2 TI variants; and (h) finally, we established that all 7 tested children with IgE-negative milk allergy had, in fact, detectable sIgE against the cationic fraction of whey, which is enriched in TI proteins that are typically in low abundance in standard milk extracts used for ELISA.
from the canonical translation of one of the isoforms of the major allergen ovalbumin (36). Finally, whether or not a specific TI pattern similar to that found in allergens is also present on helminth transcripts whose translated proteins also induce sIgE production needs to be investigated (37).

In this paper, we introduced several applications that could be clinically relevant for human allergic patients. First, with the availability of next-generation sequencing and recently developed procedures to better distinguish bona fide RDDs from mere sequencing errors (14), a simple method is now available to estimate the potential allergenicity of different sources and the allergenic potential of individual proteins. This approach might prove useful for evaluating the allergenic risk of new foods or ingredients. Second, the combination of a high-density gap together with a low hydrophobicity of TI peptides provided a simple method for rapidly identifying previously undescribed candidate allergens and explaining unusual situations of allergy. For example, our NPA23, which we believe to be previously unreported, has an aa sequence homology with 4 nonpeanut allergens, including a Fel d 2 isoform.

With such information, we found it no longer surprising that one of our peanut-tolerant atopic children (no. 19) with a detectable sIgE against NPA23 was, in fact, clinically allergic to cats. This observation raised the hypothesis of a possible clinically relevant crossreactivity between the NPA23 and Fel d 2 in some patients, a topic deserving of further investigation. Third, we presented data that suggested that one could potentially reduce the allergenicity of known allergens. Indeed, using the example of whey, we showed that the removal of cationic proteins based on their elevated isoelectric point reduces the stimulus for sIgE production against the entire food allergen. Such a strategy, if expandable industrially, could markedly reduce food allergenicity while preserving over 95% of the entire dietary (and standard) protein content. This strategy would be distinct from the current hydrolysis processes that prevent clinical manifestations of the disease by reducing the size of allergens to a point where the aggregation of IgE anchored to the FcεRI would no longer be possible (38). Before the industrial expansion of a cationic protein–depleted hypoallergenic milk, a deeper understanding of milk TI variants, including those of the

Figure 4. Effect of the addition of a TI peptide to a nonallergenic protein. Levels of specific Igs directed against ARR17L-TIP (TI peptide) and ARR17L-CAN (corresponding canonical sequence) in mice sensitized by intraperitoneal injection on days 0, 7, and 14. Single experiment with 5 mice per sensitization group. Paired t test and Wilcoxon’s signed rank test with Holm-Bonferroni correction: *P < 0.05. Box-and-whisker plot: box, interquartile range; bar, median; whiskers, 95% CI.

Figure 5. Comparison of the effect of combined or sequential injections of TI peptides. Ara h 2 antibody levels in control mice and those injected with a combined dose of 400 μg (100 μg of each of 4 TI proteins) or the sequential injection of 400 μg of each protein at a 1-week interval. Single experiment with 10 mice per sensitization group; t test, Welch’s t test, and Mann-Whitney U test with Holm-Bonferroni correction. *P < 0.05; **P < 0.01. Box-and-whisker plot: box, interquartile range; bar, median; whiskers, 95% CI.
casein allergens, is needed. Finally, our positive detection of IgE directed against the cationic fraction of whey — a casein-free milk product — permits a logical explanation of why some children with IgE-negative milk allergy do not have detectable sIgE when tested with standard milk extracts containing low amounts of low-abundance TI proteins. If one were to expand these observations to other clinical allergies in which patients do not have detectable IgE against the relevant allergens, one could envision the development of TI-enriched allergen extracts that would have a higher sensitivity for detecting early or uncommon IgE-negative allergies in human or animal patients.

In conclusion, the results of the above experiments introduce what we believe is a novel mechanism that contributes to the allergenicity of proteins. Further research is underway to determine the precise aa composition(s) and sequence(s) and respective molecular recognition patterns that control IgE iso- type switching. One also needs to explain why individuals react differently upon exposure to allergenic TI variants. In fine, we speculate that not only genetic (39, 40), but also microbiota (41) and environmental (42), factors contribute with TI in allergen to the onset and progression of allergic diseases.

Methods

Human patient characteristics

The children whose sera were included in the current study had been referred to the outpatient clinic of the Internal Medicine, Clinical Immunology and Allergology Unit of the Nancy University Hospital; the Allergology Center from the Saint Vincent de Paul Hospital; and the Pneumology and Allergology Department of the Angers University Hospital. The clinical and biological characteristics of patients and controls are summarized in Table 7 for peanut (21) and Table 9 for milk. All peanut-allergic children were diagnosed based on both a positive peanut double-blinded placebo-controlled food challenge (DBPCFC), performed as previously described (21), and a detectable sIgE reactivity toward major peanut allergens (Thermo Fisher Scientific). The control groups included 38 atopic and 6 healthy children. The 38 atopic patients had either a negative peanut DBPCFC with 10 g of peanut or documented regular absorption of peanuts without the trigger of clinical symptoms; the peanut-sIgE reactivity of these atopic children is also shown in Table 7.

Cow’s milk–allergic children were diagnosed because of the development of clinical symptoms after the ingestion of milk and their improvement or resolution after cow’s milk avoidance; they were subdivided based on detectable (i.e., IgE-mediated milk allergy) or undetectable (i.e., non–IgE-mediated milk allergy) serum sIgE reactivity toward cow’s milk extract (Thermo Fisher Scientific). Among the 7 controls with other allergies, 2 were allergic to peanuts, 2 to house dust mites, and 3 to birch or grass pollen.

Bioinformatic analyses. To identify TI events, we performed bioinformatic annotations of transcriptomes of several allergens using Illumina sequencing data, as previously described (15), but with modifications appropriate to curate short reads (see Supplemental Methods). The sequencing coverage of RNA-derived sequences was used as a proxy for transcript abundance. To detect RDDs, we examined each mRNA position for its concordance with the DNA reference sequence. Not only were the 12 possible base substitutions found in all 4 allergen sources, but insertions and gaps in the mRNA that were not present in the respective reference DNA were also detected. We herein only report the RDD gaps that occurred within the open reading frame (i.e., those that might be translated into a protein). To estimate the overall TI burden in any given source, we used the RDD rate, which we calculated as the sum of events at each position divided by the sum of reads at any given position.
phate, 150 mM NaCl, pH 7.4) containing 500 mM imidazole and 20% glycerol. Imidazole was subsequently removed using a Zeba Desalt Spin Column (Thermo Fisher Scientific) to enable the storage of rAra h 2 and its variants in a native PBS containing 10% glycerol. The recombinant proteins ARR17L-CAN, ARR17L-TIP, NPA23, and NPA27 (see below for their characterization) were eluted from the columns in a denaturing PBS containing 500 mM imidazole, 6 M urea, and 2 mM β-mercapto-ethanol. Imidazole was removed using a Zeba Desalt Spin Column to enable the storage of ARR17L-CAN and ARR17L-TIP in a denaturing PBS containing 2 M urea and that of NPA23 and NPA27 in a denaturing PBS containing 6 M urea and 2 mM β-mercapto-ethanol. Protein concentrations were determined.

**Protein preparations**

**Production of recombinant proteins.** Recombinant proteins, except for Ara h 2, were produced as N-terminal-hexa-histidine–tagged proteins in BL21 *E. coli* according to standard procedures, with the modification that clones were engineered toward *E. coli* codon usage to increase their expression levels. The canonical Ara h 2 was produced with the same method, except that the histidine-tag was positioned at the C-terminus. The proteins were then purified from inclusion bodies by immobilized metal ion affinity chromatography (His Trap FF column, GE Healthcare). Recombinant Ara h 2 and its variants were refolded on the column by a decreasing linear gradient from 8 to 0 M urea and eluted in native PBS (30 mM sodium phosphate, 150 mM NaCl, pH 7.4) containing 500 mM imidazole and 20% glycerol. Imidazole was subsequently removed using a Zeba Desalt Spin Column (Thermo Fisher Scientific) to enable the storage of rAra h 2 and its variants in a native PBS containing 10% glycerol. The recombinant proteins ARR17L-CAN, ARR17L-TIP, NPA23, and NPA27 (see below for their characterization) were eluted from the columns in a denaturing PBS containing 500 mM imidazole, 6 M urea, and 2 mM β-mercapto-ethanol. Imidazole was removed using a Zeba Desalt Spin Column to enable the storage of ARR17L-CAN and ARR17L-TIP in a denaturing PBS containing 2 M urea and that of NPA23 and NPA27 in a denaturing PBS containing 6 M urea and 2 mM β-mercapto-ethanol. Protein concentrations were determined.
using a protein assay kit (Bio-Rad), and the final molecular weight
of the recombinant purified proteins was verified by 12% SDS-PAGE
(NuPAGE; Invitrogen) after Coomassie blue staining. Endotoxin
concentrations were measured for all recombinant proteins — except
for NPA23 and NPA27, which were used solely for in vitro testing —
using the gel-clot Limulus Amebocyte Lysate (LAL) method and
ENDOSAFE LAL reagents (Charles River Endotoxin Microbial
Detection Europe). Endotoxin concentrations were less than 0.005
EU per microgram of protein.

Whey fractionation. Whey powder (Euroserum) was dissolved in
phosphate buffer (10 mM sodium phosphate, pH 7.4); then the pH
was checked, and the protein concentration was determined using a
protein assay kit (Bio-Rad). A blank gradient was performed on the
cation-exchange-column (HiTrap SP FF, GE Healthcare) connected
to a chromatography apparatus (Akta Xpress, GE Healthcare) using
10 mM sodium phosphate, pH 7.4, as a binding buffer and 10 mM
sodium phosphate, 1 M NaCl, pH 7.4, as elution buffer. A protein
amount corresponding to 50% of the binding capacity of the column
was injected, and the flow-through (containing proteins with isoelec-
tric point ≤ 7.4) was collected. After a washing step with the binding
buffer, the column was eluted with the elution buffer, and the frac-
tions (containing proteins with isoelectric point > 7.4) were collected.
The flow-through and elution pools were buffer exchanged against
PBS (30 mM sodium phosphate, 150 mM NaCl, pH 7.4), and the pro-
tein concentration was determined as above. The flow-through (non-
cationic fraction) and elution (cationic fraction) pools were aliquoted
before storage at –20°C.

Mouse studies

Mouse anaphylaxis model. Female BALB/cByJ mice were purchased
from the Charles River Breeding Centre. Blood samples were collected,
and serum was obtained and stored at –20°C until analysis.

Four-week-old mice were sensitized by intraperitoneal injection
of 400 μg of recombinant rAra h 2 and its TI variants (8 per group)
in PBS with 10% glycerol (VWR) once a week for 4 weeks (Figure 3).
After sensitization, mice were challenged by intraperitoneal injec-
tions of 350 μg of their sensitizing protein. Naive mice injected with
the buffer were used as negative controls. Rectal temperature changes
and scratching were measured to evaluate clinical allergic reactions,
as described previously (22). In another experiment, 10- to 14-week-
old mice received injections of a mixture of rAra h 2 and its TI variants

**Table 7. Characteristics of the human patient sera**

<table>
<thead>
<tr>
<th></th>
<th>Peanut allergic</th>
<th>Atopic without a peanut allergy</th>
<th>Nonatopic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Min</td>
</tr>
<tr>
<td>Females</td>
<td>18</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>8.5</td>
<td>3.4</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Clinical and biological characteristics of all peanut-allergic patients and their controls. The diagnosis of peanut allergy was established based on a positive oral provocation and detectable serum sIgE against peanut allergens; atopic controls were proven not to be allergic to peanuts. Min, minimum; Max, maximum.

**Figure 8. sIgE against canonical and TI variants of Ara h 2.**

(A) Serum sIgE reactivity of 52 peanut-allergic children toward Ara h 2 and its 3 TI variants measured by ELISA. (B) Percentage of inhibition by canonical Ara h 2 of the IgE reactivity to the canonical and v36, v38, and v40 Ara h 2 variants in 10 peanut-allergic patients. Paired t test and Wilcoxon’s signed rank test with Holm-Bonferroni correction. **P < 0.01.
v36, v38, and v40 (100 μg of each) in PBS containing 10% glycerol once a week for 4 weeks. Meanwhile, and as a comparator, 10 other mice were injected sequentially with 400 μg each of Ara h 2-v36 (week 1), Ara h 2-v38 (week 2), Ara h 2-v40 (week 3), and the canonical rAra h 2 at week 4 (Figure 5).

We also injected 5- to 7-week-old mice with the TI peptide-conjugated nonallergenic protein ARR17L-TIP and 5 others with this protein conjugated with the corresponding canonical peptide (ARR17L-CAN), both at 400 μg in PBS containing 2 M urea, once a week for 3 weeks (Figure 4). Finally, we administered intragastrically native or its cationic or noncationic whey fractions, along with 4 μg of cholera toxin (Servibio), to 4-week-old mice once a week for 6 weeks (see Figure 6, C and D).

**Mouse ELISAs.** We assayed the specific Igs to canonical and TI variants of rAra h 2 and ARR17L by ELISA. Plates (MaxiSorp, Dutscher) were coated for 1 hour at 37°C with 1 μg protein per well. After each step, plates were washed 3 times with PBS containing 0.1% Tween-20 (PBST) (MilliporeSigma). Plates were then blocked for 1 hour and incubated with diluted samples in PBS containing 2% milk and 0.05% Tween-20 (PBST) for 2 hours (200 rpm, shaking frequency of 5 Hz, 37°C). Allergen-specific IgM and IgG were detected by alkaline phosphatase–labeled goat anti-mouse IgM and fragment crystallizable (Fc) IgG (catalog a9688 and a1418, MilliporeSigma), respectively. Specific isotypes of IgG were detected by alkaline phosphatase–labeled goat anti-mouse IgG1, IgG2a, IgG2b, IgG2c, and IgG3 (catalog 1070-04, 1080-04, 1090-04, 1079-04, and 1100-04, Clinisciences), respectively. Bound IgE was detected by HRP-labeled goat anti-mouse IgE (catalog 1110-05, Clinisciences). Development was performed with TMB (3,3′,5,5′-tetramethylbenzidine, Fisher Scientific) for HRP and p-nitrophenyl phosphate (pNPP) (Millipore-Sigma) for alkaline phosphatase.

**Western blot inhibition.** One microgram of Ara h 2, Ara h 2-v36, and Ara h 2-v40 and 3 micrograms of Ara h 2-v38 were separated by 12% SDS-PAGE (NuPAGE) under denaturing and reducing conditions and transferred to polyvinylidene difluoride membrane (0.45 μm, GE Healthcare). After blocking with TBST (100 mM Tris [pH 7.5], 154 mM NaCl, 0.1% [v/v] Tween 20) containing 5% (w/v) defatted milk (TBSTM). membranes were incubated with a pool of sera from mice immunized with Ara h 2, Ara h 2-v36, Ara h 2-v38, or Ara h 2-v40 diluted 1:90 in TBSTM and preincubated overnight with 200 μg/mL of Ara h 2. Membranes were washed with TBST and incubated with HRP-labeled polyclonal anti-mouse IgE (catalog 1110-05 diluted 1:3000 in TBSTM, Southern Biotech). After washing, IgE-reactive bands were revealed by chemiluminescence (ECL Advance, GE Healthcare).

**Table 8. Percentages of IgE directed toward the TI segment of each TI variant in 10 peanut-allergic children measured by ELISA**

<table>
<thead>
<tr>
<th>Patient</th>
<th>Ara h 2-v36 (%)</th>
<th>Ara h 2-v38 (%)</th>
<th>Ara h 2-v40 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21.7</td>
<td>13.1</td>
<td>13.3</td>
</tr>
<tr>
<td>2</td>
<td>16.5</td>
<td>8.5</td>
<td>11.3</td>
</tr>
<tr>
<td>3</td>
<td>29.2</td>
<td>24.6</td>
<td>23.0</td>
</tr>
<tr>
<td>4</td>
<td>20.6</td>
<td>12.5</td>
<td>21.2</td>
</tr>
<tr>
<td>5</td>
<td>27.2</td>
<td>7.9</td>
<td>7.1</td>
</tr>
<tr>
<td>6</td>
<td>22.2</td>
<td>17.4</td>
<td>25.1</td>
</tr>
<tr>
<td>7</td>
<td>20.2</td>
<td>13.2</td>
<td>15.4</td>
</tr>
<tr>
<td>8</td>
<td>34.0</td>
<td>25.3</td>
<td>20.5</td>
</tr>
<tr>
<td>9</td>
<td>23.6</td>
<td>23.7</td>
<td>29.8</td>
</tr>
<tr>
<td>10</td>
<td>20.9</td>
<td>29.7</td>
<td>18.0</td>
</tr>
</tbody>
</table>

**Table 9. Characteristics of milk-allergic patients and their controls**

<table>
<thead>
<tr>
<th></th>
<th>IgE– milk allergy</th>
<th>IgE+ milk allergy</th>
<th>Other allergies</th>
<th>Nonatopic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean SD Min Max</td>
<td>Mean SD Min Max</td>
<td>Mean SD Min Max</td>
<td>Mean SD Min Max</td>
</tr>
<tr>
<td>Females</td>
<td>5 4 0.0 4.0</td>
<td>1 1.3 2.0 6.0</td>
<td>6.7 2.3 4.0 8.0</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>2 3 1.0 0.0</td>
<td>6 1.0 0.0 8.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (yr)</td>
<td>1.7 1.4 0.0 4.0</td>
<td>1.7 1.4 0.0 4.0</td>
<td>4.4 1.3 2.0 6.0</td>
<td></td>
</tr>
</tbody>
</table>

Here again, the diagnosis was made based on milk provocation/restriction and the presence or absence of milk-specific IgE. Of the 7 controls with other allergies, 2 were allergic to peanuts, 2 to house dust mites, and 3 to birch or grass pollen.
Human studies
ELISA and ELISA inhibition. We determined the sIgE against NPA23, NPA27, Ara h 2, and its variants v36, v38, and v40 and against the anionic and cationic whey fractions by ELISA. Plates (MaxiSorp) were coated overnight at 4°C with 2.5, 5.0, 1.0, 2.5, 4.0, 1.0, 0.5, and 0.5 μg protein per well, respectively. After each step, plates were washed 3 times with PBST. Plates were then blocked for 2 hours at 37°C and incubated with diluted human serum samples in PBSTM or PBST–fish gelatin 1% for whey fractions for 2 hours at 37°C. Bound sIgE was detected by HRP-labeled polyclonal anti-human IgE (catalog 074-1004 diluted 1:5000, KPL SeraCare) and developed with TMB. The results of this ELISA had been shown to have a high correlation to those of the commercially available ImmunoCap, even though the former required a 20-fold lower serum amount than the latter. The peanut sIgE ELISA inhibition assays were performed as described above, except that the diluted sera were preincubated overnight with large excesses (50, 100, or 200 μg/mL) of the canonical r Ara h 2.

Immunoblotting. Four micrograms of the peanut allergens NPA23 and NPA27 were separated by 12% SDS-PAGE (NuPAGE) under denaturing and reducing conditions and stained with Coomassie blue or transferred to a polyvinylidene difluoride membrane (0.45 μm, GE Healthcare) for immunoblotting. After blocking with TBST containing TBSTM, membranes were incubated with the children’s sera diluted 1:50 in TBSTM. Membranes were then washed with TBST and incubated with HRP-labeled polyclonal anti-human IgE (catalog 074-1004 diluted 1:6000 in TBSTM, KPL SeraCare), as described previously (28). After washing, IgE-reactive bands were revealed by chemiluminescence (ECL Advance, GE Healthcare). Negative control immunoblotting was carried out with the anti-human IgE antibody alone.

Statistics
The χ² and Fisher’s exact tests were used to determine the statistical significance of differences in proportions. A Shapiro-Wilk normality test was used to assess the distribution of data. When a normal distribution was present and there was no difference between variances with an F-test of equality of variance, 2-sample t-tests were used to compare differences between the means of the 2 populations. In the case of different variances between populations, we used Welch’s t-test with unpaired samples. When nonnormal distributions were encountered, Mann–Whitney U test was used with unpaired samples and a nonparametric Wilcoxon’s signed-rank test with paired samples. In the case of a need for multiple testing, P values were adjusted with a Holm-Bonferroni procedure to control the family-wise error rate. P < 0.05 was considered statistically significant. Statistics were calculated with R 3.6.1 and RStudio 1.1.463.

Study approval
Written, informed consent for use of children’s serum for research purposes was obtained from their parents. Regulatory approval for the collection and the storage of human samples for research purposes was obtained from France’s Ministry of Health (CCITIRS 15.616 bis). Rodent studies were approved by the Comité d’Étique Lorrain en Matière d’Expérimentation Animale CELMEA-66, Université de Lorraine (Nancy, France) and were performed in an animal facility accredited and controlled by France’s regulatory agencies (approval C54-547-16).

Author contributions
BT and BH conducted the bioinformatics studies and BT cowrote the paper. OR produced the recombinant proteins and natural extracts. JT conducted the animal studies. SJ and CR developed and performed the human immunoassays. FCM, MM, GK, EB, and CDS made the diagnosis in patients and provided blood samples. CF and TO contributed to the experimental design and wrote the paper, while BEB formulated the hypothesis experimentally verified in this manuscript and participated in the experimental design. All authors have seen and approved the final version of this paper.

Acknowledgments
We greatly appreciate the expert technical assistance of Lionel Bonnard, Christelle Gonet, Frédéric Colin, Abdelouahed Babnou, Marie-Christine Sencier, Pierre VATIER, Gaël Douchin, and Thibaut Matelski. We are indebted to the late Denise-Anne Babnou, Marie-Christine Sencier, Pierre VATIER, Gaël Douchin, and Thibaut Matelski. We are indebted to the late Denise-Anne Babnou, Marie-Christine Sencier, Pierre VATIER, Gaël Douchin, and Thibaut Matelski.

Address correspondence to: Bernard E. Bihain, Genclis SA, 15 rue du bois de la champelle, 54500 Vandoeuvre-lès-Nancy, France. Phone: 33.3.83.67.82.11; Email: bbihain@genclis.com.
The Journal of Clinical Investigation

2015;8 Suppl 2:58.