Transcriptional frameshifts contribute to protein allergenicity

Benoit Thouvenot,[1] Olivier Roitel,[1] Julie Tomasina,[1] Benoit Hilselberger,[1] Christelle Richard,[1] Sandrine Jacquenet,[1] Françoise Codreanu-Morel,[2] Martine Morisset,[3] Gisèle Kanny,[4] Etienne Beaudouin,[5] Christine Delebarre-Sauvage,[6] Thierry Olivry,[7] Claude Favrot,[8] and Bernard E. Bihain[1]

[1]Genclis SA, Vandoeuvre-lès-Nancy, France. [2]Unité Nationale d'Immunologie et d'Allergologie, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg. [3]Unité d'Allergologie, Département de Pneumologie, Centre Hospitalier Universitaire Angers, Angers, France. [4]Service de Médecine Interne, Immunologie Clinique et Allergologie, Hôpitaux de Brabois, Centre Hospitalier Universitaire de Nancy, Vandœuvre-lès-Nancy, France. [5]Unité d'Allergologie, Centre Hospitalier Régional de Metz, Metz, France. [6]Allergology Center Saint-Vincent de Paul Hospital, Groupe Hospitalier de l'Institut Catholique de Lille, Lille, France. [7]Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina, USA. [8]Clinic 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/chemokine pattern that defines the Th2 immune response (6). Allergens binding to IgE anchored to the FcεRI high-affinity IgE receptors on mast cells and basophils cause their aggregation and trigger a signaling cascade, leading to the release of potent proinflammatory mediators (7).

▶ **Related Commentary: p. 5118**

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 lactoserum. 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
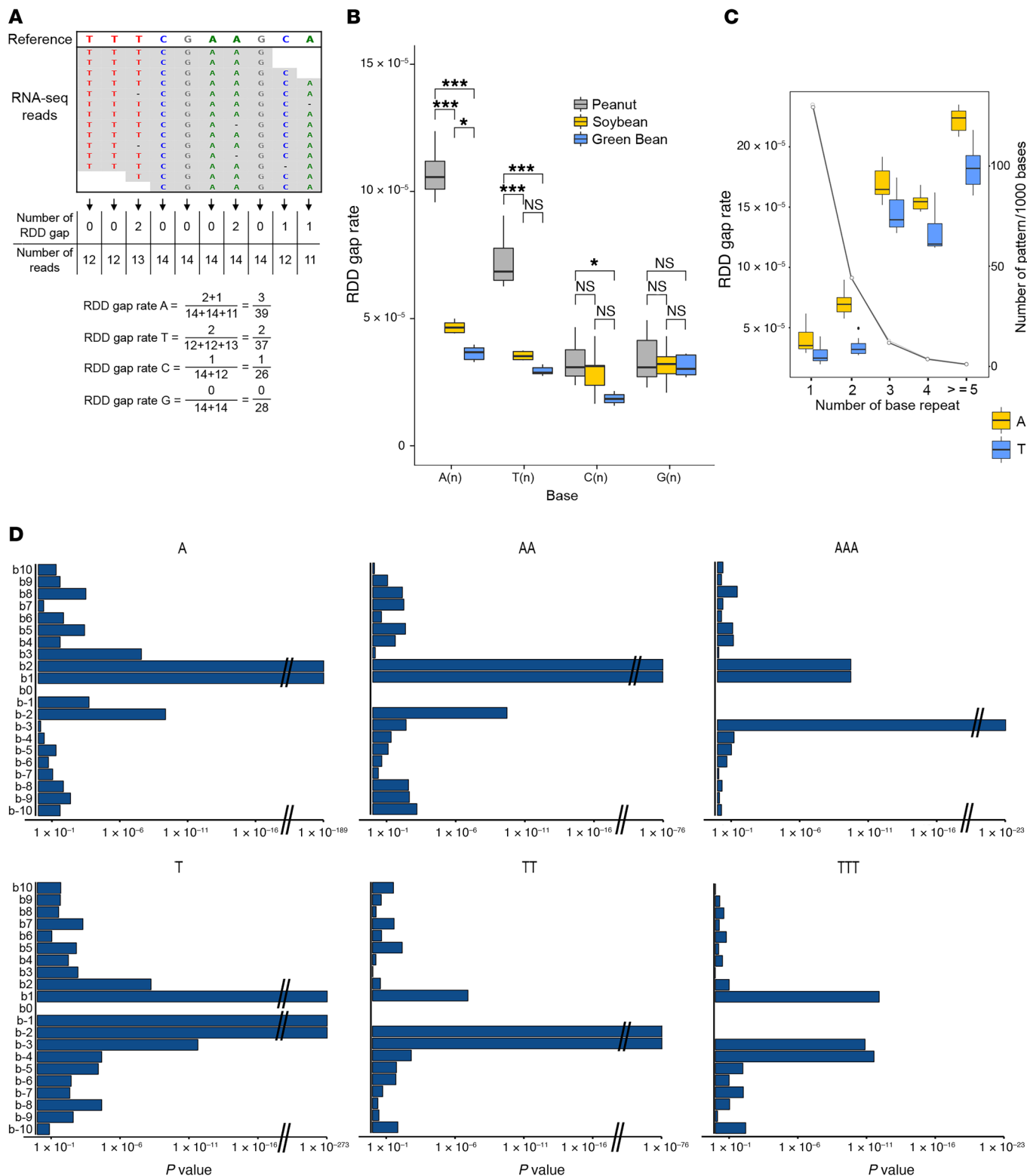
Figure 1. RDD gap rates in legumes. (**A**) Method of calculation of the RDD gap rates. (**B**) RDD gap rates related to the numbers (indicated as *n*) of A and T repeats for peanut, soybean, and green bean. (**C**) Peanut RDD gap rates related to the numbers of A and T repeats. (**D**) Base composition 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.
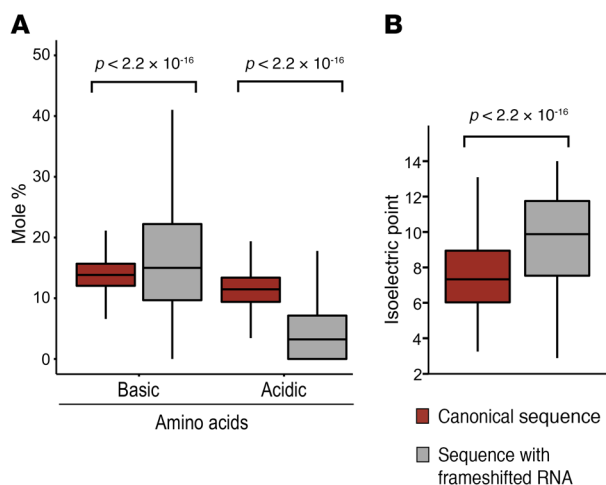
Figure 2. Theoretical and observed aa sequence compositions in peanut TI variants. (**A**) The in silico translation of peanut RNA with identified gaps showed that peptides translated from frameshifted RNA contained 4.4% acidic aa and 16% basic aa. Mole percentages were calculated using Pepstats from EMBOSS. (**B**) Impact of aa composition on the isoelectric point of TI peptides. Welch's *t* test with Holm-Bonferroni correction. (**B**) Box-and-whisker plot: box, interquartile range; bar, median; whiskers, 95% CI.

constructions (Figure 3B). These 4 purified proteins were injected without adjuvant, at a 1-week interval, into groups of 8 BALB/cByJ mice (400 μg each) (Figure 3C); control rodents received only buffer vehicle injections. We then measured the levels of IgM, IgG, and IgE directed against either the canonical (Figure 3D) or TI variants (Figure 3E) of Ara h 2 at various time points. All 4 proteins induced significant increases in Ara h 2 IgM and IgG levels relative to controls (Figure 3D). The increased subclasses of IgG were mainly IgG1 (Supplemental Figure 2). Only the injection of the 3 TI variants increased IgE reactivity against both the canonical rAra h 2 (Figure 3D) and its 3 variants (Figure 3E). In contrast, similar injections of the canonical rAra h 2 caused no detectable increase in Ara h 2 sIgE reactivity (Figure 3D). Seventy-six days after the first injection, we challenged mice with their respective sensitizing proteins. We evaluated symptomatic allergic reactions, as we previously described, by a total allergy score, the level of scratching during the first 30 minutes after injection, and the decrease in body temperature (22). All 3 TI variants of Ara h 2 induced clinical allergic reactions after their respective sensitization, which were significantly more pronounced than those recorded in control or canonical rAra h 2–injected mice (Figure 3F).

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

aa	Probability of gain	Probability of loss	Probability of conservation
Acidic	0.036	0.120	0
Basic	0.116	0.111	0.029

Acidic aa are D and E and Basic aa are R, K, and H.

Although sensitizations with the canonical rAra h 2 did not induce a 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 sIgE 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 sIgE 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 sIgE production and clinical signs of allergy in sensitized mice; furthermore, we confirmed that the sIgE 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 sIgE in mice. We next tested to determine whether the induction of sIgE 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 sIgM and sIgG against the respective chimera, but sIgE 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 sIgE 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 sIgE. 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 sIgE.

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 2. Predicted and observed aa composition in peanut canonical and TI peptides

Peanut	Canonical	TI peptides	
	Observed	Observed	Theoretical
Acidic aa %	11.5	4.4	0
Basic aa %	13.9	16.0	14.7

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-

Table 3. Comparison of allergen and nonallergen transcripts for peanut

	Allergen genes (n = 11)		Nonallergen genes (n = 10,224)		P value (allergens vs. non allergens)
	Mean	SD	Mean	SD	
Protein size	231	173	424	293	0.029
Transcript expression levels	37.06	35.87	0.36	1.15	0.021
Gap density	0.13	0.10	0.03	0.04	0.022
Earliest occurrence of TI event (0 NH2 → 1 COOH)	0.07	0.06	0.21	0.22	0.0002
Latest occurrence of TI event (0 NH2 → 1 COOH)	0.93	0.06	0.81	0.22	0.0002
Acidic aa (%)	6.01	5.36	4.57	5.51	2.117×10^{-29}
Basic aa (%)	15.62	9.07	13.72	10.35	8.942×10^{-05}
Hydrophobic aa (%)	40.47	12.17	49.99	14.68	0.003
Hydrophilic aa (%)	37.91	13.35	31.61	14.21	9.285×10^{-13}

Number of aa indicates protein size; transcript expression is expressed in percentages of reads per kilobase per million mapped reads (RPKM), with the highest expression serving as a reference; gap diversity corresponds to the number of gaps, taking into account the protein size. 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 Welsh's t test with Holm-Bonferroni correction.

Table 4. Comparison of allergen and nonallergen transcripts for soybean

	Allergen genes (n = 18)		Nonallergen genes (n = 24,238)		P value (allergens vs. nonallergens)
	Mean	SD	Mean	SD	
Protein size	334	178	500	361	0.003
Transcript expression levels	2.31	2.31	0.06	0.29	0.003
Gap density	0.06	0.02	0.02	0.03	5.18×10^{-07}
Earliest occurrence of TI event (0 NH2 -> 1 COOH)	0.07	0.08	0.26	0.26	2.024×10^{-07}
Latest occurrence of TI event (0 NH2 -> 1 COOH)	0.87	0.28	0.74	0.25	0.027
Acidic aa (%)	5.72	6.23	4.58	5.57	0.003
Basic aa (%)	16.20	10.69	13.98	10.98	0.001
Hydrophobic aa (%)	43.07	15.5	49.04	15.31	1.287×10^{-11}
Hydrophilic aa (%)	35.01	14.39	32.4	15.01	0.003

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 Welsh's t test with Holm-Bonferroni correction.

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 *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 *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 *Arachis ipaensis*, the LOC107629116 gene encodes the NPA27. We produced recombinant NPA27 in *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

Table 5. Comparison of allergen and nonallergen transcripts for sesame

	Allergen genes (n = 9)		Nonallergen genes (n = 17,285)		P value (allergens vs. non allergens)
	Mean	SD	Mean	SD	
Protein size	316	179	494	344	0.09
Transcript expression levels	70.86	29.57	0.68	1.39	0.001
Gap density	0.04	0.01	0.03	0.04	0.09
Earliest occurrence of TI event (0 NH2 → 1 COOH)	0.18	0.16	0.19	0.22	1
Latest occurrence of TI event (0 NH2 → 1 COOH)	0.95	0.02	0.80	0.22	4.248×10^{-07}
Acidic aa (%)	5.72	5.17	4.64	5.61	0.09
Basic aa (%)	17.14	10.97	14.84	11.20	0.09
Hydrophobic aa (%)	44.98	12.47	48.51	14.49	**0.009**
Hydrophilic aa (%)	31.80	12.04	32.00	14.31	1

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 Welsh's t test with Holm-Bonferroni correction.

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

	Allergen genes (n = 17)		Nonallergen genes (n = 5,587)		P value (allergens vs. nonallergens)
	Mean	SD	Mean	SD	
Protein size	420	406	650	606	0.05
Transcript expression levels	16.49	13.22	1.97	4.57	0.002
Gap density	0.05	0.03	0.02	0.02	0.003
Earliest occurrence of TI event (0 NH2 → 1 COOH)	0.16	0.22	0.34	0.25	0.007
Latest occurrence of TI event (0 NH2 → 1 COOH)	0.87	0.10	0.71	0.24	8.955×10^{-05}
Acidic aa (%)	3.54	5.31	4.51	5.52	0.003
Basic aa (%)	13.95	9.20	12.43	9.01	0.007
Hydrophobic aa (%)	51.78	13.44	54.3	14.25	0.003
Hydrophilic aa (%)	30.73	12.09	28.77	12.51	0.007

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 Welsh'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 ($P < 10^{-6}$) 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 directed against these 2 lactoserum fractions. 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.
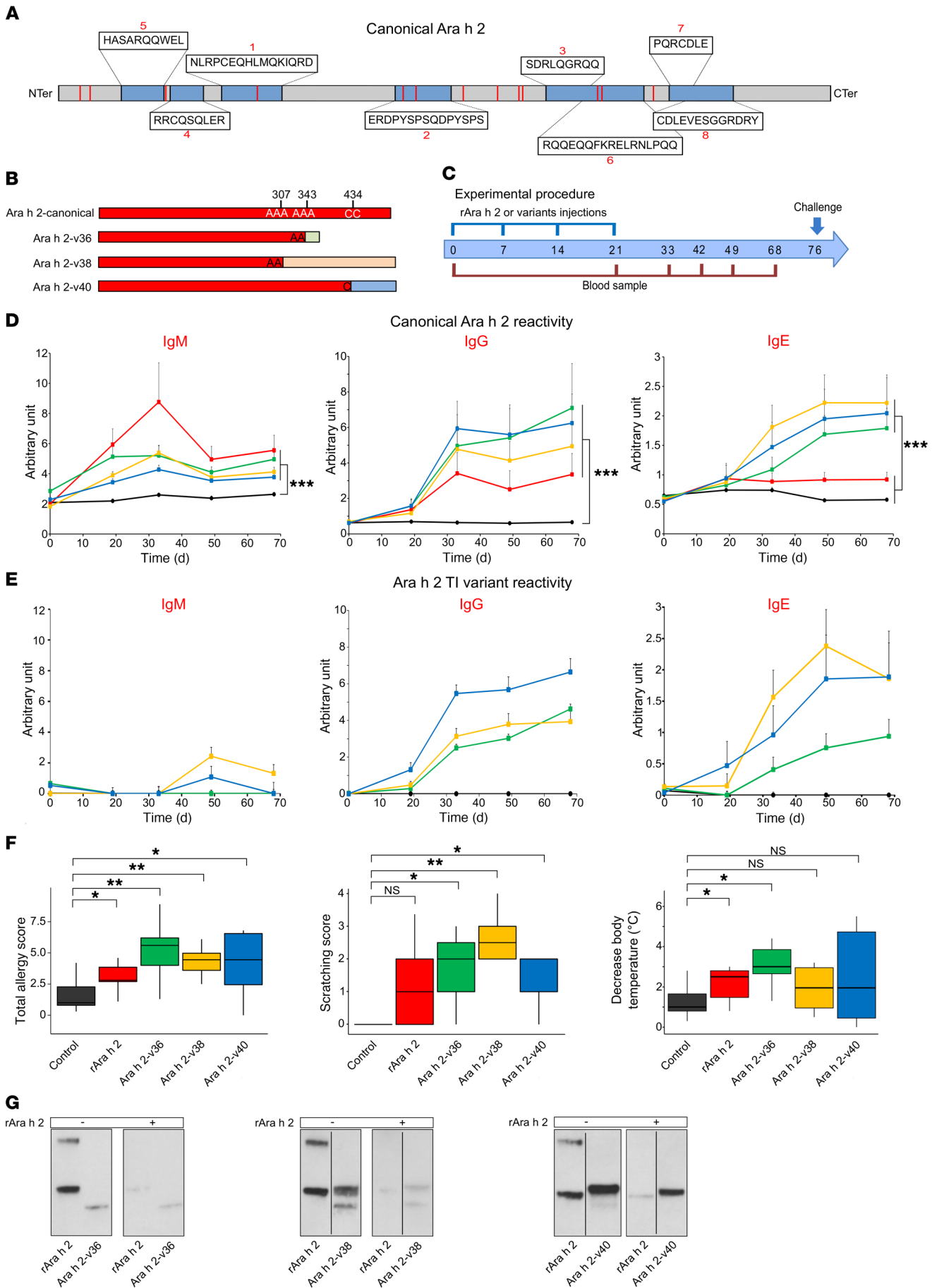
Figure 3. Evaluation of the immunogenicity of canonical and TI variants of Ara h 2. (**A**) Distribution of 13 TI gap events over the Ara h 2 mRNA (red lines); the major Ara h 2 epitopes are shown in blue boxes, and they are numbered according to the recommended nomenclature (43). The aa sequences of Ara h 2 epitopes are shown in white boxes. (**B**) Proteins injected. The same color for each respective protein is used in **B**, **D**, **E**, and **F**. (**C**) Immunization protocol. (**D**) Canonical Ara h 2–specific antibodies. (**E**) Ara h 2 TI variant–specific antibodies. Black, control mice; red, mice sensitized with Ara h 2; green, Ara h 2–v36; yellow, Ara h 2–v38; and blue, Ara h 2–v40 (**C** and **D**). Data shown in **D** and **E** are represented as mean + SEM, and lines in **E** correspond to means minus those of Ara h 2–sensitized mice. (**F**) Total allergy, body temperature, and scratching scores after challenge in sensitized mice. Data are represented as mean + SEM (**G**) IgE Western immunoblotting of rAra h 2 and the 3 TI variants with sera of mice immunized with Ara h 2–v36 (left), v38 (middle), and v40 (right). Inhibition by the canonical rAra h 2 is indicated by +. Single experiment with 8 mice per sensitization group; *t* test, Welch's *t* test and Mann-Whitney *U* tests with Holm-Bonferroni correction: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. (**F**) Box-and-whisker plot: box, interquartile range; bar, median; whiskers, 95% CI.

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.

Herein, we report for what we believe is the first time that a single base gap is a frequent TI event that causes the translation of cationic peptides because of the asymmetric degeneracy of the universal genetic codes favoring the translation of basic aa. Furthermore, we show that a high gap density and an overall increase in the RDD gap rate are both associated with allergenic-

ity and that these characteristics are not restricted to the peanut, as they are also present in all 4 tested allergen sources, irrespective of their ontogeny. Together, these observations indicate that the aa motifs that initiate allergen-sIgE production are likely absent in proteins translated from their canonical RNA; this is a possible explanation for the long-lasting, yet fruitless, quest for aa patterns that would distinguish allergens from nonallergenic proteins. While abundant canonical allergens might not initiate sIgE production sensu stricto, they can, however, provoke clinical manifestations of the disease. Indeed, we have shown previously that the canonical Ara h 2 can induce positive skin prick tests in peanut-allergic patients (27).

Previous reports have shown that transcriptional substitutions occurring only at the RNA level occur commonly in human subjects and translate into variants of canonical proteins (8, 11). For example, we established earlier that RDD substitutions occur more commonly in cancerous than in normal healthy cells (15). Furthermore, frameshifted RNA resulting in GA deletions are also present in neurological disorders (28–30). Indeed, RNA polymerase II is error prone (for a review, see refs. 31, 32), and its rate of error increases with the level of gene expression (13). Interestingly, we established herein that transcripts encoding allergens are more abundant than those coding for nonallergenic proteins, and this could increase the rate of transcription errors.

Furthermore, our data revealed that RNA frameshifts resulting from single base gaps are a rather common phenomenon that does not occur randomly, but preferentially at sites of A and T repeats and only when the composition of surrounding bases is permissive for slippage of the RNA polymerase II. This motif of 6 influential bases is similar to that previously described for TI substitutions, and it corresponds to the portion of DNA melted by the RNA polymerase II during the transcription process (15). With the advent of CRISPR/Cas9 that allows for genomic editing (33), a modification of these DNA motifs that are conducive to base deletion — perhaps by changing a base without altering the aa sequence itself — would, at least theoretically, reduce the transcriptional slippage. Such reduction would then decrease sIgE production and could help with designing foods of lower allergenicity. From the same perspective, the construction of artificial genes not permissive to TI and coding for therapeutic recombinant proteins could reduce their allergenic potential.

The data reported herein could help further the understanding of the mechanisms underlying the development of sensitization to allergens. Indeed, we provide evidence that, in both rodents and humans, cationic TI peptides relatively depleted in hydrophobic aa play a critical role in the induction of sIgE production. These molecular patterns do not require an adjuvant, thereby suggesting that they, by themselves, carry one of the hypothetical "danger signals" for an IgE production switch. Herein, we provide simple biochemical probes that might enable future dissections into the mechanism involved in regulating the complex biology of IgE. Nevertheless, our findings do not rule out that other mechanisms (i.e., protease activity, autoadjuvant properties, and others) might also contribute to IgE production (34, 35).

Furthermore, not all cationic peptides inducing the production of sIgE are the result of the translation of frameshifted RNA. For example, there is a cluster of positively charged residues resulting
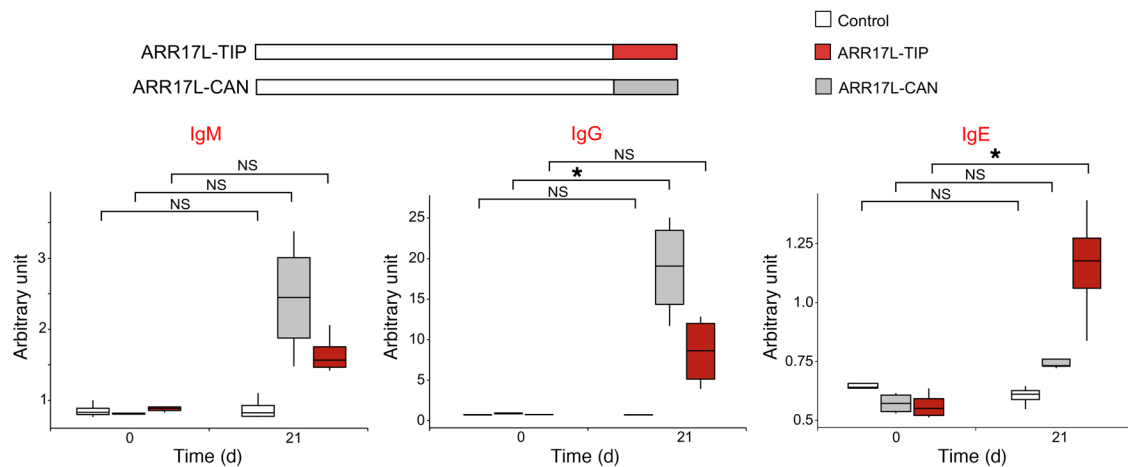
Figure 4. Effect of the addition of a TI peptide to a nonallergenic protein. Levels of specific Igs directed against ARR17L-TIP (TI peptide) and ARR17L-CAN (corresponding canonical sequence) in mice sensitized by intraperitoneal injection on days 0, 7, and 14. Single experiment with 5 mice per sensitization 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.

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
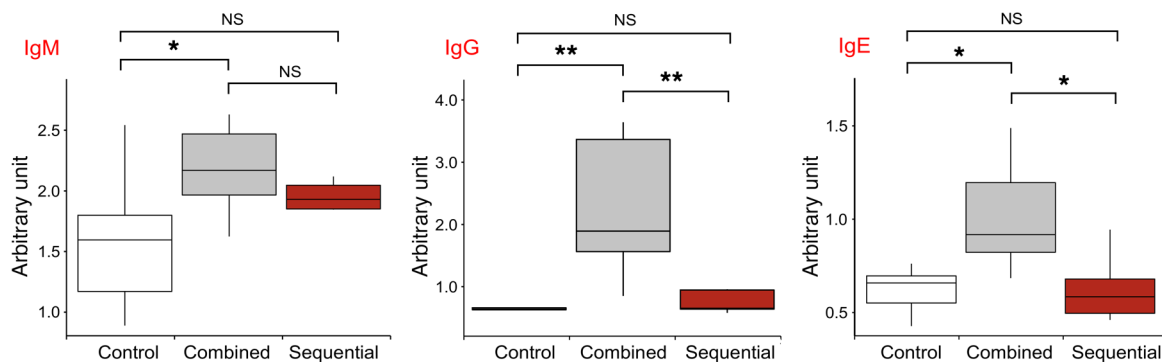
Figure 5. Comparison of the effect of combined or sequential injections of TI peptides. Ara h 2 antibody levels in control mice and those injected with a combined dose of 400 μg (100 μg of each of 4 TI proteins) or the sequential injection of 400 μg of each protein at a 1-week interval. Single experiment 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.
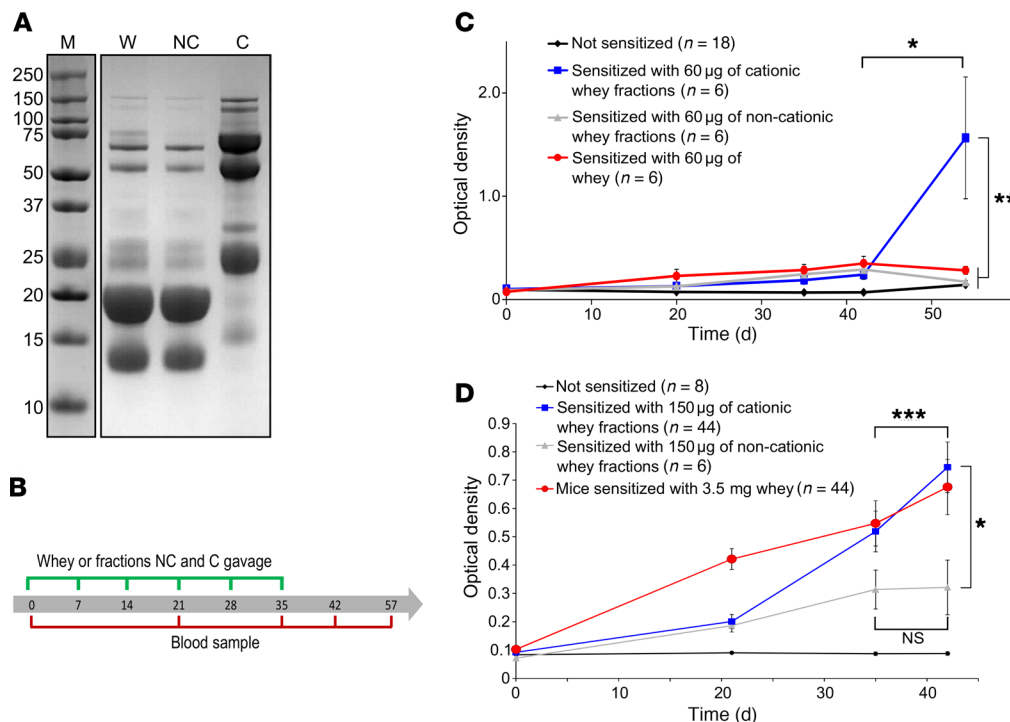
Figure 6. Differential allergenicity of whey fractions separated by their charges. (**A**) Coomassie staining of SDS-PAGE with 10 μg of either total, noncationic (isoelectric point < 7.4), or cationic (isoelectric point > 7.4) whey fractions. W, whey; C, cationic; NC, noncationic. (**B**) Gastric administration and blood sampling schedule of whey, cationic, and noncationic fractions. (**C**) Serum sIgE response in BALB/cByJ mice that had ingested the same 60 μg of either total, cationic, and noncationic whey fractions. (**D**) IgE response of mice given either 3.5 mg of total whey or 150 μg of cationic and noncationic whey fractions. Single experiment with 36 mice (**C**) and 106 mice (**D**). Mann-Whitney *U* paired and unpaired tests. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$

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 isotype 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, but with modifications appropriate to curate short reads (see Supplemental Methods) (15). 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.
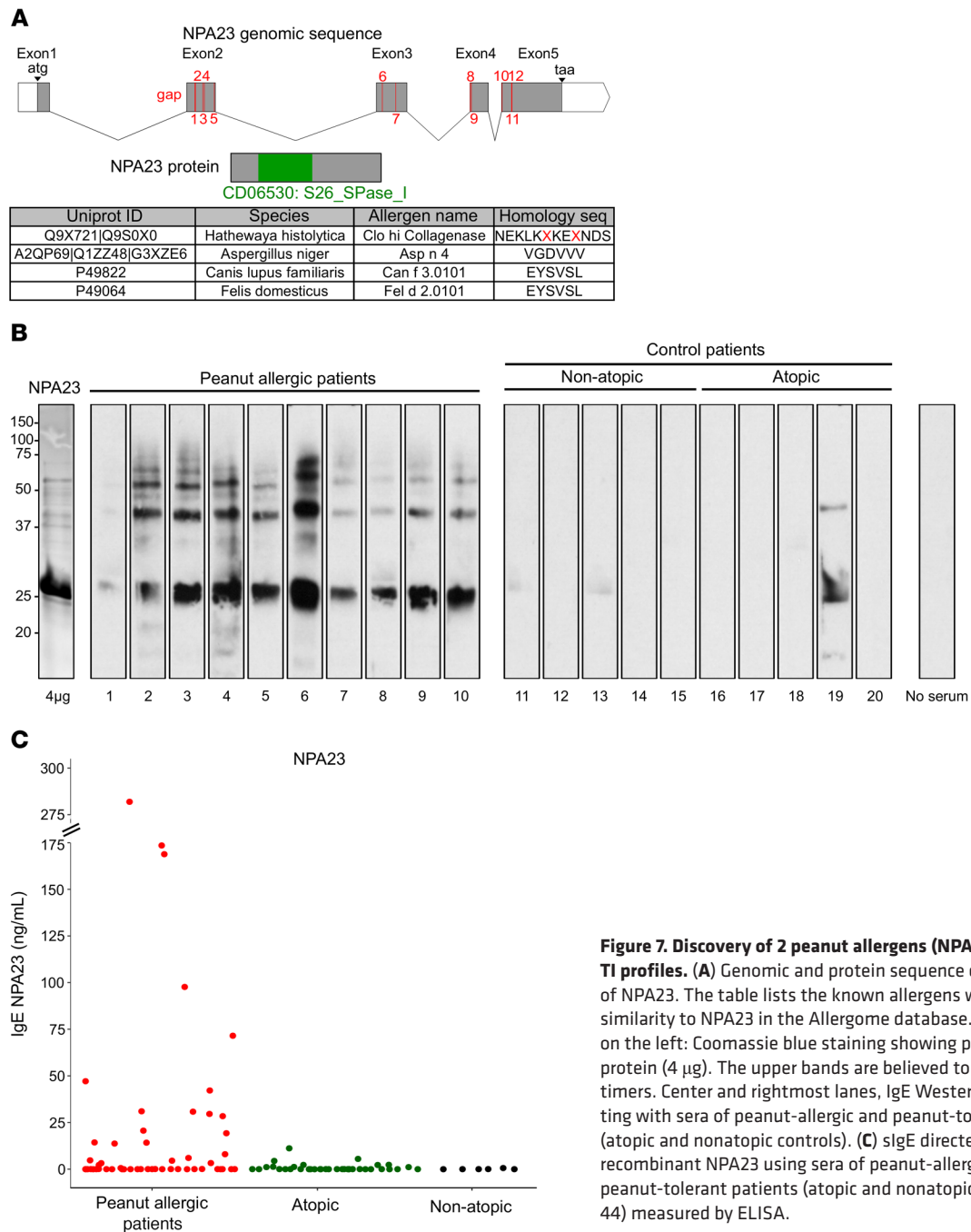
Figure 7. Discovery of 2 peanut allergens (NPA) using their TI profiles. (A) Genomic and protein sequence characteristics of NPA23. The table lists the known allergens with sequence similarity to NPA23 in the Allergome database. **(B)** First lanes on the left: Coomassie blue staining showing purified NPA23 protein (4 μg). The upper bands are believed to be NPA23 multimers. Center and rightmost lanes, IgE Western immunoblotting with sera of peanut-allergic and peanut-tolerant children (atopic and nonatopic controls). **(C)** sIgE directed toward recombinant NPA23 using sera of peanut-allergic (*n* = 52) and peanut-tolerant patients (atopic and nonatopic subjects; *n* = 44) measured by ELISA.

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 phos-

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

Table 7. Characteristics of the human patient sera

	Peanut allergic				Atopic without a peanut allergy				Nonatopic			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Females	18				12				2			
Males	34				26				4			
Age (yr)	8.5	3.4	3.0	16.0	7.2	3.3	2.0	18.0	10.7	5.1	4.0	18.0
sIgE (kU/L)												
Ara h 1	31.6	51.1	<0.1	296	<0.1	<0.1	<0.1	0.12	<0.1	<0.1	<0.1	<0.1
Ara h 2	46.5	47.3	<0.1	204	<0.1	<0.1	<0.1	0.10	<0.1	<0.1	<0.1	<0.1
Ara h 3	13.0	21.6	<0.1	100	<0.1	<0.1	<0.1	0.14	<0.1	<0.1	<0.1	<0.1
Ara h 6	11.5	16.2	<0.1	65.1	<0.1	<0.1	<0.1	0.10	ND	ND	ND	ND
Ara h 7	4.0	9.5	<0.1	48.1	<0.1	<0.1	<0.1	0.12	ND	ND	ND	ND
Ara h 8	2.3	7.0	<0.1	43.4	11.5	20.6	<0.1	71.4	<0.1	<0.1	<0.1	<0.1
Ara h 9	<0.1	0.10	<0.1	0.6	<0.1	0.2	<0.1	0.9	ND	ND	ND	ND

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.

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

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. **$P < 0.01$.
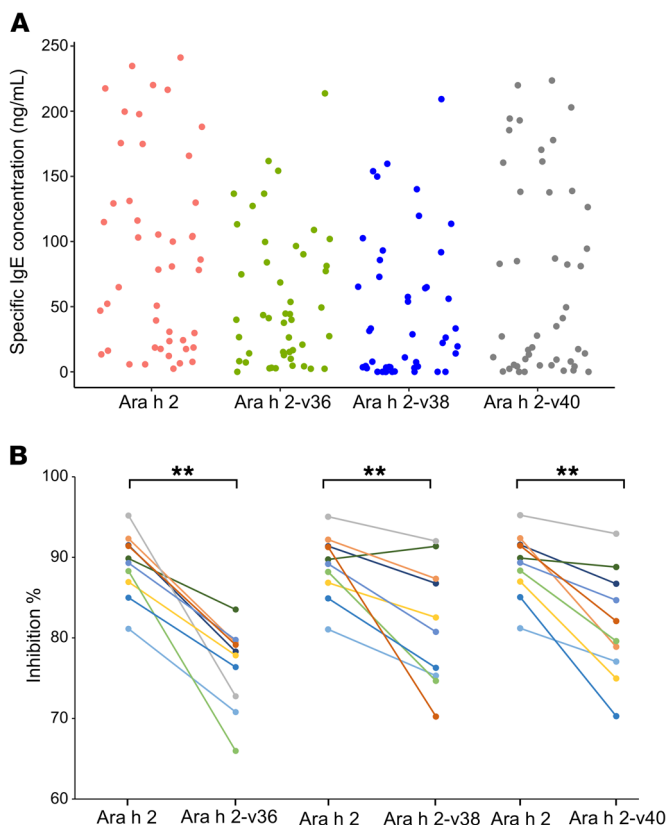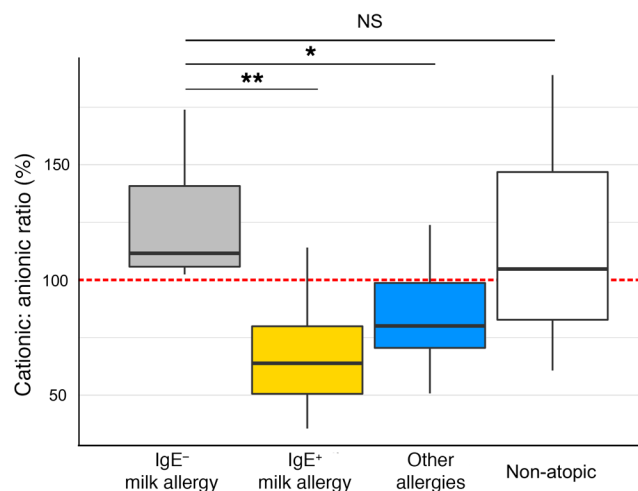
for NPA23 and NPA27, which were used solely for in vitro testing — using the gel-clot Limulus Amebocyte Lysate (LAL) method and ENDOSAFE LAL reagents (Charles River Endotoxin Microbial Detection Europe). Endotoxin concentrations were less than 0.005 EU per microgram of protein.

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 isoelectric point ≤ 7.4) was collected. After a washing step with the binding buffer, the column was eluted with the elution buffer, and the fractions (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 protein concentration was determined as above. The flow-through (noncationic 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 injections 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 8. Percentages of IgE directed toward the TI segment of each TI variant in 10 peanut-allergic children measured by ELISA

IgE reactivity with TI segment of variants (%)	Ara h 2–v36	Ara h 2–v38	Ara h 2–v40
Patient 1	21.7	13.1	13.3
Patient 2	16.5	8.5	11.3
Patient 3	29.2	24.6	23.0
Patient 4	20.6	12.5	21.2
Patient 5	27.2	7.9	7.1
Patient 6	22.2	17.4	25.1
Patient 7	20.2	19.2	15.4
Patient 8	34.0	25.3	20.5
Patient 9	23.6	23.7	29.8
Patient 10	20.9	29.7	18.0

Figure 9. IgE reactivity of milk-allergic children to charge-separated whey fractions. Sera from children ($n = 7$) with a clinical allergy to milk but undetectable milk-specific IgE using standard milk extracts (IgE⁻ milk allergy) were found to have, in fact, IgE that reacted more to the TI protein–enriched cationic fraction of whey than to its anionic fraction containing the canonical proteins. The reverse was seen in children with milk allergy and detectable milk-specific IgE ($n = 7$) and children with other allergies ($n = 7$); ratios from nonatopic children ($n = 3$) were inconclusive. t test with Holm-Bonferroni correction. $*P < 0.05$; $**P < 0.01$. Box-and-whisker plot: box, interquartile range; bar, median; whiskers, 95% CI.

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 (PBSTM) 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, Cliniseiences). 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 9. Characteristics of milk-allergic patients and their controls

	IgE⁻ milk allergy				IgE⁺ milk allergy				Other allergies				Nonatopic			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Females	5				4				1				2			
Males	2				3				6				1			
Age (yr)	1.7	1.4	0.0	4.0	1.7	1.4	0.0	4.0	4.4	1.3	2.0	6.0	6.7	2.3	4.0	8.0

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 beforehand to match 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 rAra 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 χ^2 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 not any 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 2-sample *t* test. When nonnormal distributions were observed, Mann-Whitney *U* test was used with unpaired samples and 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 (CCTIRS 15.616 bis). Rodent studies were approved by the Comité d'Etique 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).

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.

1. Chaplin DD. Overview of the immune response. *J Allergy Clin Immunol.* 2010;125(2 Suppl 2):S3–23.
2. Palm NW, Rosenstein RK, Medzhitov R. Allergic host defences. *Nature.* 2012;484(7395):465–472.
3. Gupta RS, et al. Prevalence and severity of food allergies among US adults. *JAMA Netw Open.* 2019;2(1):e185630.
4. Dunlop JH, Keet CA. Epidemiology of food allergy. *Immunol Allergy Clin North Am.* 2018;38(1):13–25.
5. Xiong H, Curotto de Lafaille MA, Lafaille JJ. What is unique about the IgE response? *Adv Immunol.* 2012;116:113–141.
6. Poulsen LK, Hummelshoj L. Triggers of IgE class switching and allergy development. *Ann Med.* 2007;39(6):440–456.
7. Wu LC. Immunoglobulin E receptor signaling and asthma. *J Biol Chem.* 2011;286(38):32891–32897.
8. Li M, et al. Widespread RNA and DNA sequence differences in the human transcriptome. *Science.* 2011;333(6038):53–58.
9. Bahn JH, Lee JH, Li G, Greer C, Peng G, Xiao X. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* 2012;22(1):142–150.
10. Ju YS, et al. Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals. *Nat Genet.* 2011;43(8):745–752.
11. Peng Z, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat Biotechnol.* 2012;30(3):253–260.
12. Wang IX, et al. RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. *Cell Rep.* 2014;6(5):906–915.
13. Wons E, Furmanek-Blaszk B, Sektas M. RNA editing by T7 RNA polymerase bypasses InDel mutations causing unexpected phenotypic changes. *Nucleic Acids Res.* 2015;43(8):3950–3963.
14. Reid-Bayliss KS, Loeb LA. Accurate RNA consensus sequencing for high-fidelity detection of transcriptional mutagenesis-induced epimutations. *Proc Natl Acad Sci USA.* 2017;114(35):9415–9420.
15. Brulliard M, et al. Nonrandom variations in human cancer ESTs indicate that mRNA heterogeneity increases during carcinogenesis. *Proc Natl Acad Sci USA.* 2007;104(18):7522–7527.
16. Lee SY, Joung JG, Park CH, Park JH, Kim JH. RCARE: RNA sequence comparison and annotation for RNA Eediting. *BMC Med Genomics.*

2015;8 Suppl 2:S8.

17. Gott JM, Emeson RB. Functions and mechanisms of RNA editing. *Annu Rev Genet*. 2000;34:499–531.

18. Pérez-Ortín JE, Alepuz P, Chávez S, Choder M. Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression. *J Mol Biol*. 2013;425(20):3750–3775.

19. Cabanillas B, Jappe U, Novak N. Allergy to peanut, soybean, and other legumes: recent advances in allergen characterization, stability to processing and IgE cross-reactivity. *Mol Nutr Food Res*. 2018;62(1):1700446.

20. Dodo HW, Viquez OM, Maleki SJ, Konan KN. cDNA clone of a putative peanut (Arachis hypogaea L.) trypsin inhibitor has homology with peanut allergens Ara h 3 and Ara h 4. *J Agric Food Chem*. 2004;52(5):1404–1409.

21. Codreanu F, et al. A novel immunoassay using recombinant allergens simplifies peanut allergy diagnosis. *Int Arch Allergy Immunol*. 2011;154(3):216–226.

22. Proust B, et al. A single oral sensitization to peanut without adjuvant leads to anaphylaxis in mice. *Int Arch Allergy Immunol*. 2008;146(3):212–218.

23. Beutier H, et al. IgG subclasses determine pathways of anaphylaxis in mice. *J Allergy Clin Immunol*. 2017;139(1):269–280.e7.

24. Khodoun MV, et al. Rapid desensitization of mice with anti-FcγRIIb/FcγRIII mAb safely prevents IgG-mediated anaphylaxis. *J Allergy Clin Immunol*. 2013;132(6):1375–1387.

25. Miyajima I, Dombrowicz D, Martin TR, Ravetch JV, Kinet JP, Galli SJ. Systemic anaphylaxis in the mouse can be mediated largely through IgG1 and Fc gammaRIII. Assessment of the cardiopul-

monary changes, mast cell degranulation, and death associated with active or IgE- or IgG1-dependent passive anaphylaxis. *J Clin Invest*. 1997;99(5):901–914.

26. Stanley JS, et al. Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. *Arch Biochem Biophys*. 1997;342(2):244–253.

27. Astier C, et al. Predictive value of skin prick tests using recombinant allergens for diagnosis of peanut allergy. *J Allergy Clin Immunol*. 2006;118(1):250–256.

28. van Leeuwen FW, et al. Frameshift proteins in autosomal dominant forms of Alzheimer disease and other tauopathies. *Neurology*. 2006;66(2 Suppl 1):S86–S92.

29. van Leeuwen FW, et al. Molecular misreading: a new type of transcript mutation expressed during aging. *Neurobiol Aging*. 2000;21(6):879–891.

30. van Leeuwen FW, Kros JM, Kamphorst W, van Schravendijk C, de Vos RA. Molecular misreading: the occurrence of frameshift proteins in different diseases. *Biochem Soc Trans*. 2006;34(Pt 5):738–742.

31. Sydow JF, et al. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell*. 2009;34(6):710–721.

32. Gamba P, Zenkin N. Transcription fidelity and its roles in the cell. *Curr Opin Microbiol*. 2018;42:13–18.

33. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014;346(6213):1258096.

34. Seo GY, et al. Mechanism underlying the suppressor activity of retinoic acid on IL4-induced IgE

synthesis and its physiological implication. *Cell Immunol*. 2017;322:49–55.

35. Cayrol C, Girard JP. Interleukin-33 (IL-33): A nuclear cytokine from the IL-1 family. *Immunol Rev*. 2018;281(1):154–168.

36. Da Silva M, et al. The family secrets of avian egg-specific ovalbumin and its related proteins Y and X. *Biol Reprod*. 2015;93(3):71.

37. Fitzsimmons CM, Falcone FH, Dunne DW. Helminth allergens, parasite-specific IgE, and its protective role in human immunity. *Front Immunol*. 2014;5:61.

38. Lowe AJ, Dharmage SC, Allen KJ, Tang ML, Hill DJ. The role of partially hydrolyzed whey formula for the prevention of allergic disease: evidence and gaps. *Expert Rev Clin Immunol*. 2013;9(1):31–41.

39. Portelli MA, Hodge E, Sayers I. Genetic risk factors for the development of allergic disease identified by genome-wide association. *Clin Exp Allergy*. 2015;45(1):21–31.

40. Ferreira MA, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet*. 2017;49(12):1752–1757.

41. Molloy J, Allen K, Collier F, Tang ML, Ward AC, Vuillermin P. The potential link between gut microbiota and IgE-mediated food allergy in early life. *Int J Environ Res Public Health*. 2013;10(12):7235–7256.

42. Platts-Mills TA. The allergy epidemics: 1870-2010. *J Allergy Clin Immunol*. 2015;136(1):3–13.

43. Deak PE, Vrabel MR, Kiziltepe T, Bilgicer B. Determination of crucial immunogenic epitopes in major peanut allergy protein, Ara h2, via novel nanoallergen platform. *Sci Rep*. 2017;7(1):3981.