

Common Physical–Chemical Properties Correlate with Similar Structure of the IgE Epitopes of Peanut Allergens

CATHERINE H. SCHEIN,* OVIDIU IVANCIUC, AND WERNER BRAUN

Sealy Center for Structural Biology, Department of Human Biological Chemistry and Genetics,
 University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-0857

Although many sequences and linear IgE epitopes of allergenic proteins have been identified and archived in databases, structural and physicochemical discriminators that define their specific properties are lacking. Current bioinformatics tools for predicting the potential allergenicity of a novel protein use methods that were not designed to compare peptides. Novel tools to determine the quantitative sequence and three-dimensional (3D) relationships between IgE epitopes of major allergens from peanut and other foods have been implemented in the Structural Database of Allergenic Proteins (SDAP; <http://fermi.utmb.edu/SDAP/>). These peptide comparison tools are based on five-dimensional physicochemical property (PCP) vectors. Sequences from SDAP proteins similar in their physicochemical properties to known epitopes of Ara h 1 and Ara h 2 were identified by calculating property distance (PD) values. A 3D model of Ara h 1 was generated to visualize the 3D structure and surface exposure of the epitope regions and peptides with a low PD value to them. Many sequences similar to the known epitopes were identified in related nut allergens, and others were within the sequences of Ara h 1 and Ara h 2. Some of the sequences with low PD values correspond to other known epitopes. Regions with low PD values to one another in Ara h 1 had similar predicted structure, on opposite sides of the internal dimer axis. The PD scale detected epitope pairs that are similar in structure and/or reactivity with patient IgE. The high immunogenicity and IgE reactivity of peanut allergen proteins might be due to the proteins' arrays of similar antigenic regions on opposite sides of a single protein structure.

KEYWORDS: Food allergy; Structural Database of Allergenic Proteins (SDAP); property distance (PD) scale; MPACK; GETAREA; peptide similarity index

INTRODUCTION

Considerable experimental and computational work has been dedicated to characterizing the sequences and substructures of food allergens that could account for cross-reactivity in sensitive individuals (1–5). This is necessary for two reasons. First, whereas the avoidance of foods related to those known to trigger symptoms may be necessary (6, 7) (but see also ref 8), eliminating all those that contain homologous allergenic proteins may result in an unnecessarily restricted diet (9, 10). Second, more accurate definition of the IgE reactive areas of these proteins can aid in identifying the allergenic potential of novel proteins and in selecting (11) or designing proteins with reduced allergenicity (2, 12–17) for specific immunotherapy.

In this paper we compare the structural and physical–chemical properties of epitopes of the major peanut allergens, Ara h 1 and Ara h 2 (18–20). Exposure to even minute amounts of peanuts can lead to an IgE-triggered cascade, ending in anaphylaxis and even death in allergic individuals (21–23). There are, with a few exceptions, no obvious patterns of

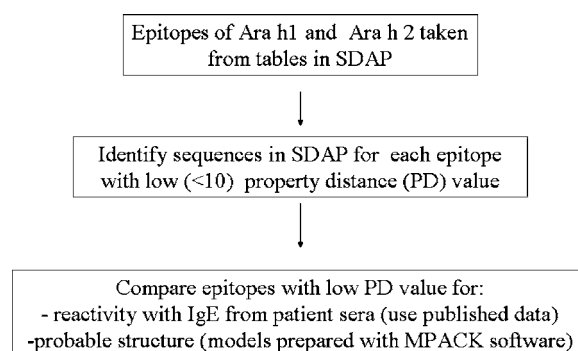


Figure 1. SDAP-based methodology used to identify sequences similar to known epitopes of the peanut allergens.

similarity at the amino acid level among the sequences that reacted with patient IgE identified in these proteins (20). Scanning mutagenesis of the IgE-reactive peptides revealed that changes in single amino acids could eliminate IgE binding (12, 18, 20). This and other experimental results suggest that we may be able to modify foods to contain proteins with slightly altered sequences that have a reduced tendency to induce allergy

* Corresponding author [telephone (409) 747-6843; fax (409) 747-6000; e-mail chschein@utmb.edu].

Table 1. Sequences in SDAP That Have PD Values of <10 to a Known Epitope from the Prodomain of Ara h 1^a

Allergen	Type/source	PD	z(PD,all)	Sequence
Ara h 1	Vicilin/peanut	0.00	8.3304	101 PRREEGGRWG 110
Ses i 3	Vicilin/sesame	6.81	5.4499	377 SRHEEGGIWP 386
Cor a 11	Vicilin/hazelnut	7.16	5.3011	435 KQQEEGGRGG 444
Amb a 2	Antigen k/ragweed	8.32	4.8116	76 GKATHGGKWG 85
Ber e 2	11S globulin/Brazil nut	9.13	4.4680	195 SQKQRGERYG 204
Gly m conglycinin	soybean	9.27	4.4113	583 PPKKEEGNKG 592
Gly m glycinin G1	soybean	9.30	4.3969	195 YQQEQGGHQS 204
Ses i 3	Vicilin/sesame	9.49	4.3171	575 PRQQQQGRAD 584
Gly m conglycinin	soybean	9.56	4.2876	617 PQQKEEGNKG 626
Gly m conglycinin	soybean	9.61	4.2662	358 FSRREGQQQG 367
Ara h 1	Vicilin/peanut	9.65	4.2495	172 PSRRFSTRYG 181
Sola t 1	Patatin/potato	9.73	4.2154	270 YTAEETAKWG 279
Jug n 2	Vicilin like/black walnut	9.75	4.2049	53 PRRESKGREE 62
Hev b 10.0101	rubber/latex	9.77	4.1984	112 PVREGGGELP 121
Hom s 1	tropomyosin/lobster	9.79	4.1913	13 RSRERGGERG 22
Ara h 4	Glycinin/peanut	9.81	4.1802	334 GRRGRGSRGG 343
Gly m conglycinin	soybean	9.90	4.1446	417 PQQKEEGSKG 426
Jug r 2	Vicilinlike/English walnut	9.91	4.1403	165 PRRESRGREE 174
Jug r 2	Vicilinlike/English walnut	9.94	4.1268	378 GRRSSGGPIS 387
Ses i 3	Vicilin/sesame	10.04	4.0858	162 KYREQQREG 171

^aOne sequence in glycinin G1 that is very similar to a known epitope of glycinin G2 (40) and another epitope found for Ara h 1 are in **bold**. Note that the epitope has a low PD value [and high z(PD) scores] to many sequences in the mature regions of other known plant vicilin allergens (see **Table 2** for the overall sequence similarity). The average value for the 829 "best fit" peptides for all SDAP allergens was 12.24 (SD = 1.58) and for all 190000 peptide windows in SDAP was 18.63 (SD = 2.36).

(24, 25). However, we must be sure in this case that all sequences that bind IgE have been identified. In a different approach, classification of the properties of sequences that have been identified as binding IgE in patient sera may enable us to predict similar regions that may also have a high propensity to be allergenic epitopes. Data on areas that bind IgEs from patient sera are available for only a small fraction of the allergens listed in SDAP. We are developing methods to compare the physical-chemical properties (PCPs) and structures of these peptides with sequences in other allergenic proteins. The hope is that eventually such methods will reveal the areas of allergenic proteins that are most likely to mediate their ability to function as allergens. We show here that some of the epitopes of Ara h 1 and Ara h 2 are similar to one another, on the basis of their PCPs as measured by a "property distance" (PD) value (26–28). We found other peptides that had low PD values to most of the epitopes within the same peanut allergen or other peanut allergenic proteins. Furthermore, epitopes and peptides similar to them were found on opposite sides of the internal dimer axis of the Ara h 1 protein. Major epitopes, that is, peptides that bound IgE from the sera of many peanut allergic patients, had several PCP homologues in their own sequence or in other peanut allergens. The results suggest that the high reactivity of peanuts may be their projection of several similar and dissimilar potential binding sites for IgE on the same protein. The PD scale provides a novel way to classify the sequences of epitopes that correlate with structural and physicochemical similarities that are not clear from the linear amino acid sequence alone.

METHODS

Calculation of PD Values. PD values measure the similarity of two peptides based on the PCPs of corresponding residues when they are aligned with each other. Each residue is represented by five numerical descriptors E1–E5, which were derived by multidimensional scaling of a set of 237 PCPs of the amino acid side chains (29). SDAP contains cross-referenced lists of epitopes, which are listed as part of the sequence details of the allergen. Here, peptides with low PD values to

a given IgE epitope were calculated automatically using the software tool provided on the Website of the structural database of allergenic proteins (SDAP) (26–28). The windows (protein regions) with the lowest PD values are returned as an output file on the SDAP Website (<http://fermi.utmb.edu/SDAP/>).

Significance Level of PD Values. Previous results have shown that sequences with PD values below 4 are quite similar, and peptides with scores up to 8–10 in homologous proteins typically have similar structure (27, 28). In addition, a z(PD) value, a determinant of significance of the PD value according to the average value of scores for the peptide in the database as a whole, is now automatically calculated in SDAP. PD searches in SDAP are followed by two histograms, one for the PD scores of the "best matching" (lowest PD value) peptide in the 829 sequences in SDAP and the second summarizing the scores for all windows (protein regions) of a given size in all of the SDAP allergen sequences. In general, the average value of lowest PD values is around 12, with a standard deviation (σ) of 2; thus, PD values below 8 are statistically significant (average value – 2 σ), and segments with PD values below 10 might be relevant if they are supported by additional evidence, such as significant solvent-accessible surface area or conformation similar to that of the search epitope. Depending on peptides, the average PD value, indicating essentially random matching, of the second histogram ranges from 17 to 26. When data from a PD search are presented in this paper, the average value of both histograms is given (**Tables 1** and **3**), as is the z(PD) score:

$$z(\text{PD}) = \frac{|\text{PD}_{\min} - \text{PD}_{\text{av}}|}{\text{SD}(\text{PD})}$$

The z(PD,all) score given in the tables in this paper is calculated from the complete PD distribution for each number. In contrast to the PD score itself, the higher the z(PD) score, the more significant the match between the peptides.

FASTA Search. A FASTA search method, implemented in SDAP, which complies with previously determined rules for determining the relationship of proteins to known allergens (30), was used to determine the overall sequence relationship of allergenic proteins (**Table 2**).

Peptide-Protein Comparison Tool. This new method in SDAP determines the PD value for two peptide sequences supplied by the

user. The shorter sequence (with fewer or equal number of residues) is used as the first sequence. The first sequence is moved along the second sequence one residue at a time, and PD values are reported for each window.

MPACK Models of Ara h 1. The MPACK modeling suite (31–33) was used to prepare a model of Ara h 1 (SwissProt sequence P43238; 626 residues, with residues 1–25 forming the signal peptide). Vicilins and related plant albumins (34) generally have a pro sequence of up to 180 amino acids, and structures are available only for the mature region that follows this. The template for homology modeling was canavalin from jack bean (PDB file 2CAV_A, resolution 2 Å), which is 47% identical (162/346) with most of Ara h 1 (172–586). At this high degree of identity, the structural model of Ara h 1 will be quite reliable (35).

As anticipated from the high degree of sequence similarity, the backbone root-mean-square deviation (RMSD) between the target and template was very low, 0.77 Å for the Ara h 1 model (from PROSUP, http://lore.came.sbg.ac.at:8080/CAME/CAME_EXTERN/PROSUP/index.html), indicating a very close structural match. The Ara h 1 model is deposited in SDAP (<http://fermi.utmb.edu/SDAP/index.html>). The GETAREA program (36) (http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl) was used to determine surface exposure of epitopes in the modeled structure.

Determining Major Epitopes. The linear epitopes of Ara h 1 and Ara h 2 were previously determined, in the groups of Drs. A. W. Burk and H. Sampson, using IgE in sera from patients with allergy to peanuts in combined molecular biology/clinical immunology studies. Data from papers that summarized this research (18, 20, 37) were used to determine which epitopes reacted with the most patient sera in protein dot spots.

RESULTS

Using the PD Scale To Find Sequences Similar to IgE Epitopes of Ara h 1 and Ara h 2. Starting from the 31 epitopes for Ara h 1 and Ara h 2 determined by other groups (18–20, 37) listed in SDAP, we used the automatic PD tool and analyzed the peptides with low PD values according to their predicted structures and IgE reactivities (Figure 1). A typical PD search starting from a major Ara h 1 epitope (Table 1) indicates that peptides with the lowest PD value come from allergens from other foods that may trigger response in peanut allergic individuals, such as tree nuts (38), soy (39, 40), and legumes (41, 42); these sequences also have the highest (i.e., most significant) z (PD) scores. The results are consistent to some extent with the overall similarity of these allergens according to expectation (E) values from an automatic FASTA search (Table 2; the lower the E value, the more significant the match). Although the IgE binding sites of most of these proteins have not been characterized, a peptide (bold in Table 1) from soybean Gly m glycinin G1, equivalent to epitope 5 of soybean G2 glycinin (40), has a low PD value. Note there are several other peptides in the sequences of both of the major allergens that have PD values below 10 to this epitope. This pattern of repeats of similar sequences was not seen for epitopes of other allergens, such as those from cedar pollen Jun a 3 (28) and Jun a 1 (43), which are less likely to cause anaphylaxis (44, 45).

Similarly, a PD search starting from a major epitope of Ara h 2 (Table 3) found another area in this protein that was a known epitope of Ara h 2. Parenthetically, a known epitope from the grass pollen allergen, Par j 1, is also similar to this epitope. Furthermore, a portion of this Ara h 2 epitope is also similar to a major epitope of Ara h 1 (Table 4).

Low PD Values Indicate Similar Structures. There are also sequences from unrelated allergens in Tables 1 and 3 that have low PD values to the test epitope. Thus, it is necessary to add other, structural criteria to determine whether a given match is a potential IgE epitope. We used our MPACK suite to prepare a reliable model structure of Ara h 1, based on a template with

Table 2. FASTA Searches Starting from the Three Major Peanut Allergens, Ara h 1, Ara h 2, and Ara h 3 (Top to Bottom) Reveal Interrelationships between These Allergens and Other Seed, Nut, and Legume Allergenic Proteins^a

allergen	source	sequence length	bit score	E score
Ara h 1	peanut/vicilin-like	614	642.0	1.7e–185
Gly m conglycinin	soybean	639	186.4	2.5e–48
Ses i 3	sesame/7S vicilin-like	585	92.8	3.6e–20
Jug r 2	English walnut/vicilin	593	90.9	1.3e–19
Cor a 11	hazelnut/7S vicilin	448	89.7	2.2e–19
Jug n 2	black walnut/vicilin-like	481	87.0	1.6e–18
Ara h 2	peanut/conglutin	156	146.1	2.2e–37
Ara h 6	peanut/conglutin	129	59.2	2.6e–11
Ara h 7	peanut/conglutin	160	55.5	4.1e–10
Ses i 2	sesame/2S albumin	148	35.8	3.1e–04
Ber e 1	Brazil nut/2S albumin	154	34.9	6.2e–04
Ses i 1	sesame/2S albumin	153	34.7	7.3e–04
Jug n 1	black walnut/2S albumin	161	33.2	2.1e–03
Jug r 1	English walnut/2S albumin	139	31.7	5.1e–03
Ric c 1	castor bean/2S albumin	258	32.2	6.9e–03
Ara h 3	peanut/glycinin	507	558.8	1.3e–160
Ara h 4	peanut/glycinin	530	506.9	5.9e–145
Gly m glycinin G1	soybean	495	186.4	1.7e–48
Gly m glycinin G2	soybean	485	183.3	1.3e–47
Cor a 9	hazelnut/11S globulin-like	515	144.8	5.8e–36
Fag e 1	buckwheat/13S globulin	565	98.4	5.8e–22
Ber e 2	Brazil nut/11S globulin	465	87.9	6.8e–19
Jug n 2	black walnut/vicilin-like	481	36.6	1.9e–03
Jug r 2	English walnut/vicilin	593	35.9	3.8e–03

^a The proteins most closely related to the top sequence have the lowest E score.

47% identity (see Methods), to demonstrate that the indicated physicochemical property similarity correlated with structural similarity.

Ara h 1 has two approximately equivalent structural regions (“internal dimer”) that differ in overall sequence, a property common to many plant allergens (4). As shown in Figure 2a, an epitope of Ara h 1, 539–548, can be related by PD score to a sequence that contains part of another known epitope, 299–308 (the identified epitope extends from 294 to 303; residues 301 and 302 were essential for IgE binding). As the figure shows, the two regions are similar in their surface-exposed residues and predicted structure and are located on symmetric positions on each side of this internal dimeric structure. Furthermore, two peptides in Ara h 1 that are similar to another major epitope (Table 4) are located on the opposite side of the internal dimer axis (Figure 2b). These symmetric peptides with similar structures would not be recognized just by comparing amino acid sequences alone, as is especially obvious when the sequences of the two epitopes in Figure 2a are compared.

Epitope Homologies in Other Peanut Allergens. For 24 of the 31 epitopes, PD searches revealed closely related sequences (PD value < 8–10) within the Ara h 1/Ara h 2 sequences or that of other peanut allergens, Ara h 3/4 and Ara h 6/7. In some cases, a single epitope had multiple related sequences (e.g., Table 4). Figure 3 summarizes the data for sequences with low PD value to the Ara h 1 and Ara h 2 epitopes in Venn diagram form. A table of the raw data is available as Supporting Information.

PD Searches Reveal a Repeat Pattern in the N-Terminal Propeptide Region of Ara h 1. In many plant vicilins, the N-terminal prodomain contains direct repeats of a “cysteine-rich” sequence. Although such a repeat pattern is not obvious in this region of Ara h 1, the PD search with epitopes from the N terminus detected many similar sequences in the pro region.

Table 3. SDAP Similarity Search, Beginning with a Major Epitope of Ara h 2, Finds Similar Sequences That Are Repeated in the Allergens from English Walnut (Jug r 2), Mustard (Sin a 1), Castor Bean (Ric c 1), and Brazil Nut (Ber e 1)^a

Allergen	Source	PD value	z(PD,all)	Sequence
Ara h 2	peanut	0.00	7.4613	27 DRRCSQLER 36
Jug r 2	English walnut	5.87	4.9586	140 QRCCQQRCEER 149
Sin a 1	yellow mustard	7.50	4.2616	92 KQQVRQQLEQ 101
Ric c 1	castor bean	7.98	4.0589	46 SQQCRQEVQR 55
Jug r 2	English walnut	8.00	4.0492	76 YECCQQQCEER 85
Jug r 2	English walnut	8.12	4.0004	49 DQRSQEERER 58
Ani s 3	nematode	8.21	3.9614	29 VRQMTDKLER 38
Jug r 2	English walnut	8.22	3.9549	101 QRRQQEERER 110
Api m 2	bee	8.27	3.9343	178 DQRVEQEAER 187
Hev b 6	latex	8.34	3.9062	51 DHNCQSNCKD 60
Jug r 2	English walnut	8.65	3.7754	532 QNNIINQLER 541
Fel d 2	dander	8.67	3.7656	310 KSHCISEVER 319
Jug r 2	English walnut	8.67	3.7655	127 QRRCCIIEQS 136
Bet v 4	birch pollen	8.75	3.7310	3 DDHPQDKAER 12
Ara h 2	known epitope	8.94	3.6510	122 GRQEQQFKR 131
Jug r 2	English walnut	8.98	3.6340	39 QIRCEERLEE 48
Ber e 1	brazil nut	9.05	3.6039	37 QEECREQMQR 46
Jug r 2	English walnut	9.13	3.5671	96 RRRCEQRRQQ 105
Jug r 2	English walnut	9.17	3.5523	144 QQRCEERQYKE 153
Par j 1	Grass pollen; known epitope	9.18	3.5460	154 DPAHKARLER 163

^a The list includes other known epitopes (**bold**). For this search, the average value for the 829 "best fit" peptides for all SDAP allergens was 11.27 (SD = 1.54) and for all 190309 peptide windows in SDAP was 17.5 (SD = 2.35).

Table 4. Major IgE Epitopes of Ara h 1 and Ara h 2 Have Several Homologues (PD Value of <10) in Their Own Sequence and Those of Other Peanut Allergens^a

Allergen	PD value	z(PD,all)	Sequence
Ara h 1	0.00	7.7515	134 EDWRRPSHQQ 143
Ara h 1	7.20	4.5811	127 EDWRQPREDW 136
Ara h 1	9.60	3.5263	102 DRRQPRREE 111
Ara h 1	9.64	3.5095	465 EQQQRGRRE 474
Ara h 3	8.48	4.0194	117 SQQQRDSHQK 126
Ara h 7	9.55	3.5470	59 EHMRRRVEQE 68
Ara h 1	0.00	7.8944	498 RRYTARLKEG 507
Ara h 3	8.07	4.3835	412 RVYDEELQEG 421
Ara h 4	8.07	4.3835	435 RVYDEELQEG 444
Ara h 1	8.60	4.1511	351 RRWSTRSSEN 360
Ara h 1	8.98	3.9859	252 NRKSFNLDEG 261
Ara h 2	9.66	3.6923	28 RRCQSQLERA 37*
Ara h 3	9.82	3.6224	124 HQKVHRFDEG 133
Ara h 2	0.00	7.0279	65 SQDPYSPS 72
Ara h 2	3.37	5.6387	58 ERDPYSPS 65
Ara h 1	6.30	4.4342	591 SQSPSSPE 598
Ara h 3	6.89	4.1924	199 PYSPYSPQ 206
Ara h 4	7.29	4.0268	222 PYSPYSPH 229

^a See also **Table 3** and **Figure 2**. Known epitopes or parts thereof are **bold**.

^b The asterisk indicates that this is the same epitope used to start the PD search of **Table 3**.

Some of the sequences with low PD values in this region also bind patient IgE (18, 20). For example, there are four different sequences in Ara h 1 with PD values of <10 to the 134 EDWRRPSHQQ 143 epitope, including two that contain parts of other known epitopes, and one in Ara h 3 (**Table 4**).

Major IgE Epitopes Have Similar Peptides in Their Own Sequence and Those of Other Peanut Allergens. We found that "major IgE epitopes", as defined by other groups, using spots analysis with individual or pooled sera, were also those with many PCP homologues in the peanut proteins. Only 7 of the 31 epitope sequences of Ara h 1 and Ara h 2 are truly unique

by our criteria, which is that there was no sequence in any of the peanut allergens with a PD score of <10 to them (Supporting Information Table E1). All of these sequences have relatively low reactivity with IgE, as determined from figures in the papers in which these epitopes are described (18, 20, 37). Three are in the first 78 amino acids of Ara h 1, which is cleaved during maturation in the peanut (46).

Conversely, the epitopes that had many PCP homologues in their own sequence and that of the other peanut allergens were among the peptides with the highest IgE reactivity. Examples of this (see **Table 4** for details) are as follows:

The 498–507 epitope of Ara h 1, considered to be a major epitope according to the number of sera that contained IgEs recognizing it [peptide 17 (18)], has two similar peptides in its own sequence, two in both Ara h 3 and Ara h 4, and is also similar to a major epitope of Ara h 2 (see **Table 3**). As **Figure 2b** shows, the peptides similar to this epitope in Ara h 1 are on the opposite side of the internal dimer axis; one is virtually identical with regard to predicted structure and location.

The two epitopes of Ara h 2 with low PD value to one another (Table 3) also have very similar, strong reactivity with IgE in protein dotspots (37).

The major Ara h 2 epitopes 59–64 and 65–72 are very similar to each other (PD values of 1.44 and 3.37) and to several sequences in Ara h 1, Ara h 3, and Ara h 4 (PD values between 5.09 and 7; **Table 4**).

DISCUSSION

In this work, we used the peptide and sequence comparison tools and information that are readily accessible in SDAP, combined with structural modeling, to detect relationships between IgE binding peptides of the major peanut allergen proteins, Ara h 1 and Ara h 2. The physicochemical similarities in epitopes indicated by low PD values correlated with predicted structure and surface exposure. Our method detects some similarities between known IgE binding sequences that could not be distinguished by amino acid similarity alone, as is especially clear when the sequences in the tables and the two epitopes in **Figure 2a** are compared. We also found many

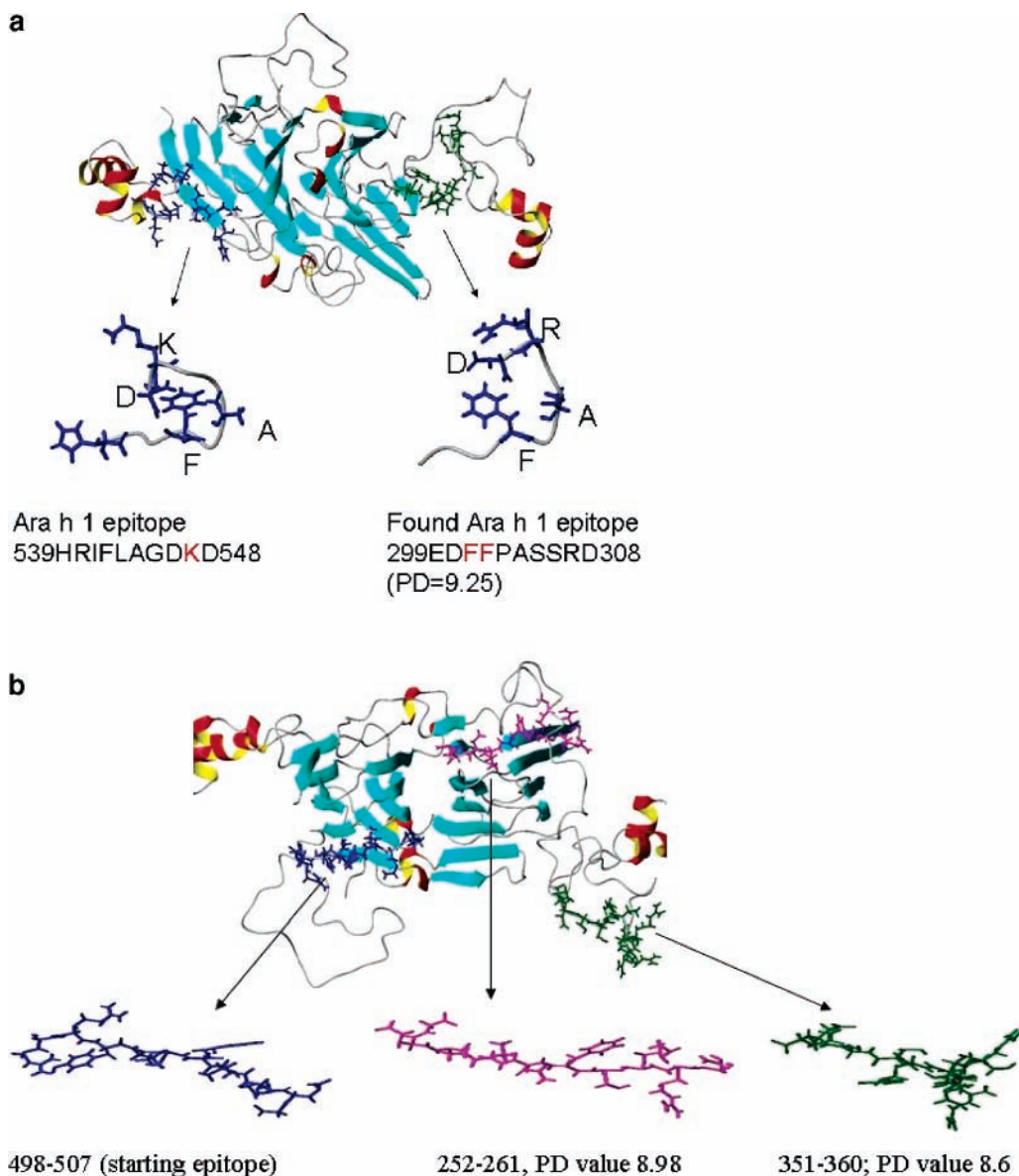


Figure 2. Epitopes of Ara h 1 and sequences with low PD value to a major epitope have similar structures and surface exposure. (a) The Ara h 1 protein has two domains of nearly identical structure but different sequence. The epitope (blue) has a low PD value to a sequence containing part of another known epitope (green) that is on the opposite side of the protein in a comparable location. The two regions have nearly identical predicted structure, with similar exposure of the large hydrophobic residues (only the surface-exposed residues are shown), despite the quite diverse amino acid sequences. (b) Two regions with a low PD value to a major epitope (blue) of Ara h 1 (see Table 4 for sequence details) are structurally similar. One (252–261, purple) is a mirror image on the opposite side of the internal dimer axis.

peptides similar to those that bind IgE in Ara h 1 and Ara h 2 in other peanut allergens, but do not mean to suggest that the same IgE would bind to all of these sequences in one individual. Rather, we suggest that peanut allergenicity may be a property of repeated sequences with similar physical–chemical properties in individual proteins and in the ensemble of albumin and seed storage proteins. This could explain why the IgE response to Ara h 1, 2, 3 and 6 develops simultaneously during induction of peanut allergy in a rodent model (47) and also why very similar proteins in other foods might not provoke the same response as peanuts do (48).

Developing New Ways To Compare Epitopes. Although an increasing number of allergens and linear IgE epitopes of allergens have been identified and archived in databases, the structural characteristics that define IgE epitopes are to a large extent unknown. Regulatory agencies for food safety and biotechnology companies producing recombinant protein prod-

ucts have sought to define what makes a protein an allergen (30). Among the problems in defining this are that most of the data is collected for binding of IgE in patient sera, which are polyclonal in nature. Another problem is that dotspots can detect only the strongest of binding phenomena, missing weaker interactions that may also mediate receptor binding in vivo. Yet another problem is the length of peptides used in the spot, and the possibility that the conformation on the spot may have little relationship to that in the intact protein. Currently, the approved guidelines for predicting the potential allergenicity of a novel protein product are limited to assessing short stretches of sequence identity or homology of longer stretches using methods such as FASTA, which are designed to detect global similarities between proteins in large databases. More recently, methods that use other criteria to quantify relationships between allergens have been suggested by us and other groups (5, 14, 26, 27, 44, 49–51). As we show here, our PD search, coupled with

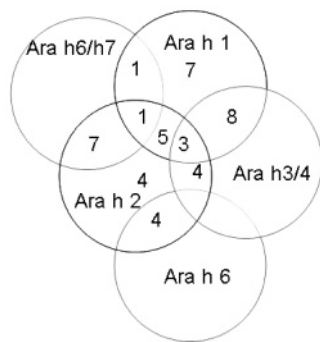


Figure 3. Venn diagram illustration of sequences in peanut allergen proteins that have a low PD value to epitopes identified in Ara h 1 and Ara h 2. Numbers in the intersections of the circles are epitopes that have peptides with low PD values in two, three, or four proteins. The numbers in the Ara h 1 and Ara h 2 circles in the nonoverlapping areas refer to amino acid stretches in their sequences that are similar to their own epitopes. Ara h 3 and Ara h 4 grouped together as the sequences with low PD value to the epitope are, for the most part, either similar or identical, consistent with the high similarity of both of these proteins to each other and to trypsin inhibitors (66). Ara h 6 and Ara h 7 also have significant overall sequence similarity to each other and to Ara h 2. Overall relationships between the peanut allergens indicated by a FASTA search (using the SDAP tool) are summarized in **Table 2**. The data used to derive this diagram are given as Supporting Information (Table E1) with selected results in **Table 4**.

structural analysis, found meaningful similarities between known epitopes of the peanut proteins. Ongoing experimental work should aid in determining whether the method can efficiently recognize potential IgE binding segments.

Correlation of PD Values with Structure. Whereas it would be expected that homologous proteins have areas of related sequences and structures that might correspond to known epitopes (28, 52), we did not anticipate the structural homologues we found in different but related domains of the same protein, Ara h 1. Previous efforts to compare these epitopes that depended solely on amino acid conservation (20) did not detect the pairs of epitopes that were on symmetric sides of the Ara h 1 molecule, such as those shown in **Figure 2a**. Furthermore, PD values (**Tables 1, 3, and 4; Figures 2 and 3**) show that most of the epitope sequences could be related to other sequences in the same protein or other peanut allergens (Supporting Information Table E1).

Our data suggest that FASTA searches (**Table 2**) can miss important relationships between immunogenic peptides. For example, Ara h 1 and Ara h 2 are classified in two different groups by FASTA searches, but their major epitopes have similar PCPs, according to our analysis (**Table 4; Supporting Information Table E1**). Although we do not suggest that this means that the same IgEs recognize both proteins, we do suggest that both proteins project similar areas that may be ideal for IgE binding. In support of this, a large-scale (77 patients) study showed that 60% of peanut allergic individuals reacted to both proteins (53); only 5–6% of the sera were specific for Ara h 1 or Ara h 2. We are now testing other methods to correlate these surface properties with IgE binding capability.

Role of Multivalency in Stimulating the Allergic Response. Multivalency (i.e., the presence of many similar and dissimilar IgE epitopes on the same protein) within the major peanut allergen proteins, echoed in other storage proteins of the peanut, may contribute to the potency of peanuts to induce an anaphylactic response in sensitive individuals. The similar peptides identified by the PD searches, even if they bind IgE more weakly

than the originally identified epitopes, could augment the ability of the protein to bind multiple IgE molecules on the surface of cells. One possible reason for these similar sequences not being detected by serum IgE is that the peanut allergens occur as multimers. It is possible that one of the domains of Ara h 1 is more involved in forming such multimers than the other, which could affect the relative surface exposure of the similar sequences. Indeed, the symmetric peptides of **Figure 2** are consistent with previous hypotheses of how an ideal, large, multivalent allergen would induce receptor aggregation on basophils (54–59). Furthermore, recent experimental results demonstrated that patient sera that contained IgE reacting with many different peptides from the sequences of the three major peanut allergens were better able to stimulate IgE mediated lysis (53).

Role of the Prodomain of Vicilins in Allergenicity. These results and those from peptide IgE immunodots suggest that the prodomain of Ara h 1 is a potent mediator of the allergenic response. The basic repeat nature of the Ara h 1 sequence was detected by the low PD scores of prodomain peptides to known epitopes (e.g., **Table 4**, top). It is tempting to hypothesize that generation of highly allergenic, soluble fragments of this region, either during normal maturation or induced by processing (60–62) might also enhance the IgE reactivity of peanuts. Judging from the IgE reactivity of patient sera, we [and others (46)] assume that some proteolytic fragment(s) of the prodomain must remain in the whole peanut. Prodomains of mammalian proteins can often be detected as independent species, whereby the best studied example would be the “C-peptide” of insulin (63). Mammalian prodomains also have regulatory roles, such as the propeptide of IL-1 α , which can induce apoptosis (64), and those of some proteases inhibit the activity of the mature proteins (65).

In conclusion, we have shown that the PD scale efficiently identifies structurally related areas in allergens. By our criteria, peanut allergens contain many repeats of sequences with similar physicochemical properties on their surfaces. Using published data, we could demonstrate that the major epitopes have several PCP homologues in their own or other peanut protein sequences.

We conclude that peanuts may be strong triggers of allergic symptoms for sensitive individuals because they contain several proteins with multiple similar regions that could bind IgE or induce cross-reactivity. However, we should note that our conclusions are based on the available IgE binding data, which have been gathered for only a small fraction of known food allergens. Furthermore, much of the data used pooled serum samples, which can further cloud the question of what part of the peptides is involved in the actual response of one individual. As more information becomes available about the specific IgE binding sites of allergens and about the degree of cross-reactivity between closely related proteins (48), we should be able to define properties of major epitopes in a more computational fashion. This would aid in predicting the allergenic potential of novel proteins and possibly in directing the design of hypoallergenic foods (12, 13, 15).

Supporting Information Available: Table (E1) listing selected results of the PD searches individually for the 31 epitopes. Some of these data are excerpted in Tables 1, 3, and 4. This material is available free of charge via the Internet at <http://pubs.acs.org>.

LITERATURE CITED

- (1) Fedorov, A. A.; Ball, T.; Mahoney, N. M.; Valenta, R.; Almo, S. C. The molecular basis for allergen cross-reactivity: crystal structure and IgE epitope mapping of birch pollen profilin. *Structure* **1997**, *5*, 33–45.
- (2) Ferreira, F.; Ebner, C.; Kramer, B.; Casari, G.; Briza, P.; Grimm, R.; Jahn-Schmid, B.; Breiteneder, H.; Kraft, D.; Breitenbach, M.; Rheinberger, H. J.; Scheiner, O. Modulation of IgE reactivity of allergens by site-directed mutagenesis: potential use of hypoallergenic variants for immunotherapy. *FASEB J.* **1998**, *12*, 231–242.
- (3) Brusic, V.; Petrovsky, N. Bioinformatics for characterisation of allergens, allergenicity and allergic crossreactivity. *Trends Immunol.* **2003**, *24*, 225–228.
- (4) Breiteneder, H.; Clare Mills, E. N. Molecular properties of food allergens. *J. Allergy Clin. Immunol.* **2005**, *115*, 14–23.
- (5) Jenkins, J.; Griffiths-Jones, S.; Shewry, P.; Breiteneder, H.; Clare Mills, E. N. Structural Relatedness of plant food allergens with specific reference to cross-reactive allergens: an *in silico* analysis. *J. Allergy Clin. Immunol.* **2005**, *115*, 163–170.
- (6) Sampson, H. A. Food allergy. Part 2: diagnosis and management. *J. Allergy Clin. Immunol.* **1999**, *103*, 981–989.
- (7) Perry, T. T.; Conover-Walker, M. K.; Pomes, A.; Chapman, M. D.; Wood, R. A. Distribution of peanut allergen in the environment. *J. Allergy Clin. Immunol.* **2004**, *113*, 973–976.
- (8) Fleischer, D.; Conover-Walker, M.; Christie, L.; Burks, A.; Wood, R. Peanut allergy: recurrence and its management. *J. Allergy Clin. Immunol.* **2004**, *114*, 1195–201.
- (9) Sicherer, S. H. Clinical implications of cross-reactive food allergens. *J. Allergy Clin. Immunol.* **2001**, *108*, 881–890.
- (10) Sampson, H. Food allergy: When mucosal immunity goes wrong. *J. Allergy Clin. Immunol.* **2005**, *115*, 139–41.
- (11) Pons, L.; Ponnappan, U.; Hall, R. A.; Simpson, P.; Cockrell, G.; West, C. M.; Sampson, H. A.; Helm, R. M.; Burks, A. W. Soy immunotherapy for peanut-allergic mice: modulation of the peanut-allergic response. *J. Allergy Clin. Immunol.* **2004**, *114*, 915–921.
- (12) Rabjohn, P.; West, C.; Connaughton, C.; Sampson, H.; Helm, R.; Burks, A.; Bannon, G. Modification of peanut allergen Ara h 3: effects on IgE binding and T cell stimulation. *Int. Arch. Allergy Immunol.* **2002**, *128*, 15–23.
- (13) Bannon, G.; Cockrell, G.; Connaughton, C.; West, C. M.; Helm, R.; Stanley, J. S.; King, N.; Rabjohn, P.; Sampson, H. A.; Burks, A. W. Engineering, characterization and *in vitro* efficacy of the major peanut allergens for use in immunotherapy. *Int. Arch. Allergy Immunol.* **2001**, *124*, 70–72.
- (14) Buhot, C.; Chenal, A.; Sanson, A.; Pouvelle-Moratille, S.; Gelb, M. H.; Menez, A.; Gillet, D.; Maillere, B. Alteration of the tertiary structure of the major bee venom allergen Api m 1 by multiple mutations is concomitant with low IgE reactivity. *Protein Sci.* **2004**, *13*, 2970–2978.
- (15) Dodo, H.; Konan, K.; Viquez, O. A genetic engineering strategy to eliminate peanut allergy. *Curr. Allergy Asthma Rep.* **2005**, *5*, 67–73.
- (16) Larsen, J. N.; Sparholt, S. H.; Ipsen, H. Epitope structure of recombinant isoallergens of Bet v 1. *Adv. Exp. Med. Biol.* **1996**, *409*, 243–249.
- (17) Smith, A. M.; Chapman, M. D. Reduction in IgE binding to allergen variants generated by site-directed mutagenesis: contribution of disulfide bonds to the antigenic structure of the major house dust mite allergen Der p 2. *Mol. Immunol.* **1996**, *33*, 399–405.
- (18) Burks, A. W.; Shin, D.; Cockrell, G.; Stanley, J. S.; Helm, R. M.; Bannon, G. A. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur. J. Biochem.* **1997**, *245*, 334–339.
- (19) Rabjohn, P.; Helm, E. M.; Stanley, J. S.; West, C. M.; Sampson, H. A.; Burks, A. W.; Bannon, G. A. Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *J. Clin. Invest.* **1999**, *103*, 535–542.
- (20) Shin, D. S.; Compadre, C. M.; Maleki, S. J.; Kopper, R. A.; Sampson, H.; Huang, S. K.; Burks, A. W.; Bannon, G. A. Biochemical and structural analysis of the IgE binding sites on Ara h 1, an abundant and highly allergenic peanut protein. *J. Biol. Chem.* **1998**, *273*, 13753–13759.
- (21) Pomes, A.; Helm, R. M.; Bannon, G. A.; Burks, A. W.; Tsay, A.; Chapman, M. D. Monitoring peanut allergen in food products by measuring Ara h 1. *J. Allergy Clin. Immunol.* **2003**, *111*, 640–645.
- (22) Sicherer, S.; Munoz-Furlong, A.; Sampson, H. A. Prevalence of peanut and tree nut allergy in the United States determined by means of a random digit dial telephone survey: a 5-year follow-up study. *J. Allergy Clin. Immunol.* **2003**, *112*, 1203–7.
- (23) Bock, S.; Munoz-Furlong, A.; Sampson, H. A. Fatalities due to anaphylactic reactions to foods. *J. Allergy Clin. Immunol.* **2001**, *107*, 191–193.
- (24) Beardslee, T. A.; Zeece, M. G.; Sarath, G.; Markwell, J. P. Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3. *Int. Arch. Allergy Immunol.* **2000**, *123*, 299–307.
- (25) Li, X.; Srivastava, K.; Grishin, A.; Huang, C. K.; Schofield, B.; Burks, W.; Sampson, H. A. Persistent protective effect of heat-killed *Escherichia coli* producing “engineered,” recombinant peanut proteins in a murine model of peanut allergy. *J. Allergy Clin. Immunol.* **2003**, *112*, 159–67.
- (26) Ivanciuc, O.; Schein, C. H.; Braun, W. Data mining of sequences and 3D structures of allergenic proteins. *Bioinformatics* **2002**, *18*, 1358–1364.
- (27) Ivanciuc, O.; Schein, C. H.; Braun, W. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res.* **2003**, *31*, 359–362.
- (28) Ivanciuc, O.; Mathura, V.; Midoro-Horiuti, T.; Braun, W.; Goldblum, R. M.; Schein, C. H. Detecting potential IgE-reactive sites on food proteins using a sequence and structure database, SDAP-Food. *J. Agric. Food Chem.* **2003**, *51*, 4830–4837.
- (29) Venkatarajan, M. S.; Braun, W. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. *J. Mol. Model.* **2001**, *7*, 445–453.
- (30) WHO. *Evaluation of Allergenicity of Genetically Modified Foods*; report of a joint FAO/WHO expert consultation; World Health Organization: Geneva, Switzerland, 2001.
- (31) Soman, K. V.; Schein, C. H.; Zhu, H.; Braun, W. Homology modeling and simulations of nuclease structures. In *Nuclease Methods and Protocols, Methods in Molecular Biology*; Schein, C. H., Ed.; Humana Press: Totowa, NJ, 2001; pp 263–286.
- (32) Ivanciuc, O.; Oezguen, N.; Mathura, V.; Schein, C. H.; Xu, Y.; Braun, W. Using property based sequence motifs and 3D modeling to determine structure and functional regions in CASP5 targets. *Curr. Med. Chem.* **2004**, *11*, 583–593.
- (33) Murtazina, D.; Puchkaev, A. V.; Schein, C. H.; Oezguen, N.; Braun, W.; Nanavati, A.; Pikuleva, I. A. Membrane protein interactions contribute to efficient 27-hydroxylation of cholesterol by mitochondrial cytochrome P450 27A1. *J. Biol. Chem.* **2002**, *277*, 37582–37589.
- (34) Mills, E. N.; Jenkins, J.; Marigheto, N.; Belton, P. S.; Gunning, A. P.; Morris, V. J. Allergens of the cupin superfamily. *Biochem. Soc. Trans.* **2002**, *30*, 925–929.
- (35) Abagyan, R. A.; Batalov, S. Do aligned sequences share the same fold? *J. Mol. Biol.* **1997**, *273*, 355–368.
- (36) Fraczkiwicz, R.; Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **1998**, *19*, 319–333.

- (37) Stanley, J. S.; King, N.; Burks, A. W.; Huang, S. K.; Sampson, H.; Cockrell, G.; Helm, R. M.; West, C. M.; Bannon, G. A. Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. *Arch. Biochem. Biophys.* **1997**, *342*, 244–253.
- (38) de Leon, M. P.; Glaspole, I. N.; Drew, A. C.; Rolland, J. M.; O'Hehir, R. E.; Suphioglu, C. Immunological analysis of allergenic cross-reactivity between peanut and tree nuts. *Clin. Exp. Allergy* **2003**, *33*, 1273–1280.
- (39) Eigenmann, P. A.; Burks, A. W.; Bannon, G. A.; Sampson, H. A. Identification of unique peanut and soy allergens in sera adsorbed with cross-reacting antibodies. *J. Allergy Clin. Immunol.* **1996**, *98*, 969–978.
- (40) Helm, R. M.; Cockrell, G.; Connaughton, C.; Sampson, H. A.; Bannon, G. A.; Beilinson, N. C.; Burks, A. W. A soybean G2 Glycinin allergen. *Int. Arch. Allergy Immunol.* **2000**, *123*, 213–219.
- (41) Lopez-Torres, G.; Salcedo, G.; Martin-Esteban, M.; Diaz-Perales, A.; Pascual, C. Y.; Sanchez-Monge, R. Len c 1, a major allergen and vicilin from lentil seeds: protein isolation and cDNA cloning. *J. Allergy Clin. Immunol.* **2003**, *112*, 1208–1215.
- (42) Wensing, M.; Knulst, A. C.; Piersma, S.; O'Kane, F.; Knol, E. F.; Koppelman, S. J. Patients with anaphylaxis to pea can have peanut allergy caused by cross-reactive IgE to vicilin (Ara h 1). *J. Allergy Clin. Immunol.* **2003**, *111*, 420–424.
- (43) Midoro-Horiuti, T.; Mathura, V. S.; Schein, C. H.; Braun, W.; Yu, S.; Watanabe, M.; Lee, J. C.; Brooks, E. G.; Goldblum, R. M. Major linear IgE epitopes of mountain cedar pollen allergen Jun a 1 map to the pectate lyase catalytic site. *Mol. Immunol.* **2003**, *40*, 555–562.
- (44) Soman, K. V.; Midoro-Horiuti, T.; Ferreon, J. C.; Goldblum, R. M.; Brooks, E. G.; Kurosky, A.; Braun, W.; Schein, C. H. Homology modeling and characterization of IgE epitopes of mountain cedar allergen Jun a 3. *Biophys. J.* **2000**, *79*, 1601–1609.
- (45) Midoro-Horiuti, T.; Goldblum, R. N.; Kurosky, A.; Wood, T. G.; Schein, C. H.; Brooks, E. G. Molecular cloning of the mountain cedar (*Juniperus ashei*) pollen major allergen, Jun a 1. *J. Allergy Clin. Immunol.* **1999**, *104*, 613–617.
- (46) Wichers, H. J.; De Beijer, T.; Savelkoul, H. F.; Van Amerongen, A. The major peanut allergen Ara h 1 and its cleaved-off N-terminal peptide; possible implications for peanut allergen detection. *J. Agric. Food Chem.* **2004**, *52*, 4903–4907.
- (47) van Wijk, F.; Hartgring, S.; Koppelman, S. J.; Pieters, R.; Knippels, L. M. Mixed antibody and T cell responses to peanut and the peanut allergens Ara h 1, Ara h 2, Ara h 3 and Ara h 6 in an oral sensitization model. *Clin. Exp. Allergy.* **2004**, *34*, 1422–1428.
- (48) Astwood, J. D.; Silvanovich, A.; Bannon, G. A. Vicilins: a case study in allergen pedigrees. *J. Allergy Clin. Immunol.* **2002**, *110*, 26–27.
- (49) Breiteneder, H.; Ebner, C. Molecular and biochemical classification of plant-derived food allergens. *J. Allergy Clin. Immunol.* **2000**, *106*, 27–36.
- (50) Ganglberger, E.; Gruenberger, K.; Sponer, B.; Radauer, C.; Breiteneder, H.; Boltz-Nitulescu, G.; Schreiner, O.; Jensen-Jarolim, E. Allergen mimotopes for 3-dimensional epitope search and induction of antibodies inhibiting human IgE. *FASEB J.* **2000**, *14*.
- (51) Huby, R. D. J.; Dearman, R. J.; Kimber, I. Why are some proteins allergens? *Toxicol. Sci.* **2000**, *55*, 235–246.
- (52) Xiang, P.; Beardslee, T. A.; Zeece, M. G.; Markwell, J.; Sarath, G. Identification and analysis of a conserved immunoglobulin E-binding epitope in soybean G1a and G2a and peanut Ara h 3 glycinins. *Arch. Biochem. Biophys.* **2002**, *408*, 51–57.
- (53) Shreffler, W. G.; Beyer, K.; Chu, T. H.; Burks, A. W.; Sampson, H. A. Microarray immunoassay: association of clinical history, in vitro IgE function, and heterogeneity of allergenic peanut epitopes. *J. Allergy Clin. Immunol.* **2004**, *113*, 776–782.
- (54) Holowka, D.; Baird, B. Antigen-mediated IGE receptor aggregation. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 79–112.
- (55) Dombrowicz, D.; Brini, A. T.; Flamand, V.; Hicks, E.; Snouwaert, J. N.; Kinet, J. P.; Koller, B. H. Anaphylaxis mediated through a humanized high affinity IgE receptor. *J. Immunol.* **1996**, *157*, 1645–1651.
- (56) Dombrowicz, D.; Lin, S.; Flamand, V.; Brini, A. T.; Koller, B. H.; Kinet, J. P. Allergy-associated FcRbeta is a molecular amplifier of IgE- and IgG-mediated in vivo responses. *Immunity* **1998**, *8*, 517–529.
- (57) Kinet, J. P. The high-affinity IgE receptor (Fc epsilon RI): from physiology to pathology. *Annu. Rev. Immunol.* **1999**, *17*, 931–972.
- (58) Garman, S. C.; Wurzburg, B. A.; Tarchevskaya, S. S.; Kinet, J. P.; Jardetzky, T. S. Structure of the Fc fragment of human IgE bound to its high-affinity receptor Fc epsilon RIalpha. *Nature* **2000**, *406*, 259–266.
- (59) Garman, S. C.; Kinet, J. P.; Jardetzky, T. S. The crystal structure of the human high-affinity IgE receptor (Fc epsilon RI alpha). *Annu. Rev. Immunol.* **1999**, *17*, 973–976.
- (60) Chung, S. Y.; Butts, C. L.; Maleki, S. J.; Champagne, E. T. Linking peanut allergenicity to the processes of maturation, curing and roasting. *J. Agric. Food Chem.* **2003**, *51*, 4273–4277.
- (61) Maleki, S. J.; Hurlburt, B. K. Structural and functional alterations in major peanut allergens caused by thermal processing. *J. AOAC Int.* **2004**, *87*, 1475–1479.
- (62) Chung, S. Y.; Maleki, S. J.; Champagne, E. T. Allergenic properties of roasted peanut allergens may be reduced by peroxidase. *J. Agric. Food Chem.* **2004**, *52*, 4541–4545.
- (63) Steiner, D. F.; Rubenstein, A. H. Proinsulin C-peptide—biological activity? *Science* **1997**, *277*, 531–532.
- (64) Pollock, A. S.; Turck, J.; Lovett, D. H. The prodomain of interleukin 1alpha interacts with elements of the RNA processing apparatus and induces apoptosis in malignant cells. *FASEB J.* **2003**, *17*, 203–213.
- (65) Groves, M. R.; Taylor, M. A.; Scott, M.; Cummings, N. J.; Pickersgill, R. W.; Jenkins, J. A. The prosequence of procaricain forms an alpha-helical domain that prevents access to the substrate-binding cleft. *Structure* **1996**, *4*, 1193–1203.
- (66) Dodo, H. W.; Viquez, O. M.; Maleki, S. J.; Konan, K. N. 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*, 1404–1409.

Received for review May 18, 2005. Revised manuscript received July 25, 2005. Accepted August 18, 2005. This work was supported by a grant from the U.S. Food and Drug Administration (FD-U-002249-01), by an Advanced Technology Program Grant—Texas Higher Education Coordinating Board (004952-0036-2003), and by the Sealy Center for Structural Biology.

JF051148A