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

CATHERINE H. S CIEH,* 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 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].
(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 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 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 — 2σ), 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:

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

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 epitopes.

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).

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

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

<table>
<thead>
<tr>
<th>Allergen</th>
<th>Type/source</th>
<th>PD</th>
<th>z(PD,all)</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ara h 1</td>
<td>Viculn/peanut</td>
<td>0.00</td>
<td>8.3304</td>
<td>101 PRREEGGREGWG</td>
</tr>
<tr>
<td>Ses i 3</td>
<td>Vicilin/sesame</td>
<td>6.81</td>
<td>5.4499</td>
<td>377 SRRREGGIGW</td>
</tr>
<tr>
<td>Cor a 11</td>
<td>Vicilin/hazelnut</td>
<td>7.16</td>
<td>5.3011</td>
<td>435 KOOQKROGKR</td>
</tr>
<tr>
<td>Amb a 2</td>
<td>Antigen K/ragweed</td>
<td>8.32</td>
<td>4.8116</td>
<td>76 GKKTHRGKG</td>
</tr>
<tr>
<td>Ber e 2</td>
<td>11S globulin/Brazil nut</td>
<td>9.13</td>
<td>4.4680</td>
<td>195 SQKVRGREGY</td>
</tr>
<tr>
<td>Gly m conglycinin</td>
<td>soybean</td>
<td>9.27</td>
<td>4.4113</td>
<td>583 KEPQBERGKG</td>
</tr>
<tr>
<td>Gly m glycinen G1</td>
<td>soybean</td>
<td>9.30</td>
<td>4.3969</td>
<td>195 YQQQEQGGQG</td>
</tr>
<tr>
<td>Ses i 3</td>
<td>Vicilin/sesame</td>
<td>9.49</td>
<td>4.3171</td>
<td>575 PQQQQGRG</td>
</tr>
<tr>
<td>Gly m conglycinin</td>
<td>soybean</td>
<td>9.56</td>
<td>4.2876</td>
<td>617 PQKQERGKG</td>
</tr>
<tr>
<td>Gly m conglycinin</td>
<td>soybean</td>
<td>9.61</td>
<td>4.2682</td>
<td>358 FSKERQGQ</td>
</tr>
<tr>
<td>Ara h 1</td>
<td>Vicilin/peanut</td>
<td>9.65</td>
<td>4.2495</td>
<td>172 PVRPFRYNG</td>
</tr>
<tr>
<td>Sola t 1</td>
<td>Patatin/potato</td>
<td>9.73</td>
<td>4.2154</td>
<td>270 YTAEBATK</td>
</tr>
<tr>
<td>Jug n 2</td>
<td>Vicilin like/black walnut</td>
<td>9.75</td>
<td>4.2049</td>
<td>53 PRSEQKREE</td>
</tr>
<tr>
<td>Hev b 10.0101</td>
<td>rubber/latex</td>
<td>9.77</td>
<td>4.1984</td>
<td>112 QVRKGRG</td>
</tr>
<tr>
<td>Hom s 1</td>
<td>tropomysosin/lobster</td>
<td>9.79</td>
<td>4.1913</td>
<td>13 RSRREGERG</td>
</tr>
<tr>
<td>Ara h 4</td>
<td>Glycinin/peanut</td>
<td>9.81</td>
<td>4.1802</td>
<td>334 GRGRGSKRG</td>
</tr>
<tr>
<td>Gly m conglycinin</td>
<td>soybean</td>
<td>9.90</td>
<td>4.1446</td>
<td>417 PQKQERGKG</td>
</tr>
<tr>
<td>Jug r 2</td>
<td>Vicilinlike/English walnut</td>
<td>9.91</td>
<td>4.1403</td>
<td>165 PRRSEQKEK</td>
</tr>
<tr>
<td>Jug r 2</td>
<td>Vicilinlike/English walnut</td>
<td>9.94</td>
<td>4.1268</td>
<td>378 GRRGSOGP</td>
</tr>
<tr>
<td>Ses i 3</td>
<td>Vicilin/sesame</td>
<td>10.04</td>
<td>4.0858</td>
<td>162 KVRQGRG</td>
</tr>
</tbody>
</table>

\(^{a}\) One sequence in glycinen G1 that is very similar to a known epitope of glycinen 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).
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 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.scbi.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 47% identity (see Methods), to demonstrate that the indicated physicochemical property similarity corresponded 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.

### 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

<table>
<thead>
<tr>
<th>allergen</th>
<th>source</th>
<th>sequence length</th>
<th>bit score</th>
<th>E score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ara h 1</td>
<td>peanut/vicilin-like</td>
<td>614</td>
<td>642.0</td>
<td>1.7e–185</td>
</tr>
<tr>
<td>Gly m conglycinin</td>
<td>soybean</td>
<td>639</td>
<td>186.4</td>
<td>2.5e–48</td>
</tr>
<tr>
<td>Ses i 1</td>
<td>sesame7s vicilin-like</td>
<td>585</td>
<td>92.8</td>
<td>3.6e–20</td>
</tr>
<tr>
<td>Jug r 2</td>
<td>English walnut/vicilin</td>
<td>593</td>
<td>90.9</td>
<td>1.3e–19</td>
</tr>
<tr>
<td>Cor a 11</td>
<td>hazelnut7s vicilin</td>
<td>448</td>
<td>89.7</td>
<td>2.2e–18</td>
</tr>
<tr>
<td>Jug n 2</td>
<td>black walnut/vicilin-like</td>
<td>481</td>
<td>87.0</td>
<td>1.6e–18</td>
</tr>
<tr>
<td>Ara h 2</td>
<td>peanut/conglutin</td>
<td>156</td>
<td>146.1</td>
<td>2.2e–37</td>
</tr>
<tr>
<td>Ara h 6</td>
<td>peanut/conglutin</td>
<td>129</td>
<td>59.2</td>
<td>2.6e–11</td>
</tr>
<tr>
<td>Ara h 7</td>
<td>peanut/conglutin</td>
<td>160</td>
<td>55.5</td>
<td>4.1e–10</td>
</tr>
<tr>
<td>Ses i 2</td>
<td>sesame2s albumin</td>
<td>148</td>
<td>35.8</td>
<td>3.1e–04</td>
</tr>
<tr>
<td>Ber e 1</td>
<td>Brazil nut2s albumin</td>
<td>154</td>
<td>34.9</td>
<td>6.2e–04</td>
</tr>
<tr>
<td>Ses i 1</td>
<td>sesame2s albumin</td>
<td>153</td>
<td>34.7</td>
<td>7.5e–04</td>
</tr>
<tr>
<td>Jug n 1</td>
<td>black walnut2s albumin</td>
<td>161</td>
<td>33.2</td>
<td>2.1e–03</td>
</tr>
<tr>
<td>Jug r 1</td>
<td>English walnut2s albumin</td>
<td>139</td>
<td>31.7</td>
<td>5.1e–03</td>
</tr>
<tr>
<td>Ric c 1</td>
<td>castor bean2s albumin</td>
<td>258</td>
<td>32.2</td>
<td>6.9e–03</td>
</tr>
<tr>
<td>Ara h 3</td>
<td>peanut/glycinin</td>
<td>507</td>
<td>558.8</td>
<td>1.3e–160</td>
</tr>
<tr>
<td>Ara h 4</td>
<td>peanut/glycinin</td>
<td>530</td>
<td>506.9</td>
<td>5.9e–145</td>
</tr>
<tr>
<td>Gly m glycinin G1</td>
<td>soybean</td>
<td>495</td>
<td>186.4</td>
<td>1.7e–48</td>
</tr>
<tr>
<td>Gly m glycinin G2</td>
<td>soybean</td>
<td>485</td>
<td>183.3</td>
<td>1.3e–47</td>
</tr>
<tr>
<td>Cor a 9</td>
<td>hazelnut11s globulin-like</td>
<td>515</td>
<td>144.8</td>
<td>5.8e–36</td>
</tr>
<tr>
<td>Fag e 1</td>
<td>buckwheat13s globulin</td>
<td>565</td>
<td>98.4</td>
<td>5.6e–22</td>
</tr>
<tr>
<td>Ber e 2</td>
<td>Brazil nut11s globulin</td>
<td>465</td>
<td>87.9</td>
<td>6.8e–19</td>
</tr>
<tr>
<td>Jug n 2</td>
<td>black walnut/vicilin-like</td>
<td>481</td>
<td>36.6</td>
<td>1.9e–03</td>
</tr>
<tr>
<td>Jug r 2</td>
<td>English walnut/vicilin</td>
<td>593</td>
<td>35.9</td>
<td>3.8e–03</td>
</tr>
</tbody>
</table>

* The proteins most closely related to the top sequence have the lowest E score.
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 Ar a h 1 with PD values of <10 to the 134 EDWRRSPHQ 143 epitope, including two that contain parts of other known epitopes, and one in Ar ah3 (Table 4).

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 Ar a 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 Ar a h 3 and Ar a h 4, and is also similar to a major epitope of Ar a h 2 (see Table 3). As Figure 2a shows, the peptides similar to this epitope in Ar a 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 Ar a 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 Ar a 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 Ar a h 1, Ar a h 3, and Ar a 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, Ar a h 1 and Ar a h 2. The physicochemical similarities in epitopes indicated by low PD values correlate 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
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 products 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
structured 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 homologies 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 epitopes 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 allergic 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. Domains 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-1α, 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


Peanut Allergens Have Similar IgE Epitopes


5789


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.