NIH Public Access
Author Manuscript
Mol Immunol. Author manuscript; available in PMC 2009 August 1.

Published in final edited form as:
Mol Immunol. 2008 August ; 45(14): 3740–3747. doi:10.1016/j.molimm.2008.05.026.

Comprehensive 3D-modeling of allergenic proteins and amino acid composition of potential conformational IgE epitopes

Numan Oezguen[1], **Bin Zhou**[2], **Surendra S. Negi**[1], **Ovidiu Ivanciuc**[1], **Catherine H. Schein**[1,3], **Gilles Labesse**[4,5], and **Werner Braun**[1*]

1*Department of Biochemistry and Molecular Biology and Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, Texas 77555-0857, USA*

2*Current address: Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive San Diego, CA 92121, USA*

3*Department of Microbiology and Immunology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0857, USA*

4*CNRS-Universités Montpellier 1 & 2, UMR5048, Centre de Biochimie Structurale, 29, Rue de Navacelles, F-34090 Montpellier Cedex, France*

5*INSERM U554, Centre de Biochimie Structurale, 29, Rue de Navacelles, F-34090 Montpellier Cedex, France*

Abstract

Similarities in sequences and 3D structures of allergenic proteins provide vital clues to identify clinically relevant IgE cross-reactivities. However, experimental 3D structures are available in the Protein Data Bank for only 5% (45/829) of all allergens catalogued in the Structural Database of Allergenic Proteins (SDAP, http://fermi.utmb.edu/SDAP). Here, an automated procedure was used to prepare 3D-models of all allergens where there was no experimentally determined 3D structure or high identity (95%) to another protein of known 3D structure. After a final selection by quality criteria, 433 reliable 3D models were retained and are available from our SDAP Website. The new 3D models extensively enhance our knowledge of allergen structures. As an example of their use, experimentally derived “continuous IgE epitopes” were mapped on 3 experimentally determined structures and 13 of our 3D-models of allergenic proteins. Large portions of these continuous sequences are not entirely on the surface and therefore cannot interact with IgE or other proteins. Only the surface exposed residues are constituents of “conformational IgE epitopes” which are not in all cases continuous in sequence. The surface exposed parts of the experimental determined continuous IgE epitopes showed a distinct statistical distribution as compared to their presence in typical protein-protein interfaces. The amino acids Ala, Ser, Asn, Gly and particularly Lys have a high propensity to occur in IgE binding sites. The 3D-models will facilitate further analysis of the common properties of IgE binding sites of allergenic proteins.

Keywords

SDAP; allergens; structural database; linear epitopes

*Corresponding author: Werner Braun, Department of Biochemistry & Molecular Biology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0857, Phone : +1 409 747 6810, Fax : +1 409 747 6000, E-Mail : webraun@utmb.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Introduction

Allergy or type I hypersensitivity is an inflammatory systemic response, characterized by high levels of specific immunoglobulin E (IgE) antibodies to normally innocuous environmental substances. Allergic diseases affect a substantial portion of the population, with as many as two million school age children in US allergic to some food type (Nowak-Wegzyn, 2007). Symptoms are generally mild and treatable with over-the-counter antihistamines (Sampson, 1999a; Sampson, 1999b; Sampson, 2005), but in some cases, such as peanut (Bock et al., 2001; Maleki and Hurlburt, 2004; Sicherer, 2003; Teuber and Beyer, 2004; Wensing et al., 2003) or shrimp (Samson et al., 2004) allergies, ingestion can lead to life-threatening anaphylactic shock (Teuber et al., 2006). For individuals allergic to a common food or pollen, life can become severely proscribed, as they must avoid ingesting or breathing even minute amounts of the proteins to which they are sensitive. Individuals with severe allergy to one protein will often react to similar proteins that may be present in quite different plants or animals (Aalberse and Stadler, 2006; Schein et al., 2007). Thus there is considerable interest in identifying the molecular characteristics that correlate with IgE binding by proteins, so as to distinguish proteins that could cause cross-reactivities (Breiteneder and Mills, 2005; Jenkins et al., 2005).

The “Structural Database of Allergenic Proteins” (SDAP, http://fermi.utmb.edu/SDAP) (Ivanciuc et al., 2003; Ivanciuc et al., 2002) was created to allow rapid analysis of closely related allergens, globally (by FASTA and BLAST searching) or at the local sequence level, using a physicochemical-property distance (PD value) to compare continuous IgE epitope sequences (Schein et al., 2005; Schein et al., 2006a). Most of the information about how IgE antibodies in the sera from atopic individuals bind to these proteins comes from comparing the reactivity of discrete, often overlapping peptides from the protein sequence. However, without structural information, the actual sequence that constitutes the IgE binding site of these “continuous epitopes” cannot be clarified, as all the amino acids are rarely 100% exposed in a folded protein. Further, we cannot predict or test “conformational IgE epitopes”, i.e., those formed from several areas of the protein sequence. Only about 19% (45 Protein Data Bank (PDB) structures + 114 close homologs) of the allergen sequences deposited in SDAP have an experimentally determined 3D structure in the PDB, or have >95% identity to a homologous protein of known structure. Further, there are no experimental structures for the most potent allergenic proteins, including those from nuts and many fungi, where continuous IgE epitopes have been defined. In this study, we set out to determine the probable structures of the remaining 81% of known allergens, by identifying suitable templates in the Protein Database (PDB) using the TOME metaserver (Douguet and Labesse, 2001) (http://bioserv.cbs.cnrs.fr/HTML_BIO/frame_meta.html). We obtained reliable 3D models for 433 sequences, including the major allergens of peanuts, tree nuts, weed and tree pollens, fungi and insects. Our approach also indicated which allergens were not good targets for modeling with these methods, and could be recommended as candidates for the structural genomics initiative. Allergens that belong to protein families (Finn et al., 2006) (http://pfam.sanger.ac.uk/) for which there is no representative experimental structure known are particularly highlighted for further study. We then used 3D-models of allergens for which the IgE epitopes had been mapped, using peptide series, to determine which amino acids had the highest surface exposure. We were thus able to do 5 the first structure based, statistical survey of the amino acids that were likely to be involved in IgE binding.

Methods

Template selection

The target sequences were submitted to the TOME metaserver (http://bioserv.cbs.cnrs.fr/HTML_BIO/frame_meta.html) (Douguet and Labesse, 2001) which

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

distributed query sequences to three fold recognition servers (FUGUE(Shi et al., 2001), mGenThreader (Jones, 1999) and 3DPSSM (Kelley et al., 2000)), collected the results of these servers, reformatted and returned the results to the user. We classified the top hits from each server as 0 for “reliable”, 1 for “medium” and 2 for “difficult” according to the E-value or the Z-score (see Table I for cutoff values). When all three fold recognition servers recognized and returned the same template, we added the scores from each of the three servers to obtain a “confidence score” between 0 and 6. For confidence scores <3, a 3D-model was generated for the longest alignment in the first round if 1) the matching region of the template with the target sequence was longer than 30 amino acids, 2) there was no gap in the alignment greater than 20 amino acids and 3) the fold classifications according to SCOP (Andreeva et al., 2004; Lo Conte et al., 2002; Murzin et al., 1995) of the top hits were the same. In the second round we analyzed the targets which did not pass the SCOP classification filter of round one. For these targets, we determined whether the regions of the suggested templates that were aligned with the target sequences had a similar fold, with the program CE (Shindyalov and Bourne, 1998). If the template regions had a root mean square deviation (RMSD) of less than 3Å to one another, they were considered to have the same fold and again the longest alignment was chosen to generate a 3D-model.

MPACK modeling

The main programs of our modeling procedure were EXDIS, DIAMOD (Mumenthaler and Braun, 1995; Sanner et al., 1989) and FANTOM (Schaumann et al., 1990). EXDIS generated geometrical constraints (lower and upper distance and dihedral angle constrains) out of the template structure for the aligned regions. DIAMOD used these geometrical constraints to generate a 3D-model structure for the target sequence which had the lowest violations of these constraints. It used rotamer libraries and generated in the past reliably good geometries with good bond distances and bond angles (Oezguen et al., 2002; Ravindranath et al., 2003; Schein et al., 2001; Schein et al., 2006b; Xu et al., 2001; Xu et al., 1999a; Xu et al., 1999b). FANTOM energies for the minimized 3D-model structures were generated, with constraints, using the ECEPP/2 (Nemethy et al., 1983) force field. The information flow to and between these programs and other programs for quality evaluation was controlled by a PERL script. Other programs used for quality control of the models were PROFIT (http://www.bioinf.org.uk/software/profit) and PROCHECK (Laskowski et al., 1988). PROFIT was used to calculate the RMSD between the target 3D-model and template for the aligned regions and PROCHECK was used to check the geometry of the 3D-models.

Residue surface accessibility

The GETAREA(Fraczkiewicz and Braun, 1998) (http://pauli.utmb.edu/cgi-bin/get_a_form.tcl) program was used to determine the solvent exposure of the residues in IgE binding peptide sequences determined for 16 allergens. Residues with >25% solvent exposure were considered to be on the surface. Propensities were calculated as ratio of probabilities (p_i/P_i). For example the propensity for Ala was calculated as the ratio of the probability to find an Ala in epitopes (on the surfaces) (p_{Ala}) and the probability to find an Ala on the whole surfaces (P_{Ala}).

Results

An analysis of the allergen sequences in the SDAP database indicated that 25 sequences were very short, 45 had experimental structures in the PDB, and another 114 were nearly identical (95% sequence identity) to other proteins of known structure (Table II). We set out to generate 3D models of the remaining 645 SDAP sequences. These target sequences were submitted to the Fold Recognition (FR) server FUGUE (Shi et al., 2001), mGenThreader (Jones, 1999) and 3DPSSM (Kelley et al., 2000) via the metaserver TOME (Douguet and Labesse, 2001). The

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

FR servers returned alignments with high confidence level for 501 sequences. The remaining 144 allergen sequences did not have clearly identifiable homolog templates at the time of the study. We generated 3D models of the aligned regions for the 501 target sequences, which had a sequence identity between 10 and 94% (Fig. 1) to the selected templates and considered only those which fulfilled three quality criteria. These were: 1) the overall conformational energy after FANTOM (Schaumann et al., 1990) minimization was negative, which indicated favorable local packing of the side chains, 2) the RMSD to the template for the aligned regions was less than 1.8Å (Fig. 1) and 3) not more than 5% of the φ/ψ dihedral angles were in the disallowed region of a Ramachandran plot(Ramachandran et al., 1990) (Fig. 2). Only 68 of the 501 target models failed to meet these criteria and we remained with good quality 3D-models for 433 sequences (37 from the second phase, see methods section). Most (396) of the 3D-models had a backbone RMSD to the template lower than 1 Å, and for those with sequence identity >60 % to the template, the backbone RMSDs were <0.7 Å. Given a good alignment the modeling procedure is obviously capable of generating 3D-models which are structurally very close to the templates. Only 3 % of the 3D-models have more than 3 % of the residues outside of the allowed region and 44 % of the 3D-models have all residues in the core regions of the Ramachandran plot.

The fourth criterion was that the PROCHECK (Laskowski et al., 1988) overall g-factor (a combination value related to proper stereochemistry, that includes terms for torsion angles and covalent geometry) was above −0.5. The g-factor was above this threshold for all the 3D-models, although many had g-factors worse than the template (above the diagonal in Figure 3). The modeling procedure obviously also corrects major flaws, as some of the 3D-models had better g-factors than the template. The lower the sequence identity, the higher the difference in the local packing should be, and hence the worse the g-factor.

The models provide structural information about peptide epitopes

As noted, our 3D-models gave structures for some of the most important and immunologically characterized allergens from many different sources. For example, we obtained a good 3D-model for 94/139 amino acids of Par j 1, one of the major allergens of *Parietaria Judaica* pollen (Asturias et al., 2003), the main cause of allergy in Mediterranean countries. The template was a non-specific lipid transfer protein from rice (PDB code 1RZL; all alpha helix), that was 31% (29/94 positions) identical to Par j 1 according to the mGenTheader alignment. The resulting 3D-model (Figure 4) has an RMSD of 0.6 Å to the template, and there were no residues in the disallowed region of the Ramachandran Plot. “Continuous IgE epitopes”, that had been previously characterized experimentally, mapped to one face of the protein. (Fig. 4c–f). In contrast, the epitope sequences of the peanut allergen Ara h 1 (Schein et al., 2005) and the fungal allergen Asp f 13 (Chow et al., 2000)map to various areas of the protein (Fig. 5). Further, many of the amino acids in the reactive peptides have no surface exposure, and thus are probably not part of an IgE binding epitope in the intact protein.

Statistical survey of amino acid propensities in IgE binding sequences

Most of the epitopes of allergens in SDAP were determined using synthetic peptides corresponding to segments of the protein sequence, and measuring the reactivity to IgE in patient sera by immunoblotting or protein dot-spots. A number of studies have shown that substituting individual amino acids in these peptides can totally abrogate IgE binding, but have failed to show a clear pattern for which amino acids are most likely to be important in forming the actual epitope surface (Cocco et al., 2003). Our 3D-models allowed us to determine which amino acids in these peptides would be surface exposed, and thus most likely to be involved in binding.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Comprehensive peptide studies are costly and time-consuming, and have only been done in detail for a small group of allergens. Here we analyzed the statistical distributions of amino acids for all allergens with experimentally known continuous IgE epitopes and known 3D-structure, either experimental or modeled structures. Experimental structures were available for only 3 of these well studied allergens (the fungal allergen Asp f 1(Yang and Moffat, 1996) (1AQZ), Jun a 1 (1PXZ) from cedar pollen(Czerwinski et al., 2005) and Ves v 5 (1QNX) from yellowjacket(Henriksen et al., 2001)). Our 3D-models provided structural information for another 13 proteins for which the IgE epitopes had been characterized, including Ara h 1 and Ara h 2 of peanuts, Asp f 13, Asp f 2, Asp f 3 from the fungus *Aspergillus fumigatus*, Cha o 1, Cry j 1, Jug r 1 from cedar pollens, Par j 1 and Par j 2 from weed pollen, Gal d 1 from chicken egg white, and Pen a 1 from shrimp. We mapped the linear peptides on the 16 protein 3D-structures, and used GETAREA (Fraczkiewicz and Braun, 1998) (http://pauli.utmb.edu/cgi-bin/get_a_form.tcl) to determine which residues in the IgE binding peptides had significant surface exposed area. As shown in the examples of Fig.5, the GETAREA results defined a subset of the residues of these peptides that could form the IgE binding site. The statistical propensity of residues to occur in these binding sites was then compared with the amino acid propensities for occurrence in the interface of 72 protein-protein complexes (Negi and Braun, 2007).

The surface propensities for allergens were similar to those for the proteins that formed the complexes, with the exception that there were somewhat more charged amino acids (Figure 6a). However, comparison of the propensities for amino acids to occur in the potential IgE binding sites with that of interface regions in other types of protein complexes revealed surprising differences (Fig. 6b). The large hydrophobic residues, such as Phe, Trp, Tyr, Ile, Leu and Met, that characterize protein interfaces, were much less likely to be in the epitope interfaces. While most of the amino acids were less likely to occur in epitopes, compared to their overall interface propensities, five amino acids were more likely to be in epitopes: Ala, Ser, Asn, Gly and most particularly, Lysine. These findings will allow us to formulate testable hypotheses about IgE binding sites on other allergenic proteins.

Discussion

We describe here the first systematic attempt to model all allergenic proteins (Table II, Fig. 1–3), and the first systematic comparison of the properties of known IgE epitopes based on both their sequence and probable structure. We observed a distinct pattern of preferred amino acids in the antibody binding sites (Figure 4 and 6), an unexpected result that is rendered more significant since the survey covered fungal, pollen and food allergens from many different PFAM classes. As most of the information about IgE binding to allergenic proteins is based on continuous peptide studies, reliable 3Dstructures were essential to designate the surface exposed residues that are most likely to form the binding site for IgE.

These 3D structures for allergens can help in the annotation of biochemical function or in predicting cross-reactivity among homologous proteins that goes beyond overall sequence similarity(Chapman et al., 2007). For example, the cockroach allergen Bla g 2 was considered first as an aspartic protease based on sequence similarity to proteins of this family. The X-ray crystal of Bla g 2 showed a zinc-binding cleft, and that the conformations of the residues in the suggested active site are distorted in such a fashion as to preclude catalytic function (Gustchina et al., 2005).

The mapping of conserved residues in a family of related allergens on the protein surface can explain observed clinical cross-reactivity. For example, the venoms of insects show considerable cross-reactivity among many insect species(Caruso et al., 2007). Ves v 5 represents a family of venom allergens produced by insects ranging from wasps to fire ants

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(King, 1996; King et al., 2001). The Ves v 5 homologues of the allergens from the *Vespula* genus the *Vespula* and the *Vespa/Dolichovespula* genera. are all serologically cross-reactive, yet not all generate cross-reactions in sensitive individuals. This may be due to dramatic differences in conserved surface areas of the allergens, as mapped on the 3D structure of Ves v 5 (Henriksen et al., 2001). Almost all the surface residues are conserved among the *Vespula*, but only 5 conserved patches are shared with the *Vespa/Dolichovespula* genera.. Most of these patches conserved across both genera are smaller than the critical size below the expected size of antibody binding sites of 800 to 1000 $Å^2$.

The importance of conformational epitopes and their impact on our understanding of the molecular basis for cross-reactivity was also experimentally demonstrated for pollen allergens (Bonds et al., 2008). Four linear IgE epitopes of Jun a 1, the dominant allergen in mountain cedar pollen, were mapped on the crystal structure of Jun a 1(Czerwinski et al., 2005), and denaturation experiments gave evidence for the conformational nature of some of these epitopes(Varshney et al., 2007). The IgE in many sera from Japanese patients detected Cry j 1 from Japanese cedar, and also Jun a 1 from Texas mountain cedar pollen extracts by ImmunoCAP. Mapping the regions of the epitopes on a 3D model of Cry j 1 could explain the extent to which these epitopes are responsible for the cross-reactivity between Jun a 1 and Cry j 1 (Midoro-Horiuti et al., 2006). We anticipate that future analysis of the 3D models, now available on our SDAP Website, will facilitate a similar analysis for other allergens, and further define the structural basis of cross reactivity.

Our modeling method, which combined automatic methods with a minimum of human intervention, proved to be a robust way to model the 3D structures of allergens (Fig. 1–3). The automatic method was made possible by rapidly identifying likely templates with the TOME metaserver. We also introduced standard classifiers to test reliability, so that the 3D-models now catalogued in SDAP have been vetted to remove unlikely 3D-models, and those with sub-optimal stereochemical properties. The 3D-models (Fig. 4,5), coupled with previously determined experimental structures, can now be used to make comprehensive comparisons of their properties and likely IgE epitopes (Fig. 6).

While many groups have used individual homology 3D-models of allergenic proteins (for example (Bannon, 2001; Dodo et al., 2005; Gehlhar et al., 2006; Ivanciuc et al., 2003; Schein et al., 2005; Soman et al., 2000)), the only previous source of automatically generated 3D-models with quality criteria for allergenic proteins were those included in the online database, MODBASE (Pieper et al., 2004), (http://salilab.org/modbase). However, only 9 of the proteins with known IgE epitopes were present in MODBASE, compared to 13 in our study. We were also able to manually inspect all our 3D-models, something that is not possible in larger databases. The existence of additional structures will aid in the comparison of known IgE epitopes, and allow testing of potential conformational epitopes. For example, we can combine methods that detect sequences with a high degree of similarity to known epitopes, such as the PD score in the SDAP database, with comparison of structure of the areas in the 3D model. Initial tests of this methodology revealed that several epitopes of the peanut allergens are very similar to one another in their physicochemical properties and structure (Schein et al., 2005).

Amino acid propensities in IgE binding sites

Mapping the previously determined epitopes on the 3D-models indicated that only a small subset of the amino acid residues had sufficient surface exposure to be involved in binding IgE in the intact protein. Our statistical survey of these sites indicated that certain amino acids, particularly lysines, had a higher propensity to occur in IgE epitope sites. The potential importance of surface lysines in binding IgE's has been experimentally observed for the linear epitopes of Phl p 5b (Gehlhar et al., 2006). We suggest that this observation is true for a broader range of allergens. Although the number of allergens with known IgE epitopes is currently

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

limited to 16 allergens, those allergens belong to 9 different PFAM families (Finn et al., 2006) (http://pfam.sanger.ac.uk/), and thus represent a diverse sample. Overall, the binding sites for the IgE molecules were considerably more hydrophilic than protein-protein interfaces for other complexes. Our 3D-model of the Par j 1 protein (Fig. 4) illustrates this in detail: the surface to which the continuous epitopes map is quite hydrophilic, and marked by highly exposed lysine side chains.

Towards a 3D structural classification of all allergens

All generated 3D models have a reliable template with a well defined classification of the three-dimensional protein structure of the allergen. This structural classification, defined according to SCOP classes (Andreeva et al., 2004), divides the allergens in a hierarchical way according to their similarity in the protein fold. This also provides another parameter for predicting cross-reactivity. Antibodies bind to surface patches of folded proteins, so allergens with a similar 3D structure are more likely to bind to the same IgE antibody(Aalberse, 2007; Aalberse and Stadler, 2006). However, this prerequisite is not sufficient for binding and future research is needed to incorporate other structural information for a successful prediction of cross-reactivity. The SCOP number for all 3D models is available from our SDAP website.

To obtain a complete structural classification of all allergens, we have prepared a list of candidates to be recommended to the structural genomics initiative. The proteins that we were not able to model reliably (supplementary data), belong to 71 different PFAM (Finn et al., 2006) (http://pfam.sanger.ac.uk/) families (supplementary data). For 17 there is no representative experimental structure in the PDB. These proteins are likely to have a novel fold, as they are not similar to any protein of known structure. Therefore it would be highly beneficial if the structural genomics projects or others would experimentally determine the 3D structures of representatives of those 17 PFAM families (highlighted in the supplementary data).

Conclusion

Our aim was to generate reliable homology 3D-models for those allergens whose structures are not solved experimentally or do not have very close homologs with known structure. We have generated good quality homology 3D-models for 67 % (433/645) of the allergens in this category. All reliable 3D-models are available via appropriate links through the SDAP web pages. There are still 212 allergen sequences without a clear template. Selected sequences from the list of these “difficult modeling targets”, which could represent novel folds, are good candidates for experimental structure determination. Analysis of the surface exposed areas of known linear IgE epitopes indicated a distinct propensity of finding certain amino acid types in epitopes as compared to protein-protein complex interfaces. The propensity to find Lys in the epitopes is significantly higher and the propensities for Phe, Trp, Met and Ile significantly lower. This reflects the properties of the IgE binding partner and/or the binding dynamics. The binding process might be guided via electrostatic funneling which would explain the net positive charge or at least the high density of Lys at the epitope region.

By mapping known continuous IgE epitopes on the surface of the 3D-models, we showed that only select residues are surface exposed. This is especially the case for long peptides (larger than ca. 10 amino acids). The 3D-models can be very useful in refining the sequences of these peptides to better identify the real site of IgE binding. This will facilitate the design of apo-allergenic proteins (Bannon, 2001; Dodo et al., 2005).

The 3D-models and known experimental structures in combination with the findings of the amino acid distribution on the epitopes can be used to develop new methods and/or to increase

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the predictive power of existing ones for prediction of allergenicity (Aalberse and Stadler, 2006) and cross-reactivity (Bonds et al., 2008; Goodman, 2006).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by a contract from the U.S. Food and Drug Administration (HHSF223200710011I), and grants from the National Institute of Health (R01 AI 064913), and the U.S. Environmental Protection Agency under a STAR Research Assistance Agreement (No. RD 833137). The article has not been formally reviewed by the EPA, and the views expressed in this document are solely those of the authors.

References

Aalberse RC. Assessment of allergen cross-reactivity. Clin Mol Allergy 2007;5:2. [PubMed: 17291349]

Aalberse RC, Stadler BM. In silico predictability of allergenicity: From amino acid sequence via 3-D structure to allergenicity:. From amino acid sequence via 3-D structure to allergenicity. Molecular Nutrition & Food Research 2006;50:625–627. [PubMed: 16764015]

Andreeva A, Howorth D, Brenner SE, Hubbard TJP, Chothia C, Murzin AG. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Research 2004;32:D226–D229. [PubMed: 14681400]

Asturias JA, Gomez-Bayon N, Eseverri JL, Martinez A. Par j 1 and Par j 2, the major allergens from Parietaria judaica pollen, have similar immunoglobulin E epitopes. Clinical and Experimental Allergy 2003;33:518–524. [PubMed: 12680870]

Bannon G, Cockrell G, Connaughton C, West CM, Helm R, Stanley JS, King N, Rabjohn P, Sampson HA, Burks AW. Engineering, characterization and in vitro efficacy of the major peanut allergens for use in immunotherapy. Int. Arch. Allergy Immunol 2001;124:70–72. [PubMed: 11306930]

Bock SA, Munoz-Furlong A, Sampson HA. Fatalities due to anaphylactic reactions to foods. J Allergy Clin Immunol 2001;107:191–193. [PubMed: 11150011]

Bonds RS, Midoro-Horiuti T, Goldblum R. A structural basis for food allergy: the role of cross-reactivity. Curr Opin Allergy Clin Immunol 2008;8:82–86. [PubMed: 18188023]

Breiteneder H, Mills ENC. Plant food allergens - structural and functional aspects of allergenicity. Biotechnology Advances 2005;23:395–399. [PubMed: 15985358]

Caruso B, Bonadonna P, Severino MG, Manfredi M, Dama A, Schiappoli M, Rizzotti P, Senna G, Passalacqua G. Evaluation of the IgE cross-reactions among vespid venoms. A possible approach for the choice of immunotherapy. Allergy 2007;62:561–564. [PubMed: 17441797]

Chapman MD, Pomes A, Breiteneder H, Ferreira F. Nomenclature and structural biology of allergens. J Allergy Clin Immunol 2007;119:414–420. [PubMed: 17166572]

Chow LP, Liu SL, Yu CJ, Liao HK, Tsai JJ, Tang TK. Identification and expression of an allergen Asp f 13 from Aspergillus fumigatus and epitope mapping using human IgE antibodies and rabbit polyclonal antibodies. Biochem J 2000;346(Pt2):423–431. [PubMed: 10677362]

Cocco RR, Jarvinen KM, Sampson HA, Beyer K. Mutational analysis of major, sequential IgE-binding epitopes in alpha(s1)-casein, a major cow's milk allergen. Journal of Allergy and Clinical Immunology 2003;112:433–437. [PubMed: 12897753]

Czerwinski EW, Midoro-Horiuti T, White MA, Brooks EG, Goldblum RM. Crystal structure of Jun a 1, the major cedar pollen allergen from Juniperus ashei, reveals a parallel beta-helical core. J Biol Chem 2005;280:3740–3746. [PubMed: 15539389]

Dodo H, Konan K, Viquez O. A genetic engineering strategy to eliminate peanut allergy. Curr. Allergy Asthma Rep 2005;5:67–73. [PubMed: 15659267]

Douguet D, Labesse G. Easier threading through web-based comparisons and cross-validations. Bioinformatics 2001;17:752–753. [PubMed: 11524382]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A. Pfam: clans, web tools and services. Nucleic Acids Research 2006;34:D247–D251. [PubMed: 16381856]

Fraczkiewicz 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.

Gehlhar K, Rajashankar KR, Hofmann E, Betzel C, Weber W, Werner S, Bufe A. Lysine as a critical amino acid for IgE binding in Phl p 5b C terminus. International Archives of Allergy and Immunology 2006;140:285–294. [PubMed: 16735798]

Goodman RE. Practical and predictive bioinformatics methods for the identification of potentially cross-reactive protein matches. Mol Nutr Food Res 2006;50:655–660. [PubMed: 16810734]

Gustchina A, Li M, Wunschmann S, Chapman MD, Pomes A, Wlodawer A. Crystal structure of cockroach allergen Bla g 2, an unusual zinc binding aspartic protease with a novel mode of self-inhibition. J Mol Biol 2005;348:433–444. [PubMed: 15811379]

Henriksen A, King TP, Mirza O, Monsalve RI, Meno K, Ipsen H, Larsen JN, Gajhede M, Spangfort MD. Major venom allergen of yellow jackets, Ves v 5: structural characterization of a pathogenesis-related protein superfamily. Proteins 2001;45:438–448. [PubMed: 11746691]

Ivanciuc O, Mathura V, Midoro-Horiuti T, Braun W, Goldblum RM, Schein CH. Detecting potential IgE-reactive sites on food proteins using a sequence and structure database, SDAP-Food. J. Agric. Food Chem 2003;51:4830–4837. [PubMed: 14705920]

Ivanciuc O, Schein CH, Braun W. Data mining of sequences and 3D structures of allergenic proteins. Bioinformatics 2002;18:1358–1364. [PubMed: 12376380]

Jenkins JA, Griffiths-Jones S, Shewry PR, Breiteneder H, Mills ENC. Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: An in silico analysis. Journal of Allergy and Clinical Immunology 2005;115:163–170. [PubMed: 15637564]

Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999;287:797–815. [PubMed: 10191147]

Kelley LA, MacCallum RM, Sternberg MJE. Enhanced genome annotation using structural profiles in the program 3D-PSSM. Journal of Molecular Biology 2000;299:499–520. [PubMed: 10860755]

King TP. Immunochemical studies of stinging insect venom allergens. Toxicon 1996;34:1455–1458. [PubMed: 9028002]

King TP, Jim SY, Monsalve RI, Kagey-Sobotka A, Lichtenstein LM, Spangfort MD. Recombinant allergens with reduced allergenicity but retaining immunogenicity of the natural allergens: hybrids of yellow jacket and paper wasp venom allergen antigen 5s. Journal of Immunology 2001;166:6057–6065.

Laskowski BC, Yoon DY, McLean D, Jaffe RL. Chain Conformations of Polycarbonate from Abinitio Calculations. Macromolecules 1988;21:1629–1633.

Lo Conte L, Brenner SE, Hubbard TJP, Chothia C, Murzin AG. SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Research 2002;30:264–267. [PubMed: 11752311]

Maleki SJ, Hurlburt BK. Structural and functional alterations in major peanut allergens caused by thermal processing. J. AOAC Int 2004;87:1475–1479. [PubMed: 15675461]

Midoro-Horiuti T, Schein CH, Mathura V, Braun W, Czerwinski EW, Togawa A, Kondo Y, Oka T, Watanabe M, Goldblum RM. Structural basis for epitope sharing between group 1 allergens of cedar pollen. Mol Immunol 2006;43:509–518. [PubMed: 15975657]

Mumenthaler C, Braun W. Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry. J Mol Biol 1995;254:465–480. [PubMed: 7490763]

Murzin AG, Brenner SE, Hubbard T, Chothia C. Scop - a Structural Classification of Proteins Database for the Investigation of Sequences and Structures. Journal of Molecular Biology 1995;247:536–540. [PubMed: 7723011]

Negi SS, Braun W. Statistical analysis of physical-chemical properties and prediction of protein-protein interfaces. Journal of Molecular Modeling 2007;13:1157–1167. [PubMed: 17828612]

Nemethy G, Pottle MS, Scheraga HA. Energy Parameters in Polypeptides .9. Updating of Geometrical Parameters, Nonbonded Interactions, and Hydrogen-Bond Interactions for the Naturally-Occurring Amino-Acids. Journal of Physical Chemistry 1983;87:1883–1887.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Nowak-Wegzyn A. New perspectives for use of native and engineered recombinant food proteins in treatment of food allergy. Immunology and Allergy Clinics of North America 2007;27:105–127. [PubMed: 17276882]

Oezguen N, Adamian L, Xu Y, Rajarathnam K, Braun W. Automated assignment and 3D structure calculations using combinations of 2D homonuclear and 3D heteronuclear NOESY spectra. Journal of Biomolecular Nmr 2002;22:249–263. [PubMed: 11991354]

Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, Webb B, Greenblatt D, Huang CC, Ferrin TE, Sali A. MODBASE, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Research 2004;32:D217–D222. [PubMed: 14681398]

Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of Polypeptide-Chain Configurations. Current Science 1990;59:813–817.

Ravindranath G, Xu Y, Schein CH, Rajarathnam K, Painter SD, Nagle GT, Braun W. NMR solution structure of Aplysia attractin - pheromone protein from the mollusk Aplysia californica. Biochemistry 2003;42:9970–9979. [PubMed: 12924946]

Sampson HA. Food allergy. Part 1: Immunopathogenesis and clinical disorders. Journal of Allergy and Clinical Immunology 1999a;103:717–728. [PubMed: 10329801]

Sampson HA. Food allergy. Part 2: Diagnosis and management. Journal of Allergy and Clinical Immunology 1999b;103:981–989. [PubMed: 10359874]

Sampson HA. Food allergy: When mucosal immunity goes wrong. Journal of Allergy and Clinical Immunology 2005;115:139–141. [PubMed: 15637560]

Samson KTR, Chen FH, Miura K, Odajima Y, Iikura Y, Rivas MN, Minoguchi K, Adachi M. IgE binding to raw and boiled shrimp proteins in atopic and nonatopic patients with adverse reactions to shrimp. International Archives of Allergy and Immunology 2004;133:225–232. [PubMed: 14976390]

Sanner M, Widmer A, Senn H, Braun W. GEOM, a new tool for molecular modeling based on distance geometry calculations with NMR data. J. Comp. Aided Mol. Des 1989;3:195–210.

Schaumann T, Braun W, Wuthrich K. A program, FANTOM, for energy refinement of polypeptides and proteins using a Newton-Raphson Minimizer in the torsion angle space. Biopolymers 1990;29:679–694.

Schein CH, Ivanciuc O, Braun W. Common physical-chemical properties correlate with similar structure of the IgE epitopes of peanut allergens. Journal of Agricultural and Food Chemistry 2005;53:8752–8759. [PubMed: 16248581]

Schein, CH.; Ivanciuc, O.; Braun, W. Structural Database of Allergenic Proteins (SDAP). In: Maleki, SJ.; Burks, AW.; Helm, RM., editors. Food Allergy. Washington, D.C: ASM Press; 2006a. p. 257-283.

Schein CH, Ivanciuc O, Braun W. Bioinformatics approaches to classifying allergens and predicting cross-reactivity. Immunology and Allergy Clinics of North America 2007;27:1–27. [PubMed: 17276876]

Schein CH, Nagle GT, Page JS, Sweedler JV, Xu Y, Painter SD, Braun W. Aplysia attractin: biophysical characterization and modeling of a water-borne protein pheromone. Biophys. J 2001;81:463–472. [PubMed: 11423429]

Schein CH, Oezguen N, Volk DE, Garimella R, Paul A, Braun W. NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. Peptides 2006b;27:1676–1684. [PubMed: 16540201]

Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 2001;310:243–257. [PubMed: 11419950]

Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Engineering 1998;11:739–747. [PubMed: 9796821]

Sicherer S, Munoz-Furlong A, Sampson HA. 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;112:1203–1207. [PubMed: 14657884]

Soman KV, Midoro-Horiuti T, Ferreon JC, Goldblum RM, Brooks EG, Kurosky A, Braun W, Schein CH. Homology modeling and characterization of IgE epitopes of mountain cedar allergen Jun a 3. Biophys. J 2000;79:1601–1609. [PubMed: 10969020]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Teuber, S.; Beyer, K.; Comstock, S.; Wallowitz, M. The big eight foods: clinical and epidemiological overview. In: Malecki, S., editor. Food Allergy. Washington DC: ASM Press; 2006. p. 49-79.

Teuber SS, Beyer K. Peanut, tree nut and seed allergies. Current Opinion in Allergy and Clinical Immunology 2004;4:201–203. [PubMed: 15126942]

Varshney S, Goldblum RM, Kearney C, Watanabe M, Midoro-Horiuti T. Major mountain cedar allergen, Jun a 1, contains conformational as well as linear IgE epitopes. Mol Immunol 2007;44:2781–2785. [PubMed: 16423400]

Wensing M, Knulst AC, Piersma S, O'Kane F, Knol EF, Koppelman SJ. 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. [PubMed: 12589366]

Xu Y, Jablonsky MJ, Jackson PL, Braun W, Krishna NR. Automatic assignment of NOESY cross peaks and determination of the protein structure of a New World scorpion neurotoxin using NOAH/DIAMOD. J. Mag. Res 2001;148:35–46.

Xu, Y.; Schein, CH.; Braun, W. Combined automated assignment of NMR spectra and calculation of three-dimensional protein structures. In: Krishna; Berliner, editors. Biological Magnetic Resonance: Structure Computation and Dynamics in Protein NMR. 17. New York: Kluwer Academic/Plenum Publishers; 1999a. p. 37-79.

Xu Y, Wu J, Gorenstein D, Braun W. Automated 2D NOESY assignment and structure calculation of crambin (S22/I25) with the self-correcting distance geometry based NOAH/DIAMOD programs. J. Mag. Res 1999b;136:76–85.

Yang X, Moffat K. Insights into specificity of cleavage and mechanism of cell entry from the crystal structure of the highly specific Aspergillus ribotoxin, restrictocin. Structure 1996;4:837–852. [PubMed: 8805570]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

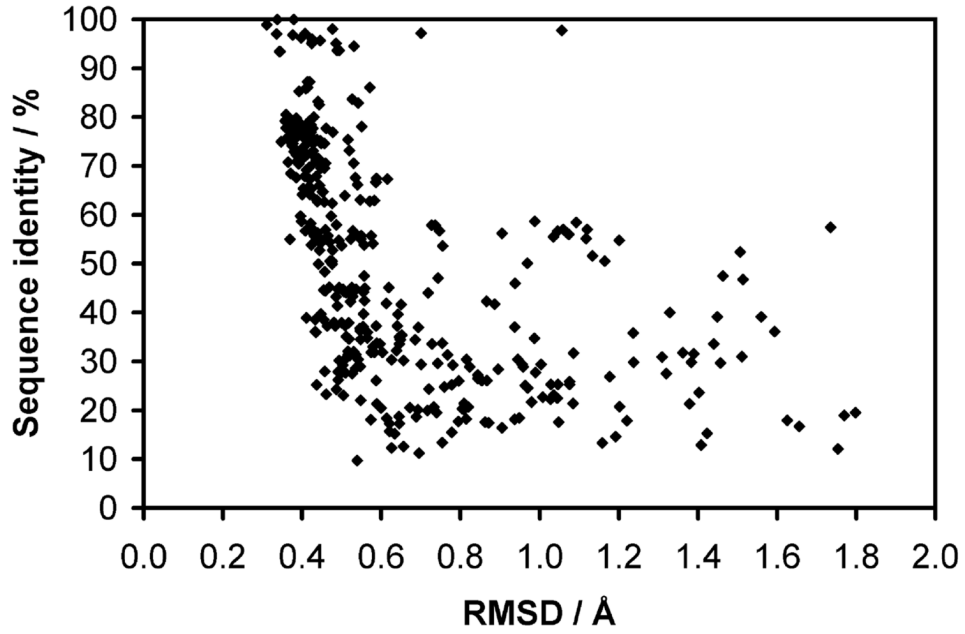

Figure 1.
Correlation plot of sequence identity and RMSD values between the 3D-models and their templates in the aligned regions.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

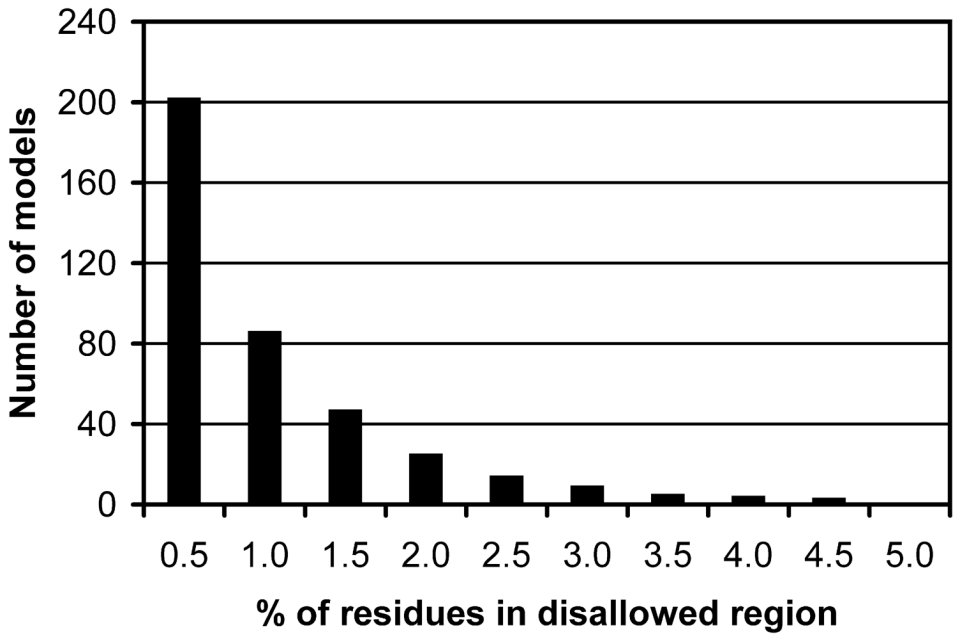

Figure 2.
Most of the 3D-models have >98% of their residues in allowed areas of the Ramachandran Plots.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

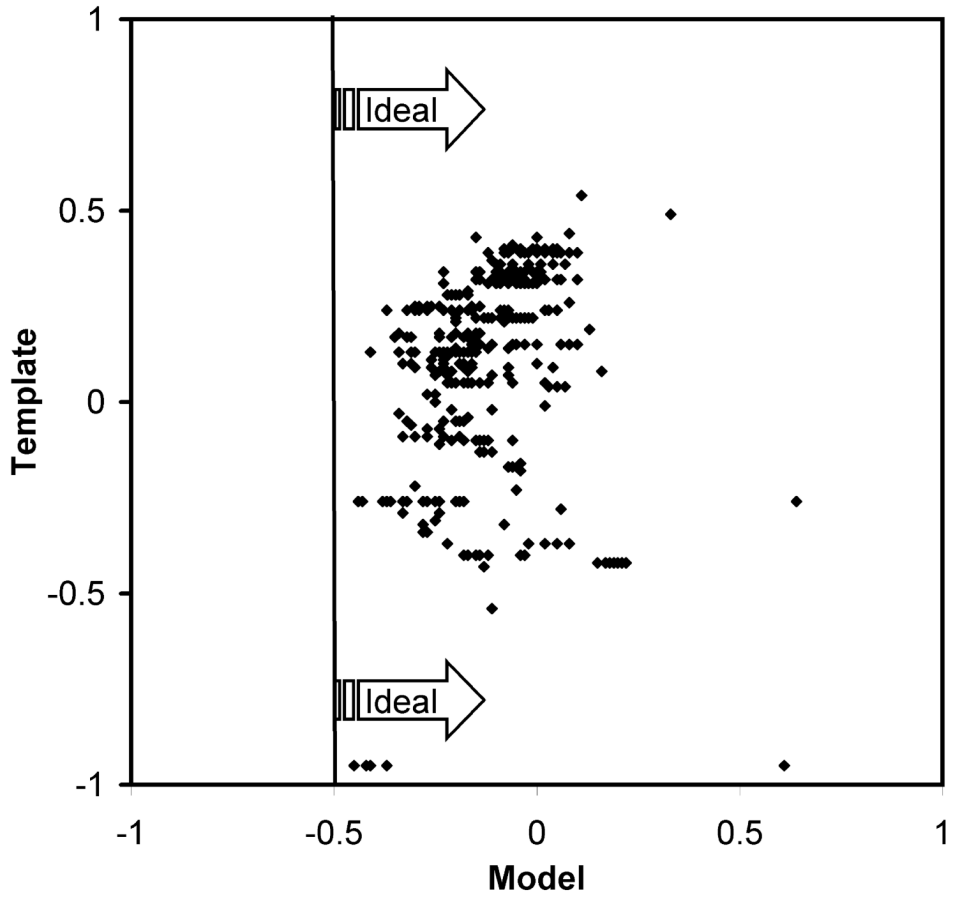

Figure 3.
Distribution of the PROCHECK overall g-factor of the 3D-models and their corresponding templates.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

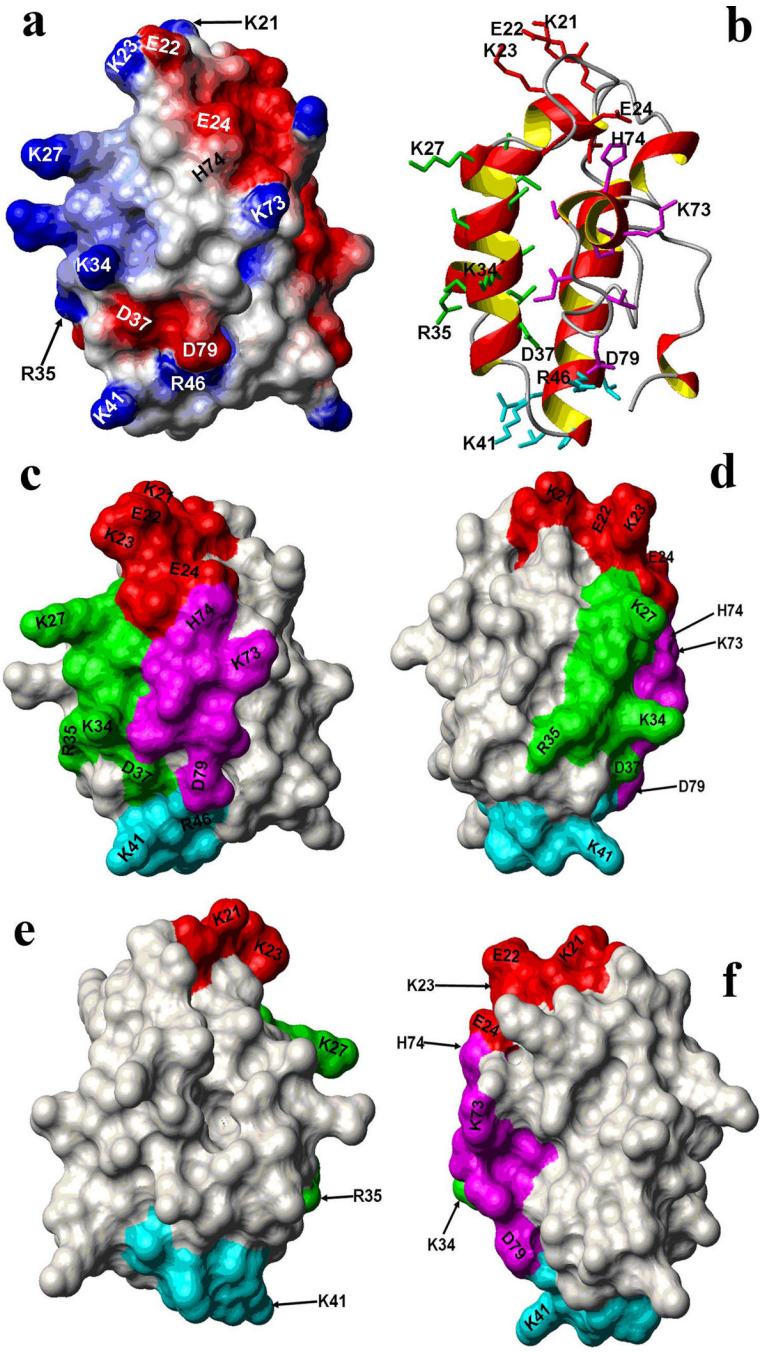

Figure 4.

Linear epitopes map to a hydrophilic, lysine rich face of the 3D-model of the weed pollen allergen Par j 1. Charged residues on the surface of linear epitopes are labeled. a) surface electrostatic potential, b) ribbon plot showing the surface exposed side chains of the linear epitopes, colored to indicate the different epitopes; c–f: rotations around the y-axis, starting from the orientation of panel a, showing the epitopes colored coded as in b). This face is positively charged while the back side is predominantly negative.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

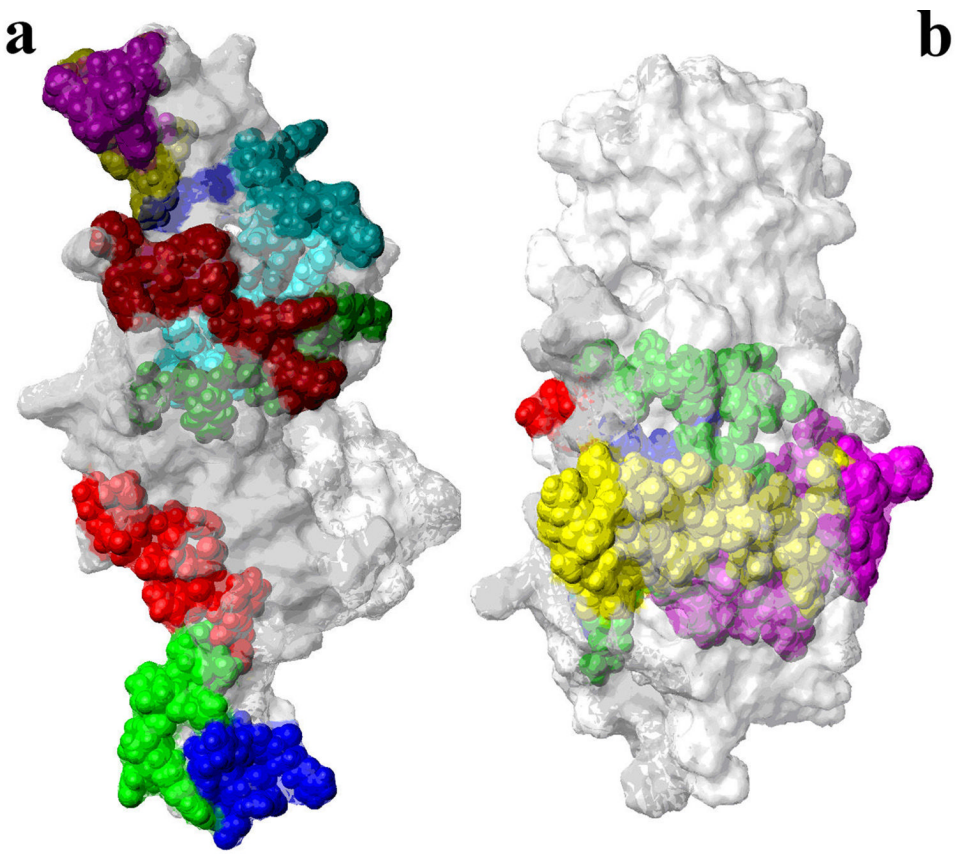

Figure 5.
Linear epitopes of the allergens a) Ara h 1 from peanut and b) Asp f 13 from a fungus are only partially surface exposed. The surface is transparent and the different epitopes are demarcated by different colors.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

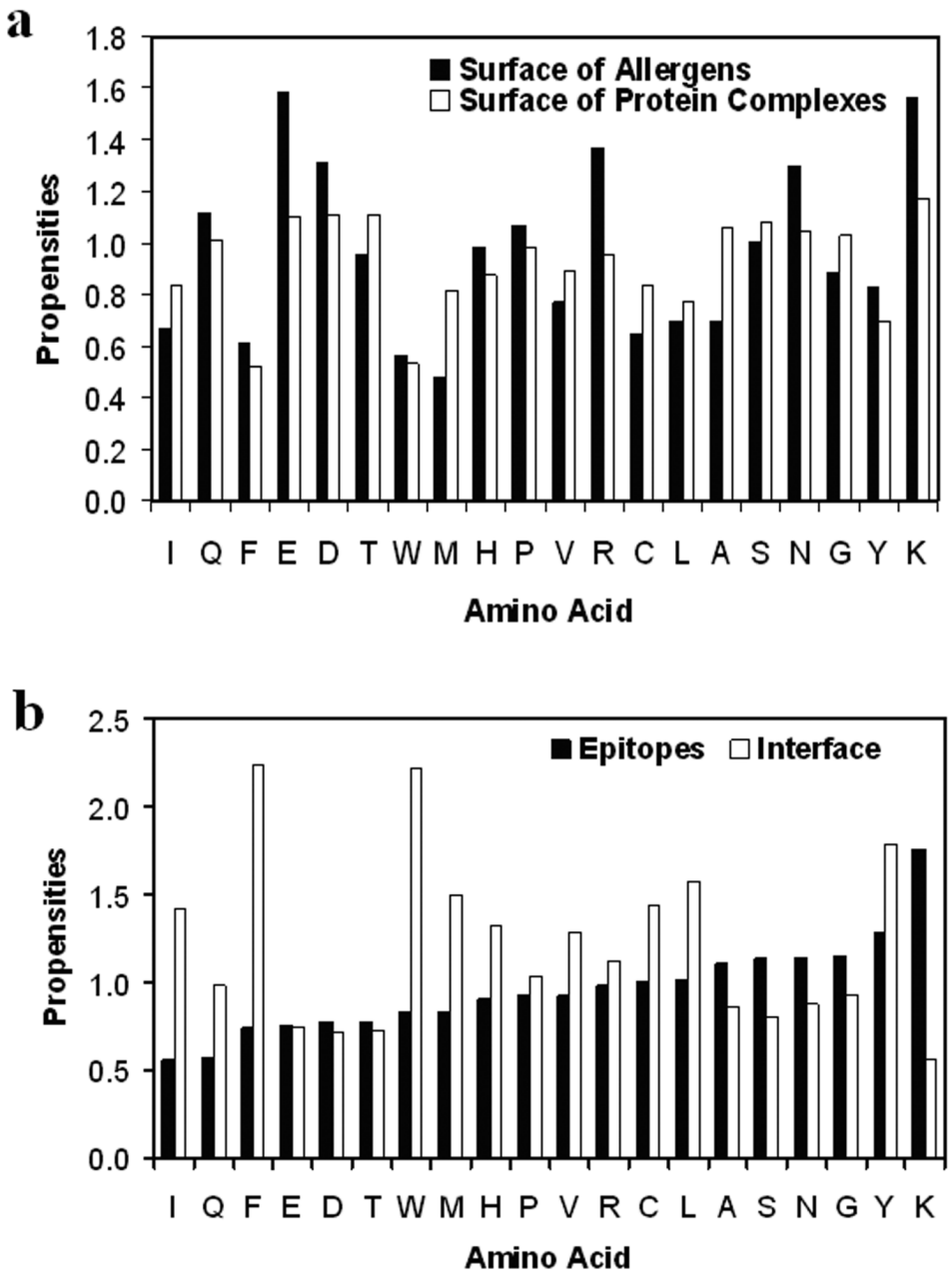

Figure 6.

The propensities of the amino acids, relative to their overall occurrence in the protein sequence, to occur on the surfaces of proteins (a) or in the epitope or interface residues (b), The propensities for occurrence in IgE binding sites are based on the surface exposed residues of linear IgE epitopes determined for 16 allergens (3 experimental structures, 13 3D-models). The propensities for occurrence in the interface are based on a previous study of 72 protein complexes (Negi and Braun, 2007).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table I

Cutoffs for the classification of fold recognition server results

Server	Category 0	Category 1	Category 2
3D- PSSM	$\log E \leq -2$	$-2 < \log E \leq 0$	$\log E > 0$
mGenThreader	$\log E \leq -3$	$-3 < \log E \leq -1$	$\log E > -1$
FUGUE	$Z \geq 6$	$6 > Z \geq 3$	$3 > Z$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table II

Summary of the modeling results

Sequences in SDAP	**829**
PDB structures	45
Very close homologs to PDB structures	114
Short sequences (< 30 amino acids)	25
Sequences to be modeled	645
FR alignments classified reliable	501
Good 3D-models	433
3D-Models did not pass quality filters	68
Difficult targets	144