

Research Article

Combination of two epitope identification techniques enables the rational design of soy allergen Gly m 4 mutantsHeide Havenith¹Karolin Kern²Paul Rautenberger²Holger Spiegel^{1*}Michael Szardenings²Elke Ueberham²Jörg Lehmann²Matthias Buntru¹Simon Vogel¹Regina Treudler³Rainer Fischer^{1,4}Stefan Schillberg^{1,5}

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/biot.201600441.

Submitted: 22-Jul-2016
Revised: 29-Nov-2016
Accepted: 30-Nov-2016

This article is protected by copyright. All rights reserved.

¹ Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Forckenbeckstraße 6, 52074 Aachen, Germany

² Fraunhofer Institute for Cell Therapy and Immunology IZI, Perlickstraße 1, 04130 Leipzig, Germany

³ Dermatology, Venereology and Allergology Clinic, University of Leipzig, Philipp-Rosenthal-Straße 23, 04130 Leipzig, Germany

⁴ RWTH Aachen University, Institute for Molecular Biotechnology, Worringerweg 1, 52074 Aachen, Germany

⁵ Institute for Phytopathology and Applied Zoology, Justus-Liebig University Giessen, Heinrich-Buff-Ring 26-32, 35392 Gießen Germany

***Correspondence:** Holger Spiegel, Fraunhofer IME, Forckenbeckstraße 6, 52074 Aachen, Germany.

E-mail: holger.spiegel@ime.fraunhofer.de

Keywords: Bet v 1 homolog, epitope mapping, peptide microarray, phage display, soy allergy

Abstract

Background: Detailed IgE-binding epitope analysis is a key requirement for the understanding and development of diagnostic and therapeutic agents to address food allergies.

Methods: We combined an IgE-specific linear peptide microarray with random phage peptide display for the high-resolution mapping of IgE-binding epitopes of the major soybean allergen Gly m 4, which is a homologue to the birch pollen allergen Bet v1.

Results: Three epitopes were identified and mapped to a resolution of four key amino acids, allowing the rational design and the production of three Gly m 4 mutants with the aim to abolish or reduce the binding of epitope-specific IgE. In ELISA, the binding of the mutant allergens to polyclonal rabbit-anti Gly m 4 serum as well as IgE purified from Gly m 4-reactive soybean allergy patient sera was reduced by up to 63% compared to the wild-type allergen. Basophil stimulation experiments using RBL-SX38 cells loaded with patient IgE showed a decreased stimulation from 25% for the wild-type Gly m 4 to 13% for one mutant.

Conclusion: The presented approach demonstrates the feasibility of precise mapping of allergy-related IgE-binding epitopes, allowing the rational design of less allergenic mutants as potential therapeutic agents.

1 Introduction

Soybean is one of eight foods that account for 90% of all food allergies. However, it is a popular vegan protein source and an important nutritional ingredient [1, 2]. The growing prevalence of allergies, currently affecting more than 25% of the European population, has increased the demand for allergens as diagnostic, therapeutic and research reagents [3].

The three major soybean allergens are Gly m 4, Gly m 5 and Gly m 6, and all three are being used as diagnostic reagents. Gly m 4 is of special interest because it is homologous to the major birch pollen allergen Bet v 1, and 70% of patients allergic to birch pollen produce cross-reactive, Gly m 4-specific IgE [4].

The detailed analysis of IgE-binding epitopes is often hampered by the challenge of obtaining sufficient amounts of IgE from allergic patients [5]. Nevertheless, detailed IgE-binding epitope maps are needed for the precise diagnosis of sensitization patterns and to modify allergens in a directed manner for immunotherapeutic applications. Several groups have recently presented data on food allergen epitope prediction [6] and mapping using different methodology including protein fragments [7] peptide array [8] and phage display [9]. Reducing the reactivity of patient sera to allergens in which potential IgE-binding epitopes have been eliminated means that the mutated allergens can be used as a therapeutic strategy based on the upregulation of allergen-specific IgG, thus downregulating effector cell responses either by competing for IgE-binding epitopes or suppressing inflammatory cascades by signaling through FcγRIIB (CD32b), which in turn promotes the elimination of IgE from the mast cell surface and blocks mast cell activation [10, 11].

Allergen-based vaccines for immunotherapy can be generated using diverse strategies. For example, wild-type allergens from different sources were mixed and administered by subcutaneous injection in a double-blind, placebo-controlled study to reduce the symptoms of grass pollen allergy [12]. Wild-type allergens can also be chemically or physically modified to randomly change their antigenic properties, producing so-called allergoids. A rational alternative is the use of genetically modified allergens, such as differentially-folded conformers, protein chimeras of different allergen sources [13] or site directed mutagenesis [14].

Here we describe the mapping of allergy-related Gly m 4 epitopes for sera derived from soy allergy patients by combining an IgE specific peptide array technology and a phage display approach with a random peptide library. We identified three distinct epitopes facilitating the targeted substitution of single amino acids with biochemically similar replacements bearing larger side chains to modify the epitope surface, reducing the allergenicity while maintaining the allergen structure.

2 Materials and methods

2.1 Peptide micro array

The Gly m 4 sequence was subdivided into 15-mer peptides with four-residue overlaps and a chip was designed bearing duplicates of every peptide. The microarrays were purchased from PEPperPRINT (Heidelberg, Germany) which printed the peptides one residue at a time directly onto a glass slide using solid-phase Fmoc chemistry.

Each slide was washed in staining buffer (PBS containing 0.05% (v/v) Tween-20 and 0.1% (w/v) BSA, pH 7.4) for 10 min at room temperature and blocked for 30 min in blocking buffer (PBS containing 0.05% (v/v) Tween-20 and 1% (w/v) BSA, pH 7.4). The slides were then incubated in sera from soy allergy patients diluted 1:50 in staining buffer overnight at 4°C. After three washes in standard buffer (PBS containing 0.05% (v/v) Tween-20, pH 7.4) the slides were shaken at 200 rpm for 30 min with a goat-anti human IgE primary detection antibody (Thermo Fisher Scientific, Waltham, MA, USA) diluted to 200 ng/mL in staining buffer. After three further washes, bound goat antibodies were detected by secondary Cy5-labeled donkey anti-goat antibody (Abcam, Cambridge, UK) (200 ng/mL in staining buffer, 30 min, room temperature). Fluorescence signals on the array were detected in a microarray reader at 635 nm and the data were evaluated using PepSlide Analyser software (PEPperPRINT) to determine the signal intensity of each peptide spot. The software also subtracted the local background signal for every spot, thus normalizing for the different background levels in each serum sample.

2.2 Epitope discovery by peptide phage display

A trinucleotide-based, 16-mer peptide phagemid library covering $>10^9$ clones (ENTE-1;[15]) was used for two rounds of panning against soy allergic patient's serum. We diluted the amount serum to 24 μ g of serum protein in 2.5 mL PBS and used this to coat protein-binding ImmunoTMTubes with MaxisorpTM (Nunc, Wiesbaden, Germany) for 1 h at 4°C (18 rpm). The coated tubes were washed with 2.5 mL blocking buffer (5% (v/v) nonfat dry milk in PBS) and incubated with 4 mL blocking buffer at 4°C for at least 15 min (18 rpm). The tubes were washed three times with 4 mL wash buffer I (PBS) and 1 mL blocking buffer containing 30 μ g of non-allergic serum protein and the phage library ENTE1 was added to each tube. In the first selection round, 4×10^{11} colony forming unit (cfu) of the ENTE1 library was used per experiment, whereas the second selection used only 1000 x the recovered phages (typically 10^9 – 10^{10} cfu). The tubes were incubated for 2 h at 4°C (18 rpm) then washed five times with 1 mL 0.1% (v/v) Tween-20 in PBS, and subsequently with 0.5% (w/v) Tween-20 in PBS. Then the bound phages were eluted in 1 mL of elution buffer (0.1 M glycine-HCl, pH 2.2). The eluate was immediately neutralized with 200 μ L neutralization buffer (1 M Tris-HCl, pH 8.2). Eluted phages were amplified, spread on large agar plates (~ 10 clones per mm^2), incubated overnight, washed off the plate with a small amount of dYT and an aliquot was incubated with helper phage in dYT overnight for phage production. Phagemid DNA was prepared from another aliquot, and the library insert of the phagemid DNA was amplified with single-end primers comprising all necessary elements for direct transfer to an Illumina MiSeqTM system (Illumina, San Diego, USA). At least 170,000 sequences were analyzed using LibDB software (PolyQuant, Bad Abbach, Germany; Delaroque et al., manuscript in preparation), which removes sequences with potential sequencing errors by comparison to the design of the degenerate trinucleotide and adjacent sequences. LibDB then generates a statistical database of short amino acid motifs from all sequences (trimers, tetramers and pentamers). The design and quality of the trinucleotide based peptide phage display library allowed the statistical comparison of the observed and theoretical frequencies of sequence motifs to generate an enrichment score. Plotting the enrichment scores of the motifs against the Gly m 4 protein sequence allows to identify the position of potential epitopes. Further analyses of sequences with such a motif was carried out to

identify additionally enriched amino acids in their neighbourhood. This served to confirm the likelihood of a potential epitope.

2.3 Visualization of IgE epitopes on Gly m 4

Putative IgE-binding epitope sequences were visualized using available structural information for soybean Gly m 4 (PDB entry 2K7H) and the Protean 3D software (DNASTar, Madison, WI, USA). The 3D structure was also used to support the selection of solvent-exposed residues within the epitopes suitable for substitution.

2.4 Cloning of allergen constructs

The sequences of wild-type Gly m 4 (PR-10, GenBank X60043) and the three mutants V₄₁L/V₄₄L, N₄₃Q and V₄₁L/V₄₄L/A₇₇-L/N₇₈-Q were obtained as synthetic genes from Geneart (Thermo Fisher Scientific). To generate constructs for cell-free protein expression, the synthetic genes were introduced into the expression vector pIVEX_GAA_Omega, carrying the T7 promoter and the sequence encoding for the His₆ tag [16], using the restriction sites NcoI/BamHI, resulting in the vectors pIVEX_GAA_Omega_PR-10, pIVEX_GAA_Omega_V₄₁-L/V₄₄-L, pIVEX_GAA_Omega_N₄₃-Q and pIVEX_GAA_Omega_V₄₁-L/V₄₄-L/A₇₇-L/N₇₈-Q, respectively.

2.5 Cell-free recombinant protein expression in tobacco cell lysate

For *in vitro* translation of the three Gly m 4 mutants, 200 µL plasmid DNA representing each construct (1 µg/mL) was added to the tobacco BY-2 cell lysate (BYL) [17]. The coupled transcription-translation reactions were carried out for 48 h in 5-mL reaction volumes.

2.6 Purification of recombinant allergen mutants

After centrifugation of the cell-free reaction (40,000 x g, 10 min, 4°C) the clear supernatant (4 mL) was loaded onto a 1-mL Sepharose HiTrap Column (GE Healthcare, Germany) and the resin was freshly charged with 1 mL of 200 mM nickel solution. The column was washed with 20 column volumes (cv) of washing buffer (PBS, pH 7.4) and the allergen mutants were eluted in PBS containing a gradient of 0–250 mM imidazole (pH 8.0) over 30 cv. The 1-mL elution fractions were collected and analyzed by Coomassie stained SDS-PAGE for the presence, integrity and purity of the protein. Sample concentrations were quantified using the Bradford assay [18].

2.7 Purification of serum IgE

To reduce the nonspecific background of the immunological assays and increase the IgE antibody concentration, 10 serum samples from patients with a documented soy allergy were pooled and purified by affinity chromatography. Due to the limited amounts of sera, it was not possible to refer to the sera used for the epitope analysis, and a new serum pool of patients with documented soy allergy was used. The IgE-specific antibody Xolair (Novartis, Basel, Switzerland) was coupled to an NHS-activated Sepharose (GE Healthcare) according to the manufacturer's instructions. Purification was carried out using 2.5 mg antibody per mL matrix.

2.8 ELISA

ELISA plates were coated with 100 µl per well of the Gly m 4 variants (2 µg/mL in PBS). After blocking nonspecific binding sites with 1% (w/v) BSA, serial dilutions of either rabbit anti-Gly m 4 serum with a starting dilution of 1:200 (generated by immunization with recombinant, plant-derived Gly m 4), or serial dilutions of the purified IgE antibody fraction with a starting dilution of 1:5 were added. The rabbit serum was detected with a horse-radish-peroxidase HRP-conjugated goat-anti-rabbit polyclonal antibody (Jackson ImmunoResearch), whereas the IgE antibody was detected with rabbit-anti-human IgE (Thermo Fisher Scientific) and HRP-conjugated goat-anti-rabbit polyclonal antibody (Jackson

Immunoresearch). The color reaction used the chromogen 2'2'-azino-di-(3-ethylbenzthiazoline-6-sulfonic acid) in the appropriate buffer (Roche, Mannheim, Germany) according to the manufacturer's instructions and absorbance was measured at 405 nm. The binding data were evaluated by determining the reactivity by linear regression in a signal intensity range from 0.3 to 0.6 intensity units.

2.9 Determination of the melting temperature (T_M)

The potential impact of the mutations on protein folding and stability was addressed by comparative melting curve analysis using allergen concentrations of 300 $\mu\text{g/mL}$. A volume of 12.5 μL of each sample was mixed with 7.5 μL SyproOrange™ (Thermo Fisher Scientific) and the mixture was heated from 20°C to 90°C at a rate of 0.5°C/s. Fluorescence produced by SyproOrange™ interacting with hydrophobic protein regions after enfolding the protein was measured at an excitation wavelength of 520 nm and an emission wavelength of 548 nm. The signal intensity was displayed over the temperature gradient, and the steepest slope was defined as the T_M .

2.10 Mast cell stimulation assay

RBL SX-38 cells (kindly provided by Dr. Jean-Pierre Kinet, Harvard Institute of Medicine, Boston, MA, USA) were cultured as previously described [19]. The cells were seeded in 96-well plates (1×10^5 cells/well) and incubated overnight (37°C, 5% CO_2) to allow the growth of adhered cells. After washing the cells with PBS, we added 100 μL of affinity-purified IgE antibody fraction from Gly m 4-reactive soy allergy patient sera, diluted 1:5 in culture medium without fetal calf serum, and the mixture was incubated overnight. After another washing step with Na-PIPES buffer (25 mM disodium-PIPES, 100 mM NaCl, 5 mM KCl, 0.4 mM MgCl_2 , 1 mM CaCl_2 , 5.6 mM glucose, 0.1% (w/v) BSA, pH 7.0) the cells were stimulated with 50 μL of the antigen diluted in Na-PIPES buffer (concentration 1 $\mu\text{g/mL}$). Alternatively, we used soy aqueous extract (1:50 dilution), prepared by grinding soybean

flakes (kindly provided by Pia Meinschmidt, Fraunhofer IVV, Munich) in three volumes of water, centrifuging at 16,000 x g for 15 min at 4°C and diluting the supernatant in Na-PIPES buffer.

Serially-diluted IgE antibody (starting concentration 1.25 µg/mL, Abcam) was used as a positive control for cell sensitization, whereas 2.5 µg/mL rabbit anti-human IgE polyclonal antibody (Thermo Fisher Scientific) was used for FcεRI-mediated stimulation of mast cells. Complete degranulation of the cells was provoked by addition of 50 µL of a 0.1% (v/v) Triton-X-100 solution. A β-hexoaminidase color reaction was used as an indicator for cell degranulation. Supernatant (30 µL) was incubated with 50 µL 4-nitrophenyl N-acetyl-β-D-glucosaminide (1.3 mg/mL in citrate buffer, pH 4.5) for 1 h at 37°C. After neutralization with 100 µL 50 mM carbonate buffer (pH 10) the color density was determined by absorbance reading at 405 nm. Relative degranulation was calculated as ratio of sample signal intensity and the average of signal intensity from the Triton X-100 treatment and multiplied by 100 to generate percentage.

2.11 Ethical statement

The human blood samples from soy allergy patients used in this study were provided by the University of Leipzig (Department of Dermatology), Philipp-Rosenthal-Straße 23, 04130 Leipzig. The use of the samples for this study was authorized by the responsible ethical committee (EK 428/12-17122012, 21.1.2013). The donors were informed that the blood samples would be used for research purposes only. For each donor, a statement of informed consent has been signed.

3 Results

3.1 Identification of IgE-binding epitopes within the allergen Gly m 4

IgE-binding epitopes in the major soybean allergen Gly m 4 were analyzed in 15 serum samples from patients with soy allergy as identified by positive prick test to soy drink using an IgE-specific peptide microarray. The Gly m 4 peptides were presented as C-terminal bound 15-mers with four-residue overlap and all sera analyzed reacted with at least one peptide of Gly m 4 in the peptide microarray. In total 19 of the 37 displayed peptides were recognized by the patients, and two sequence regions from AA₃₇₋₆₃ and AA₆₄₋₈₃ showed the highest reactivity (Fig. 1A). Region AA₃₇₋₆₃ was detected by IgE in seven patients' sera and region AA₆₄₋₈₃ even by IgE from eleven of the 15 patients' sera analyzed. Furthermore for the peptides AA₁₁₇₋₁₃₅ we observed reactivity for six patients' sera. Figure 1B shows the position of these three sequence areas identified by peptide micro array in the Gly m 4 3D-structure (PDB #2K7H) in correspondence to the color-code used in Figure 1A, confirming their location on the allergen surface.

The 15 patients' sera analyzed by peptide micro array as well as 32 additional sera from soy allergic patients were screened against a random 16-mer peptide phage display library featuring a complexity of 5×10^9 phages. The phages were enriched in two rounds of panning on serum coated to Immuntubes® and analyzed by next generation sequencing, yielding in a minimum of 170,000 peptide sequences for each selection experiment. The LibDB databases' statistical data obtained from each experiment and data from the naïve library (control) were used to identify potential epitopes. Several fold enrichment of motifs in the data sets especially in the second selection round over the calculated value and that observed in the naïve library revealed motifs as fragments of six potential epitopes (Fig 1A). Each of these epitopes could be found in data sets from several different patients. Presence in at least two investigated data sets was a criterion for the candidate sequence to be a soy allergy related epitope. Figure 1C shows the position of the six sequence motifs identified by phage display analysis in the Gly m 4 3D-structure (PDB #2K7H) in correspondence to the color-code used in Figure 1A, confirming their location on the allergen surface. To illustrate the phage sequence alignment and the discovery of consensus sequences ten phage peptides representing two motifs are

shown in Figure 2A. Bold letters indicate homologous amino acids to the Gly m 4 sequence and colored letters indicate the consensus motif of the phages. Figure 2B graphically represents in a WebLogo [20] the consensus sequence for all phages enriched for the NVEG (781 phages identified in four patients) or IDEA (1047 phages identified in six patients) motif. This kind of visualization makes clear that the NVEG and the IDEA motif are highly conserved amino acids for the antibody binding whereby the surrounding amino acids can be variable and may have only minor impact on the antibody binding.

Although the phage display analysis reveals detailed information about epitope amino acids essential for antibody binding, the technique is not able to discriminate between different antibody isotypes and always represents the sum of B-cell epitopes. Therefore we used the combination of data from the IgE-specific peptide array and the high accuracy specification of the epitope's amino acids by phage display analysis to describe the IgE-epitopes of Gly m 4. The two sequence regions (AA₃₇₋₅₅ and AA₆₄₋₈₃) recognized by highest patient number in the peptide micro array were also identified in the phage display analysis and sequences could be narrowed down to the consensus motifs of NVEG (AA₄₃₋₄₆) for the first region and IDEA (AA₇₄₋₇₇) and NLGYSY (AA₇₈₋₈₃) for the second region.

3.2 Design of the Gly m 4 mutants

To confirm whether the sequence motifs NVEG (AA₄₃₋₄₆) and IDEA (AA₇₄₋₇₇)/NLGYSY (AA₇₈₋₈₃) are really recognized by substantial proportions of serum IgE Gly m 4 mutants were constructed. The idea was to increase the surface area of the epitope by insertion of an additional methyl group, to generate sterical hindrance for antibody binding. To maintain the three dimensional structure of the allergen we substituted only one or two amino acids per motif and to retain the chemical properties only uncharged amino acids were substituted. Following these considerations the only possible substitutions were Valin to Leucin, Asparagine to Glutamine or Alanine to Valine.

For the NVEG (AA₄₃₋₄₆) motif a N₄₃-Q mutation was designed and to maximize the effect of surface increase a second mutant where the V₄₄ and additionally the V₄₁ in front of the consensus motif were exchanged with leucine (V₄₁-L/V₄₄-L). To address both consensus motifs in the second sequence region

by one mutant a double amino acid substitution was designed for the IDEA (AA₇₄₋₇₇)/NLGYSY (AS₇₈₋₈₃) motif resulting in an A₇₇-V/N₇₈-Q variant of Gly m 4. To visualize the intended surface increase by the addition of methyl groups the 3D-structures of the three mutations (N₄₃-Q, V₄₁-L/V₄₄-L, A₇₇-V/N₇₈-Q) were modeled using SWISS-MODEL [21] based on the PDB wild-type model (#2K7H) and aligned with the wild-type model using DNASTar Software (Fig. 3).

Figure 3A/B shows the epitopes at atomic resolution and the epitope surface in the three-dimensional structural model. Substituted amino acids are highlighted in blue to show the greater area occupied by the additional methyl group. The V₄₁-L/V₄₄-L replacement increased the surface area by 14 Å² compared to the wild-type epitope, whereas the N₄₃-Q replacement increased the surface area by 32 Å². Surface area analysis confirmed the visual impression that the mutants increased the space occupied by the epitope, although interestingly the single amino acid substitution N₄₃Q increased the spatial dimensions of the epitope more than the double substitution V₄₁-L/V₄₄-L. Figure 3C shows the A₇₇-L/N₇₈-Q substitutions in green, and the overall increase in surface area from 25 Å², as predicted by homology modelling.

3.3 Expression and immunological characterization of Gly m 4 Mutants

To verify the influence of the in silico modeled surface increase on the recognition by Gly m 4-specific antibodies (from immunized rabbits and patient sera) expression constructs were generated for three Gly m 4 mutants. The first mutant bearing the N₄₃-Q mutation (N₄₃-Q), the second the V₄₁-L/V₄₄-L mutation (V₄₁-L/V₄₄-L) and the third bearing the V₄₁-L/V₄₄-L mutation in the first sequence region and the A₇₇-V/N₇₈-Q mutation in the second region (V₄₁-L/V₄₄-L/A₇₇-V/N₇₈-Q). The wild-type Gly m 4 allergen and the three mutated versions were expressed in 5 mL batches in the cell-free BYL system and purified by IMAC with gradient elution, resulting in a yield of 500 µg recombinant protein with a purity greater than 95% as verified by SDS-PAGE (Fig. 4A). Melting point analysis was carried out to compare the structural integrity of the mutants compared to the recombinant wild-type Gly m 4. As shown in Fig. 4B, the T_M of the wild-type Gly m 4 was 60.0°C, in good agreement with the 60–61°C range previously reported [22, 23]. The T_M of mutants V₄₁-L/V₄₄-L and V₄₁-L/V₄₄-L/A₇₇-L/N₇₈-Q was

slightly increased to wild-type Gly m 4 (60.5°C) but N₄₃-Q was less thermodynamically stable, with a T_M of 57.0°C.

Next we analyzed the binding of Gly m 4-specific polyclonal rabbit serum or a pool of IgE antibodies purified from soy allergy patient sera to the three mutant variants. The binding of Gly m 4-specific rabbit serum to mutants V₄₁-L/V₄₄-L, N₄₃-Q or V₄₁-L/V₄₄-L/A₇₇-L/N₇₈-Q was reduced by 14%, 15% or 40%, respectively, compared to the wild-type allergen (Fig. 5A). The binding of IgE antibodies purified from patient sera to mutant N₄₃-Q was reduced by 54% and for the mutants V₄₁-L/V₄₄-L and V₄₁-L/V₄₄-L/A₇₇-L/N₇₈-Q by 43% and 63%, respectively, so the IDEA (AA₇₄₋₇₇)/NLGYSY (AS₇₈₋₈₃) motif was also confirmed as a relevant target for IgE binding (Fig. 5A).

In addition the immunological effect of the mutated IgE-binding epitopes was functionally compared to the wild-type allergen using a mast cell degranulation assay, revealing the same trend as the binding assays. The wild-type allergen induced 25% degranulation, whereas for the V₄₁-L/V₄₄-L, N₄₃-Q and V₄₁-L/V₄₄-L/A₇₇-L/N₇₈-Q mutants this was reduced to 18%, 22% and 13%, respectively (Fig. 5B).

4. Discussion

Epitope analysis using an IgE-specific peptide array and refined by data from epitope mapping using phage peptide display allowed the high-resolution mapping of three IgE-binding epitopes in the soybean allergen Gly m 4. Visualization of the epitopes in the context of the Gly m 4 three-dimensional structure confirmed that the three epitopes are located on the surface of the allergen and are thus accessible for antibody binding. The epitopes were further confirmed by the generation of epitope-specific Gly m4 mutants followed by functional analysis.

The here identified consensus motif NVEG (AA₄₃₋₄₆) is located in the so-called P-loop, a highly conserved region among PR-10 proteins from different species and is highly homologous to the Bet v 1 sequence NVIG (AA₄₃₋₄₆). The importance of this region for IgE binding was described for the birch pollen allergen by a co-crystallization with the mouse antibody BV-16, which is able to inhibit IgE binding to Bet v 1 by 40%. Bet v 1 allergen revealed binding of BV-16 to a conformational epitope in

this region [24]. Furthermore, it has been shown that a site directed mutagenesis of Glu₄₅ resulted in reduced IgE binding for the allergen Bet v 1 and the cheery allergen Pru v 1 [5, 25]. Even though the epitope was described in detail, they could not compete for binding with a homologous peptide and concluded, that it is a conformational epitope [5].

In our study we have identified the dominant linear epitope motif NVEG (AA₄₃₋₄₆) as part of a larger sequence region described in literature as discontinuous epitope [24]. One or two amino acid substitutions in this region (N₄₃-Q, V₄₁-L/V₄₄-L) caused a 50% reduction in IgE binding, in agreement with the earlier report discussed above for the birch pollen allergen Bet v 1 [5].

Additionally we identified a second Gly m 4 sequence motif IDEA (AA₇₄₋₇₇)/NLGYSY (AS₇₈₋₈₃), which was recognized by 12 out of 15 patients. Its relevance was confirmed in the V₄₁-L/V₄₄-L/A₇₇-L/N₇₈-Q mutant, which reduced the binding activity of IgE antibodies even further than the two NVEG mutants (N₄₃-Q and V₄₁-L/V₄₄-L) alone. This sequence motif shows low homology to the Bet v 1 sequence (AA₇₄₋₈₂ EVDHTNFKYN), which has only minor impact on IgE binding compared to the P-Loop region [26]. In another study monoclonal anti-Bet v 1 antibodies binding to a peptide from AA₇₃₋₁₀₃ inhibited the basophil activation with Bet v 1 in birch pollen allergic patients [26]. Furthermore epitope mimotopes were identified by phage display screening where the phage elution was specific for different allergens [27]. They subdivided the Bet v 1 sequence into relevant amino acids stretches and rated the frequency of mimotope peptides identified for this regions. Besides a strong recognition of phages mimicking the P-loop region relevance for the N-terminus, the C-terminus and the region from AA₇₃₋₈₃ was shown [27]. Further authors identified for three out of five patients phages that mimic the sequence from AA₁₂₀₋₁₂₉, which is in good agreement with the results from our study, as we identified the ETKGD motif (AA₁₂₁₋₁₂₅) either in the peptide array and the phage display analysis. As this region was not included in the mutagenesis experiments, it might be reasonable for the incomplete inhibition of IgE binding for the V₄₁-L/V₄₄-L/A₇₇-L/N₇₈-Q mutant.

The C-terminus of Bet v 1 has been identified in several studies as a conformational region for IgE binding of patients with birch pollen allergy [13, 27]. As the C-terminus of Gly m 4 (AA₁₄₆₋₁₅₈) shows 70% sequence match to Bet v 1 the epitope might be also relevant for Gly m 4, but was not identified using

our approach. This might be reasoned by the structure of the C-terminus, which adopts an α -helix and could with regard to the experiments performed for the peanut allergen Ara h 2 where the allergen showed a stronger signal on peptide arrays comprising 20-mer rather than 15-mer peptides, not be displayed in the peptide micro array [28].

The potential impact of the mutations on the allergen structure was investigated by melting point analysis, revealing a melting point of 60°C for wild-type Gly m 4 and the V₄₁-L/V₄₄-L mutant, a slightly higher T_M of 60.5°C for the V₄₁-L/V₄₄-L/A₇₇-L/N₇₈-Q mutant, but a lower T_M of 57°C for the N₄₃Q mutant. The melting point of wild-type Gly m 4 was previously reported as 60°C, determined by circular dichroism spectroscopy [22, 23]. Therefore, the NVEG (AA₄₃₋₄₆) V₄₁-L/V₄₄-L and IDEA (AA₇₄₋₇₇)/NLGYSY (AS₇₈₋₈₃) A₇₇-L/N₇₈-Q mutations do not affect the structure and/or thermodynamic stability of Gly m 4 and maintain the three-dimensional surface structure of the allergen, facilitating the presentation of authentic B-cell epitopes to ensure a specific IgG response towards the native allergen during allergen-based immunotherapy. These findings are in agreement with the data reported by Holm et al [29] in the context of the generation of engineered Bet v 1 mutants featuring up to nine point mutations while retaining their protein folding pattern. Our mast cell degranulation assay showed that the Gly m 4 mutants were less efficient than expected in terms of inhibiting the release of β -hexoaminidase given the clear reduction of IgE antibody binding activity shown in the ELISA experiments. A possible explanation for this observation is provided by a study, showing that one high-affinity antibody combined with another high-affinity or even a low-affinity antibody can be sufficient to crosslink receptors, resulting in cell degranulation [30]. Therefore, a low-affinity antibody probably not detectable stringent in ELISA could have a major impact in the mast cell degranulation assay. It thus seems reasonable to include cell stimulation assays as selection criteria for protein variants with low allergenicity, to exclude ELISA-specific effects and to include effects based on the epitope distribution on the allergen surface and the interactions within the Fc ϵ RI-IgE-antigen complex.

In summary, we used a powerful combination of epitope analysis techniques and a cell-free expression system derived from tobacco cells [16] for the rapid production, purification and characterization of recombinant allergens. In contrast to bacterial systems plant-based production platforms are

advantageous for plant-derived food allergens because native signaling motifs can be kept in the sequence and lead to an authentic protein processing and post-translational modifications such as disulfide bridges and plant-specific glycosylation profiles [17]. Although IgE binding was not completely inhibited, indicating that not all existing epitopes were modified, our approach was successfully applied to identify different Gly m 4 epitopes recognized by IgE antibodies and map them due to the combination of IgE specific peptide array and a random phage display library at high resolution. Especially the combination of both techniques enables the identification of linear core epitope regions formerly described as conformational epitopes [13, 26]. In medical diagnostic applications, the identification of small and specific epitopes within larger protein sequences can be valuable to understand reactions in individual patients and to distinguish between IgE and IgG antibody binding. Furthermore, the high-resolution epitope mapping facilitated the creation of mutants as potential allergen-based immunotherapy candidates, offering reduced allergenicity while maintaining structural homology to the wild-type allergen.

Acknowledgments

We acknowledge Dr. Stefan Rasche for assistance with protein purification, Michael Küpper for melting point analysis and Christoph Kühn for cell culture work. We thank Dr Richard M Twyman for manuscript editing. This work was supported by Fraunhofer Gesellschaft through the MAVO project LowAllergen.

5 References

- [1] Klemola, T., Vanto, T., Juntunen-Backman, K., Kalimo, K., *et al.*, Allergy to soy formula and to extensively hydrolyzed whey formula in infants with cow's milk allergy: a prospective, randomized study with a follow-up to the age of 2 years. *J Pediatr* 2002, *140*, 219-224.
- [2] Savage, J. H., Kaeding, A. J., Matsui, E. C., Wood, R. A., The natural history of soy allergy. *J Allergy Clin Immunol* 2010, *125*, 683-686.

- [3] Schmidt, G., Gadermaier, G., Pertl, H., Siegert, M., *et al.*, Production of recombinant allergens in plants. *Phytochemistry reviews : proceedings of the Phytochemical Society of Europe* 2008, 7, 539-552.
- [4] Berneder, M., Bublin, M., Hoffmann-Sommergruber, K., Hawranek, T., Lang, R., Allergen Chip Diagnosis for Soy-Allergic Patients: Gly m 4 as a Marker for Severe Food-Allergic Reactions to Soy. *Int Arch Allergy Immunol* 2013, 161, 229-233.
- [5] Spangfort, M. D., Mirza, O., Ipsen, H., Van Neerven, R. J., *et al.*, Dominating IgE-binding epitope of Bet v 1, the major allergen of birch pollen, characterized by X-ray crystallography and site-directed mutagenesis. *J Immunol* 2003, 171, 3084-3090.
- [6] Mishra, A., Jain, A., Arora, N., Mapping B-cell epitopes of major and minor peanut allergens and identifying residues contributing to IgE binding. *J Sci Food Agric* 2016, 96, 539-547.
- [7] Longo, V., Costa, M. A., Cibella, F., Cuttitta, G., *et al.*, Multiple IgE recognition on the major allergen of the Parietaria pollen Par j 2. *Mol Immunol* 2015, 63, 412-419.
- [8] Hansen, C. S., Dufva, M., Bogh, K. L., Sullivan, E., *et al.*, Linear epitope mapping of peanut allergens demonstrates individualized and persistent antibody-binding patterns. *J Allergy Clin Immunol* 2016.
- [9] Christiansen, A., Kringelum, J. V., Hansen, C. S., Bogh, K. L., *et al.*, High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum. *Sci Rep* 2015, 5, 12913.
- [10] Uermösi, C., Zabel, F., Manolova, V., Bauer, M., *et al.*, IgG-mediated down-regulation of IgE bound to mast cells: a potential novel mechanism of allergen-specific desensitization. *Allergy* 2014, 69, 338-347.
- [11] Ferreira, F., Ebner, C., Kramer, B., Casari, G., *et al.*, Modulation of IgE reactivity of allergens by site-directed mutagenesis: potential use of hypoallergenic variants for immunotherapy. *Faseb J* 1998, 12, 231-242.
- [12] Jutel, M., Jaeger, L., Suck, R., Meyer, H., *et al.*, Allergen-specific immunotherapy with recombinant grass pollen allergens. *J Allergy Clin Immunol* 2005, 116, 608-613.
- [13] Gepp, B., Lengger, N., Bublin, M., Hemmer, W., *et al.*, Chimeras of Bet v 1 and Api g 1 reveal heterogeneous IgE responses in patients with birch pollen allergy. *J Allergy Clin Immunol* 2014, 134, 188-194.

- [14] Ma, Y., Gadermaier, G., Bohle, B., Bolhaar, S., *et al.*, Mutational analysis of amino acid positions crucial for IgE-binding epitopes of the major apple (*Malus domestica*) allergen, Mal d 1. *Int Arch Allergy Immunol* 2006, 139, 53-62.
- [15] Szardenings, M., *J Recept Sig Transd* 2003, pp. 307-349.
- [16] Buntru, M., Vogel, S., Stoff, K., Spiegel, H., Schillberg, S., A versatile coupled cell-free transcription-translation system based on tobacco BY-2 cell lysates. *Biotechnol Bioeng* 2015, 112, 867-878.
- [17] Buntru, M., Vogel, S., Spiegel, H., Schillberg, S., Tobacco BY-2 cell-free lysate: an alternative and highly-productive plant-based in vitro translation system. *Bmc Biotechnol* 2014, 14, 1-11.
- [18] Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976, 72, 248-254.
- [19] Wiegand, T. W., Williams, P. B., Dreskin, S. C., Jouvin, M. H., *et al.*, High-affinity oligonucleotide ligands to human IgE inhibit binding to Fc epsilon receptor I. *J Immunol* 1996, 157, 221-230.
- [20] Crooks, G. E., Hon, G., Chandonia, J. M., Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res* 2004, 14, 1188-1190.
- [21] Arnold, K., Bordoli, L., Kopp, J., Schwede, T., The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006, 22, 195-201.
- [22] Guhsl, E. E., Hofstetter, G., Hemmer, W., Ebner, C., *et al.*, Vig r 6, the cytokinin-specific binding protein from mung bean (*Vigna radiata*) sprouts, cross-reacts with Bet v 1-related allergens and binds IgE from birch pollen allergic patients' sera. *Mol Nutr Food Res* 2014, 58, 625-634.
- [23] Petersen, A., Rennert, S., Kull, S., Becker, W. M., *et al.*, Roasting and lipid binding provide allergenic and proteolytic stability to the peanut allergen Ara h 8. *Biol Chem* 2014, 395, 239-250.
- [24] Mirza, O., Henriksen, A., Ipsen, H., Larsen, J. N., *et al.*, Dominant epitopes and allergic cross-reactivity: complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. *J Immunol* 2000, 165, 331-338.
- [25] Neudecker, P., Lehmann, K., Nerkamp, J., Haase, T., *et al.*, Mutational epitope analysis of Pru av 1 and Api g 1, the major allergens of cherry (*Prunus avium*) and celery (*Apium graveolens*): correlating IgE reactivity with three-dimensional structure. *Biochem J* 2003, 376, 97-107.

[26] Gieras, A., Cejka, P., Blatt, K., Focke-Tejkl, M., *et al.*, Mapping of conformational IgE epitopes with peptide-specific monoclonal antibodies reveals simultaneous binding of different IgE antibodies to a surface patch on the major birch pollen allergen, Bet v 1. *J Immunol* 2011, *186*, 5333-5344.

[27] Mittag, D., Batori, V., Neudecker, P., Wiche, R., *et al.*, A novel approach for investigation of specific and cross-reactive IgE epitopes on Bet v 1 and homologous food allergens in individual patients. *Mol Immunol* 2006, *43*, 268-278.

[28] Shreffler, W. G., Lencer, D. A., Bardina, L., Sampson, H. A., IgE and IgG4 epitope mapping by microarray immunoassay reveals the diversity of immune response to the peanut allergen, Ara h 2. *J Allergy Clin Immun* 2005, *116*, 893-899.

[29] Holm, J., Gajhede, M., Ferreras, M., Henriksen, A., *et al.*, Allergy vaccine engineering: epitope modulation of recombinant Bet v 1 reduces IgE binding but retains protein folding pattern for induction of protective blocking-antibody responses. *J Immunol* 2004, *173*, 5258-5267.

[30] Christensen, L. H., Holm, J., Lund, G., Riise, E., Lund, K., Several distinct properties of the IgE repertoire determine effector cell degranulation in response to allergen challenge. *J Allergy Clin Immunol* 2008, *122*, 298-304.

Figure legends

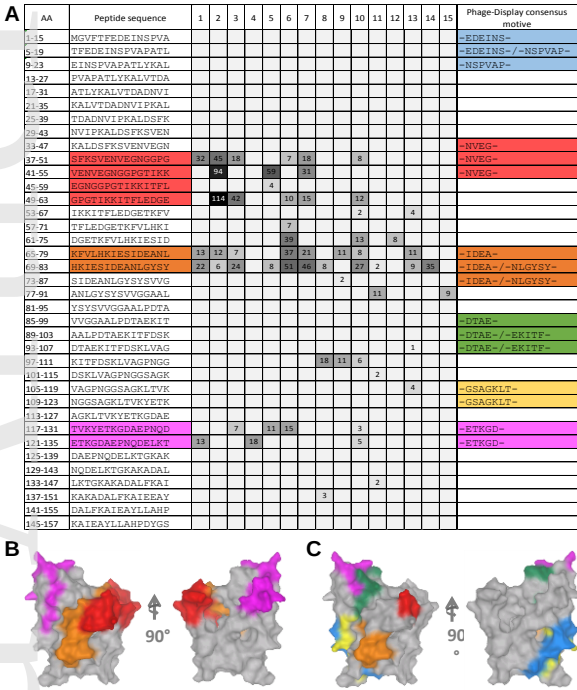


Figure 1. Results of the IgE specific peptide micro array and the random peptide phage display analysis.

A: Peptides coupled to the peptide micro array are listed on the left side of the table. Serum antibody reactivities to the peptides in the micro array are shown with relative signal intensity in the middle of the table. Most relevant peptide sequences are marked with colored background. In the right column those consensus motifs identified from peptide phage display data are listed according to their presence in the micro array peptide amino acid sequences in the left column. Most relevant peptide sequences are marked with colored background. B: Three-dimensional structure of Gly m 4 (PDB #2K7H) showing the three regions identified by peptide microarray analysis in different colors: region AA₃₇₋₆₃ is red, AA₆₄₋₈₃ is orange and AA₁₁₇₋₁₃₅ is purple. C: Consensus motifs identified by phage display are shown in the three-dimensional Gly m 4 structure (PDB #2K7H) using the same color code as in panel A.



Figure 2. Two examples for the identification of consensus motifs by phage display analysis with a random phage peptide library

A: Each example shows an alignment of selected Gly m 4-related phage sequences, which were enriched during panning. Bold letters indicate sequence homology to the Gly m 4 sequence and colored letters indicate the consensus motif for the group of phages. Amino acids used for single point mutation are marked with an asterisk and the name of the corresponding mutant is indicated. B: Sequence logo generated by the WebLogo Software [20] to display the consensus sequence for 781 phage sequences (257 unique sequences) for the NVEG motif and 1047 phages (375 unique sequences) for the IDEA motif. The height of the symbols indicates the relative frequency of the corresponding amino acid in the set of analyzed sequences. Basic amino acids are displayed in blue, acidic amino acids in red, neutral amino acids in purple and green and hydrophobic ones in black.

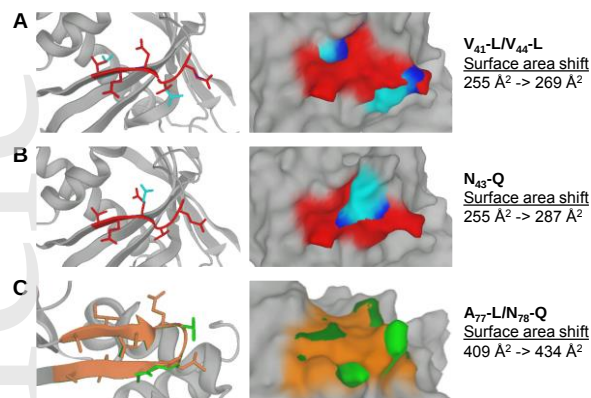


Figure 3. Visualization of the putative impact of epitope mutations on the surface morphology and area of wild-type Gly m 4 using overlays and homology models of the mutant allergens.

The left panel shows the backbone sections and the side chains of each epitope, the central panel shows the predicted surface of the epitope, and the right panel provides the calculated surface area shifts. A: The VENVEGN epitope (AA₄₁₋₄₇) is shown in red and the V₄₁-L/V₄₄-L mutations are highlighted in light blue. B: The N₄₃-Q mutation is shown in blue in the same context as described in panel A. C: Motif IDEA (AA₇₄₋₇₇)/NLGYSY (AS₇₈₋₈₃) is shown in orange, and amino acid substitutions in the A₇₇-L/N₇₈-Q mutations are highlighted in green.

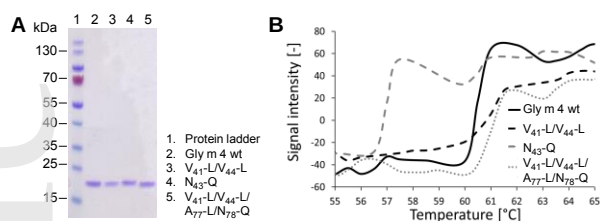


Figure 4. Coomassie stained SDS-Page of purified protein fractions and T_M values of wild-type Gly m 4 and the three mutant variants.

A: Coomassie stained SDS-Page under reducing conditions of the cell-free produced wild-type Gly m 4 protein and the three mutant variants purified by IMAC. B: The temperature gradient induced unfolding was monitored using SyproOrange™ fluorescence. The fluorescence signal intensity was plotted against the temperature. T_M was determined from the steepest slope.

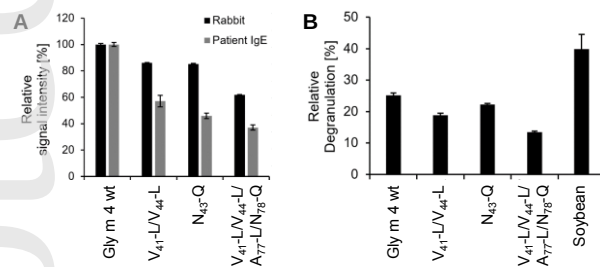


Figure 5. Immunological analysis of wild-type Gly m 4 and the three mutants by ELISA and mast cell degranulation assay.

A: The reactivity of a rabbit anti-Gly m 4 serum is displayed as relative values in black. The relative reactivity of soy allergy patient IgE is plotted in grey. Data represent means and standard deviations ($n = 3$). B: Mast cell degranulation in RBL-SX 38 cells sensitized with soy allergy patient IgE and stimulated with wild-type Gly m 4, the three mutants and an aqueous soy extract. Data show values relative to total degranulation induced by the combination of IgE and anti-IgE antibodies. Data represent means and standard deviations ($n = 3$).