

Identifizierung potentiell allergener Soja-Epitope mittels Peptid  
Phage Display

Dissertation  
zur Erlangung des akademischen Grades  
„doctor rerum naturalium“  
(Dr. rer. nat.)

vorgelegt dem Rat der Fakultät für Biowissenschaften  
der Friedrich-Schiller-Universität Jena

von Karolin Kern (M.Sc.)

geboren am 11.07.1986 in Lutherstadt Wittenberg

Gutachter:

Prof. Dr. Stefan Lorkowski

Friedrich Schiller Universität Jena

Prof. Dr. Stefan Dübel

Technische Universität Braunschweig

Prof. Dr. med. Regina Treudler

Universitätsklinikum Leipzig Klinik und Poliklinik für Dermatologie,  
Venerologie und Allergologie

Datum der Verteidigung

07.01.2020

## Abstract

This thesis was carried out as part of the MAVO project „*LowAllergen*“, which has dealt with detailed identification of soy related epitopes. The precise mapping of antibody epitopes allows a more detailed and differentiated understanding of immune-related diseases. It may lead to the development of novel therapies and diagnostic tools. Here, we describe the identification of new or more confined epitopes in food allergy-associated soy proteins. By the combination of an optimized peptide-phage display library with next-generation sequencing, sophisticated *in silico* data analysis and subsequent peptide microarray analysis it was possible to identify 405 potential epitope motifs in 14 soybean proteins. More than 60% of them have not yet been described as potential allergens. Epitopes are described in soy proteins, which have not been known as potentially allergenic before. Ten µl of patient serum is sufficient for epitope characterization followed by validation using microarray. This approach to applying the peptide phage display is a successful method for the identification of large numbers of valid epitopes, as is shown here. Eighty-three peptides, representing the 42 most frequently found epitope candidates, were validated by microarray analysis using 49 sera from patients who had been tested positive in skin prick test (SPT). Of these, 56 were bound by antibodies either by serum IgE or by serum IgG antibodies or by both. Individual epitope patterns were found for each patient and protein. The combination of a few peptides meets the criteria for the characterization of patient sera. Microarray analysis is the method of choice for the detection of individual patient's epitope patterns as well as for the sensitization pattern, for example for geographical variations of patients. As seen in the example of Bet v 1 and Cor a 1, an advantage of this method is the identification of cross-reactive antibodies and multi sensitization. The epitope resolved analyses reveal a high prevalence of IgE binding to certain epitopes from patients with clinical symptoms. Evaluation of individual immune profiles of patients with soy sensitization allows the identification of peptides with the potential to broadly study individual IgE/IgG binding to epitopes for the first time.

## Zusammenfassung

Mit der detaillierten Identifizierung von Epitopen der Sojabohne befasste sich das MAVO-Projekt *LowAllergen*. Die präzise Charakterisierung der Antikörperpitope erlaubt ein besseres und differenziertes Verständnis von Immunerkrankungen. Die hier vorgestellten Ergebnisse werden das Verständnis der Allergie an sich und die Nahrungsmittelallergie-Diagnostik zu verbessern.

In dieser Arbeit wird die Identifizierung von neuen oder bereits publizierten Epitopen in hoher Auflösung in Nahrungsmittelallergie-assoziierten Sojaproteinen beschrieben. Durch die Kombination einer optimierten Peptide Phage Display Bibliothek mit Next Generation Sequencing (NGS), speziell entwickelten *in silico* Datenanalyse und anschließenden Microarray Analyse war es möglich insgesamt 405 Epitope in 14 Sojaproteinen zu identifiziert. Mehr als 60 % von diesen werden noch nicht als potentielle Allergene beschrieben. Es wurden auch Epitope in Sojaproteinen identifiziert, die vorher noch nicht als Allergene beschrieben werden. Für die Charakterisierung der Epitope und dessen spätere Validierung im Microarray wurden lediglich ca. 10 µl Patientenserum benötigt.

Dieses ist eine effiziente Methode zahlreiche Epitope zu beschreiben. Mit 49 Seren, die ein positives Ergebnis im Prick-Test hatten, wurden 83 Peptide, die 42 der häufigsten gefundenen Epitope repräsentieren, im Microarray getestet. Von diesen wurden 56 Peptide von Antikörpern erkannt, entweder nur von Serum IgE, nur von Serum IgG oder von beiden Antikörpern. Es stellte sich heraus, dass eine Kombination von wenigen Peptiden ausreicht, um die Allergien genauer zu charakterisieren. Die Ergebnisse zeigen, dass der Microarray für die Detektion von individuellen Epitopmustern der Patienten, sowie das Sensibilisierungsmuster für beispielsweise geografische Varianten bei den Patienten, die Methode der Wahl ist. Wie am Beispiel von Bet v 1 und Cor a 1 gezeigt werden konnte, ist ein großer Vorteil die Identifizierung von kreuzreaktiven Antikörpern und Multisensibilisierung. Die epitop-aufgelöste Analyse zeigt eine hohe Prävalenz der IgE-Bindung zu bestimmten Epitope der Sojabohne mit klinischen Reaktionen. Die Evaluierung des individuellen Profils des Immunsystems der Patienten mit Sensibilisierung gegen Sojabohnen erlaubt zum ersten Mal die Identifizierung von Peptiden mit dem Potenzial das breite individuelle IgE/IgG Bindungsmuster von Epitopen zu studieren.

## Inhaltsverzeichnis

I.	Abkürzungsverzeichnis .....	5
II.	Verzeichnis eingetragener Warenzeichen.....	6
III.	Abbildungsverzeichnis.....	7
IV.	Tabellenverzeichnis .....	7
1.1	Allergie Allgemein.....	9
1.1.1	Die Familie der Immunglobulinproteine .....	9
1.1.1.1	Antikörper Epitope .....	11
1.1.2	Epitopidentifizierung .....	11
1.1.3	Entstehung einer Allergie .....	12
1.1.4	Kreuzreaktionen .....	13
1.1.5	Pflanzenbasierte Allergene.....	14
1.1.5.1	Prolamin Familie.....	14
1.1.5.2	Cupin Familie .....	15
1.1.5.3	Profilin Familie.....	15
1.1.5.4	Bet v 1 / PR10 Familie.....	15
1.1.6	Allergene tierischen Ursprungs .....	16
1.2	Nahrungsmittelallergien.....	16
1.3	Das Orale Allergiesyndrom (OAS).....	18
1.3.1	Diagnose einer Nahrungsmittelallergie.....	18
1.3.1.1	Nahrungsmittelprovokation .....	18
1.3.1.2	Bestimmung Allergen-spezifischer Serum IgE-Antikörper .....	19
1.3.1.3	Prick-Test .....	<b>Fehler! Textmarke nicht definiert.</b>
1.3.2	Behandlung einer Nahrungsmittelallergie .....	21
1.4	Soja .....	22
1.4.1	Allergene in Soja .....	22
1.4.2	Kreuzreaktionen bei Sojaallergie.....	24
1.5	Peptid Phage Display .....	25
1.5.1	Der Bakteriophage M13 .....	26
1.5.2	Biopanning.....	26
1.5.3	Peptide Phage Display Bibliotheken.....	27
1.5.4	Epitopidentifizierung mittels Phage Display.....	30
1.6	Zielstellung .....	31
2	Material und Methoden .....	32
2.1	Material .....	32

2.1.1	Mikrobiologisches Material.....	32
2.1.1.1	<i>E. coli</i> -Stämme.....	32
2.1.1.2	Plasmid .....	32
2.1.2	Seren.....	33
2.1.3	Synthetische Peptide.....	35
2.1.4	ENTE-1 Phagenbibliothek .....	37
2.1.5	Chemikalien .....	37
2.1.6	Geräte-Liste .....	38
2.1.6.1	Zentrifugen .....	38
2.1.6.2	Photometer .....	38
2.1.6.3	Micro Plate Reader .....	38
2.1.6.4	Inkubatoren .....	38
2.1.6.5	ArraysScanner .....	38
2.1.7	Computersoftware .....	38
2.1.8	Lösungen .....	39
2.1.8.1	Medien .....	39
2.1.8.2	Puffer.....	39
2.1.8.3	Antibiotika Stammlösungen .....	40
2.1.9	Antikörper .....	40
2.1.10	Polymerase.....	41
2.1.11	Reinigungskits.....	41
2.2	Methoden.....	41
2.2.1	Kultivierung der Bakterien.....	41
2.2.2	Selektionszyklus.....	41
2.2.2.1	„Panning“.....	41
2.2.2.1.1	Immobilisierung des Zielproteins .....	41
2.2.2.1.2	Infektion von TG1 ( $\lambda$ ) F' mit eluierten Phagemid-Partikel.....	42
2.2.2.1.3	Reinigung der Phagemid-Partikel .....	42
2.2.2.2	Titerbestimmung.....	42
2.2.3	Bestimmung der Proteinkonzentration in humanen Serum .....	43
2.2.4	Behandlung der DNA – Extraktion bis Sequenzierung .....	43
2.2.4.1	Isolation und Aufreinigung der DNA .....	43
2.2.4.2	Messung der DNA-Konzentration.....	43
2.2.4.3	Primerhybridisierung für Sequenzierung .....	43
2.2.5	Agarose-Gel-Elektrophorese .....	45
2.2.6	Aufreinigung und Konzentrationsbestimmung der PCR-Produkte .....	45

2.2.7	DNA-Sequenzierung .....	45
2.2.8	Bestimmung der optischen Dichte von Kulturlösungen.....	45
2.2.9	Generierung der Sequenzdaten .....	45
2.2.10	<i>In silico</i> Analyse zur Identifizierung von Epitopen.....	46
2.2.11	Alignment mit den Proteinsequenzen.....	47
2.2.12	Microarray .....	47
2.2.12.1	Messung des Microarray .....	48
3	Ergebnisse .....	49
3.1	Projektübersicht .....	49
3.2	Peptid-Peptid Phage Display .....	50
3.3	Valide Sequenzen der Datensätze.....	51
3.4	<i>In silico</i> Analyse zur Identifizierung von Epitopen.....	54
3.5.4	Aminosäureverteilung innerhalb der Epitope der Sojabohne .....	57
3.4	Untersuchung auf Kreuzreaktionen .....	58
3.5	Validierung der Epitopkandidaten im Microarray.....	60
4	Diskussion.....	70
4.1	Experimentelles Design .....	70
4.1.1	Phagenbibliothek ENTE-1 .....	70
4.1.2	Selektionen .....	72
4.1.3	Next Generation Sequencing.....	73
4.2	Identifizierung der Epitope über Vierermotive .....	73
4.3	Array Datenauswertung .....	76
4.4	Vergleich mit den in der Literatur beschriebenen Soja-Epitopen.....	80
4.4.1	Gly m 5 Epitope .....	80
4.4.2	Gly m 6 Epitope .....	82
4.4.3	Epitope anderer Sojaproteine .....	82
4.5	Kreuzreaktionen .....	83
5	Literaturverzeichnis.....	87
6	Anhang .....	101

I.	Abkürzungsverzeichnis
AS	Aminosäuren
CCDs	cross-reactive carbohydrate determinants, Kreuzreaktive Kohlenhydratepitope
DBPCFC	double-blind, placebo-controlled food challenge doppelblinde, placebokontrollierte orale Provokation
DNA	deoxyribonucleic acid Desoxyribonukleinsäure
ELISA	Enzyme-linked Immunosorbent Assay
FAO	Food and Agriculture Organization Nahrungsmittel und Landwirtschaft Organisation
HBsAg	Hepatitis B Virus surface antigen Hepatitis B Virus -Oberflächenantigen
IgE	Immunglobulin E
IgG	Immunglobulin G
IgM	Immunglobulin M
IL-13-	Interleukin 13
IL-4	Interleukin 4
kDa	Kilodalton
LFDs	lateral flow devices Lateral Flow Test
MW	Molekulargewicht
NGS	Next-Generation-Sequencing Nächste Generation der Sequenzierung
nsLTPs	non-specific lipid transfer proteins nicht spezifische Lipid-Transferproteine
OAS	orales Allergiesyndrom
OFC	offene Nahrungsmittelprovokation
PCR	Polymerasekettenreaktion
pI	isoelektrischer Punkt
PR10	pathogenesis-related proteins 10 Pathogenese-zusammenhängende Proteine 10
PVDF	Polyvinylidenfluorid
RAST	Radio-Allergo-Sorbent-Test



SBPCFC	einfachblinde, placebokontrollierte orale Nahrungsmittelprovokation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis Natriumdodecylsulfat-Polyacrylamidgelelektrophorese
T <sub>H</sub> -Zelle	T <sub>H</sub> Helfer Zelle

II. Verzeichnis eingetragener Warenzeichen

Agilent Turbo-MP

Avanti<sup>®</sup> J-26 XP

Beckman Coulter<sup>®</sup>

DNASTAR<sup>®</sup>

Geneprix<sup>®</sup>

Immulite 2000

Immuno<sup>™</sup> Tubes

Maxisorp<sup>™</sup>

MiSeq<sup>™</sup>

Phadia ImmunoCAP<sup>®</sup>

## III. Abbildungsverzeichnis

Abbildung 1: Zweidimensionales Model eines Immunglobulinmoleküls.....	10
Abbildung 2: Einteilung der Nahrungsmittelunverträglichkeit .....	17
Abbildung 3: Selektionszyklus eines Peptid Phage Display Experimentes.....	27
Abbildung 4: Die absolute Häufigkeit der Aminosäuren in der naiven ENTE-1 Bibliothek.....	29
Abbildung 5: Vektorkarte des pPepPr3A.....	32
Abbildung 6: Ablauf der Arbeiten.....	50
Abbildung 7: Epitopidentifizierung mittels Vergleich der Anreicherungen.....	54
Abbildung 8: Erweiterung des Vierermotivs.....	55
Abbildung 9: Zusammenfassung des Antikörperbindungsmusters.....	67
Abbildung 10: Strukturelles Epitop in Gly m TI.....	71
Abbildung 11: Anreicherung des Motivs „PRAV“ in zwei Selektionsexperimente.....	76
Abbildung 12: Epitopkarte von Gly m 5 .....	81
Abbildung 13: Epitopkarte des Proteins Gly m 2.....	118
Abbildung 14: Epitopkarte von Gly m 4 .....	120

## IV. Tabellenverzeichnis

Tabelle 1: Allergene der Sojabohne .....	23
Tabelle 2: Häufigkeit der Sojaallergie auf einzelne Protein aufgelöst.....	24
Tabelle 3: Liste der humanen Seren, die verwendet worden.....	33
Tabelle 4: Peptidliste .....	35
Tabelle 5: PCR – Vorschrift für einen 50 µl Ansatz.....	44
Tabelle 6: PCR Programm .....	44
Tabelle 7: Sequenzen der verwendeten Primer der PCR für NGS.....	44
Tabelle 8: Übersicht der verwendeten Primer .....	44
Tabelle 9: Übersicht der untersuchten Proteinsequenzen .....	47
Tabelle 10: Übersicht der Konzentration der selektierten Phagen in den Peptid Phage Display Experimenten .....	51
Tabelle 11: Übersicht der erhaltenen Sequenzen nach der Datenprozessierung.....	53
Tabelle 12: Anzahl der identifizierten Epitope .....	56
Tabelle 13: Die relative Häufigkeit der Aminosäuren in den identifizierten Epitopen, in den Proteinsequenzen und der Bibliothek.....	57
Tabelle 14: Identifizierte potentiellen Epitope des Proteins Bet v 1.....	59
Tabelle 15: Zusammenfassung der Microarray Experimente.....	62
Tabelle 16: Liste der identifizierten, potentiellen Epitopkanidaten.....	101

## 1. Einleitung

Nahrungsmittelallergien stellen eine wachsende Bedrohung für viele Menschen dar und vermindern die Lebensqualität durch die notwendigen Einschränkungen bei der Auswahl an Lebensmitteln. 3,6 % der deutschen Bevölkerung reagieren mit klinischen Symptomen auf mindestens ein Lebensmittel. Am häufigsten reagieren die Deutschen auf Haselnüsse und Äpfel [1]. Mittlerweile müssen laut EU-Verordnung Lebensmittelzutaten von 14 Allergenen gekennzeichnet werden [2]. Eine dieser deklarierungspflichtigen Lebensmittelzutaten ist Soja. In Asien wird die Sojabohne schon viele Jahrhunderte verzehrt, aber auch in den westlichen Ländern wird Soja aufgrund seiner lebensmitteltechnologischen und ernährungsphysiologischen Eigenschaften vielfältig eingesetzt. Sojazutaten sind allgegenwärtig. Das aus den Samen gewonnene Lecithin wird beispielsweise als Emulgator in der Schokoladenherstellung verwendet, allerdings spielt die Sojabohne auch für die Ölproduktion eine wichtige Rolle. Weiterhin finden verschiedenste Sojapräparate als lebensmitteltechnologische Beimischung in Backerzeugnissen Anwendung, werden sie Fleischerzeugnissen wie Wurst und Hamburgern zugesetzt und bilden die Grundlage für Säuglingsmilchnahrung und Diätprodukte [3, 4]. Allerdings besitzt die Sojabohne ein großes allergenes Potenzial. Die Food and Agriculture Organization (FAO) zählt sie zu den „big eight“. Hierbei handelt es sich um acht Lebensmittel (Milch, Eier, Erdnüsse, Nüsse, Getreide, Soja, Fisch und Krustentiere), die für 90 % der Nahrungsmittelallergien verantwortlich sind [5]. Warum Lebensmittelproteine eine allergische Reaktion verursachen und andere wiederum nicht, ist noch wenig bekannt. Allerdings gibt es eine Vielzahl von pflanzlichen und tierischen Proteinen, die in ihre Struktur und/oder ihrer Funktion konserviert sind. Diese Proteine scheinen im Vergleich zu anderen Proteinen vermehrt allergen zu wirken.

Bis zu 35 % der Bevölkerung in westlichen Ländern klagen über eine selbstdiagnostizierte Nahrungsmittelallergie [6]. Dies zeigt das Ausmaß des Problems und die Notwendigkeit einer geeigneten Diagnostikmethode. Eine präzise Diagnose ist wichtig, um lebensbedrohliche Reaktionen zu vermeiden und um individuelle Risiken abzuschätzen aber ebenso um unsinnige Diäten zu vermeiden, da der Verzicht auf das Allergen die einzig wirksame Therapie darstellt. Mittlerweile gibt es zwar viele Methoden eine Allergie zu diagnostizieren, aber nur die doppelt verblindete Nahrungsmittelprovokation erlaubt mit Sicherheit eine Nahrungsmittelallergie zuverlässig zu bestätigen.

1934 wurde die erste allergische Reaktion auf Soja beschrieben [7]. Seitdem wurden viele Fortschritte bei der Identifikation und Charakterisierung der Sojabohnenallergene erzielt. Allerdings sind bisher nur teilweise bekannt, welche allergenen Strukturen direkt für die Auslösung der Allergie verantwortlich sind. Die immunreaktiven Stellen, auch Epitope genannt, können überall im Molekül vorliegen. Bisher wurden Epitope mittels Peptide Array, im Immunoblot oder durch bioinformatische

Methode identifiziert. Diese „Epitope“ umfassen 10–30 oder mehr Aminosäuren [8–10]. Epitope bestehen im Sinne der vom Antikörper erkannten Aminosäuren aber aus fünf bis sieben Aminosäuren oder drei bis vier Zuckerresten [11]. Mit der in unserem Labor entwickelten Methode des Peptid Phage Displays in Kombination mit dem Next Generation Sequencing (NGS) können wir das Epitop auf die einzelne Aminosäure genau beschreiben. Im folgenden Kapitel gebe ich einen Überblick über Allergie im Allgemeinen sowie Nahrungsmittelallergie im Speziellen. Zum Schluss werde ich kurz auf die aktuelle Situation der Sojaallergie eingehen.

## 1.1 Allergie Allgemein

Der Begriff Allergie wurde von Kinderarzt Clemens von Pirquet [12] 1906 geprägt und beschreibt eine Immunantwort gegen Substanzen aus der Umwelt, die für die meisten Leute harmlos sind. 1963 unterteilten Gell und Coombs die Überempfindlichkeitsreaktion des Immunsystems in 4 Gruppen [13].

Die Typ 1 Allergie ist eine IgE vermittelte Reaktion, deren Symptome innerhalb von Sekunden bis Minuten auftreten können. Freie Antigene werden von IgE Antikörpern, die auf der Oberfläche von Mastzellen präsentiert werden, gebunden und führen zu einer Freisetzung von Entzündungsmediatoren.

Die Typ 2 Allergie ist eine IgM oder IgG vermittelte Reaktion. Diese Antikörper binden an zellgebundenen Antigene und markieren diese für den Abbau. Diese werden auch als Autoantikörper bezeichnet. Autoantikörper, die gegen Körperzellen gerichtet sind, verursachen Krankheiten wie z.B. hämolytische Anämie.

Die Typ 3 Allergie ist eine IgG vermittelte Reaktion, bei der Antikörper lösliche Antigene binden. Dabei werden Komplexe gebildet, welche vom Komplementsystem erkannt werden und zur Entzündungsreaktion führen. Diese Komplexe können sich in Kapillaren ablagern, wodurch es zur Schädigung der Niere kommen kann.

Die Typ 4 Allergie ist eine zellbasierte Reaktion. Die  $T_{\text{Helfer}}1$ -Zellen ( $T_{\text{H}}1$ -Zellen) werden durch ein Antigen, welches durch eine Antigen-präsentierende Zelle präsentiert wird, aktiviert. Tritt das Antigen in Zukunft erneut auf, aktivieren die  $T_{\text{H}}1$ -Zellen Makrophagen und lösen eine Entzündungsreaktion aus.

### 1.1.1 Die Familie der Immunglobulinproteine

Die Immunglobuline, auch Antikörper genannt, werden zum ersten Mal 1890 von Behring und Kitasato als ein Bestandteil des Blutes beschrieben, der das Diphtherie-Toxin neutralisieren konnte [14]. Antikörper bestehen aus mehreren Komponenten: zwei schweren Ketten (H) und zwei leichten Ketten (L), wogegen die L-Kette sich aus einer  $\kappa$  oder einer  $\lambda$ -Kette zusammensetzen kann. Sowohl

die H-, als auch die L-Kette beinhaltet eine N-terminale, variable (V) Domäne und eine oder mehrere C-terminale, konstante (C)-Domänen [15]. Sowohl die variablen als auch die konstanten Domänen sind ca. 12-13 kDa groß und enthalten etwa 110-130 Aminosäuren. Eine typische leichte Kette und schwere Kette sind ca. 25 kDa bzw. 55 kDa groß [14].

Schon seit längerer Zeit ist bekannt, dass Antikörper durch den Verdau von Enzymen in Fragmente gespalten werden können. Das Enzym Papain separierte das Antikörpermolekül in 2 F<sub>ab</sub>- und ein F<sub>c</sub>-Fragment. Jedes F<sub>ab</sub>-Fragment kann ein Molekül binden. Das Enzym Pepsin spaltete das Antikörpermolekül in ein F<sub>c</sub>- und ein dimeres F<sub>(ab)2</sub>-Fragment, welches kreuzvernetzen und Antigene binden kann. Das F<sub>ab</sub>-Fragment besteht aus einer vollständigen leichten Kette und aus einer C<sub>H1</sub>-Domäne, der V-Domäne der schweren Kette (Abbildung 1). Das F<sub>ab</sub>-Fragment kann weiter in F<sub>v</sub>-Fragmente unterteilt werden, welche in V<sub>H</sub> und V<sub>L</sub>-Domänen, sowie in konstante Fragmente strukturiert sind. Die konstanten Fragmente bestehen aus C<sub>L</sub> und C<sub>H1</sub>-Domänen [16].

Es gibt fünf verschiedene Antikörperklassen (IgG, IgA, IgD, IgE und IgM), welche unterschiedliche Komponenten und Funktionen haben. Die C-Proteindomäne der schweren Kette bestimmt die Funktion und die V-Proteindomänen reglementieren die antigenbindende Spezifität des Immunglobulins. Bei IgG-, IgA- und IgD-Antikörpern beinhaltet die schwere Kette des Antikörpers drei C-Domänen (C<sub>H1</sub>, C<sub>H2</sub>, C<sub>H3</sub>). Die schwere Kette des IgM und IgE-Antikörper besitzt noch eine zusätzliche, vierte C-Domäne (C<sub>H4</sub>). Das F<sub>c</sub>-Fragment definiert die Antikörperklasse (Isotyp) und Subklasse des Immunglobulins [14].

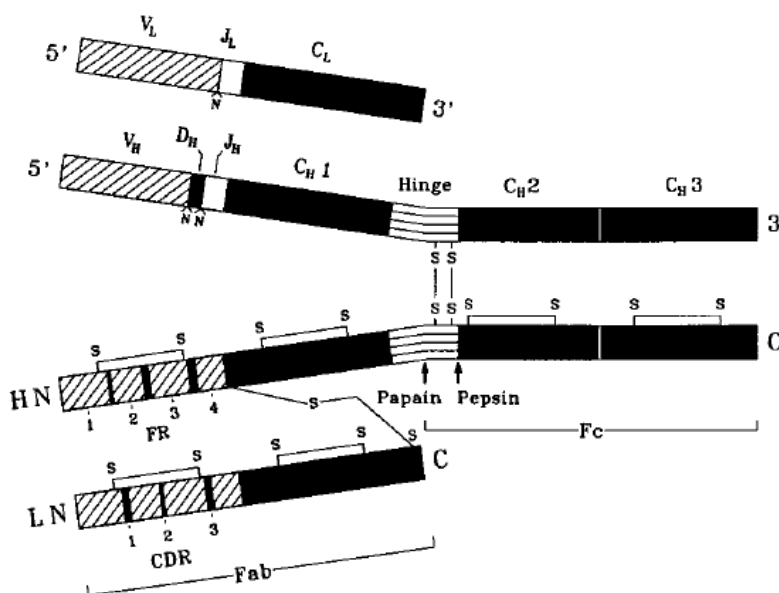


Abbildung 1: Zweidimensionales Model eines Immunglobulinmoleküls bestehend aus drei konstanten Domänen. Die oberen Dimere repräsentieren die Nukleotid- Struktur der schweren (H)- und leichten (L)-Kette [16].

### 1.1.1.1 Antikörper Epitope

Eine Antikörper-Antigen Interaktion findet zwischen dem Paratop, welches die Antigenbindungsstelle des Antikörpers ist, und dem Epitop, welches die Antikörperbindungsstelle im Protein ist, statt [17].

Man unterscheidet kontinuierliche und diskontinuierliche Epitope. Die kontinuierlichen Epitope sind linear, werden durch eine zusammenhängende Aminosäuresequenz gebildet und können nicht durch Denaturierung zerstört werden. Diskontinuierliche Epitope sind in der Regel mit der dreidimensionalen Struktur des Proteins assoziiert. Diese Art von Epitopen kann während der Denaturierung des Proteins modifiziert bzw. zerstört werden. Die klinische Signifikanz der Epitope ist von ihrer Struktur und Lokalisation im Protein abhängig. Zum Beispiel sind in Allergenen mit hoher klinischer Relevanz wie Milch oder Erdnuss kurze IgE-bindende Epitope im hydrophoben Teil des Proteins lokalisiert [18].

Es gibt leider keine strukturellen Eigenschaften, die die Vorhersage des allergenen Potentials eines Proteins erleichtern könnten. Dennoch gehören Allergene zu Proteinfamilien, die strukturell konservierte Eigenschaften aufgrund ihrer biologischen Aktivitäten aufweisen. Sie haben zum Beispiel eine globuläre, kompakte Struktur, die durch Hydrogen- oder Disulfidbrücken stabilisiert werden. Allergene sind meist glykosyliert und resistent gegenüber Proteolyse oder verdauenden Enzymen [19].

Um eine allergene Reaktion auszulösen ist zudem die relative Lage der Epitope wichtig, damit sie einen Antikörper/Rezeptor auch so binden, dass es zu einem biologischen Signal kommt.

### 1.1.2 Epitopidentifizierung

Es gibt mehrere Ansätze neue Epitope zu identifizieren, z.B. im ELISPOT, mittels Western Blot, im Microarray, mit NMR-Analysen, mittels Massenspektrometrie oder die Vorhersage mit bioinformatischer Software (*in silico*).

Sun et al. führten eine Studie zur Identifizierung von Epitopen in Gly m 5.01 ( $\alpha$ -Untereinheit von  $\beta$ -Conglycinin) durch. Mit Hilfe von bioinformatischen Programmen untersuchten sie unter anderem die Sekundärstruktur und die Allergenität des Proteins. Sie konnten so 15 Peptide in der Länge von 20 – 30 Aminosäuren identifizieren. Im Dot-Blot konnten elf dieser Peptide als IgE-Antikörper bindende Peptide mit Seren von Soja-Allergikern bestätigt werden [9]. Andere Studien nutzten auch *in silico* Ansätze, um Epitope zu identifizieren. Mit Hilfe von Strukturanalysen und bioinformatischen Werkzeugen konnten ähnliche Strukturmerkmale zwischen einzelnen Proteinfamilien identifiziert und so ähnliche Epitope in verschiedenen Proteinen gefunden werden [20]. Sehr zuverlässig aber auch sehr aufwendig bezüglich Materialverbrauch und Kosten sind NMR-Methoden [21]. Darüber hinaus gibt es einige Programme, die eine Epitop-Vorhersage zulassen [22, 23]. Einen anderen Ansatz verfolgten Schulten et al., in dem sie Wiesen-Lieschgras Extrakt mittels eines 2D-Gel auftrennten. Das

Gel wurde auf eine Nitrozellulosemembran geblottet und mit einem Serumpool von allergischen Patienten gegen Wiesen-Lieschgras inkubiert. Extraktproteine, die Serum IgE gebunden haben, wurden mittels MALDI-TOF analysiert [24]. Den gleichen Ansatz führten auch Fu et al. durch. Allergene Proteine des Sojas wurden mit einem Pool von Schweineseren, die mit Sojamehl gefüttert worden waren, im Western blot identifiziert. Hauptsächlich wurde Gly m 5.01 erkannt. Um die einzelnen Epitope aus Gly m 5.01 zu identifizieren, wurde das Protein im Microarray analysiert. Dazu wurde es in 15 Aminosäure lange Peptide, mit einem Überhang von acht Aminosäuren auf eine SPOT Zellulose Membran gespottet. Es wurden so 47 Peptide, die IgG binden identifiziert [8]. Epitope können auch mittels Phage Display untersucht werden. Bøgh et al. untersuchten drei Serumproben von Erdnussallergikern mittels Phage Display, indem sie die Ph.D.-7 Phage Display Peptid Bibliothek von New England BioLabs benutzten. In dieser Publikation wurden das individuelle IgE- und das IgG4- Erkennungsmuster jedes Patienten untersucht [25]. Dieselbe Peptid Bibliothek benutzten auch Christiansen et al. um vier Patientenseren von Erdnussallergikern zu untersuchen. Diese Gruppe identifizierte patienten-spezifische Peptidcluster. Um die Echtheit dieser Epitope des Ara h 1 Proteins zu untersuchen, benutzen die Autoren ein Peptidmikroarray. In dieser Publikation wurde ein 12mer Peptid gefunden, welches von drei Seren erkannt wurde [26].

### 1.1.3 Entstehung einer Allergie

Eine IgE-basierte Immunreaktion entsteht, wenn sich die adaptive IgE-Immunantwort, die gegen Parasiten gerichtet ist, fälschlicherweise gegen Lebensmittel - oder andere Substanzen richtet. Diese können z.B. durch Inhalation, Hautkontakt oder Nahrungsaufnahme aufgenommen werden. Die Entstehung einer Allergie umfasst drei Schritte. Der erste Schritt wird Sensibilisierung genannt und beginnt zum Beispiel mit dem Übergang eines relativ intakten Moleküls durch die Darmschleimhaut. Es kommt vor, dass der Darm nicht in der Lage ist intakte Allergene auszuschließen, weil er beispielsweise verletzt oder infiziert ist, das Allergen ein geringes Molekulargewicht hat oder der Darm im Kindesalter noch nicht ausgereift ist. Die Aufnahme von intakten Allergenen kann gelegentlich auch durch einen (voll entwickelten) Darm eines Erwachsenen geschehen. Gegenüber Allergien prädisponierte Personen produzieren vermehrt antigenspezifische B-Zellen, sowie ein spezielles Set aus T<sub>Helfer</sub> Zellen (T<sub>H</sub>-Zellen). Diese regen die B-Zellen an, sich in IgE-produzierende Plasmazellen zu differenzieren. Ohne T<sub>Helfer</sub>-Zelle entwickelt sich die B-Zelle nicht in eine IgE sekretierende Plasmazelle. Sekretierte IgE-Antikörper werden schnell von hochaffinen, unter anderem auf Mastzellen befindlichen, IgE-Rezeptoren gebunden. Mastzellen enthalten große Mengen an Histamin und andere Endzündungsmediatoren. Das Ergebnis der Sensibilisierung ist, dass die Mastzellen an der Oberfläche gebundene, allergenspezifische IgE-Antikörper besitzen.

Der zweite Schritt der Allergieentstehung erfolgt einige Zeit nach der Sensibilisierung. Ein Allergen oder multivalente Allergenfragmente werden nach der Aufnahme wieder absorbiert. Auf der Mastzelle gebundene IgE-Antikörper können diese Allergene binden, sodass das Allergen mit mindestens zwei rezeptorgebundenen IgE-Antikörpern interagiert. Dafür sind mindestens zwei Epitope oder IgE Bindungsstellen pro Fragment eines Allergens notwendig, die natürlich nicht identisch sein müssen. Bei einem Molekül mit nur einer einzigen IgE-Bindungsstelle kommt es aufgrund der fehlenden Dimerisierung nicht zur Histaminausschüttung [27]. Diese löst ein Signal aus, das die Ausschüttung oder Sekretion von Histamin und anderen Entzündungsmediatoren veranlasst [28]. Gleichzeitig wird die Produktion der IgE Antikörper angeregt und es kann zur Bildung weiterer Antikörper gegen andere Epitope kommen. Der letzte Schritt ist die Ursache der klinischen Symptome. Histamin und andere Entzündungsmediatoren stimulieren eine Vielzahl von allergischen Symptomen in anderen Zellen und Organen [11].

#### 1.1.4 Kreuzreaktionen

Als immunologische Kreuzreaktion definiert man die Situation, dass ein Antikörper mehrere Antigene erkennt [29], die beispielsweise strukturell verwandt sind. Nicht alle Kreuzreaktionen sind klinisch relevant und IgE basiert. Oftmals sind auch T-Zellen involviert [30]. Kreuzreaktionen spielen bei Proteinen mit sehr hoher Sequenz- und/oder Strukturhomologie eine Rolle. Sowohl kontinuierliche, als auch diskontinuierliche Epitope können Kreuzreaktionen auslösen. Bislang wurden 28 Gruppen von allergierelevanten, kreuzreaktiven Polypeptiden identifiziert: sechs Gruppen der Pathogen - verwandten Proteinen (PR), elf Gruppen von diversen Enzymen (zum Beispiel Proteasen und Glykolyse-Enzyme) und Proteinen, wie Transportproteine und Protease Inhibitoren, sowie regulierende und Strukturproteine [31]. Polypeptide, wie Profilin oder „Nicht-Spezifische Lipid Transfer Proteine“ (nsLTPs) besitzen großes Potenzial für Kreuzreaktionen. Sie werden daher auch als Panallergene bezeichnet [29]. Panallergene sind homologe Proteine, welche in vielen verschiedenen Organismen vorkommen und eine Immunglobulin-basierte Reaktion zwischen evolutionär unverwandten Polypeptiden hervorrufen.

Kreuzreaktionen werden häufig bei inhalierten Allergenen, z.B. Pollen beobachtet. Eine große Anzahl von Patienten, welche gegen Birke allergisch reagieren, reagiert auch auf Apfel, Haselnuss, Sellerie und Möhre [32]. In mediterranen Ländern reagieren Leute mit Allergien, welche durch nsLTPS verursacht werden, auch auf Rosengewächse, zum Beispiel Birne, wobei in Nordeuropa die Leute vermehrt an einer Birkenpollenallergie leiden. Aufgrund der Serumantikörper der Patienten kann nicht unterschieden werden, mit welchem Allergen der Patient sensibilisiert wurde [33]. Weitere Beispiele sind Kreuzreaktionen zwischen Latex und Früchten, Hausstaubmilbe und Tropomyosin, sowie Schimmelpilzen und Spinat [34–36].



Ähnliche Kreuzreaktionen rufen auch die „cross-reactive carbohydrate determinants“ (CCDs) in Pflanzen hervor. CCDs bestehen aus Zuckerresten wie N-Glykanen und kommen in Pollen, Pflanzen und Insekten vor. Es gibt Hinweise, dass CCDs das allergische Potential einiger Lebensmittel erhöhen [37]. Freie Zuckerreste (N-Glykane) sind normalerweise nicht immunogen, es sei denn, sie sind an Proteine gekoppelt. In Säugetieren vorkommende N-Glykane sind weniger immunogen, weil sie in vielen Säugetierzellen konserviert sind und der Körper dagegen Toleranzen entwickelt. Zuckerreste von Nicht-Säugetieren, zum Beispiel von Pflanzen, unterscheiden sich von den N-Glykanen der Säugetiere und können daher eine starke Immunantwort verursachen. Auf pflanzliche Glykane reagierende Antikörper sind häufig kreuzreaktiv. Diese Antikörper können nicht klar zwischen verschiedenen pflanzlichen N-Glykanen unterscheiden [38].

#### 1.1.5 Pflanzenbasierte Allergene

Allergene Pflanzen sind phänotypisch sehr variabel, dabei sind aber ihre Proteine stark konserviert. Sie werden aufgrund ihrer Sequenzähnlichkeit, konservierte 3D Struktur und Funktion in vier Hauptfamilien eingeteilt: die Prolamin Familie, die Cupin Familie, Profilin und die Bet v 1 Familie.

##### 1.1.5.1 Prolamin Familie

Die Prolamin Familie umfasst die größte Anzahl von Pflanzenallergenen, z.B. „2S seed storage Albumin“, „cereals seed storage proteins“, „cereals  $\alpha$ -amylase/trypsin inhibitors“ und „non-specific lipid transfer proteins (nsLTPs)“ [39]. Prolamine werden aufgrund ihrer Wasser/Alkohollöslichkeit, sowie ihres Gehaltes an Prolin und Glutamin dieser Familie zugeordnet. Sie enthalten viele schwefelhaltige Aminosäuren. Typisch für Prolamine sind vier  $\alpha$ -Helixbündel, die durch Disulfidbrücken stabilisiert sind, in denen acht konservierte Cysteinreste involviert sind [40].

Die Funktion von Speicherproteinen wie die 2S Albuminen ist es, Aminosäure- bzw Stickstoffreserven für die Keimung bereitzustellen. 2S Albumine sind auch an der Abwehr von pathogenen Pilzen beteiligt. Zu dieser Familie gehören unter anderem die Hauptallergene in Nüssen, Sesam und Senfsaaten. Weitere Vertreter dieser Familie sind  $\alpha$ -Amylase und Protease Inhibitoren, die eine Resistenz in Pflanzengewebe gegen Fraßschädlinge verursachen. Allergene in Weizen, Gerste, Reis und Mais gehören ebenfalls zu dieser Familie [41]. Die nsLTPs, die auch zu der Prolamin Familie gehören, umfassen monomere Proteine niedrigen Molekulargewichts, die ca. 7-9 kDa groß sind. Sie haben eine protektive Rolle in Pflanzen und teilweise in Früchten. Sie besitzen eine kompakte und stabile Tertiärstruktur aus  $\alpha$ -Helices und Loops, die durch acht Disulfidbrücken stabilisiert werden. Zusätzlich besitzen sie eine zentrale Lipid-Bindungsstelle. Die Bindung von hydrophoben Liganden stabilisiert diese Struktur. Lipid-Transfer Proteine verursachen häufig eine starke Reaktion [42].

#### 1.1.5.2 Cupin Familie

Die Cupin Familie beinhaltet globuläre Speicherproteine. Diese allergenen Proteine sind in den Früchten von Leguminosen und Nüssen vorhanden. Der Name kommt von ihrer typischen  $\beta$  Barrel Struktur bestehend aus sechs  $\beta$ -Faltblättern, die mit sechs  $\alpha$ -Helices assoziiert sind. Zusammen bilden sie eine Bindungsstelle für hydrophobe Liganden [43]. Untergruppen werden anhand der Häufigkeit des Cupinmotivs im Protein definiert. Monocupine können Monomere, Dimere oder oligomere Proteine sein. Meist sind es Enzyme, z.B. Dioxygenasen. Zwei Cupinmotive besitzen die Globuline in Speicherproteinen, die häufig durch Salzlösung extrahiert werden können. Diese wurden aufgrund ihres Sedimentationskoeffizienten während der Ultrazentrifugation in verschiedene Gruppen eingeteilt, z.B. 7S/8S Globuline (Viciline) und 11S Globuline (Legumine). 7S/8S Globuline sind Trimere, deren Untereinheiten zwischen 50 – 60 kDa groß sind. Oft sind sie glykosyliert, was sie widerstandsfähig gegen Proteasen macht. 11S Globuline haben sechs Untereinheiten mit je 60 kDa und sind selten glykosyliert. Jede Untereinheit besteht aus einem nicht-kovalenten Dimer von zwei Polypeptidketten [44]. 7S und 11S Globuline besitzen zwar eine geringe Sequenzidentität, sind aber strukturell stark verwandt und haben ein gemeinsames 3D Motiv. Globuline sind klinisch relevante Allergene in Erdnuss, Soja, Wal- und Haselnuss [42].

#### 1.1.5.3 Profilin Familie

Profiline sind cytosolische, aktinbindende Proteine, die zwischen 12 und 15 kDa groß sind [45] und nur in Blütenpflanzen vorkommen, beispielsweise in Erdnuss, Apfel und Sellerie. Diese Proteine weisen eine kompakte Struktur bestehend aus antiparallelen  $\beta$ -Faltblättern, die von beiden Seiten durch  $\alpha$ -Helices umschlossen werden, auf [46]. Die Proteinfamilie zeichnet sich durch hohe Sequenz- und Strukturähnlichkeit aus und umfasst daher kreuzreaktive Proteine aus vielen Pflanzen [42].

#### 1.1.5.4 Bet v 1 / PR10 Familie

Die Bet v 1 Familie besteht aus acht Unterfamilien, welche als „pathogenesis-related proteins 10“ PR10 zusammengefasst werden. Diese Allergene sind homolog zu dem Hauptbirkenpollenallergen Bet v 1. Sie kommen in Früchten der Rosengewächse, z.B. Apfel, Kirsche und Birne, sowie in Früchten der Doldenblütler Familie, z.B. Sellerie und Möhren, vor [33]. Die PR10-Proteine besitzen eine stark konservierte Sequenz. Ihre Polypeptidkette besteht aus 154-160 Aminosäuren [39]. Die Bet v 1 Homologe enthalten ein GXGXXG oder ein GXG Motiv, welches für die Bindung von Phosphatgruppen verantwortlich ist [47]. Zusätzlich enthalten diese Proteine eine charakteristische Struktur, welche durch sieben  $\beta$ -Faltblätter umschlossene lange C-terminale Helix gekennzeichnet ist. Zwei zusätzliche Helices sind mit den zwei  $\beta$ -Faltblättern verbunden. Diese 3D-Struktur ist zu einer Y-förmigen hydrophoben Bindungstasche geformt [48].

### 1.1.6 Allergene tierischen Ursprungs

Allergene tierischen Ursprungs sind im Vergleich zu den Pflanzenallergenen nicht so zahlreich. Die häufigsten sind in drei strukturell ähnlichen Familien unterteilt: Tropomyosine, EF-Hand-Proteine und Caseine [49].

Tropomyosine sind in Muskeln oder anderen Zellen mit Myofibrillen enthalten [50]. Sie enthalten 40 aufeinanderfolgende Wiederholungen aus sieben Aminosäuren. Diese Proteine formen eine  $\alpha$ -helikale „coiled-coil“ dimere Struktur. Tropomyosine sind klinisch relevant für Allergien auf Weichtiere und Schalentiere [51].

Die EF-Hand-Proteine enthalten ein konserviertes Strukturmotiv, welches aus einem Kalzium bindenden, zwölf Aminosäuren langes Helix-Loop-Helix Motiv und zwei benachbarten zwölf Aminosäuren langen Helices besteht. [39]. Dieser Loop kann Calcium- oder Magnesium- Ionen mit verschiedenen Wertigkeiten binden. Dasselbe Motiv ist auch in der großen Calcium-bindenden Proteinfamilie vorhanden. Zum Beispiel enthält Parvalbumin drei EF-Hand-Motive, wobei zwei davon Calcium binden. Durch Hitzebehandlung wird das Calcium entfernt. Dabei kommt es zu einer großen Strukturveränderung innerhalb des Proteins, wobei auch diskontinuierliche Epitope zerstört werden. Die verbliebenen IgE-bindende Epitope sind aber ausreichend, um eine allergische Reaktion auszulösen [52].

Casein ist in Milch vorhanden und bindet Kalzium- Ionen durch Phosphoserine oder Phosphothreonine [42].

### 1.2 Nahrungsmittelallergien

In den letzten 30 Jahren hat sich die Häufigkeit der Nahrungsmittelallergien sowohl in Industrie- als auch in Entwicklungsländern fast verdoppelt [53]. 3,6 % der deutschen Bevölkerung reagieren auf mindestens ein Lebensmittel, wie in einer Studie mit über 1300 Teilnehmern festgestellt wurde. Weltweit leiden ca. 220 – 250 Millionen Menschen an einer Nahrungsmittelallergie [54]. Am meisten reagieren die Deutschen mit klinischen Symptomen auf Haselnüsse und Äpfel. Diese werden auch als Pollen-assoziierte Lebensmittel bezeichnet, weil deren Allergene mit Pollenproteinen kreuzreagieren [1]. Häufig tritt eine Typ 1 Nahrungsmittelallergie in den ersten zwei Lebensjahren auf. Dies betrifft ca. 6-8 % der Kleinkinder [55]. Allerdings verlieren bis zu 80 % der Kinder ihre Nahrungsmittelallergie wieder, nachdem das Immunsystem mit fünf Jahren gereift ist [56].

Nicht nur die Zahl der Menschen mit Nahrungsmittelallergien steigt an, sondern auch die Anzahl der allergenen Lebensmittel. So müssen mittlerweile seit Dezember 2014 laut EU-Verordnung (Nr. 1169/2011) 14 allergene Lebensmittel gekennzeichnet werden. Unter den kennzeichnungspflichtigen Allergenen befinden sich unter anderem Milch (einschließlich Lactose), glutenhaltiges Getreide, Nüsse und viele andere Lebensmittel und Lebensmittelzutaten [2]. Schon diese Beispiele zeigen, dass

eine Lebensmittelunverträglichkeit unterschiedlichste Gründe haben kann. Es gibt zwei verschiedene Bereiche, in die eine Lebensmittelunverträglichkeit eingeteilt werden kann: immunologisch, nicht-immunologisch. Immunologische Reaktionen können entweder durch IgE Antikörper, durch Zellen oder beides ausgelöst werden. Bei nicht-immunologischen Reaktionen spricht man von einer Lebensmittelintoleranz, die durch Enzymdefekte, pharmakologische Reaktionen oder durch unbekannte Mechanismen verursacht werden, siehe Abbildung 2 [57].

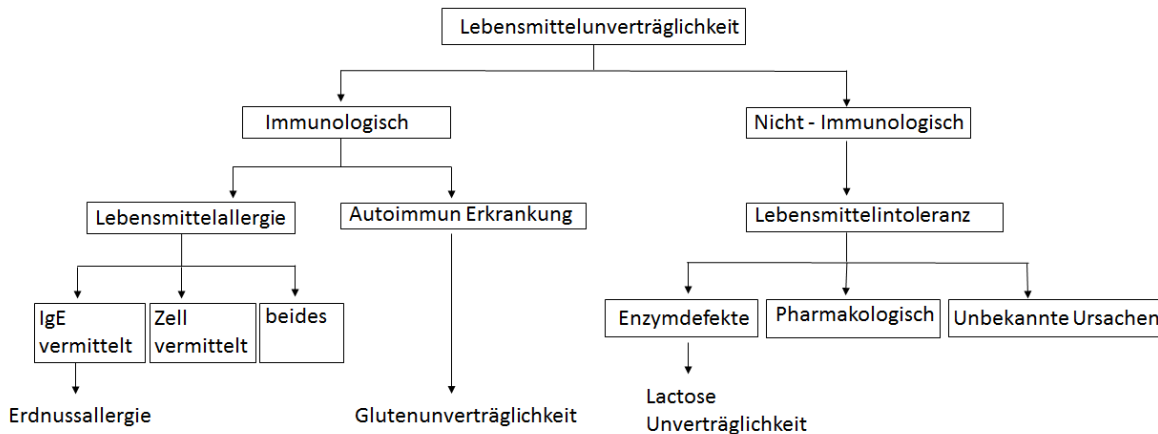


Abbildung 2: Einteilung der Nahrungsmittelunverträglichkeit [Adaptiert nach [57]]

Eine Nahrungsmittelallergie ist eine immunvermittelte Reaktion, die reproduzierbar bei Aufnahme eines bestimmten Lebensmittels auftritt und bei Verzicht auf dieses Lebensmittel ausbleibt [58]. Allerdings muss eine Unverträglichkeit eines Lebensmittels nicht gleich eine Allergie bedeuten. Viele Symptome weisen sowohl auf eine Nahrungsmittelallergie, als auch auf eine Unverträglichkeitsreaktion hin. Einige Symptome geben hingegen eindeutige Hinweise auf eine Allergie oder auf eine Unverträglichkeit. So ist IgE-basierte Immunreaktion meist eine akute, schnell einsetzende Reaktion, die oft 3-120 min nach Allergenkontakt auftritt. Die Symptome reichen von mild bis lebensbedrohlich und können in vier Kategorien eingeteilt werden: dermatologisch (Nesselsucht, akut anschwellende Gesichtsteile), gastrointestinal (Übelkeit, Erbrechen), respiratorisch (Asthma, laufende Nase) und systematisch (Anaphylaktischer Schock) [11].

Eine nicht-IgE-basierte Immunreaktion ist meist eine verspätete, nicht akute Reaktion, die 2- 48 Stunden nach Allergenkontakt auftritt. Symptome sind unter anderem atopisches Ekzem, Verstopfung oder blutiger Stuhlgang. 40 % der Patienten mit atopischem Ekzem leiden zusätzlich auch an einer Nahrungsmittelallergie [59]. Für diese Personen ist es oft möglich, kleinere Dosen des vermeintlichen Allergens zu tolerieren. Unglücklicherweise gibt es auch zahlreiche Anzeichen, welche gleichzeitig auf eine IgE-basierte und eine nicht-IgE-basierte Immunreaktion hindeuten, z.B. abdominale Schmerzen, Durchfall, Juckreiz und Hautrötung [60].

### 1.3 Das Orale Allergiesyndrom (OAS)

Symptome können nach dem Lebensmittelverzehr entweder unmittelbar oder verzögert auftreten. Krankheitssymptome, die bei einer IgE-basierten immunologischen Reaktion nach wenigen Minuten im Lippen, Mund - und Rachenraum auftreten, werden als orales Allergiesyndrom (OAS) zusammengefasst [61]. Der direkte Kontakt mit Nahrungsmitteln löst z.B. Kribbeln, Bläschen, Schwellungen und Rötungen im Mund- und Rachenraum aus. In 2 % der Fälle sind auch Organe betroffen und es kann eine lebensbedrohliche Reaktion, wie z.B. ein anaphylaktischer Schock, auftreten [62].

Das OAS ist häufig mit hitzelabilen Allergenen assoziiert, wie sie z.B. in Obst- und Gemüsesorten enthalten sind. Auch gibt es einige kreuzreaktive Proteine (z.B. Tropomyosin), welche sowohl in Aeroallergenen (Der p 10, Hausstaubmilbe) als auch in einigen Lebensmitteln (Pen a 1, Garnelen) gefunden werden. Dies erklärt ebenfalls, dass einige Patienten mit einer Nahrungsmittelallergie auch eine respiratorische Allergie haben [63]. 50-90 % der Birkenpollenallergiker leiden am OAS [64]. Eine höhere Allergenmenge führt häufig zu intensiveren Symptomen. Häufige Auslöser sind z.B. Melone, Banane, Apfel, Kiwi, Tomate und Sellerie. In vielen Fällen lösen diese Lebensmittel lediglich im rohen, jedoch selten im gekochten oder prozessierten Zustand allergische Reaktionen aus [65].

#### 1.3.1 Diagnose einer Nahrungsmittelallergie

Die Behandlung einer Nahrungsmittelallergie ist oft aufgrund der Multisensibilisierung und der subjektiven Wahrnehmung der Symptome schwierig. Objektive, klinische Anhaltspunkte fehlen häufig [58]. Mittlerweile gibt es viele Testsysteme eine Nahrungsmittelallergie zu diagnostizieren. Die Ergebnisse korrekt zu deuten ist auch heutzutage noch eine große Herausforderung. Der Schlüssel zur Diagnose ist die Patientengeschichte, sowie der zeitliche Zusammenhang zwischen Lebensmittelaufnahme und Reaktion [58]. Ein weiterer Hinweis auf eine immunbasierte Reaktion ist das Vorliegen einer Atopie in der Familiengeschichte. Hingegen kann eine Nahrungsmittelallergie nicht ausgeschlossen werden, wenn in der Familiengeschichte bisher keine Allergie auftrat [42].

Derzeit ist es zwar möglich viele Allergene gleichzeitig zu testen, jedoch steigt die Gefahr der falsch-positiven Testergebnisse mit der Anzahl. Validierte Allergietests sind Haut-Pricktest (SPT), Bestimmung Allergie-spezifischer IgE- Antikörper und Nahrungsmittelprovokation. Abgesehen von der oralen Provokation existieren keine beweisbaren Parameter einer Nahrungsmittelallergie [66].

##### 1.3.1.1 Nahrungsmittelprovokation

Es gibt verschiedene Varianten eine Nahrungsmittelprovokation durchzuführen. Es kann eine offene, eine einfach verblindete und/oder eine doppelt verblindete Nahrungsmittelprovokation ausgeführt werden. In einer offenen Nahrungsmittelprovokation (OFC) werden keine Placebos verwendet. Bei

einer einfach verblindeten Nahrungsmittelprovokation (SBPCFC) weiß der Proband nicht, ob ein Placebo oder das Nahrungsmittel verabreicht wird. Während einer doppelt verblindeten Nahrungsmittelprovokation (DBPCFC) weiß sowohl der Patient, als auch der Arzt nicht, ob ein Placebo oder das Lebensmittel verabreicht wird. Ergebnisse der OFC sind schwieriger zu deuten als die Ergebnisse der SBPCFC. Gründe dafür sind die verspätete oder subjektive Wahrnehmung von Symptomen in Abwesenheit von Placebos [42]. Der Goldstandard ist die DBPCFC, weil alle subjektiven Tendenzen entfernt werden. Bei einer DBPCFC wird eine steigende Menge des Allergens oder eines Placebos über den Tag aufgenommen und die Symptome dokumentiert. Ein paar Tage später wird der Test mit weiteren Konzentrationen wiederholt. Der Test birgt das Risiko eines anaphylaktischen Schocks und sollte daher nur in Kliniken durchgeführt werden [67]. Die DBPCFC ist kosten- und zeitintensiv. Zusätzlich ist es schwierig geeignete Placebos zu erhalten. Daher wird oft eine offene Lebensmittelprovokation anstatt einer verblindeten durchgeführt [68].

Eine Eliminationsdiät wird gelegentlich durchgeführt, um eine Nahrungsmittelallergie zu diagnostizieren. Diese wird angewendet, wenn kein Provokationstest durchgeführt werden kann oder soll.

#### 1.3.1.2 Bestimmung Allergen-spezifischer Serum IgE-Antikörper

Das Vorliegen eines allergenspezifischen-Serum-IgE-Antikörpers kann zwar die Sensibilisierung gegenüber einem bestimmten Lebensmittel bedeuten, diagnostiziert aber keine klinische Nahrungsmittelallergie und sagt nichts über die Häufigkeit und die Schwere der allergischen Reaktion nach Aufnahme des Nahrungsmittels aus [69]. Da nur am IgE-Rezeptor gebundene IgE-Antikörper eine Immunreaktion auslösen können (aber nicht müssen), ist die fehlende Unterscheidung zwischen zirkulierenden und membran-gebundenen IgE-Antikörper ein großer Nachteil dieses Tests. Es gibt mehrere Tests allergenspezifische IgE-Antikörper zu bestimmen. Der Radio-Allergo-Sorbent-Test (RAST) wird immer öfters durch quantitative immunochemische Tests ersetzt, z.B. durch Enzyme-linked Immunosorbent Assay (ELISA) oder ImmunoCAP®, Immulite® oder HYTEC\_288® [57]. Das Allergen wird im ersten Schritt auf einer Oberfläche immobilisiert. Anschließend wird mit dem Patientenserum inkubiert. Enthält dieses Immunglobuline gegen das Allergen, binden diese an das immobilisierte Protein. Nach einem Waschschriff werden die gebundenen Immunglobuline detektiert. Dies kann z.B. durch radioaktiv-markierte Sonden, wie im RAST oder durch Fluoreszenz, wie im ELISA oder ImmunoCAP® erfolgen. Die Sensitivität und Spezifität dieser Tests variiert bezüglich der Bedingungen und der verwendeten/genutzten Detektionsantikörper. Daher sind Testergebnisse von verschiedenen Laboren nicht miteinander vergleichbar [70]. Wang et al. führten eine Studie mit 50 Patienten durch und nutzte verschiedenen Systeme (Phadia ImmunoCAP, Agilent Turbo-MP und Siemens Immulite 2000), die eine Allergie gegen Milch, Ei und Erdnuss diagnostizieren

können. Die jeweils verwendete Testsubstanz war nicht standardisiert und jedes System nutzte zudem verschiedene Antigene. Beispielsweise wurde bei Milch einmal entrahmte Milch, gefriergetrocknete Milch oder „normale“ Milch verwendet. Jedes System lieferte verschiedene Testergebnisse bezüglich der IgE-Konzentration für dieselbe Serumprobe [71]. Abhängig von dem getesteten Lebensmittel ist ein hoher IgE-Level ein guter Indikator, um die Allergie im oralen Provokationstest zu bestätigen [6]. Allerdings haben bis zu 40 % der Personen mit hohen IgE-Titern keine klinischen Symptome [58].

Der ImmunoCAP® ISAC ist ein spezieller Chip-basierter Bluttest mit dem man simultan bis zu 103 verschiedene Lebensmittelallergene testen kann. Die gleichzeitigen Messung der Allergene erlaubt die Analyse des Sensibilisierungsmusters und die Dauer der Sensibilisierung [63].

#### 1.3.1.3 Haut-Pricktest

Zur Diagnose von Nahrungsmittelallergien wird am häufigsten der preiswerte und schnell durchzuführende Haut-Pricktest (SPT) angewendet [72]. Dieser wird zum ersten Mal im Jahr 1867 beschrieben [73]. Der größte Nachteil des Haut-Pricktest ist, wie schon bei der spezifischen IgE-Messung erwähnt, dass diese Tests nicht zwischen freizirkulierenden und zellgebundenen IgE-Antikörpern unterscheiden können [74]. Ein weiterer Nachteil ist die Subjektivität des Ergebnisses und die fehlende Standardisierung der Reagenzien und des Verfahrens. Allerdings bleibt dieser Test aus zahlreichen Gründen der am häufigste verwendete Test, um eine Allergie zu bestätigen. Zum einen ist der SPT wenig invasiv und man kann gleichzeitig mehrere Allergene testen. Bei einem Haut-Pricktest werden Lebensmittelextrakte und Kontrollen meist auf der Haut des Unterarms oder Rückens aufgebracht und mit einer Lanzette leicht angestochen. Die Hautstelle, an der der Test durchgeführt wird, kann die Reaktivität beeinflussen [73]. Nach ca. 15 min wird die Reaktion sowohl mit einer Positivkontrolle (Histamin), als auch mit einer Negativkontrolle verglichen. Ist der Durchmesser der Quaddel 3 mm größer als die Negativkontrolle, ist das Testergebnis positiv [75, 76]. Die Größe der Quaddeln korreliert mit der Wahrscheinlichkeit einer Allergie. Die Sensitivität und Spezifität des Haut-Pricktest ist relativ schlecht, wenn man nur diese zwei Kontrollen (Positivkontrolle und Negativkontrolle) verwendet [77]. Dadurch werden im klinischen Alltag Nahrungsmittelallergien häufig überdiagnostiziert. Ein positives Testergebnis zeigt eine Sensibilisierung gegenüber dem getesteten Lebensmittel, beweist aber nicht das Vorliegen einer echten Nahrungsmittelallergie [69]. Negative Testergebnisse können zwar mit 95 % Wahrscheinlichkeit eine IgE-basierte Reaktion ausschließen, aber in nur 60 % der Fälle weist der Haut-Pricktest auf eine IgE-basierte Reaktion hin [78]. Allerdings ist der SPT bei diagnostizierten Lebensmittelallergikern sehr gut geeignet, um das Lebensmittel, welches die Allergie hervorruft, zu identifizieren [58]. Die Sensitivität, d.h. die Wahrscheinlichkeit falsch negativer Ergebnisse, variiert bezüglich der genutzten

Lebensmittelextrakte und ist ein wenig höher als bei der Messung der spezifischen Serum IgE-Antikörper, welche eine höhere Spezifität, weniger falsch positive Ergebnisse, hat [73].

Es gibt noch weitere Tests, unter anderem die Bestimmung von IgG-Antikörper oder deren Subklassen. Sie verwenden einzelne Proteine oder Proteingemische. Diese Tests haben keine weitere Bedeutung für die Praxis, da deren Aussagekraft zweifelhaft ist [79].

### 1.3.2 Behandlung einer Nahrungsmittelallergie

Die einzige Therapie, welche weitestgehend erfolgsversprechend ist, ist die strikte Vermeidung des Allergens. Aufgrund der vielen versteckten Zutaten in Lebensmitteln kann sich dies als schwierig erweisen. Ein weiterer Therapieansatz ist die Immuntherapie. Dabei können Patienten erreichen, dass sie eine höhere Konzentration des Allergens tolerieren und so das Risiko einer allergischen Reaktion minimieren. Bei dieser Therapie wird ein abgestufter Allergenkontakt genutzt, um eine Toleranz zu entwickeln. Eine Immuntherapie kann drei bis fünf Jahre dauern. Bei der Immuntherapie werden sowohl rohe, als auch prozessierte bzw. denaturierte Lebensmittel verwendet [80]. Häufig wird die Immuntherapie oral durchgeführt, sie kann aber auch sublingual erfolgen. Eine sublinguale Immuntherapie wird häufig bei Umweltallergenen [81] und seltener bei Nahrungsmittelallergien [82] durchgeführt. Die Behandlung erfolgt in drei Phasen. Phase I besteht aus der klinischen Bestimmung der Dosisgrenze, bei der eine Reaktion auftritt, gefolgt von einer langsamen Steigerung der Dosis, meist im zweiwöchigen Rhythmus. Die zweite Phase besteht aus dem Verabreichen einer konstanten Dosis über Wochen bis Jahre. Anschließend erfolgt eine Untersuchung mittels Nahrungsmittelprovokation [83]. Die Immuntherapie kann die Schwere der Symptome reduzieren. Der Erfolg der Immuntherapie wird viel diskutiert. In einer Studie konnten 62 % der Erdnussallergiker desensibilisiert werden [84]. Andere Studien belegen aber, dass bei 20 % der Patienten keine Verbesserung eintrat [83]. Eine Immuntherapie kann nur mit einem Lebensmittel durchgeführt werden. Eine multiple Nahrungsmittelallergie, die häufig auftritt, wird mit rekombinanten, monoklonalen Antikörpern gegen IgE behandelt. Omalizumab ist ein rekombinanter Antikörper, der an der konstanten  $\epsilon$ -Domäne des IgE-Antikörpers bindet und somit die Bindung des IgE-Antikörpers verhindert [85], wodurch die allergische Reaktion, nicht aber ihre Ursache unterbunden wird.

Ein Patient mit einer Nahrungsmittelallergie kann diese im Laufe seines Lebens meist wieder verlieren. Bis zu 70 % der Kinder mit einer Allergie gegen Milch oder Eier entwickeln von alleine mit der Zeit eine Toleranz. Dies ist auch bei Soja, Weizen, vielen Früchten und Gemüse sehr wahrscheinlich [55]. Die Toleranz entwickelt sich meist schrittweise. Zuerst ist die prozessierte Form des Lebensmittels symptomfrei verzehrbar, bevor man die rohe Form des Nahrungsmittels vertragen kann. So scheint es, dass der Körper auch Toleranzen gegenüber diskontinuierlichen Epitopen entwickelt, wenn er zuerst Toleranzen für prozessierte Lebensmittel ausbildet [86].



## 1.4 Soja

Soja, *Glycine max*, gehört zu der Familie der Hülsenfrüchte. Es gibt mehr als 200 Sorten der Sojabohne und die Samen variieren stark in ihrer Oberflächentextur und ihrem Aussehen [87]. Die Samen enthalten ca. 20 % Öl und 38 % Protein. Aus der Sojabohne werden viele verschiedene Produkte gewonnen, z.B. Sojaöl, Sojamehl, Sojamilch und fermentierte Sojaprodukte, wie Sojasauce, Miso und Tofu, um nur einige zu nennen. Auch in der Lebensmittelindustrie wird die Sojabohne vielseitig eingesetzt, um beispielsweise die Textur zu beeinflussen, den Proteingehalt zu vergrößern oder als Emulgator. Soja ist eine preiswerte Proteinquelle und Sojazutaten werden daher in Soßen, Fleischprodukten, Backwaren, Schokolade und viele anderen Produkten eingesetzt. Die Hülsenfrucht wird auch als Muttermilchersatz verwendet. Ein weiterer Verwendungszweck besteht als Ersatzstoff bei Kuhmilchallergie, Lactose- oder Galaktoseunverträglichkeit. Allerdings kann Soja und Sojaprodukte IgE- und nicht IgE-basierte Nahrungsmittelallergie hervorrufen [42].

In Deutschland leiden ca. 1,7 % der Bevölkerung an einer Sojaallergie, die mittels Prick-Test diagnostiziert wurde [88]. Allergische Reaktionen können schon bei der Aufnahme von 88 mg Sojaprotein auftreten [89]. Allerdings hat jeder Patient einen individuellen Schwellenwert. Die Anzahl der Sojaallergiker steigt stetig an. Die Häufigkeit der Soja-Sensibilisierung stieg von 2 % bei Zweijährigen auf 7 % bei Zehnjährigen in einer deutschen Multi-Center Allergie Studie mit 1314 Kindern [90]. Eine Sojaallergie scheint häufiger bei atopischen Patienten als in der „normalen“ Population aufzutreten. In einer US-Studie litten 13 % der atopischen Kinder zusätzlich an einer Sojaallergie [91].

Bei den meisten Kindern, die an einer Sojaallergie leiden, „verwächst“ sich diese Allergie. In einer Studie mit 133 Kindern litten nur noch 75 % der Kinder nach vier Jahren, 55 % der Kinder nach sechs Jahren und 30 % der Kinder nach 10 Jahren an einer Sojaallergie. Diese identifizierte auch Kinder, bei denen die allergischen Symptome nach der Toleranz einer Soja - Diät auftraten. Es wurde hier eine Kreuzreaktion mit Erdnuss- oder Birken-spezifischen IgE vermutet [92].

Symptome der Sojaallergie sind generell mild. Meist ist diese Form der Nahrungsmittelallergie mit dem OAS assoziiert. Allerdings sind auch einige schwere gastrointestinale Symptome und Anaphylaxie bekannt [93, 94].

### 1.4.1 Allergene in Soja

Die erste allergische Reaktion gegen Soja in Menschen wurde 1934 beschrieben [7]. Soja enthält ca. 38 % Protein. Wie viele Proteine der Sojabohne IgE reaktiv sind, wird von Quelle zu Quelle unterschiedlich angegeben. Mindestens 16 IgE-bindende Proteine wurden identifiziert und stehen in der Allergome Datenbank (<http://www.allergome.org>). Acht davon stehen in Datenbank der

internationalen Vereinigung der immunologischen Gesellschaften (IUIS) (<http://www.iuisonline.org>), (Tabelle 1). Xiang et al. sprechen sogar von 34 IgE reaktiven Proteinen [95].

Tabelle 1: Allergene der Sojabohne [96]

Allergen	Protein	Molekulargewicht	Familie
Gly m 1	Hydrophobic Seed Protein	7	Hydrophobic seed protein
Gly m 2	Defensin	8	Defensin
Gly m 3	Profilin	14	Profilin
Gly m 4	PR-10 protein	17	Bet v 1 verwandtes Protein
Gly m 5	$\beta$ -Conglycinin (7S lobulin, vicilin)		Cupin
	Subunit $\alpha$	63	
	Subunit $\alpha'$	65	
	Subunit $\beta$	50,5	
Gly m 6	Glycinin (11S globulin, legumin)		Cupin
	Subunit Gy1	55,7	
	Subunit Gy2	54,4	
	Subunit Gy3	54,2	
	Subunit Gy4	63,8	
	Subunit Gy5	58,2	
Gly m 7	Seed biotinylated protein	76,2	
Gly m 8	2S Albumin	18,5	Prolamin

Die Sojabohne besteht aus zwei Hauptspeicherproteinen, Gly m 6, auch Glycinin (11S) genannt, und Gly m 5, auch  $\beta$ -Conglycinin (7S) genannt. Diese Hauptspeicherproteine gehören zur Cupinfamilie. Diese machen 70 % des gesamten Proteingehalts aus. Gly m 5 ist ca. 180 kDa groß und besteht aus drei Untereinheiten,  $\alpha$  (~67kDa),  $\alpha'$  (~71kDa) und  $\beta$  (~50kDa) [97], die alle glykosyliert sind [98]. Nur die  $\alpha$ -Untereinheit ist allergen, allerdings weisen die  $\alpha'$  - und  $\beta$ -Untereinheit jeweils 90,14 % bzw. 76,2 % Sequenzidentität auf.

Gly m 6 ist ein Hexamer und ist 360 kDa groß. Jede Untereinheit besteht aus einer sauren und basischen, durch Disulfidbrücken verlinkten, Polypeptidkette [99]. Es gibt fünf Untereinheiten, die in verschiedenen Kombinationen der sauren und basischen Polypeptidkette zusammengesetzt sind [99, 100]. Die Untereinheit Gly m 6.01 und Gly m 6.05 enthalten die meisten Epitope für dieses Protein [101].

Andere Sojaproteine, welche als Allergen charakterisiert wurden, sind die Thiolprotease (Gly m Bd 30k) [102], der Kunitz Trypsin Inhibitor [94] und Gly m Bd 28K [95]. In Tabelle 2 sind *in vitro* IgE-Bindung Studien zusammengefasst, in denen die Anzahl der Patienten bestimmt worden, welche an das genannte Allergen litten.

Tabelle 2: Häufigkeit der IgE für einzelne Proteine bei Sojaallergie.

Protein	Patient	Quelle
Gly m 1	>90 %	[103]
Gly m 3	69 %	[97]
Gly m 4	86 %	[104]
Gly m 5	5-67 %	[105]
Gly m 6	5-58 %	[106]

Über die Schwere der Sojaallergie wurden einige Studien durchgeführt. Unter schwerer Sojaallergie leidende Kinder sind gegen Gly m 5 und Gly m 6 sensibilisiert [106]. Jedoch äußert sich ein hoher IgE-Level gegen Gly m 5 und Gly m 6 bei Erwachsenen in milden Symptomen. Hohe IgE Titer gegen Gly m 4 werden oft mit Sojamilchallergie assoziiert [107]. Soja ist auch ein durch die Luft übertragbares Allergen. Die Pathologie und die allergenen Reaktionsprofile sind für die Aufnahme und Inhalation der Sojabohne unterschiedlich. So sind z. B. die Hüllproteine der Bohne in Sojaproteinisolaten nicht vorhanden. Daher scheinen jeweils unterschiedliche Sojaproteine bezüglich einer Allergie dominant zu sein [108].

#### 1.4.1.1 Sojaproteine als Diagnostischer Marker

Die Proteine wurden daraufhin untersucht, welches sich als bester Diagnostikmarker für eine Sojaallergie eignet. So wurden von Holzhauser Gly m 5 und Gly m 6 als potentielle Marker vorgeschlagen, da diese Proteine lineare Epitope enthalten [105]. Allerdings deutet, wie bereits erwähnt, ein hoher IgE Titer gegen Gly m 4 auf eine Anaphylaxie bei Sojamilchkonsum hin [109]. Klemans dagegen publizierte, dass Gly m 8 die größte Genauigkeit als Diagnosemarker aufweist, weil Gly m 5 und Gly m 6 auch mit milden Symptomen assoziiert sind und Gly m 4 nur ein guter Marker bei Sojamilch ist [107]. Dieses wasserlösliche Protein ist jedoch in Tofu und anderen Sojaerzeugnissen stark unterrepräsentiert.

#### 1.4.2 Kreuzreaktionen bei Sojaallergie

Serologische Kreuzreaktionen gegen andere Hülsenfrüchte von sojaallergischen Patienten gibt es mit Erdnuss [110] und der grünen Erbse [111].

Oft kommt es bei Birkenpollenallergikern zu einer Kreuzreaktion nach dem Konsum von Lebensmitteln, die das Sojaprotein Gly m 4 enthalten. Birkenpollenspezifische IgEs kreuzreagieren mit Gly m 4 [112], dessen Sequenz 48 % identische Aminosäuren mit Birkenpollenallergen Bet v 1 hat [113]. In einer Studie mit 50 Birkenpollenallergiker reagierten 8 Patienten auch auf Gly m 4 [114]. Das Erdnussprotein Ara h 3 kreuzreagiert mit dem Sojaprotein Gly m 6 [115]. Erdnussallergische Kinder haben einen höheren IgE-Titer gegen die Erdnussproteine Ara h 1-3 und Sojaprotein Gly m 5-6, als asymptotische Kinder, die gegen Erdnuss sensibilisiert sind [110]. Allerdings ist die klinische

Kreuzreaktion im Vergleich zur häufig in Prick-Test und IgE-basierten Test beobachteten Kreuzsensibilisierung relativ selten [116]. Eine klinische Kreuzreaktion zu Soja wird in 1 - 6,5 % der Fälle von Antikörpern aus Erdnussallergikern in einer Placebo kontrollierten Studie beschrieben [117]. Von 140 Erdnussallergikern reagierten 7 % der Patienten auch auf Soja, was durch die Patientengeschichte, Serum IgE-Level, Prick-Test und Nahrungsmittelprovokation in einer weiteren Studie festgestellt wurde [118].

Wie bereits erwähnt kann die Kreuzreaktivität auf nah verwandten oder fast identischen Strukturen beruhen. Proteine der Sojabohne können Proteinfamilien zugeordnet werden, welche ein in vielen Pflanzen vorkommendes Strukturmotiv enthalten. Eine weitgehende Übereinstimmung von Sequenzen ist ein guter Indikator für eine Kreuzreaktivität [119]. Die Sojaproteine Gly m 1 –Gly m 6 können großen Proteinfamilien zugeordnet werden, die jeweils konservierte Struktur motive enthalten. Dadurch könnte die Kreuzreaktivität der Sojabohne erklärt werden. So gehört Gly m 1 zu der großen Familie der Prolamine. Dieses Protein enthält eine funktionell konservierte Domäne, welche in Lipid-Transfer-Proteinen, Speicherproteinen und Trypsin- $\alpha$ -Amylase-Inhibitoren vorkommt [40]. Das Strukturmotiv von Gly m 2 beinhaltet eine doppel- $\psi$   $\beta$ -Fassstruktur. Gly m 3 enthält das EF-Hand Strukturmotiv [120]. Gly m 4 gehört zu den PR10 Proteinen der Bet v 1 Familie. Gly m 5 und 6 gehören zu der Cupin Familie [121]. Um diese Kreuzreaktionen besser charakterisieren zu können, ist es wichtig, die Epitope genau zu charakterisieren.

### 1.5 Peptid Phage Display

Das Prinzip, welches dem Phage Display zugrunde liegt, wurde zuerst von G. Smith 1985 publiziert und seine Arbeiten 2018 mit dem Nobelpreis honoriert. Er beschrieb eine Methode um Gene von Peptiden oder Proteinen und Bakteriophagen-Hüllproteinen so zu verbinden, dass Fusionsproteine auf der Hülle der Bakteriophagen daraus resultieren [122]. Mit der Phage-Display-Technologie als Selektionsmethode ist es möglich ein einzelnes Peptid oder Protein mit einem gewünschten Merkmal aus einem Pool von Milliarden von (Gen-)Varianten zu identifizieren, ohne vorab detaillierte Informationen über die Beziehung zwischen Funktion und Struktur zu besitzen [123]. Die Phage Display Technologie gekoppelt mit der *in vitro* Selektion erlaubt eine schnelle Identifizierung und Optimierung von Peptiden oder Proteinen basierend auf ihren strukturellen und funktionalen Eigenschaften [124]. Die Peptide Phage Display Bibliotheken können Strukturen von natürlichen Epitopen imitieren [125, 126]. Allerdings sind Peptide, die auf Phagemid-Partikel dargestellt werden und von Anti-Allergenen Antikörper erkannt werden, nicht notwendigerweise natürliche B-Zell-Epitope. Sie imitieren nur die physiologischen Eigenschaften der Aminosäuren und sind daher nur Mimotope [127].

### 1.5.1 Der Bakteriophage M13

Zu den normalerweise verwendeten Bakteriophagen gehört M13 [128], welcher zu den filamentösen und nicht lytischen Bakteriophagen gehört. Diese enthalten einzelsträngige DNA (ssDNA) und infizieren das Bakterium *Escherichia coli* (*E.coli*) durch den F-Pilus [129]. Die infizierten *E.coli* Zellen produzieren kontinuierlich Phagemid-Partikel in das Medium, ohne dabei durch Lyse zerstört zu werden. Das Prinzip des Phagen-Display beruht darauf, dass Peptide auf der Oberfläche der Phagemid-Partikel als Fusionsproteine eines der Hüllproteine exprimiert und damit nach außen präsentiert werden. Durch Bindung an Zielmoleküle können die Phagemid-Partikel angereichert werden [128]. Auch für Bibliotheken von peptid- oder proteincodierenden Genen werden Bakteriophagenhüllproteine als Fusionspartner verwendet. Diese werden durch randomisierte Synthese [130] oder als cDNA [126] einer Proteinfamilie (z.B. Antikörper mRNAs) hergestellt. Diese auf der Oberfläche präsentierten exogenen Peptide oder Proteine können beispielsweise von Antikörpern erkannt und durch affine Anreicherung isolieren werden. Diesen Prozess nennt man auch „Biopanning“.

### 1.5.2 Biopanning

Das „Biopanning“ beinhaltet die Bindung der auf den Phagemid-Partikel präsentierten Peptide an immobilisierten Zielmolekülen (Abbildung 3). Während des Selektionsprozesses binden spezifische Antigene, die auf der Phagemid-Partikeloberfläche dargestellt werden, an beispielsweise einen korrespondierenden Antikörper, der auf einer festen Phase, z.B. Reaktionsgefäß oder in einer 96-Lochplatte, immobilisiert wurde. Die nicht-bindenden Phagemid-Partikel werden im Idealfall in nachfolgenden Waschschrritten entfernt. In einem letzten Schritt werden die gebundenen Phagemid-Partikel durch beispielsweise Absenkung des pH-Wertes oder durch Salzpufferlösungen eluiert und anschließend durch Infektion von *E.coli*-Bakterien vermehrt. Die amplifizierte Phagemid-Partikelpopulation wird in einer weiteren Selektionsrunde eingesetzt. Jeder Zyklus reichert theoretisch spezifische Phagemid-Partikel um einen Faktor von ca. 10<sup>3</sup> an. Ein typisches Selektionsexperiment besteht aus mehreren aufeinanderfolgenden Zyklen, bevor einzelne, selektive Klone analysiert werden [131].

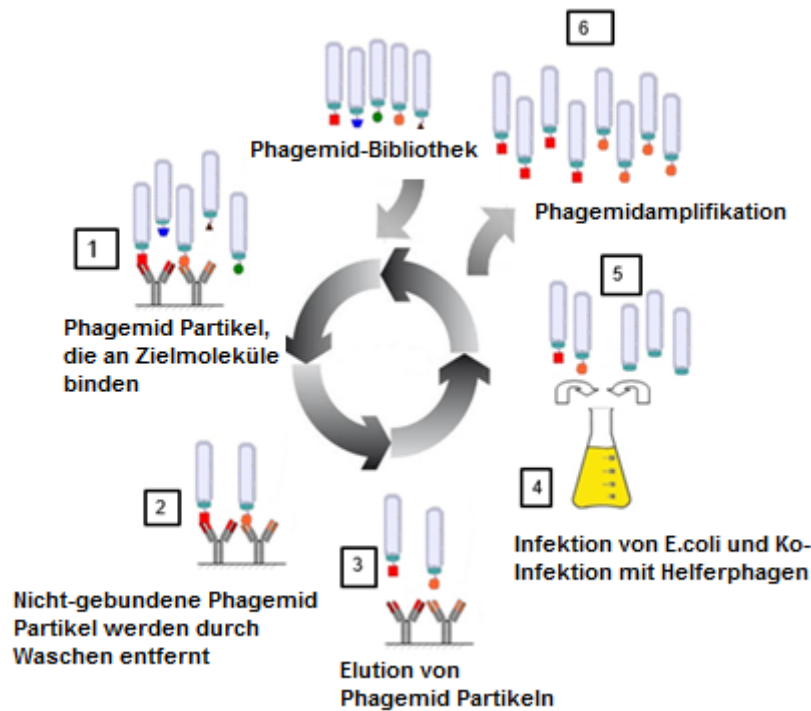


Abbildung 3: Selektionszyklus eines Peptid Phage Display Experimentes. Zielmoleküle werden auf einer Oberfläche immobilisiert (1). Nicht-gebundene Phagemid-Partikel werden durch stringentes Waschen entfernt (2). Gebundene Phagemid-Partikel werden eluiert (3) und infizieren *E.coli* Bakterien (4). Durch Amplifikation (5) entsteht eine Subbibliothek mit angereicherten, spezifischen an das Zielmolekül bindenden Phagemid-Partikeln (6).

Das Peptid Phage Display lässt sich vielseitig einsetzen. Peptide oder Proteine mit speziellen Bindungseigenschaften für ein Zielmolekül können identifiziert werden. Diese Methode stellt zusätzlich einen vielversprechenden Ansatz für das Protein – und Wirkstoffdesign dar. Eine weitere Anwendung ist die Produktion und Verbesserung der Immunogenität von synthetischen Peptid-Impfstoffen [131]. Im Rahmen der Epitopcharakterisierung stellt das Phage-Display einen neuen Ansatz zur Krankheitsdiagnostik dar [132, 133]. Neben Antikörper-Epitopen können auch Tumormarker identifiziert werden [128].

### 1.5.3 Peptide Phage Display Bibliotheken

Phage Display Bibliotheken können Antikörpergene [134, 135], cDNA [136, 137], randomisierte Oligonukleotide [138] oder sogar Gene ganzer Enzyme [139] enthalten. Häufig genutzte Bibliotheken sind Peptidgenbibliotheken, welche durch die Klonierung eines randomisierten Oligonukleotids zwischen der kodierenden Sequenz des Signalpeptids und dem 5'-Ende des Hüllproteingens hergestellt werden [140]. Umgangssprachlich werden Peptid-Phage-Display-Bibliotheken, die auf dem M13-Phagen basieren, als Phagen Bibliothek bezeichnet, unabhängig ob sie als Phage oder

Phagemid konstruiert wurden. Ein Phagemid ist ein Plasmid aus Phagen- und Plasmidgenen. Dieses enthält sowohl einen bakteriellen (ColE1) als auch den Replikationsursprung und Verpackungssignal der fd Bakteriophagen und bei Phage Display Phagemiden ein Gen eines Phagenhüllproteins, in dem die Bibliothek konstruiert wird, zum Beispiel pIII [123]. Einer der ersten Anwendungen von Oligonukleotiden zur Herstellung von DNA-Molekülen, die repliziert wurden, wurde genutzt, um neue Restriktionsstellen in Moleküle einzubauen. Dies wurde für die Konstruktion mehrerer Klonierungsstellen für pUC und verwandte Vektoren angewendet [141]. Eine Peptid Phage Display Bibliothek kann beispielsweise durch Klonierung von randomisierten synthetischen Oligonukleotiden oder Exonshuffling hergestellt werden [128, 140, 142].

Eine typische Bibliothek, welche auf der NNK-Synthese basiert, enthält  $32^n$  ( $n$  = Anzahl der Codons) Genvarianten, die aber trotzdem nur  $20^n$  verschiedene Peptide codieren. N steht für alle vier DNS-Nukleotide und K für alle außer Adenosin. Dadurch werden unter anderem unerwünschte Stopp-Codons vermieden. Wenn man mit dieser Bibliothek ein 6mer identifizieren möchte, also  $32^6 = 1 \cdot 10^9$ , würde nur eine Sequenz in der Bibliothek dieses 6mer Motiv enthalten. Dieses in einer Selektion zu „fischen“ würde einem Lotteriegewinn gleichkommen.

Es gibt einige Vorteile einer Phage Display Bibliothek über andere Expressionsvektoren. Es gibt eine physische Verknüpfung zwischen exprimierten Peptid oder Protein und der DNA, die die Peptide/Proteine kodiert. Diese DNA ist im selben Partikel verpackt. Die Phagenpartikel sind sehr klein, ein Milliliter Lösung enthält bis zu  $10^{14}$  Partikel. Dies erlaubt ein Screening von einer großen Anzahl verschiedener Partikel pro Runde. Außerdem erlaubt der Selektionsprozess eine relativ einfache Identifizierung von spezifischen Phagen [143].

Die zufällige Gensynthese resultiert aufgrund der Codonredundanz in einer statistisch ungleichen Verteilung der Aminosäuren. Ein Nachteil der zurzeit verfügbaren zufälligen Peptide Bibliotheken ist Überpräsentation einiger Aminosäuren. Einige Aminosäuren, wie beispielsweise Leucin, Arginin und Serin werden, im Vergleich zu anderen Aminosäuren, durch mehr als einem Codon kodiert. Zum Beispiel sind in der Ph.D.-7 Bibliothek, welche die am häufigsten genutzte, kommerzielle Phagen Bibliothek für Peptid-Ligand Screening ist [144], Prolin, Lysin und Alanin im Vergleich zu anderen Aminosäuren, wie beispielsweise Tryptophan und Cystein deutlich häufiger vorhanden [144]. Ein weiterer Nachteil der Ph.D.-7 Bibliothek ist die Länge der hypervariablen Region mit sieben Aminosäuren. Die Bibliothek ist nicht so komplex wie Bibliotheken mit größeren hypervariablen Regionen.

Die ENTE-1 Bibliothek nutzt die trinukleotidbasierte Synthese. Die randomisierte ENTE-1 Bibliothek hat eine Komplexität von  $10^9$  individuellen Phagemid-Partikel. Diese Bibliothek ist sehr diverse. In einem Datensatz der naiven ENTE-1 Bibliothek bestehend aus 2.191.037 Sequenzen kommen 93 % der Sequenzen nur einmal vor.

Die Peptide, welche auf der Oberfläche des Phagemid-Partikel als Fusionsprotein am N-Terminus des Hüllproteins pIII exprimiert wurden, enthalten eine Sequenz von 16 randomisierten Aminosäuren (AS). Es wurden speziell gewählte Eigenschaften bei der Auswahl der AS während einer Trinukleotid-Synthese für eine gleichmäßige Selektion berücksichtigt. Das Design der Bibliothek erlaubt maximal 18 Codons pro Position, aber praktisch alle Aminosäurekombinationen. Der hypervariable Bereich beginnt immer mit einem Glycin. Cystein ist nur in bestimmten Bereichen erlaubt, wodurch die Wahrscheinlichkeit von nahe beieinander liegenden Cystein-Resten reduziert wird, um die Bildung von ungewollten Disulfidbrücken zu vermeiden. Das instabile und oxidationsempfindliche Methionin und das stark hydrophobe Tryptophan sind unterrepräsentiert. Zudem sind keine Stopp-Codons innerhalb des 16 Aminosäuren langen hypervariablen Bereichs erlaubt. Alle diese Eigenschaften haben statistische Vorteile.

Ein weiteres Merkmal der Bibliothek ist, dass bestimmte „problematische“ Aminosäuren, wie beispielsweise Methionin und Tryptophan limitiert in ihrer Präsenz sind. Die gleichmäßige Aminosäureverteilung der ENTE-1 Bibliothek ist in Abbildung 4 gezeigt. Methionin ist nur an der vierten Position erlaubt, während Tryptophan an mehreren Stellen erlaubt ist. Damit relativ große Schleifen, die durch Ausbildung von Disulfidbrücken zwischen den Cysteinen gebildet werden, dargestellt werden können, und die Zahl direkt benachbarter Cysteine in selektierten Sequenzen verringert wird, ist das Vorhandensein von Cystein nur an einigen Stellen erlaubt.

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
A	0	0	0	0	2191037	0	34151	76520	153562	0	88729	90916	95938	100031	118955	119709	0	96036	118884	126681	127464	161028	0	0	0
C	0	0	0	0	0	0	0	0	0	133512	0	0	0	0	136517	136585	267836	121213	149378	147331	151087	156268	0	0	0
D	0	0	0	0	0	0	100365	123154	0	0	110704	107973	110750	112440	110501	110810	0	104047	90503	101216	97099	904071	0	0	0
E	0	0	0	0	0	0	221137	183688	0	0	158353	147600	144629	150969	146535	157900	310391	142220	133014	133911	131917	0	0	0	0
F	0	0	0	0	0	0	57968	78767	109056	0	101181	112129	104761	103705	113215	112570	0	113177	125253	120734	118538	379930	0	0	0
G	0	0	0	0	0	0	2191037	85376	103083	0	86405	82349	87783	88715	90944	100828	178719	88551	83765	81336	86315	0	0	0	2191037
H	0	0	0	0	0	0	169225	178908	238156	0	169814	171762	172496	174550	173301	172518	0	188708	174396	172313	174961	0	0	0	0
I	0	0	0	0	0	0	215137	130143	0	0	133170	114031	108465	112208	102369	106284	141135	106561	104087	103313	99509	264727	0	0	0
K	0	0	0	0	0	0	105385	98678	0	0	88572	79587	73915	75848	76108	80838	125274	78017	69903	66715	61683	0	0	0	0
L	0	0	0	0	0	0	66987	81841	112543	0	86132	98391	100582	98205	103137	95668	0	115485	109933	106757	117833	0	0	0	0
M	0	0	0	0	0	0	0	170014	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	200214	150796	184711	0	144163	129278	122934	120818	88246	85983	0	86683	83142	84963	80614	190511	0	0	0
P	0	0	0	0	0	0	113309	127462	153273	0	121310	122330	122191	125540	122353	117864	219843	134650	116878	116438	117499	0	0	0	0
Q	0	0	0	0	2191037	0	153597	171298	247948	0	162297	181687	185438	178011	179468	164833	348111	192752	180063	183556	190274	0	0	0	0
R	0	0	0	0	0	0	53795	90961	129483	0	87069	95480	101072	100811	101858	101037	157540	105159	104589	101254	110996	0	0	0	0
S	0	0	0	0	0	0	81294	91597	120926	855912	99376	105141	103563	101030	108958	105561	0	108991	116251	113344	112467	0	2191037	2191037	
T	2191037	0	0	0	0	0	122506	111800	140677	0	111815	107110	105643	103782	104130	102972	0	106902	103741	106902	110329	0	0	0	0
V	0	2191037	2191037	0	0	0	183783	167925	185016	0	170716	161944	161970	165212	169776	182048	0	163797	170590	174738	175603	0	0	0	0
W	0	0	0	0	0	0	72515	105421	142589	0	130258	142192	149365	137158	0	0	234317	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	130880	136705	0	0	140973	141128	139533	141998	142675	137028	207871	136998	148587	149534	150238	433710	0	0	0

Abbildung 4: Die absolute Häufigkeit der Aminosäuren in der naiven ENTE-1 Bibliothek (der Datensatz enthält 2,19 Millionen Sequenzen). FATVVQA am Anfang ist Teil der ompA Leader Sequenz und SSG ist ein Fragment des C-terminalen Peptidlinkers zu P3.

Wenn man mit dieser Bibliothek ein 6mer identifizieren möchte, also  $18^6 = 3,4 \cdot 10^7$ , enthalten viel mehr Sequenzen das gewünschte 6mer Motiv bzw. Motivvarianten und die Selektion ist wahrscheinlicher. Außerdem gewährleistet die ENTE-1 Bibliothek eine statistische Gleichverteilung der Aminosäuren. Dadurch und weil es durch den pL Promotor (siehe 4.1.1) nur geringen biologischen Selektionsdruck gibt, ist eine statistische Analyse möglich.



Die verwendete Bibliothek ist essenziell wichtig für den Erfolg der Epitopidentifizierung mittels Peptide Phage Display. Durch die Verwendung der zurzeit üblichen Bibliotheken ist es nicht möglich eine Vielzahl von Epitopen gleichzeitig zu identifizieren, wie oben an dem Beispiel mit dem 6mer Motiv gezeigt. Die ungleiche Verteilung der Aminosäuren erlaubt keine statistische Untersuchung der Epitope. Mit diesen Bibliotheken ist es nur möglich durch Anreicherung von Sequenzen Bindungspartner zu identifizieren. Durch die Sequenzanreicherung, d.h. durch Durchführung von mehreren Selektionsrunden, verliert man viele Informationen. Es werden außerdem sehr häufig unspezifische Sequenzen angereichert (siehe 4.2.1) [145]. Außerdem verliert man auch Zeit, weil mehrere Selektionsrunden durchgeführt werden müssen.

Die ENTE 1 Bibliothek mit ihren bereits beschriebenen Eigenschaften erlaubt mithilfe des Next Generation Sequencing, bei dem eine große Anzahl von Sequenzen bereitgestellt wird, eine statistische Analyse einer Vielzahl von Epitopsequenzen. Mit dieser Bibliothek ist es möglich mit wenigen Selektionsrunden eine Vielzahl von Bindungssequenzen zu identifizieren. Auch kann man längere Motive selektieren.

#### 1.5.4 Epitopidentifizierung mittels Phage Display

Peptid Phage Display wurde bereits 1988 erfolgreich für die Identifizierung von Epitopen von monoklonalen Antikörpern angewendet. de la Cruz benutzte Peptid Phage Display um Epitope von drei Antikörper, die gegen *Plasmodium falciparum* CS Protein gerichtet sind, zu identifizieren [146].

Peptid Phage Display wurde aber mit geringem Erfolg für die Identifizierung von Serumantikörper - Epitopen angewendet. Für die erfolgreiche Identifizierung von einigen wenigen Epitopen werden üblicherweise hohe Antikörperkonzentrationen benötigt [26, 147, 148]. Casey et al. [147] benutzten affinitätsgereinigte Antikörper aus Patientenserum von Epstein-Barr Virus (EBV) infizierten Patienten um Peptide anzureichern, die für die Detektion von Patientenantikörper geeignet sind. Nach sechs Selektionsrunden zeigte nur einer von 30 Klonen eine Spezifität für IgG-Antikörper gegen EBV. Folgori et al. [148] identifizierten nur 2 krankheitsspezifische Mimotope mit Patientenserum von Hepatitis B Mantelprotein (HBsAg) immunisierten Patienten. Christiansen et al. [24] benutzten die Ph.D.<sup>TM</sup> 7 Peptid Phage Display Bibliothek für die Untersuchung von vier Patientenseren von Erdnuss sensibilisierten Menschen. Sie glichen die patientenspezifischen Peptidcluster mit der Proteinsequenz von Ara h 1 ab. Für die finale Epitopidentifizierung untersuchten sie jedoch das Serum in einem Ara h 1 basierten 12mer Peptid Microarray. Nur ein 12mer Peptid von Ara h 1, welches mit den patientenspezifischen Peptidcluster übereinstimmte, wurde von IgE-Antikörpern aus drei Seren erkannt. In all diesen Beispielen wurde die Affinitätsselektion angewendet, um IgE oder IgG Serumantikörper anzureichern.

## 1.6 Zielstellung

Die Sojabohne ist aus ökonomischer Sicht eine perfekte Pflanze. Sie hat einen hohen Proteingehalt und ist gleichzeitig preiswert anzubauen. Sie stellt eine ideale Proteinquelle dar und wird daher in der Lebensmittelindustrie vielfältig eingesetzt. Die Einsatzmöglichkeiten sind unter anderem den Proteingehalt des Lebensmittels zu erhöhen, als auch die Eigenschaften des Nahrungsmittels zu verändern, um beispielsweise die Textur des Lebensmittels zu verbessern. Der universelle Einsatz der Sojabohne in Lebensmittel hat zur Folge, dass die Anzahl der Sojaallergiker ständig steigt.

Diese Arbeit wurde im Rahmen eines MAVO-Projektes angefertigt, welches von der Fraunhofer Gesellschaft gefördert wurde. Ziel dieses Projektes war es, sojabasierte Lebensmittelzutaten mit reduzierten allergischen Eigenschaften zu entwickeln. Im Rahmen dieser Promotion war es ursprünglich meine Aufgabe 5 Epitope der Sojabohne aminosäuregenau zu beschreiben, um dieses Projektziel verwirklichen zu können. Es sollte außerdem überprüft werden, ob dieselben Epitope in mehreren Seren identifiziert werden können. Basierend auf den Ergebnissen sollte untersucht werden, ob es möglich ist die Allergiediagnostik zu verbessern.

- 2. Material und Methoden
- 2.1 Material
- 2.1.1 Mikrobiologisches Material
- 2.1.1.1 *E. coli*-Stämme

*E. coli*-Stamm Genotyp/Phänotyp  
 TG1( $\lambda$ ) F [*traD36 proAB<sup>+</sup> lacI<sup>q</sup> lacZ $\Delta$ M15*]*supE thi-1*  
 $\Delta$  (*lac- proAB*)  $\Delta$ (*mcrB- hsdSM*)5, (*r<sub>K</sub>-m<sub>K</sub>*-),  $\lambda$ <sup>+</sup>

Quelle: [149]

Bakteriophage Genotyp/Phänotyp  
 M13K07 Km<sup>R</sup>-Gen (Tn903) und Plasmid Replikationsursprung (p15A)  
 integriert in Replikationsursprung  
 Konzentration 10<sup>14</sup> CFU/ml

Quelle: [150]

- 2.1.1.2 Plasmid

Plasmid Genotyp/Phänotyp  
 pPepPr3A Ampicillinresistenz-Gen (Abbildung 5)

Quelle: [151]

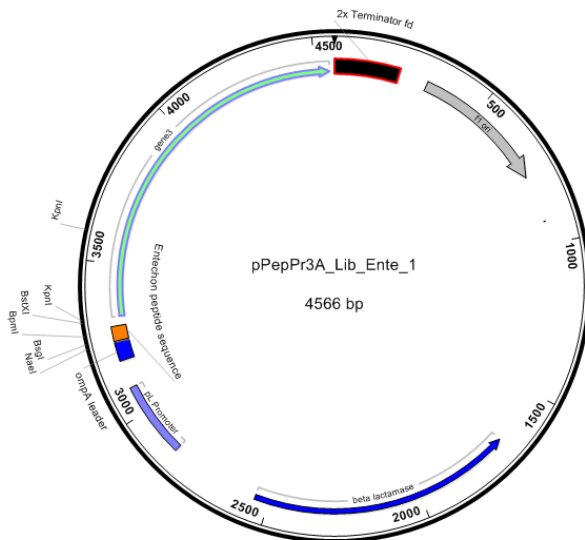


Abbildung 5: Vektorkarte des pPepPr3A

## 2.1.2 Seren

Die während der Arbeit verwendeten Seren, wurden von dem Universitätsklinikum Leipzig (UKL) – Klinik und Poliklinik für Dermatologie, Venerologie und Allergologie bereitgestellt. Die Seren wurden von 50 Personen, Median 36 Jahre (19- 66 Jahre, 26 Männer) gespendet. Alle Spender hatten einen positiven Prick-Test auf Soja (Größe > 3 mm). Der Prick-Test wurde mit kommerziell, verfügbaren Sojadrink (ALPRO, Uelzena, Deutschland) durchgeführt. Zusätzlich wurde spezifisches IgE mittels ImmunoCAP assay, (Thermo Scientific Phadia, Freiburg, Deutschland) gemessen. Alle Individuen waren positiv bei >0.35kU/l für mindestens ein Sojaallergen (Gly m 4, Gly m 5, Gly m 6). Zusätzlich wurde auch der gesamt IgE- Gehalt bestimmt. Häufig wurden auch die Birkenpolle im Prick-Test positiv getestet (Tabelle 3). Die Probe F1 spendete ein Kollege am Fraunhofer Institut für Zelltherapie und Immunologie (IZI). Die Seren wurden aliquotiert bei -80°C gelagert. Ein entsprechendes Ethikvotum (Nr. 428-12-17122012) der Universität Leipzig lag vor.

Tabelle 3: Liste der humanen Seren, die verwendet worden.

Serums- Bezeichnung	IgE Titer (KU/L)	Herkunft	Klassifikation	Alter zum Testzeitpunkt	Symptome
13001	86,4	UKL	Sojaallergie, Birkenpollenallergie	53	nicht bekannt
13002	124	UKL	Sojaallergie, Birkenpollenallergie	59	nicht bekannt
13003	88	UKL	Sojaallergie, Birkenpollenallergie	28	moderate
13004	26,9	UKL	Sojaallergie, Birkenpollenallergie	61	nicht bekannt
13005	-	UKL	Sojaallergie, Birkenpollenallergie	36	keine
13006	>5000	UKL	Sojaallergie, Birkenpollenallergie	29	nicht bekannt
13007	964	UKL	Sojaallergie, Birkenpollenallergie	36	schwere
13008	302	UKL	Sojaallergie, Birkenpollenallergie	64	keine
13009	>5000	UKL	Sojaallergie, Birkenpollenallergie	45	keine
13010	248	UKL	Sojaallergie, Birkenpollenallergie	29	keine
13011	127	UKL	Sojaallergie, Birkenpollenallergie	30	milde
13012	75,5	UKL	Sojaallergie, Birkenpollenallergie	43	milde
13013	-	UKL	Sojaallergie, Birkenpollenallergie	23	milde
13014	456	UKL	Sojaallergie, Birkenpollenallergie	23	keine

Serums- Bezeichnung	IgE Titer (KU/L)	Herkunft	Klassifikation	Alter zum Testzeitpunkt	Symptome
13016	336	UKL	Sojaallergie, Birkenpollenallergie	29	moderate
13017	272	UKL	Sojaallergie, Birkenpollenallergie	57	keine
13018	260	UKL	Sojaallergie, Birkenpollenallergie	27	nicht bekannt
13019	-	UKL	Sojaallergie, Birkenpollenallergie	48	keine
13020	78,2	UKL	Sojaallergie, Birkenpollenallergie	26	keine
13021	261	UKL	Sojaallergie, Birkenpollenallergie	31	keine
13022	>100	UKL	Sojaallergie, Birkenpollenallergie	32	nicht bekannt
13023	872	UKL	Sojaallergie, Birkenpollenallergie	28	milde
13024	220	UKL	Sojaallergie, Birkenpollenallergie	21	moderate
13025	859	UKL	Sojaallergie, Birkenpollenallergie	23	keine
14001	72,3	UKL	Sojaallergie, Birkenpollenallergie	63	keine
14002	157	UKL	Sojaallergie, Birkenpollenallergie	48	milde
14003	56	UKL	Sojaallergie, Birkenpollenallergie	54	milde
14004	650	UKL	Sojaallergie, Birkenpollenallergie	19	moderate
14005	68,1	UKL	Sojaallergie, Birkenpollenallergie	27	keine
14006	60,7	UKL	Sojaallergie, Birkenpollenallergie	33	keine
14007	-	UKL	Sojaallergie, Sojaallergie	65	moderate
14008	>5000	UKL	Sojaallergie, Birkenpollenallergie	34	nicht bekannt
14009	49,4	UKL	Sojaallergie, Birkenpollenallergie	44	schwere
14010	-	UKL	Sojaallergie, Birkenpollenallergie	23	nicht bekannt
14011	51	UKL	Sojaallergie, Birkenpollenallergie	49	keine
14012	117	UKL	Sojaallergie, Birkenpollenallergie	66	moderate
14013	2427	UKL	Sojaallergie, Sojaallergie	27	nicht bekannt
14014	31,9	UKL	Sojaallergie, Birkenpollenallergie	45	nicht bekannt
14015	>5000	UKL	Sojaallergie, Sojaallergie	62	milde
14016	309	UKL	Sojaallergie, Birkenpollenallergie	50	milde
14017	206	UKL	Sojaallergie	44	keine

Serums- Bezeichnung	IgE Titer (KU/L)	Herkunft	Klassifikation	Alter zum Testzeitpunkt	Symptome
14018	>5000	UKL	Sojaallergie	45	keine
14019	-	UKL	Sojaallergie	63	keine
14020	-	UKL	Sojaallergie	27	milde
14021	138	UKL	Sojaallergie	25	milde
14022	>5000	UKL	Sojaallergie	34	keine
14023	-	UKL	Sojaallergie	50	nicht bekannt
14024	>5000	UKL	Sojaallergie	33	keine
14025	134	UKL	Sojaallergie	54	milde
F1	n.b.	IZI	keine Allergien bekannt	50	
EU	n.b.	IZI	nicht-atopisches Serum	n.b.	
F3	n.b.	IZI	nicht-atopisches Serum	n.b.	
F4	n.b.	IZI	nicht-atopisches Serum	n.b.	
F5	n.b.	IZI	nicht-atopisches Serum	n.b.	

n.b.: nicht bestimmt

### 2.1.3 Synthetische Peptide

Die im Validierungsexperiment (hier ein *Microarray*) getesteten Peptide sind in der folgenden Tabelle 4 gelistet. Die Peptide wurden vom Institut für Biochemie der Universitätsklinik Köln/Abteilung für Chemie von Frau Prof. Dr. Neundorf synthetisiert und von sojaspezifischen AK erkannt. Sie liegen als Trifluoressigsäure (TFA) -Salze vor (ca. 30 %) und sind am N-Terminus acetyliert und besitzen am C-Terminus einen EBES-Lys(N3)-NH<sub>2</sub>-Linker.

Tabelle 4: Peptidliste, die in der Arbeit verwendet worden sind.

Motiv	Sequenz des Peptides	ID	Reinheit	pI <sup>(a)</sup>	MW <sup>(b)</sup> in Da	Kommentare
GREEGQQQ	NKVLFGREEGQQQG	DN8	>90	6,14	1970,8	Gly m 5/PrSeq
IAEG	AEGHPLSLKF	D010	>85	6,79	1522,5	Gly m TI / PhSeq
KYQ	SDKYQEEFQPR	D011	>85	4,68	1565,4	Gly m 6/PhSeq
KYQ	TEKYQQNSSG	D012	>75	5,66	1565,4	Gly m 6/PhSeq
KYQ	EFLKYQQQQQ	D013	>75	6,1	1763,7	Gly m 6/PhSeq
NHEE	GENHEEQHVI	D014	>75	4,72	1615,4	Gly m Bd30K/PhSeq
NHEE	YHNHEEEAKR	D015	>75	6,02	1736,6	Gly m Bd30K/PhSeq
CDHP	TCDHPVKGQP	D016	-	6,4	1504,4	Gly m Bd30K/PhSeq
CDHP	GEYSCDITYT	D017	>60	3,67	1574,4	Gly m Bd30K/PhSeq
OPRQ	SHHDQPRQHA	D019	>70	6,78	1636,4	Gly m 5/PhSeq
KALQK	KCQCKALQKI	D020	>85	9,39	1584,7	Gly m 8/ PrSeq
QTEGRVCESQSHGFH	QTEGRVCESQSHGFH	D023	-	6	2082,0	Gly m 2/ PrSeq
QEEDEEEE	QDQEEDEEDE	D024	>90	3,23	1689,3	Gly m Bd30K/PhSeq
FASQVVVQTEGRVC	FASQVVVQTEGRVCE	D01	>50	4,53	2120,0	Gly m 2/ PrSeq
	S					
SHGFH	SQSHGFHGLCNRDH	D02	>85	6,77	2264,0	Gly m 2/ PrSeq
	NCA					
DDEEEQ	GVYNSQVDDEEEQN	D03	>90	3,71	2406,0	Gly m 6/PhSeq
	QRD					

Motiv	Sequenz des Peptides	ID	Reinheit	pI(a)	MW(b) in Da	Kommentare
QEEDEEEE	GYNPCRQEEDEELHH KC	D04	>75	4,92	2453,3	Gly m Bd30K/PhSeq
KKNHG	GTHFSKAVLGKKHGD EF	D05	>90	8,51	2239,1	Gly m TI / PhSeq
KKYL	GPVQKKYLFDS	D06	>90	8,5	1662,5	Gly m Bd30K/PhSeq
KKYL	GEFSKKYLQAP	D07	>90	8,5	1648,5	Gly m Bd30K/PhSeq
IAEG	GCVQVIAEGHPHC	D08	>90	5,98	1730,6	Gly m TI / PhSeq
CDHP	KEQYSCDHPP	D018	>75	5,32	1583,3	Gly m Bd30K/PhSeq
LSSTE	FFLSSTEAQ	D021	>90	4	1409,2	Gly m 5/PhSeq
QKIY	DRHQKIYHF	D022	>85	8,6	1623,5	Gly m 6/PrSeq
HPEREPQQPGEKE	QHPEREPQQPGEKEE	DN1	>85	4,65	2198,9	Gly m 5/PrSeq
PRQEEHEQREEQ	PRQEEHEQREEQE W	DN2	>85	4,35	2420,1	Gly m 5/PrSeq
KEQQQE	DANIELVGLKEQQQE	DN3	>75	4	2094,9	Gly m 5/PrSeq
EGEQP, EGPR, PRPF	DEGEQPRPFPPRPR	DN4	>90	6,18	2206,1	Gly m 5/PrSeq
IKEQQQRQ	NIELVGKEQQQRQQ Q	DN5	>90	6,14	2320,2	Gly m 5/PrSeq
	EPQQKGQSSRPQDR H	DN6	>90	8,85	2158,9	Gly m 6/PrSeq
NSERDSYRNQACH	QSCNSERDSYRNQ	DN7	>85	6,06	1967,6	Gly m 5/PrSeq
GREEGQQQ, GQQQGEE	NKVLFGREEGQQQG	DN8	>90	6,14	1970,8	Gly m 5/PrSeq
GNQEQ	RRFYLAGNQEQEFL	DN9	>90	6,14	2152	Gly m 6/PrSeq
	NPDIHPETMQQ	DN10	>90	4,13	1819,6	Gly m 6/PrSeq
	HAANVVGNKESQRE AR	DN11	>75	8,75	2146,9	Gly m 7/PrSeq
CTFD	GVINCETFDQYQPIEI GC	DK3	>90	3,57	2278	PR5/ PhSeq
LSSTE	FFLSSTEAQ	DL1	>90	4	1452	Gly m 5/PhSeq
GDFQ	GVFSPSTNFHPYGDF QI	DL2	>90	5,08	2305	Gly m 5/PhSeq
FPGC	MIFPGCPS	DL3	>90	5,27	1275	Gly m 6/PhSeq
SSRPQD, PQDRHQ, IYHFRE	DRHQKIYHF	DL4	>90	8,6	1623	Gly m 6/PrSeq
QTEGRVCESQSHGFH	QTEGRVCESQSHGFH	DL5	>65	6	2083	Gly m 2/ PrSeq
SADGV	GVSSADGVGQINIDQ N	DL6	>90	3,56	1955	Gly m Bd30K/PhSeq
QEEDEEEE	GYVPCRQEEDEELHH KC	DL7	>65	4,92	2449	Gly m Bd30K/PhSeq
QEEDEEEE	QDQEEDEEDE	DL8	>90	3,23	1647	Gly m Bd30K/PhSeq
QKLD	GATSSDRQLDKESPG YVI	DL9	>90	4,56	2305	Gly m TI / PhSeq
QKLD	QKLDKESLAK	EO12	>90	8,5	1584,8	Gly m TI / PrSeq
EGEQP	TQHEGEQPFY	EO13	>85	4,5	1660,8	Gly m 5/PhSeq
NKVL	SDQNKVLQPH	EO14	>90	6,46	1590,7	Gly m 5/PhSeq
PRPQHP	GSPRPQHPIDV	EL10	>90	6,74	1681,9	Gly m 5/PhSeq
NSERDS	GNANSERDSIIQ	EL11		4,37	1685,8	Gly m 5/PhSeq
STFD	GRTNSTFDNAG	EL12	>70	5,84	1521,7	PR5/ PrSeq

Motiv	Sequenz des Peptides	ID	Reinheit	MW(b)		Kommentare
				pI(a)	in Da	
SAGDV	GYGSADGVDYW	EL13	>85	3,56	1571,7	Gly m Bd30K/PrSeq
CNRD	GFNLGNRDRPAP	EL14	>50	8,25	1741,9	Gly m 2/ PhSeq

(a): pI - isoelektrischer Punkt; (b): MW – Molekulargewicht; PrSeq: Proteinsequenz; PhSeq: Phagensequenz

#### 2.1.4 ENTE-1 Phagenbibliothek

Zur Affinitätsselektion wurde die durch *cosmix-plexing* erzeugte Phagenbibliothek ENTE-1 mit ca.  $10^9$  individuellen Klonen verwendet, bei der es sich um eine Peptid präsentierende Phagenbibliothek handelt. Diese wurde in einer Kooperation des Fraunhofer IZI mit der Firma PolyQuant GmbH synthetisiert und kloniert. Die Phagenbibliothek liegt mit einer Konzentration von  $1 \times 10^{13}$  CFU/ml vor.

#### 2.1.5 Chemikalien

Agar-Agar	Sigma, Deisenhof, D
Agarose NEEO, Ultra Qualität	Carl Roth, Karlsruhe D
Ammoniumchlorid	Carl Roth, Karlsruhe D
BSA	AppliChem D
H3BO3	VWR, Darmstadt, D
Calciumchlorid	Carl Roth, Karlsruhe, D
Glucose	Carl Roth, Karlsruhe, D
FeSO4	Merck, Darmstadt, D
Ethidiumbromid	Invitrogen, Carlsbad, CA, USA
Glycin	Carl Roth, Karlsruhe, D
Hefeextrakt	Fluka, Buchs, CH
Kaliumchlorid	Carl Roth, Karlsruhe, D
Kaliumhydrogenphosphat	Carl Roth, Karlsruhe, D
Kupfersulfat-Pentahydrat	Carl Roth, Karlsruhe, D
Cobalt(II)-nitrat	Carl Roth, Karlsruhe, D
Magnesiumchlorid	Carl Roth, Karlsruhe, D
Magnesiumsulfat	Carl Roth, Karlsruhe, D
Mangan(II)-chlorid	Sigma, Deisenhof, D
$\beta$ -Mercaptoethanol	Carl Roth, Karlsruhe, D
Natriumacetat	Carl Roth, Karlsruhe, D
Natriumchlorid	AnalaR NORMAPUR, Darmstadt D
di-Natriumhydrogenphosphat	Carl Roth, Karlsruhe, D
non fat dry milk (NFDM)/ Milchpulver	Carl Roth, Karlsruhe, D



Polyethylenglycol	Carl Roth, Karlsruhe, D
Thiamin	Carl Roth, Karlsruhe, D
Tris.HCl	Carl Roth, Karlsruhe, D
Trypton	FLUKA, Buchs, CH
Tween®20	Carl Roth, Karlsruhe, D
Zinksulfat-Hexahydrat	Sigma, Deisenhof, D
Ammoniummolybdat	Sigma, Deisenhof, D

#### 2.1.6 Geräte-Liste

##### 2.1.6.1 Zentrifugen

Avanti® J-26 XP, Beckman Coulter®

Biofuge primo R, Heraeus

Micro 200 R, Hettich Zentrifugen

##### 2.1.6.2 Photometer

BioPhotometer, Eppendorf

Spectralphotometer NanoDrop ND-2000, Thermo Scientific

##### 2.1.6.3 Micro Plate Reader

Infinite M200, Tecan

##### 2.1.6.4 Inkubatoren

Zirkulierender Luftinkubator, Heraeus Thermo Scientific

Schüttelinkubator Innovar® 40R, New Brunswick Scientific Edison

##### 2.1.6.5 Arrayscanner

Geneprix® Professional 4200A

#### 2.1.7 Computersoftware

Tecan i-control 1.4., Tecan

DNASTAR® 12.1.0.145, Lasergene

2.1.8 Lösungen

5x Bradford Reagenz

0,125 g Coomassie Blue G250 (Brilliant Blue)  
 60 ml 99,8 % Ethanol (unvergällt)  
 125 ml Phosphorsäure(85 % H<sub>3</sub>PO<sub>4</sub>)  
 mit Wasser auf 250 ml auffüllen  
 bei 4°C im Kühlschrank lagern

2.1.8.1 Medien

Für alle Zubereitungen wurde deionisiertes Wasser einer Millipore-Analge (Leitfähigkeit < 0,058 µS/cm) verwendet.

dYT-Medium:

10g/l Hefeextrakt  
 16 g/l Trypton  
 5 g/l NaCl  
 Autoklavieren

M9 Medium:

Die folgenden Lösungen wurden steril filtriert. Die entsprechenden Mengen der Lösungen wurden bei 50°C gemischt.

2x Phosphates:

1.86 g/l Na<sub>2</sub>HPO<sub>4</sub>  
 0.93 g/l KH<sub>2</sub>PO<sub>4</sub>

10,000x Traces:

26.8 g/l H<sub>3</sub>BO<sub>3</sub>  
 18.1 g/l MnCl<sub>2</sub>\*H<sub>2</sub>O  
 2.2 g/l ZnSO<sub>4</sub>\*7 H<sub>2</sub>O  
 3.9 g/l ((NH<sub>4</sub>)<sub>2</sub>MoO<sub>4</sub>)  
 0.79 g/l CuSO<sub>4</sub>\*5 H<sub>2</sub>O  
 0.49 g/l Co(NO<sub>3</sub>)<sub>2</sub>\*6H<sub>2</sub>O

10,000 x FeEDTA:

6 g/l FeSO<sub>4</sub>  
 6 g/l C<sub>6</sub>HO<sub>7</sub>H<sub>2</sub>O  
 1 g/l Dinatrium-EDTA

1,000x Thiamine:

1 g/l Thiamine  
 Glucose 50 % (w/v) 10 ml  
 Auffüllen auf 1 l H<sub>2</sub>O

2.1.8.2 Puffer

1x PBS:

8,0 g/l NaCl  
 0,2 g/l KCl  
 1,44 g/l Na<sub>2</sub>HPO<sub>4</sub>• 2 H<sub>2</sub>O

T-PBS:	0,24 g/l $\text{KH}_2\text{PO}_4$ 50 mM Tris-HCl 150 mM NaCl 1x PBS
Blockingpuffer:	5 % Non fat dry milk (NFDM) / Magermilchpulver (Affinitätsselektions Experiment), 1 % BSA/ 0,1 % Tween®20 in 1xPBS (Microarray Experiment)
Waschpuffer:	0,1 % oder 0,5 % Tween®20 in 1x PBS
Elutionspuffer:	0,1 M Glycin*HCl, pH 2,2
Neutralisationspuffer:	1M Tris*HCl, pH 8.0
PEG/NaCl:	200 g/l Polyethylenglycol 8000 (PEG) 145.25 g/l NaCl Auffüllen auf 1 l $\text{H}_2\text{O}$
50x TAE-Puffer:	242 g/l Tris 57,1 ml Essigsäure (100 %) 100 ml 0,5 M EDTA Auffüllen auf 1 l $\text{H}_2\text{O}$

### 2.1.8.3 Antibiotika Stammlösungen

Alle Stammlösungen der Antibiotika werden bei  $-20^\circ\text{C}$  aufbewahrt.

Ampicillin (100 mg/ml):	4 g Ampicillin (Natriumsalz) 40 ml $\text{H}_2\text{O}$ , sterilfiltriert (0.2 $\mu\text{m}$ )
Kanamycin (50 mg/ml):	2.5 g Kanamycin (Natriumsalz) 50 ml $\text{H}_2\text{O}$ , sterilfiltriert (0.2 $\mu\text{m}$ )

Folgende Endkonzentrationen wurden bei den Experimenten eingesetzt:

Ampicillin 200  $\mu\text{g}/\text{ml}$   
Kanamycin 50  $\mu\text{g}/\text{ml}$

### 2.1.9 Antikörper

Anti-Human IgG H&L (Cy5®), polyklonal, entwickelt in Ziege, Abcam (ab97172)

Anti-Human IgE Fc, monoclonal HP6029, entwickelt in Maus, Abcam (ab99834)

Anti-Mouse IgG H+L (Cy5®), polyklonal, entwickelt in Ziege, Thermo Fisher Scientific (A10524)

#### 2.1.10 Polymerase

Vent® DNA Polymerase, New England BioLabs (M0254S)

#### 2.1.11 Reinigungskits

QIAquick Purification (QIAGEN), Produktnummer: 28104

### 2.2 Methoden

#### 2.2.1 Kultivierung der Bakterien

Die Bakterien wurden in dYT-Medium, dem, wenn es nötig war, das entsprechende Antibiotikum zugesetzt wurde, kultiviert. F'-Episom enthaltene Stämme wurden auf M9-Platten kultiviert. Für eine längere Aufbewahrung wurden in Kryogefäßen 1 ml frisch kultivierte Übernachtskultur mit 1 ml sterilem Glycerin (50 %) gemischt, in flüssigem Stickstoff eingefroren und bei -80°C aufbewahrt.

#### 2.2.2 Selektionszyklus

##### 2.2.2.1 „Panning“

##### 2.2.2.1.1 Immobilisierung des Zielproteins

Immunotubes („Protein binding Immuno™ Tubes“ mit „maxisorp™“ Oberfläche von Nunc) wurden mit je 2,5 ml PBS, das Serum entsprechend 24 µg Serumprotein enthielt, beschichtet und eine Stunde bei 4 °C rotierend inkubiert. Die beschichteten Gefäße wurden mit 2,5 ml Blocking-Puffer gespült und anschließend mit weiteren 4 ml Blocking-Puffer 15 min rotierend bei 4 °C geblockt. Nachdem der Blocking-Puffer entfernt wurde, wurden die Immunotubes dreimal mit 4 ml Waschpuffer gewaschen. Im Anschluss wurde mit der Phagenbibliothek zwei Stunden bei 4 °C rotierend inkubiert. Die Phagenbibliothek enthielt je nach Selektionsrunde  $10^9$ -  $10^{12}$  Phagemid-Partikel in 1 ml Blocking-Puffer, der noch  $1 \cdot 10^{12}$  M13K07 und 30 µg Serum Proteine von Personen ohne diagnostizierter Sojaallergie enthielt. Es wurden zwei Selektionsrunden durchgeführt. In der ersten Selektionsrunde wurden  $10^{12}$  Phagemid-Partikel der naiven ENTE-1 Bibliothek eingesetzt. In der 2. Selektionsrunde wurde das 100-1000 fache der Zahl der eluierten Phagemid-Partikel eingesetzt. Nicht gebundene Phagemid-Partikel wurden durch fünfmaliges Waschen mit 1 ml Wasch-Puffer entfernt. In späteren „Biopanning“-Runden wurde mit höherer Detergenzienkonzentration gewaschen (PBS/0,5 % Tween®20). Die gebundenen Phagemid-Partikel wurden zur Elution mit 1 ml Elutionspuffer 5 min vortextend eluiert und in einem neuen Reaktionsgefäß mit 200 µl Neutralisationspuffer neutralisiert. Die neutralisierten Phagemid-Partikel wurden bei 4 °C aufbewahrt.

#### 2.2.2.1.2 Infektion von TG1 ( $\lambda$ ) F' mit eluierten Phagemid-Partikel

Zur Infektion wurden die eluierten Phagemid-Partikel mit einer 10 ml TG1( $\lambda$ )-Kultur Kultur mit einer O.D.<sub>600</sub> 0,5-1 vermischt und 20 min bei 37 °C schüttelnd inkubiert. Von dieser Bakterienkultur wurde ein kleines Volumen für die Titerbestimmung (2.2.3.2) abgenommen und der Rest wurde auf einem 250x250 mm Agar-Nährboden, mit 200 µg/ml Ampicillin, ausgestrichen. Nach dem Antrocknen unter der Laminar-Flow-Bank wurden die Platten über Nacht bei 30 °C inkubiert.

Am nächsten Tag wurden die Kolonien in 25 ml dYT-Medium, das 10<sup>10</sup> cfu/ml M13K07 enthielt, auf einer Schaukelplattform abgewaschen. Diese Zellen wurden in drei Proben aufgeteilt. Für die Phagemidproduktion wurde 25 ml dYT-Medium (+ 200 µg/ml Ampicillin und 10<sup>10</sup> cfu/ml M13K07) in 250 ml Kulturkolben angeimpft, sodass etwa eine O.D.<sub>600</sub> 0,5 erreicht wurde. Diese Kultur wurde eine Stunde bei 37°C und über Nacht, aber nicht länger als 13 Stunden, bei 30°C geschüttelt. 2x 1 ml der Flüssigkultur wurden als Rücklage in 50 % Glycerol bei -80 °C gelagert. 200 µl der Flüssigkultur wurden für die DNA- Reinigung aufbewahrt.

#### 2.2.2.1.3 Reinigung der Phagemid-Partikel

Die 25 ml Übernachtskultur wurde 30 min bei 20.000 x g und 4 °C in einem Zentrifugenröhrchen zentrifugiert, um die Bakterienzellen zu entfernen. Der Überstand wurde vorsichtig in ein neues Zentrifugenröhrchen dekantiert und mit ¼ Volumen kalter PEG/NaCl-Lösung versetzt und gründlich gemischt. Es wurde eine Stunde auf Eis inkubiert und anschließend 30 min bei 20.000 x g und 4 °C zentrifugiert. Nach sorgfältiger Entfernung des Überstands wurde das Pellet in 1 ml 1xPBS resuspendiert, in ein 1,5 ml Reaktionsgefäß überführt und 20 min bei 4 °C, 14.000 x g zentrifugiert, um unlösliche Substanzen zu entfernen. Der Überstand wurde in ein neues 2,0 ml Reaktionsgefäß überführt, und mit ¼ Volumen PEG/NaCl versetzt. Die resuspendierten Phagemid-Partikel wurden mindestens 15 min auf Eis inkubiert und durch 30 min bei 4 °C und 20000 x g pelletiert. Die Flüssigkeit wurde entfernt, das Pellet wurde in 200 µl PBS resuspendiert und bei 4 °C bis zur Weiterverarbeitung (maximal einige Tage) aufbewahrt. Eine Titerbestimmung (2.2.2.2) wurde im Anschluss durchgeführt.

#### 2.2.2.2 Titerbestimmung

Eine Kolonie TG1( $\lambda$ ) wurde zuvor in 10 ml dYT- Medium verdünnt und so lange bei 37 °C schüttelnd inkubiert, dass eine O.D.<sub>600</sub> von 0,5-1,0 erreichten wurde. Um die Anzahl der infektiösen Phagemid-Partikel zu bestimmen, wurden in einer 96-Loch Polystyrol Mikrotiterplatte bis zu zehn Verdünnungsstufen angefertigt. Dazu wurden 10 µl Phagemid Lösung in 90 µl sterilem dYT-Medium verdünnt. Von den letzten drei Verdünnungsstufen wurden 10 µl der verdünnten Phagemid-Partikel mit 90 µl frischer Bakterienkultur vermischt und 30 min bei 37 °C inkubiert, um die Infektion und die

Expression der Resistenz-Gene zu gewährleisten. Anschließend wurden die Lösungen gut durchmischt und viermal je 3 µl auf dYT- Agrar-Platten, denen 200 µg/ml Ampicillin zugesetzt wurden, aufgetragen. Die Platten wurden ca. 5 min unter der Laminar-Flow-Bank angetrocknet und über Nacht bei 30°C inkubiert. Am nächsten Tag wurden die Kolonien gezählt und die Anzahl der „colony forming units“ (cfu) bestimmt.

### 2.2.3 Bestimmung der Proteinkonzentration in humanen Serum

Die Proteinkonzentration der Seren wurde nach der Bradford - Methode [152] bestimmt. Das Prinzip des Bradford Assays liegt in der Bindung des Farbstoffes Coomassie Brilliant Blue G-250 an Proteine im sauren Milieu. Die Bindung verschiebt das Absorptionsmaximum des Farbstoffes von 465 auf 595 nm. Die Farbreaktion ist abhängig vom Gehalt an aromatischen und basischen AS (z.B. Arginin). Zur Kalibrierung diente eine Eichkurve (BSA-Standard) mit den Konzentrationen von 0 mg - 2 mg. Die Seren wurden mit Bradford-Reagenz versetzt, 5 min bei RT inkubiert und anschließend die Absorption bei 595 nm bestimmt.

### 2.2.4 Behandlung der DNA – Extraktion bis Sequenzierung

#### 2.2.4.1 Isolation und Aufreinigung der DNA

Die Isolierung und Aufreinigung der DNA wird nach entsprechendem Protokoll des DNA-Kits Nucleo Spin Plasmid® Kit von Macherey-Nagel (Produktnummer: 740588.10) durchgeführt.

#### 2.2.4.2 Messung der DNA-Konzentration

Die Konzentration der DNA wird mithilfe eines NanoDrop ND-2000 Spektralphotometers nach Herstelleranweisung bei einer Wellenlänge von 260 nm ermittelt. Für einen Einsatz in der Polymerasekettenreaktion (PCR) sollten diese bei etwa 100 µg/µl liegen. Die Konzentrationsmessung für die Sequenzierung wird mittels Qubit® Fluorometer der Firma Invitrogen bestimmt. Die Messung wird mit dem Qubit® dsDNA BR Assay Kit (Produktnummer: Q32850) nach Herstelleranleitung durchgeführt. Für die Sequenzierung sind Konzentrationen von > 5 ng/µl nötig.

#### 2.2.4.3 Primerhybridisierung für Sequenzierung

Für das Next-Generation-Sequencing (NGS) der Firma illumina (MiSeq™) sind spezifische Primer für jede DNA-Probe nötig, sodass eine parallele Sequenzierung möglich ist. Mittels PCR können spezifische Primer an die aufgereinigte DNA hybridisiert werden (Tabelle 7/Tabelle 8). Die verwendete besonders stabile Vent® DNA Polymerase besitzt aufgrund ihrer exo-Nukleaseaktivität eine Korrekturlese-Funktion (proofreading) eine geringe Fehlerrate (high-fidelity).

Tabelle 5: PCR – Vorschrift für einen 50 µl Ansatz

Substanz	Menge (µl)	
Puffer (10fach, mit MgSO <sub>4</sub> )	5	Mastermix
dNTPs	1	
Primer 1 (P5_N501)	1	
Polymerase	0,5	probenspezifisch
H <sub>2</sub> O	ad 50	
DNA	1	
Primer 2 (spezifisch)	1	

Tabelle 6: PCR Programm

Schritt	Temperatur	Zeit	Beschreibung	
1	95°C	5 min	Initiale Denaturierung	15x wiederholen
2	95°C	45 sec	Denaturierung	
3	61,7°C	45 sec	Hybridisierung der Primer	
4	72°C	15 sec	Synthese der DNA	

Tabelle 7: Sequenzen der verwendeten Primer der PCR für NGS

Primername	Sequenz
Primer 1	AATGATACGGCGACCACCGAGATCTACAC TCTTCCCTACACGACGCTCTTCCGATCT
P5_N501	GGCTGGTTTCGCTACCGTAG
Primer 2	CAAGCAGAAGACGGCATAACGAGAT (XXXXXX) <sup>(a)</sup>
(spezifisch)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

(a) XXXXXX: Individueller Primer Sequenzabschnitt

Tabelle 8: Übersicht der verwendeten Primer

Primer 2 (spezifisch)	Individueller Primer Sequenzabschnitt
P7_A701	CGTGAT
P7_A702	ACATCG
P7_A703	GCCTAA
P7_A704	TGGTCA
P7_A705	CACTGT
P7_A706	ATTGGC
P7_A707	GATCTG
P7_A708	TCAAGT
P7_A709	CTGATC
P7_A710	AAGCTA
P7_A711	GTAGCC
P7_A712	TACAAG
P7_A713	TTGACT
P7_A714	GGAACT
P7_A715	TGACAT
P7_A716	GGACGG

<b>Primer 2 (spezifisch)</b>	<b>Individueller Primer Sequenzabschnitt</b>
P7_A718	GCGGAC
P7_A719	TTTCAC
P7_A720	GGCCAC
P7_A721	CGAAAC
P7_A722	CGTACG
P7_A723	CCACTC
P7_A725	ATCAGT
P7_A727	AGGAAT

#### 2.2.5 Agarose-Gel-Elektrophorese

Zur Qualitätskontrolle der PCR-Produkte diente die Größenauftrennung der Probe mittels Gel-Elektrophorese nach Russel [153]. Ethidiumbromid, welches mit der DNA interkaliert, diente der Bandenerkennung im UV-Licht. Das Probenvolumen betrug 8  $\mu\text{l}$  (plus 2  $\mu\text{l}$  6x Ladepuffer). Die Einstellungen des Laufs betragen 100 V, 400 mA und 25 min. Die Banden wurden mit einem UV-Transilluminator detektiert.

#### 2.2.6 Aufreinigung und Konzentrationsbestimmung der PCR-Produkte

Die Aufreinigung der PCR-Produkte erfolgte mithilfe des QIAquick Purification Kits nach entsprechender Anleitung. Diese PCR-Produkte der Proben wurden äquimolar konzentriert (etwa 10 ng/ $\mu\text{l}$ ).

#### 2.2.7 DNA-Sequenzierung

Die aufkonzentrierten DNA-Proben wurden von einem Mitarbeiter der Fraunhofer IZI mittels MiSeq™ der Firma illumina (NGS) sequenziert. Hierbei handelte es sich um ein Verfahren, bei dem die Sequenzierung während der Synthese/Amplifikation brückenförmig an einer Festphase verläuft. Die Detektion geschieht über den Einbau fluoreszenzmarkierter Nukleotide [154].

#### 2.2.8 Bestimmung der optischen Dichte von Kulturlösungen

Die Zelldichte der Bakterienkulturen wird über die optische Dichte (O.D.) bei einer Wellenlänge von 600 nm im Photometer der Firma Eppendorf gegen dYT-Medium ermittelt.

#### 2.2.9 Generierung der Sequenzdaten

Für die Datenprozessierung werden mehrere Programme verwendet. Bei der illumina-Sequenzierung werden von jedem DNA-Molekül zwei Sequenzen gelesen, eine vorwärts und eine rückwärts gelesene Sequenz. Zuerst werden die Sequenzen mit trimmomatic und pear aus dem EMBOSS Softwarepaket bearbeitet. Trimmomatic dient dazu, die Primer-Sequenzen aus den Sequenzen der



Datensätze zu entfernen. Pear vereint die vor- und rückwärts gelesenen DNA-Sequenzen zu einer Einzigem und bereinigt dabei Sequenzierfehler. Anschließend werden die Sequenzen in einer interaktiven JAVA™ basierten Webapplikation, genannt LibDB, welche in unserer Arbeitsgruppe in Kooperation mit der Firma PolyQuant GmbH entwickelte wurde, eingelesen und translatiert. In der Software ist das Codon-Design der ENTE-1 Bibliothek hinterlegt. In den konstanten Regionen am Anfang und am Ende der Sequenz sind nur einzelne und in dem variablen Bereich nur bestimmte Codons erlaubt. Entspricht eine Sequenz nicht dem Design der Peptidbibliothek, wird diese automatisch von LibDB gelöscht. Zusätzlich generiert die Software eine Datenbank, welche alle im Datensatz enthaltenen Motive (3-mer, 4-mer und 5-mer) beinhaltet, die mit den jeweiligen Sequenzen verlinkt sind. Ebenso wird deren Statistik berechnet. Diese Statistik beinhaltet als pLog-Werte die gefundene Häufigkeit, den theoretischen Wert und deren Differenz (theoretischer – gefundener log-Wert), auch Anreicherung genannt. Alle drei Werte werden logarithmisch angegeben. Für jedes Serum werden Datensätze der 1. und 2. Selektionsrunde erfasst.

#### 2.2.10 *In silico* Analyse zur Identifizierung von Epitopen

Die *in silico* Auswertung der erhaltenen Sequenzdaten erfolgt mit Hilfe der „*sequence coverage*“-Funktion des Programms LibDB. Dabei wird die im Peptide Phage Display gebundenen Sequenzen, sowie der naiven Phagenbibliothek mit dem der Antigensequenz (siehe Abbildung 7) verglichen. Die Sequenz wird jeweils in Vierermotive unterteilt, wodurch die Analyse vereinfacht wird. Für jedes Motiv wird die Häufigkeit im Datensatz, die Anreicherung und die relative Häufigkeit berechnet (Gleichung 1 und 2). Die relative Häufigkeit mit der bestimmte Motive auftreten, berechnet sich aus der Differenz der Anzahl der gefundenen Motive, sowie der Anzahl der validen Sequenzen und ist ein statistischer empirischer Wert. Die Anreicherung beschreibt die Differenz der Frequenz der erwarteten Sequenzen sowie der relativen Frequenz und wird als negativer dekadischer Logarithmus angegeben.

$$\text{relative Häufigkeit} = -\log \left( \frac{\text{Anzahl Motive}}{\text{Anzahl valider Sequenzen}} \right) \quad \text{Gl.1}$$

$$\text{Anreicherung} = -\log \left( \frac{\text{Frequenz erwarteter Sequenzen}}{\text{relative Frequenz}} \right) \quad \text{Gl.2}$$

Die Anreicherung der Vierermotive, der im Peptide Phage Display gebundenen Sequenzen, wird mit der Anreicherung der Vierermotive aus der naiven Bibliothek verglichen. Mit diesem Vergleich können die relativen Abweichungen zwischen Anreicherung bestimmter Vierermotive im Serum während der Affinitätsselektion gegenüber der Häufigkeit in der naiven Bibliothek erkannt werden. LibDB erlaubt es, die dazugehörigen Sequenzen der potenziellen Epitop-Motive und benachbarter Motive aus der Datenbank auszulesen und weiter zu analysieren.

## 2.2.11 Alignment mit den Proteinsequenzen

Die Datensätze wurden mit den in Tabelle 9 gelisteten Proteinsequenzen, welche in der Onlinedatenbank für Nukleotidsequenzen GenBank des National Centers for Biotechnology Information [155] hinterlegt sind, im sogenannten „Sequence Coverage“ (LibDB) verglichen.

Tabelle 9: Übersicht der untersuchten Proteinsequenzen

Protein	Allergienname	Genebank ID	Uniprot ID	bp	MW
Hydrophobic Seed Protein	Gly m 1	A29385	P24337	80	8,36
Defensin-like protein	Gly m 2	z13956	Q07502	74,0	8,4
Profilin1	Gly m 3.01	AJ223981	O65810	131	14,1
Profilin2	Gly m 3.02	AJ223982	O65809	131	14,1
Stress-induced protein SAM22	Gly m 4	x60043	P26987	158,0	16,8
Beta-conglycinin, alpha chain	Gly m 5.01	X17698	P13916	605,0	70,2
Beta-conglycinin, alpha' chain	Gly m 5.02	m13759	P11827	639,0	74,3
Beta-conglycinin, beta chain	Gly m 5.03	s44893	P25974	439,0	50,5
Glycinin G1	Gly m 6.01	AB113349	P04776	495,0	55,7
Glycinin G2	Gly m 6.02	d00216	P04405	485,0	54,3
Glycinin G3	Gly m 6.03	x15123	P11828	481,0	54,2
Glycinin G4	Gly m 6.04	X02626	P02858	562,0	63,6
Glycinin G5	Gly m 6.05	m10962	P04347	516,0	58,0
Seed Biotinylated Protein	Gly m 7	U59626	Q39846	643,0	67,8
2S Albumin	Gly m 8	AAB71140.1	P19594	158,0	18,5
Trypsin inhibitor A	Gly m TI	S45092	P01070	216,0	24,0
Lectin	Gly m Agglutinin	K00821	P05046	285,0	30,9
18kD SMP		AAA33984.1	Q01417	173,0	17,6
Albumin 1		AJ011935.1	Q9ZQX0	119,0	13,0
Basic 7S globulin 2		AB084260.1	Q8RVH5	433	47,205
Major pollen allergen Bet v 1-A		X15878	P15494	160,0	17,5

Mithilfe der Computersoftware MegAlign Pro von DNASTAR® 12.1.0.145 konnten die Sequenzen des Peptid Phage Display Experiments im sogenannten Alignment mit den Proteinsequenzen verglichen und Konsensusmotive identifiziert werden. Durch Alignments aufgeführter kreuzreaktiver Pflanzenproteine wurden Homologien bezüglich vermutlicher Epitopsequenzen überprüft. Tabelle 9 listet alle untersuchten Proteinsequenzen auf.

## 2.2.12 Microarray

Die Peptide der Epitopkandidaten wurden von der Arbeitsgruppe von Ines Neundorf, Universität zu Köln, synthetisiert. Die Peptide hatten folgende Struktur: Acetyl (Ac) – N-Epitopsequenz-C-Ebes-Lys( $\epsilon$ -N3)-Amid (Ebes = Amino-3,6-Dioxaoctyl-Succinaminsäure). Alle Peptidsequenzen basierten entweder auf der naiven Sojaproteinsequenz oder auf der Mimotopsequenz aus dem Phage Display Experiment. Die Peptidlänge variierte zwischen acht und neunzehn Aminosäuren. Die Peptide

wurden an Amino-DBCO gekoppelt und in drei verschiedenen Konzentrationen (81 ng/μl, 27 ng/μl, 9 ng/μl) in zwei verschiedenen Cluster auf GLYMO ((3-glycidyoxypropyl)trimethoxysilane)-modifizierten Objektträger immobilisiert. Jeder Objektträger beinhaltet insgesamt 2160 Spots.

#### 2.2.12.1 Messung des Microarray

Der Objektträger wurde eine Stunde in Blockingpuffer bei 4°C inkubiert. Das Serum wurde 1/50 in Blockingpuffer verdünnt und über Nacht bei 4 °C inkubiert. Der Objektträger wurde 2x 5 min mit Waschpuffer (PBS/ 0.1 %Tween) gewaschen und der Anti-human IgG oder Anti-human IgE wurde 1/5000 in Blockingpuffer verdünnt und eine Stunde bei 4 °C inkubiert. Nach einem weiteren Waschschrift von 2x 5 min mit Waschpuffer wurde für die Detektion von humanen IgG der Objektträger im Scanner gescannt. Für die Detektion von humanen IgE wurde ein sekundärer Antikörper 1/5000 in Blockingpuffer verdünnt und eine Stunde bei 4 °C inkubiert. Nach einem weiteren Waschschrift wurde der Objektträger im bei 635 nm gescannt.

### 3. Ergebnisse

#### 3.1 Projektübersicht

Die Arbeit umfasst drei ineinandergreifende, hier getrennt dargestellte Teilabschnitte: Selektion, auch „Biopanning“ genannt, die *in silico* Datenanalyse der gebundenen Phagemidsequenzen und die Validierung der identifizierten Epitopkandidaten im Microarray. Bereits in vorangegangenen Arbeiten wurde die Selektionsmethode für Serumbinder optimiert [156, 157], sodass in dieser Arbeit die Seren nur noch abgearbeitet werden mussten. Es standen 50 atopische Seren, sowie zwei nicht-atopische Seren zur Verfügung. Im Rahmen dieser Arbeit wurde eine Selektion mit der naiven ENTE-1 Bibliothek auf 25 Seren durchgeführt.

Eine Konkurrenz mit unspezifischen Serumbestandteilen wurde während der Selektion durchgeführt, um nicht vor allem Binder an allgemeine Serumbestandteile, wie z.B. Serumalbumin zu erhalten. Dadurch erhielt man als Ergebnis der Selektion eine Subbibliothek aus potenziellen Bindern, die nur an Serumproteinen von sensibilisierten Personen binden.

Nach der Selektion wurden die Peptidgene aller gebundenen Klone im NGS – Verfahren sequenziert. Nach der Sequenzierung wurden die Sequenzen mittels Trimmomatic und Pear aus dem Softwarepaket EMBOSS prozessiert. Diese Programme entfernen die Primersequenzen und bereinigen Sequenzierfehler durch das Zusammenfügen der vor- und rückwärts gelesenen Sequenzen. Anschließend werden die Sequenzen in einer speziellen Software namens LibDB, die fehlerhafte Sequenzen entfernt, eingelesen. Die Software generierte aus den Sequenzen eine Datenbank mit 3mer und 4mer Motiven und berechnete deren Häufigkeit. Die Identifizierung der Epitope erfolgte dann durch den Vergleich der Sojaproteinsequenzen mit den angereicherten Motiven gebundener Phage Sequenzen.

Im letzten Abschnitt wurden die Epitopkandidaten im Microarray validiert. Dazu wurden 85 Epitopkandidaten synthetisiert und auf einen Glaträger aufgebracht. Viele davon wurden mit mehreren Varianten synthetisiert, um zu testen, ob die Mimotope aus dem Peptid Phage Display oder die native Sequenz aus dem Sojaprotein besser Serumantikörper binden können. Die Detektion erfolgt mit einem anti-human IgE - und einem anti-human IgG Antikörper, um zu überprüfen, ob und welche Immunglobuline an das Peptid binden.

Ein weiterer Microarray wurde von unserem Projektpartner (Fraunhofer IME, Aachen) durchgeführt. Der von der Firma PEPperPRINT bereitgestellte Array enthielt überlappende Peptide, welche 10 Sojaproteine präsentieren. Die Epitopkandidaten wurden auch mit den Ergebnissen dieses Microarrays abgeglichen.

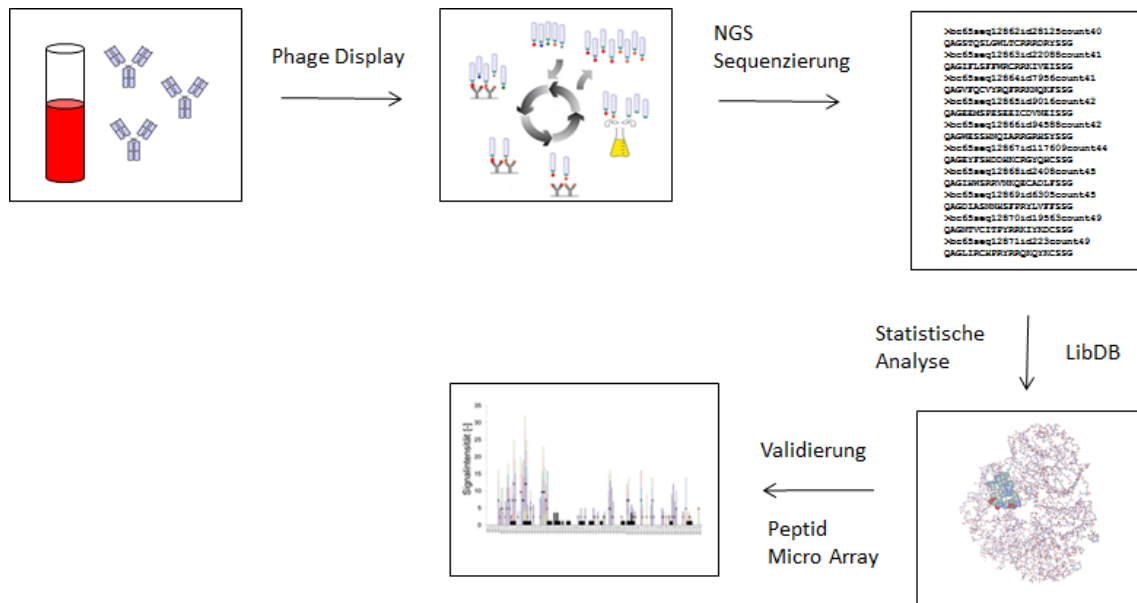


Abbildung 6: Ablauf der Arbeiten

### 3.2 Peptid-Peptid Phage Display

Ziel dieser Experimente war die Anreicherung von Phagemidklonen, deren Peptide an Serumbestandteile von sensibilisierten Personen gebunden haben. Die ENTE-1 Bibliothek wurde zur Selektion auf 25 Seren verwendet. Nach jeder Selektionsrunde wurde die Anzahl der eluierten Phagemid-Partikel durch Titration bestimmt. Um genügend Klone für die Sequenzierung bzw. für die statistische Auswertung zu haben, war eine Ausbeute von  $10^4$ - $10^6$  Klonen erforderlich. Bei einem Ergebnis von weniger als  $10^4$  Klonen nach der ersten Selektionsrunde wurde das Experiment wiederholt. Eine Ausbeute von  $5 \cdot 10^7$  und mehr Klonen deutete auf eine Kontamination oder unzureichendes Waschen hin, in diesem Fall wurde die Selektion ebenfalls wiederholt. Wurde bei der Wiederholung der Experimente das gleiche Ergebnis erzielt, wurde mit der Probe weitergearbeitet. Für die zweite Selektionsrunde wurden mindestens 1000-mal mehr Phagemid-Partikel eingesetzt, als in der ersten Selektionsrunde selektiert wurden.

Die Ergebnisse der Selektion sind in Tabelle 10 dargestellt. Der Output der ersten und zweiten Selektionsrunde lag im Durchschnitt bei  $10^6$  Klonen, sodass für die statistische Auswertung eine genügend große Anzahl von Sequenzen vorhanden war. Nach der ersten Selektionsrunde wurde eine Überschreitung von  $10^7$  Klonen in vier Experimenten (UKL-IZI-2013-003, UKL-IZI-2013-006, UKL-IZI-2014-001 und UKL-IZI-2014-015) und nach der zweiten Selektionsrunde in zwei Experimenten (UKL-IZI-2014-016 und UKL-IZI-2014-018) beobachtet. Eine Unterschreitung von  $10^4$  Klonen wurde nach der zweiten Selektionsrunde in vier Experimenten beobachtet (UKL-IZI-2013-007, UKL-IZI-2013-014, UKL-IZI-2013-015, UKL-IZI-2014-015, EU).

Tabelle 10: Übersicht der Konzentration der selektierten Phagemid-Partikel in den Peptid Phage Display Experimenten

Probe	Input 1. Panning [cfu/ml]	Output 1. Panning [cfu/ml]	Input 2. Panning [cfu/ml]	Output 2. Panning [cfu/ml]
13001	$4,0 \cdot 10^{11}$	$4,8 \cdot 10^7$	$5,8 \cdot 10^{11}$	$6,8 \cdot 10^4$
13002	$4,0 \cdot 10^{11}$	$3,0 \cdot 10^5$	$3,0 \cdot 10^{10}$	$8,0 \cdot 10^4$
13003	$4,0 \cdot 10^{11}$	$2,0 \cdot 10^7$	$2,0 \cdot 10^{11}$	$3,5 \cdot 10^5$
13006	$4,0 \cdot 10^{11}$	$3,5 \cdot 10^7$	$1,0 \cdot 10^{10}$	$7,0 \cdot 10^4$
13007	$4,0 \cdot 10^{11}$	$1,7 \cdot 10^5$	$1,7 \cdot 10^9$	$1,0 \cdot 10^3$
13009	$4,0 \cdot 10^{11}$	$2,1 \cdot 10^6$	$1,0 \cdot 10^{10}$	$5,8 \cdot 10^4$
13014	$4,0 \cdot 10^{11}$	$2,2 \cdot 10^6$	$1,0 \cdot 10^{10}$	$4,2 \cdot 10^3$
13015	$4,0 \cdot 10^{11}$	$3,1 \cdot 10^6$	$1,0 \cdot 10^{10}$	$3,8 \cdot 10^3$
13016	$4,0 \cdot 10^{11}$	$4,5 \cdot 10^5$	$4,0 \cdot 10^9$	$1,1 \cdot 10^5$
13021	$4,0 \cdot 10^{11}$	$1,8 \cdot 10^5$	$4,0 \cdot 10^8$	$1,0 \cdot 10^6$
13022	$4,0 \cdot 10^{11}$	$4,6 \cdot 10^4$	$4,6 \cdot 10^{10}$	$1,1 \cdot 10^5$
13025	$4,0 \cdot 10^{11}$	$3,3 \cdot 10^4$	$3,0 \cdot 10^8$	$1,0 \cdot 10^6$
14001	$4,0 \cdot 10^{11}$	$1,2 \cdot 10^7$	$1,0 \cdot 10^{10}$	$1,4 \cdot 10^6$
14003	$4,0 \cdot 10^{11}$	$2,3 \cdot 10^5$	$4,0 \cdot 10^{10}$	$1,1 \cdot 10^5$
14004	$4,0 \cdot 10^{11}$	$2,6 \cdot 10^6$	$5,0 \cdot 10^9$	$7,0 \cdot 10^4$
14006	$4,0 \cdot 10^{11}$	$2,6 \cdot 10^6$	$2,6 \cdot 10^9$	$2,0 \cdot 10^5$
14008	$4,0 \cdot 10^{11}$	$2,0 \cdot 10^7$	$2,0 \cdot 10^{10}$	$4,3 \cdot 10^5$
14013	$4,0 \cdot 10^{11}$	$5,8 \cdot 10^6$	$5,8 \cdot 10^{10}$	$3,2 \cdot 10^5$
14015	$4,0 \cdot 10^{11}$	$1,3 \cdot 10^7$	$1,3 \cdot 10^{10}$	$4,1 \cdot 10^3$
14016	$4,0 \cdot 10^{11}$	$3,8 \cdot 10^5$	$3,8 \cdot 10^9$	$1,1 \cdot 10^7$
14018	$4,0 \cdot 10^{11}$	$5,1 \cdot 10^5$	$5,1 \cdot 10^9$	$4,6 \cdot 10^7$
14021	$4,0 \cdot 10^{11}$	$4,0 \cdot 10^4$	$4,0 \cdot 10^{10}$	$1,0 \cdot 10^6$
14024	$4,0 \cdot 10^{11}$	$7,6 \cdot 10^5$	$7,6 \cdot 10^{10}$	$3,7 \cdot 10^5$
F1	$4,0 \cdot 10^{11}$	$5,5 \cdot 10^6$	$5,0 \cdot 10^9$	$2,5 \cdot 10^5$
EU	$4,0 \cdot 10^{11}$	$6,0 \cdot 10^5$	$5,0 \cdot 10^{10}$	$1,1 \cdot 10^3$

### 3.3 Valide Sequenzen der Datensätze

Für die Datenprozessierung wurden mehrere Programme verwendet. Aufgrund der illumina-Sequenzierung waren von jeder Sequenz eine vorwärts und eine rückwärts gelesene Sequenz vorhanden. Zuerst wurden die Sequenzen mit trimmomatic und pear aus dem EMBOS Softwarepaket bearbeitet. Trimmomatic dient dazu, die Primersequenzen aus den Sequenzen der Datensätze zu entfernen. Pear vereinigt die vor- und rückwärts gelesenen Sequenzen zu einer einzigen Sequenz und bereinigt unter Berücksichtigung der Qualität der Sequenzierdaten in den FASTQ Dateien Sequenzierfehler. Anschließend wurden die Sequenzen in einer interaktiven JAVA™ basierten Webapplikation, genannt LibDB, welche von einer Firma PolyQuant GmbH in Kooperation mit unserer Arbeitsgruppe entwickelte wurde, eingelesen. In der Software ist das Trinukleotid-Design der ENTE-1 Bibliothek hinterlegt. In der konstanten Region am Anfang und am Ende der Sequenz ist nur eine bestimmte Sequenz erlaubt. Die hypervariable Region besteht aus 16 Aminosäuren. Der variable Teil der Sequenz beginnt mit einem Glycin. An Position fünf waren die

Aminosäuren Cystein oder Serin in der Bibliothek ENTE-1 festgelegt. Ein weiterer Vorteil der Bibliothek ist die reduzierte Anzahl der Aminosäuren Methionin und Tryptophan. Methionin ist nur an der vierten Position, Tryptophan nur an sechs Positionen der Peptidsequenz erlaubt. Ebenso ist das Cystein-Codon nur an bestimmten Positionen eingebaut. Die Eigenschaften dienen u. a. statistischen Vorteilen und erweisen sich für Bakteriophagen als expressionsgünstiger. Wurde einer dieser Kriterien nicht erfüllt, wurde die Sequenz nicht in die Datenbank eingelesen. Das Programm legte die Sequenzen in Zuordnung zu den Datensätzen ab. Außerdem konnte man basierend auf einem Algorithmus für alle zusammenhängenden 3er und 4er Motive, die in den einzelnen Sequenzen vorkommen, Daten berechnen und in der Datenbank hinterlegen. Diese Daten setzten sich aus dem tatsächlichen Vorkommen, dem vorhergesagten Vorkommen und der Anreicherung zusammen. Alle drei Werte werden logarithmisch angegeben.

In Tabelle 11 ist die Anzahl der Sequenzen nach der Bearbeitung der Daten gelistet. Im Gegensatz zur Spalte „Sequenzen“, die die Anzahl alle Sequenzen angibt, zeigt die Spalte „gültige Sequenzen“ nur die Anzahl der Sequenzen, die dem Design der Bibliothek entsprachen. Die Spalte der Motive zeigt die Anzahl der 3er und 4er Motive in dem Datensatz.

Die Größe der Datensätze ist sehr heterogen. Mit 71.383 Sequenzen ist der Datensatz „2013\_S14\_1.Selektion“ der kleinste Datensatz. Der Datensatz „2013\_S3\_1.Selektion“ ist mit 2.055.262 Sequenzen der größte Datensatz. Um eine statistische Auswertung der Sequenzen vornehmen zu können sollte ein Datensatz mindesten 150.000 Sequenzen enthalten, um Motive statistisch mindestens fünf bis zehnmal-10x im Datensatz zu haben. Bestimmte Aminosäuren sind in der ENTE-1 Bibliothek unterrepräsentiert und daher sollte ein Datensatz mindestens 100.000 Motive enthalten, um diesen für die weiteren Analysen zu berücksichtigen. Von 50 Datensätzen erfüllen sieben Datensätze (13007\_2.Selektion, 13014\_1.Selektion, 13014\_2.Selektion, 13015\_2.Selektion, 13016\_2.Selektion, 14015\_2.Selektion, EU\_2.Selektion) nicht die Kriterien. Obwohl die Experimente mit diese Seren wiederholt wurde, konnte keine bessere Ausbeute erreicht werden. Diese Datensätze wurden nicht für die Identifizierung von Epitopkandidaten herangezogen. Diese Datensätze wurden allerdings dahingehend untersucht, ob bereits identifizierte Epitopkandidaten in diesen vorhanden sind.

Tabelle 11: Übersicht der erhaltenen Sequenzen nach der Datenprozessierung

Name	Sequenzen	Gültige Sequenzen	Motive	Gültige Sequenzen in %
13001_1.Selektion	1.230.904	790.090	139.469	64,2
13001_2.Selektion	1.276.371	738.780	127.035	57,9
13002_1.Selektion	419.905	279.366	139.161	66,5
13002_2.Selektion	637.574	411.556	138.596	64,6
13003_1.Selektion	3.784.343	2.055.262	139.471	54,3
13003_2.Selektion	2.112.012	1.167.605	135.158	55,3
13006_1.Selektion	2.092.746	1.078.149	139.470	51,5
13006_2.Selektion	318.469	205.045	138.414	64,4
13007_1.Selektion	1.500.280	815.314	139.457	54,3
13007_2.Selektion*	556.360	264.350	75.119	47,5
13009_1.Selektion	537.547	346.863	126.753	64,5
13009_2.Selektion	848.200	470.983	131.379	55,5
13014_1.Selektion*	160.395	115.704	135.861	72,1
13014_2.Selektion*	146.479	71.383	52.126	48,7
13015_1.Selektion	346.343	243.526	138.677	70,3
13015_2.Selektion*	329.146	128.273	61.597	39,0
13016_1.Selektion	679.029	452.253	139.423	66,6
13016_2.Selektion*	535.987	113.289	137.622	21,1
13021_1.Selektion	1.820.394	1.038.809	139.254	57,1
13021_2.Selektion	673.251	355.624	130.125	52,8
13022_1.Selektion	354.593	250.000	137.028	70,5
13022_2.Selektion	610.346	426.253	136.286	69,8
13025_1.Selektion	455.749	255.637	135.635	56,1
13025_2.Selektion	782.146	382.363	130.389	48,9
14001_2.Selektion	1.170.029	642.053	139.414	54,9
14003_1.Selektion	550.733	384.581	139.316	69,8
14003_2.Selektion	434.973	311.223	139.183	71,5
14004_1.Selektion	484.419	247.917	139.306	51,2
14004_2.Selektion	433.903	233.331	131.461	53,8
14006_1.Selektion	998.711	522.494	139.438	52,3
14006_2.Selektion	936.424	443.371	139.192	47,3
14008_1.Selektion	1.081.916	514.584	139.452	47,6
14008_2.Selektion	1.065.700	478.617	139.262	44,9
14013_1.Selektion	1.312.739	717.278	139.466	54,6
14013_2.Selektion	815.048	509.873	106.526	62,6
14015_1.Selektion	1.149.874	626.171	139.459	54,5
14015_2.Selektion*	116.706	71.634	65.366	61,4
14016_1.Selektion	614.826	403.030	139.402	65,6
14016_2.Selektion	819.148	544.317	139.438	66,4
14018_1.Selektion	1.024.527	660.057	139.446	64,4



Name	Sequenzen	Gültige Sequenzen	Motive	Gültige Sequenzen in %
14018_2.Selektion	736.622	445.325	139.286	60,5
14021_1.Selektion	481.890	315.531	139.153	65,5
14021_2.Selektion	368.251	240.691	138.638	65,4
14024_1.Selektion	350.132	168.677	138.873	48,2
14024_2.Selektion	534.828	369.992	138.022	69,2
EU_2.Selektion*	255.345	141.695	51.804	55,5
F1_1.Selektion	773.410	424.276	139.415	54,9
F1_2.Selektion	526.043	267.620	109.049	50,9

\* diese Datensätze wurden nicht zur Analyse von Epitopkandidaten herangezogen

### 3.4 *In silico* Analyse zur Identifizierung von Epitopen

Ausgehend von der theoretisch erwarteten Häufigkeit, die sich aus der Codonzusammensetzung der Bibliothek ergibt, sowie der tatsächlich beobachteten Häufigkeit, die für jedes Motiv aus seiner gefundenen Anzahl in allen Sequenzen und der Größe des Datensatzes errechnet wurde, kann die Anreicherung für jedes Motiv in der Datenbank berechnet werden. Diese Anreicherung wird gegen die Sequenz eines potenziellen Antigens zusammen mit den Daten der naiven Bibliothek dargestellt (Abbildung 7). Zur Auswertung wurden vor allem die Vierermotive in Betracht gezogen, welche sowohl in der ersten als auch in der zweiten Selektionsrunde gegenüber der naiven Bibliothek angereichert wurden.

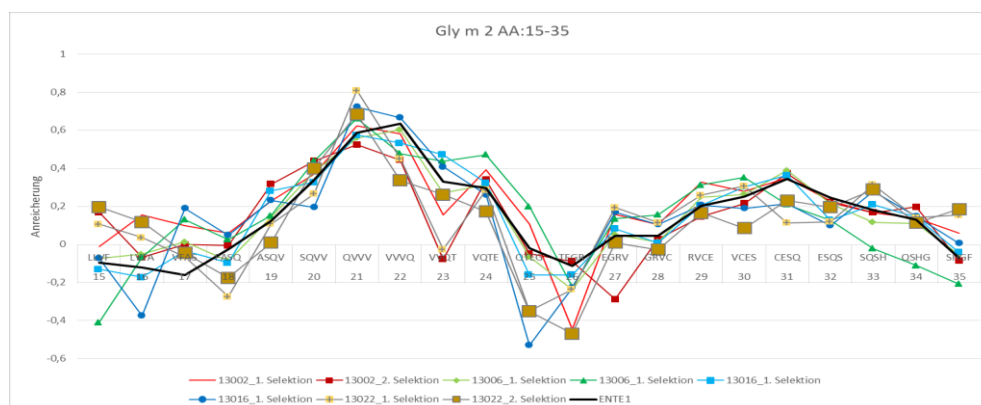


Abbildung 7: Epitopidentifizierung mittels Vergleich der Anreicherungen. Der erste Schritt ist der Vergleich der Anreicherung der 4mer Motive aus den selektierten Datensätzen mit der Anreicherung der 4mer Motive aus der naiven Bibliothek. Ist ein 4mer Motiv aus den selektierten Datensätzen stärker angereichert als ein 4mer Motiv aus der naiven Bibliothek, wurde dieses Motiv weiter untersucht.

Ein Motiv ist Teil eines potenziellen Epitops, wenn zusätzlich noch weitere Kriterien erfüllt werden. Zum einen wurde überprüft, ob es eine Steigerung der Häufigkeit in der zweiten Selektionsrunde gegenüber der ersten Selektionsrunde gab. Anschließend wurde kontrolliert, ob die Vierermotive entsprechend der Proteinsequenz oder durch Aminosäuren mit ähnlichen chemischen, physikalischen oder strukturellen Eigenschaften erweiterbar waren (Abbildung 8). So konnten 5mer

oder 6mer Motive erkannt werden. Des Weiteren sollte ein potenzielles Epitop in Datensätzen von zwei unterschiedlichen Personen zu finden sein.

In Tabelle 12 ist die Anzahl der potenziellen Epitope für jedes Sojaprotein gelistet. Die Tabelle 16 (Anhang) zeigt die Sequenz der individuellen Epitope. Entsprechend Tabelle 16 (Anhang) galt ein Epitope als identifiziert, wenn sich die angereicherten Motive auf fünf oder mehr Aminosäuren erweitern ließen.



Abbildung 8: Erweiterung des Vierermotivs. Das Vierermotiv **ASQV** kann im Vergleich zur Proteinsequenz links mit **VF** und rechts mit **VV** erweitert werden. Alle 3308 Sequenzen, die das Motiv **ASQV** enthalten, ergeben folgendes Weblogo. Die Buchstabengrößen geben die Frequenz der jeweiligen Aminosäure wieder. Man sieht sehr gut die N- und C-terminal häufigen Aminosäuren: Valin bzw Phenylalanin.

Tabelle 12: Anzahl der identifizierten Epitope

Protein	Epitop
Gly m 1	9
Gly m 2	9
Gly m 3.01	6
Gly m 3.02	2
Gly m 4	15
Gly m 5.01	59
Gly m 5.02	38
Gly m 5.03	19
Gly m 6.01	30
Gly m 6.02	31
Gly m 6.03	21
Gly m 6.04	14
Gly m 6.05	35
Gly m 7	12
Gly m 8	14
Gly m TI	3
Gly m Agglutinin	6
Gly m BD30K	23
18kD SMP	11
Albumin	10
Basic 7S globulin	38
Summe	405

Es sollten die sojaspezifischen Epitope identifiziert werden, die von möglichst vielen sojasensibilisierten Personen erkannt werden. Wurden sie/es nur in einem Serum identifiziert, galt es als „individuelles“ Epitop und wurden nicht in Tabelle 16 aufgelistet. Wurden Epitopsequenzen bereits in der IEDB als publizierte antigene Determinanten gefunden, wurde ebenfalls die ID des Epitops gelistet. Insgesamt wurden 405 Epitope aus 14 Sojaproteinen gefunden. Die meisten Antikörperbindungsstellen wurden in Gly m 6 gefolgt von Gly m 5 identifiziert. Die wenigsten antigenen Determinanten wurden in Gly m TI erfasst. Sojabohnenproteine haben im Durchschnitt pro 100 Aminosäuren 9 Epitope. Die Länge dieser variiert von Epitop zu Epitop, das Kürzeste ist fünf AS und das Längste, zumindest aus den zusammengetragenen Sequenzen verschiedener Patienten, 16 AS lang. Diese unterschiedliche Epitoplänge resultiert vermutlich aus der Überlappung der unterschiedlich erkannten Epitopen der verschiedenen, individuellen Serumantikörper. Im Durchschnitt war ein Epitop der Antikörper sieben AS lang.

## 3.5.4 Aminosäureverteilung innerhalb der Epitope der Sojabohne

Die Aminosäurezusammensetzung der Epitope wurde analysiert. Tabelle 13 zeigt das Ergebnis dieser Untersuchungen. Betrachtet man jede Aminosäure für sich, ist die Aminosäure Glutaminsäure gefolgt von Glutamin am häufigsten in den identifizierten Motiven zu finden. Glutamin ist neben Cystein einer der häufigsten Aminosäuren in der Bibliothek. Am seltensten sind die Aminosäuren Methionin und Tryptophan in den Epitopen vorhanden. Aufgrund des Designs der Bibliothek sind aber zahlreiche Methioninmotive in der Bibliothek nicht vorhanden. Zu berücksichtigen ist auch, dass Methionin, Tryptophan und Cysteine sehr selten in Proteinen vorkommen. Sortiert man jedoch die Aminosäuren anhand ihrer Eigenschaften zusammen bestehen die Epitope am häufigsten aus hydrophoben/unpolaren gefolgt von polar/neutralen und basischen sowie sauren Aminosäuren. Dies entspricht auch der Aminosäureverteilung innerhalb der Sojaproteine und der eingesetzten ENTE-1 Bibliothek.

Tabelle 13: Die relative Häufigkeit der Aminosäuren in den identifizierten Epitopen, in den Proteinsequenzen und der Bibliothek.

	Aminosäure	Epitop	Protein	Bibliothek	Epitop	Protein	Bibliothek
hydrophob/ unpolar	A	0,042	0,066	0,041	0,396	0,464	0,437
	G	0,078	0,073	0,038			
	I	0,032	0,047	0,053			
	L	0,053	0,080	0,035			
	M	0,001	0,014	0,005			
	F	0,036	0,042	0,050			
	P	0,071	0,053	0,052			
	W	0,005	0,007	0,031			
	Y	0,028	0,024	0,068			
	V	0,049	0,057	0,064			
basisch	H	0,036	0,023	0,069	0,125	0,134	0,141
	K	0,032	0,056	0,032			
	R	0,057	0,055	0,040			
polar/neutral	S	0,082	0,075	0,063	0,303	0,270	0,316
	T	0,027	0,047	0,041			
	N	0,041	0,059	0,051			
	C	0,032	0,017	0,081			
	Q	0,122	0,072	0,080			
sauer	D	0,042	0,046	0,043	0,175	0,132	0,106
	E	0,134	0,087	0,062			

### 3.4 Untersuchung auf Kreuzreaktionen

Die Proteine Gly m 4, Bet v 1 und Cor a 1 gehören zur gleichen Proteinfamilie. Es ist bekannt, dass Serumantikörper, die gegen ein Protein dieser Proteinfamilie gerichtet sind, ein oder mehrere andere Proteine dieser Proteinfamilie erkennen und kreuzreagieren können. Mithilfe der Peptide Phage Display Methode ist es theoretisch möglich alle im Serum befindlichen Antikörper zu erfassen. Daher wurde überprüft, ob sich auch Antikörper gegen das Hauptallergen der Birke Bet v 1 im Serum von sojasensitiven Personen befindet oder ob die Antikörper, die gegen das Sojaprotein Gly m 4 gerichtet sind, kreuzreagieren könnten. Es war bekannt, dass einige Personen auch gegen Haselnuss allergisch sind. Daher sollte es auch möglich sein, Antikörper zu finden, die gegen ein Haselnussprotein gerichtet sind. In Tabelle 14 sind die identifizierten Epitope des Proteins Bet v 1 und Cor a 1 gelistet. Insgesamt konnten 12 Epitope des Protein Bet v 1 und 14 Epitope des Proteins Cor a 1 aus dem Serum von sojasensitiven Personen identifiziert werden. Die Länge der identifizierten Bereiche reichte von fünf bis 16 Aminosäuren und ein Epitop ist im Durchschnitt etwa acht Aminosäuren lang.

Tabelle 14: Identifizierte potentiellen Epitope des Proteins Bet v 1

Nr.	GenBank ID	Protein	Epitop	Position	Homologie zu anderen Proteinen	Antigen Index	IEDB ID publizierter Epitope
1	X15877	Bet v 1	GVFNYETET	2-10		-0,6	22968, 41641, 22971, 41640, 174533, 68540, 17250
2			TTSVIPA	10-16		1,3	14364,014571, 17250, 22971, 41641, 46603, 63568, 66838, 68540,
3			LFPKVAPQAIS	30-40		1,5	1320, 8365, 9504, 16239, 26996, 35940, 44710, 141047, 174373,174394
4			ISSVENIEG	39-46		-0,6	2091, 3711, 17323, 35940, 48158, 48965, 141047, 141071, 174390,
5			ISFPEGFPFKY	57-65		0,9	21651, 26821, 31418, 47761, 64346, 174507, 174516, 181123, 191485,
6			YNYSVIEGGP	82-91		0,4	8578, 16486, 24948, 34550, 65315, 67978, 75243, 130830, 140995,
7			SNEIK	100-104		0,8	10392, 19127, 21660, 28621, 31427, 35551, 47901, 59815, 62156,
8			SNKYH	118-122		1	8400, 20019, 22381, 27132, 28631, 31428. 58605, 59859, 65531, 75243
9			DHEVKA	126-131		1,5	8400, 8538, 19000, 28631, 34453, 59859, 64566, 74142, 174567,
10			EQVKASKE	132-139		0,9	8400, 8538, 14740, 69174, 74142
11			LRAVESYLL	145-153		0,3	14465, 19448, 29783, 37734, 39038, 41568, 58813, 64938, 136739,
12			AHSDAY	154-159		-0,6	5308, 14319, 29783, 53253, 130739, 130848, 131027, 141099, 174589,
13	X70999	Cor a 1	GVFNYEVETPSVIPAA	2-17	VIPAA / Gly m 5.03	-0,6	
14			RLFKEY	18-23		0,3	4175, 27968, 66488
15			YVLDG	23-27		0,3	76332
16			IPKVAPQAI	31-39		0,8	69411, 76332
17			TSVENVEG	40-47	SVENVEG / Gly m 4	-0,3	
18			GEGSRYKY	60-67		0,73	
19			VDEVND	72-77		0,9	
20			FTYSYTVIEGDVL	80-92	EGDVL / Gly m 6.04, Gly m 6.05	0,83	
21			LEKVCHLK	96-104		0,9	
22			KIVAAP	104-109		-0,3	
23			ISSKFHA	117-123		0,45	
24			KGDHEINAE	124-133		1	
25			LRAVETYL	145-152		0,3	
26			HSAEYN	155-160		-0,3	

### 3.5 Validierung der Epitopkandidaten im Microarray

Der Microarray diente als Validierungsmethode für die im Peptid Phage Display identifizierten Peptide. Dafür wurden 49 Serumproben verwendet. Wurde eine Bindung an die Peptide von Antikörpern aus den Seren zumindest einiger sojasensibilisierter Personen nachgewiesen, kann man das Epitop des Peptids als bestätigt ansehen. Für den Array wurden die Peptide mittels „Click Chemie“ auf ein speziell beschichteten Glasträger aufgebracht. Der Nachweise einer Antikörper-Bindung erfolgte mittels eines fluoreszenzmarkierten Anti-human IgG bzw. Anti-human-IgE Antikörper. Am Ende des Experimentes wurden die Signale mittels eines Array-Scanners ausgelesen. Von mehr als 400 Epitopkandidaten wurden 83 ausgewählt und im Microarray getestet. Diese 83 Peptide repräsentierten 42 der am häufigsten gefundenen potenziellen Epitope aus elf Sojaproteinen. Teilweise wurden auch Sequenzen ausgewählt, bei denen das potenzielle Epitop nur drei oder vier Aminosäuren lang ist. Wurden diese Sequenzen im Array nicht bestätigt, wurden sie nicht in die Epitopliste mit aufgenommen. Die Sequenz der ausgewählten Peptide basierte entweder auf der nativen Sojaproteinsequenz (Pr) und/oder auf der Phagemidsequenz (Ph). Von 85 Peptiden basierten 36 auf der Sojasequenz und 49 auf der Peptid Phage Display Sequenz. Die Peptide wurden in vier verschiedenen Konzentrationen und jeweils als Triplikate auf den Glasträger aufgebracht. Ein Glasträger beinhaltete zwei Cluster, sodass mit einer entsprechenden Maske zwei Seren auf ein Glasträger getestet werden könnte. Insgesamt waren 2160 Peptidspots auf dem Glasträger vorhanden. Ein getestetes Peptid galt als positiv, wenn das Signal fünffach über dem Hintergrund lag, der mit PBS anstatt mit Serum bestimmt wurde. Der Wert wurde so hoch angesetzt, weil die Objektträger einen hohen Hintergrund lieferten. Bessere Ergebnisse würde man bekommen, wenn die Arrays optimiert werden würden. Würde man diesen Wert auf dreimal über Hintergrund herabsetzten, dann würden mehr Peptide als positiv getestet werden. Dabei steigt aber auch die Gefahr, dass unspezifische Bindung detektiert wird.

Tabelle 15 zeigt die Zusammenfassung der Ergebnisse der Microarray-Experimente. Insgesamt konnten 56 von 83 Peptiden bestätigt werden. 55 Peptide wurden von IgE erkannt und 43 Peptide wurden von IgG erkannt. 56 der 83 Peptide repräsentieren 34 von 42 Epitopkandidaten. 22 von diesen 34 Peptiden wurden von mindestens 5 verschiedenen Individuen erkannt. Beispielsweise wurde das 6mer Motiv FSREEG aus Gly m 5.01 von IgEs aus 21 Seren erkannt, aber nur von IgGs aus neun Seren. Vier Seren enthielten sowohl IgE als auch IgG Antikörper gegen dieses Peptid (EM14). Das Peptid (EN2), welches das 4mer Motiv KYQQ darstellte, wurde von IgE und IgG aus 12 Seren erkannt. Davon enthielten 7 Seren sowohl IgE als auch IgG Antikörper gegen dieses Peptid (EN2). Peptide aus Gly m 5.01 (EM14), Gly m 5.02 (EN17), Gly m 6.03 (EN5) und Gly m BD 30K (DO17) wurden von den Soja sensibilisierten Personen am häufigsten erkannt. Einige Beispiele zeigen die korrekte Auswahl der Epitopsequenzen. Beispielsweise enthalten die Peptidpaare EN3 (Pr)/DO14

(Ph), EN17 (Pr)/EN18 (Ph), und DN11 (Pr)/EN8 (Ph) Aminosäurekombinationen, die für die Antikörperbindung wichtig sind. In einigen Beispielen sind die Mimotopsequenzen, die aus dem Phage Display abgeleitet worden sind, besser für die Antikörperbindung geeignet, zum Beispiel die korrespondierenden Peptidpaare EN5 (Ph)/DN6 (Pr), EM20 (Ph)/EM19 (Pr) und DO19 (Ph)/EN4 (Pr). Im Gegensatz dazu ist für das 5mer Motiv LCNRD aus Gly m 2 die Proteinsequenz für die Antikörperbindung besser geeignet. Dies zeigt das korrespondierende Proteinpaar DO2(Pr)/EL14(Ph).



Tabelle 15: Zusammenfassung der Microarray Experimente. Die Peptide wurden als positiv gewertet, wenn das Signal 5x über dem Hintergrund lag; b.t. bedeutet, dass die Signalstärke mit allen Seren unter dem Schwellenwert lag. Peptide basierten entweder auf der nativen Sojaproteinsequenz (Pr) und/oder auf der Peptid Phage Display Sequenz (Ph). Unterstrichene Aminosäuren in der Peptid-Sequenz stellen das Mimotop dar. Unterstrichene Peptide sind klinisch relevante Epitope. Die Zahlen in den Spalten „Nur IgE-positiv“, „IgE+IgG positiv“, „IgE+IgG positiv“ und „Total“ bezeichnet die Anzahl der positiv getesteten Personen.

Protein	Peptid-Ursprung	Peptid - code	Peptidsequenz	Position	nur IgE-positiv	nur IgG-positiv	IgE+IgG positiv	Total	IEDB Code
Beta amylase	Pr	EN9	-----GIGDFQSYD	202-221	2	b.t.	b.t.	2	
	Ph	DL2	GVFSPPTNFHPY <u>GDFQI</u>	205-208	1	b.t.	b.t.	1	
Gly m 2	Pr	EN14	LLVFASQVVVQTEG	15-28	1	1	b.t.	2	
	Ph	EN16	EVSFASQVLIY	18-21	4	2	1	7	
	Pr	DO1	---FASQVVVQTEGRVCES	18-33	2	b.t.	1	3	
Gly m 2	Ph	EN15	SANQVVVQTHV	21-25	1	2	2	5	
	Pr	EN11	-----VQTEGRV <u>CE</u> SQSH	24-36	b.t.	b.t.	b.t.	0	
	Pr	DO23	-----QTEGRV <u>CE</u> SQSHGFH	25-39	3	b.t.	b.t.	3	
	Ph	EN13	-----TEHTEGRV <u>LV</u> N	26-30	b.t.	b.t.	b.t.	0	
	Ph	EN12	-----QLQRV <u>CE</u> SLHY	29-33	2	b.t.	b.t.	2	
Gly m 2	Pr	DO2	--SQSHGFHGL <u>LC</u> NRDHNCA	33-49	4	1	b.t.	5	
	Ph	EN10	VCHQSHGF <u>GH</u> H	34-38	2	2	b.t.	4	
Gly m 2	Ph	EL14	GFNLCNRDRPAP	42-45	3	3	1	7	
Gly m 5.01	Pr	EM19	KCLQSCNSERDSYRN	37-51	2	1	b.t.	3	
	Ph	EM20	AINQSCNSHPI	40-44	7	3	b.t.	10	
	Pr	DN7	---QSCNSERDSYRNQ	40-52	2	b.t.	b.t.	2	
Gly m 5.01	Ph	EL11	GNANSERDSIIQ	43-48	b.t.	b.t.	b.t.	0	
Gly m 5.01	Pr	EM18	-PRPRQHPERE	76-86	18	1	3	22	181299
	Ph	EL10	GSPPRQHPIDV	78-83	b.t.	b.t.	b.t.	0	181299, 181447
Gly m 5.01	Pr	DN1	QHPERE <u>PQ</u> PGKEKE	81-95	b.t.	b.t.	b.t.	0	181299
Gly m 5.01	Pr	EN4	-PRPQ <u>PRQ</u> EEE	107-116	b.t.	b.t.	b.t.	0	181447
	Ph	DO19	SHHDQ <u>PRQ</u> HA	110-113	9	4	1	14	181447
Gly m 5.01	Pr	DN2	PRQEEHEQREEQE <u>W</u>	111-125	b.t.	b.t.	b.t.	0	181447

Protein	Peptid- Ursprung	Peptid - code	Peptidsequenz	Position	nur IgE- positiv	nur IgG- positiv	IgE+IgG positiv	Total	IEDB Code
Gly m 5.01	Pr	EM16	KRGEKGS <del>E</del> E	131-140	b.t.	1	b.t.	1	181334
	Ph	EM17	VYAEKGS <del>E</del> AY	134-137	b.t.	b.t.	1	1	181334
Gly m 5.01	Ph	DO21	FFLSSTEAQ	324-328	6	b.t.	2	8	181316
Gly m 5.01	Pr	<u>EM14</u>	KVLF <del>S</del> REEGQQQ	355-366	16	5	5	26	
	Ph	<u>EM15</u>	HITFSREEGSI	358-363	10	2	b.t.	12	
Gly m 5.01	Pr	EM12	PFNLRSRDIY	405-415	2	1	b.t.	3	181455
	Ph	EM13	ATYLRSDQAV	408-412	3	1	b.t.	4	181455
Gly m 5.01	Pr	DN3	DANIELVGLKEQQQE	469-483	1	b.t.	b.t.	1	
	Ph	DN5	--NIELVGLKEQQQRQQQ	471-482	b.t.	b.t.	3	3	
Gly m 5.01	Pr	EL15	GLKEQQQEQQQ	476-486	b.t.	b.t.	b.t.	0	
	Ph	EM11	QEGKEQQQPNV	478-482	2	1	2	5	
Gly m 5.02	Pr	EN19	--YRNQACHARSN	49-59	1	1	b.t.	2	
	Ph	EN21	CDTYNQACHLAL	51-55	b.t.	b.t.	b.t.	0	
Gly m 5.02	Ph	EN20	CFNVIQACHANPC	52-56	b.t.	b.t.	b.t.	0	
	Pr	<u>EN17</u>	VEEEEECEEGQI	63-74	18	3	14	35	
Gly m 5.02	Ph	EN18	HIQEEEECEGDL	66-70	1	2	b.t.	3	
	Pr	DN4	--DEGEQPRPFPRPR	95-109	2	2	b.t.	4	
Gly m 5.02	Ph	EO13	TQHEGEQPFY	96-99	1	b.t.	b.t.	1	
	Pr	DN8	---NKVLFSGREEGQQQG	354-367	1	2	b.t.	3	
Gly m 5.02	Ph	EO14	SDQNKVLQPH	389-392	b.t.	b.t.	b.t.	0	
	Ph	EN5	ISQQDRHRI	127-130	19	5	9	33	538745
Gly m 6.01	Pr	<u>EN2</u>	EFLKYQQEQG	191-200	3	b.t.	1	4	13815, 37068
	Ph	DO13	EFLKYQQQQQ	192-197	b.t.	1	b.t.	1	13815, 37068
Gly m 6.01	Ph	<u>DO11</u>	--SDKYQEEFQPR	194-196	7	b.t.	2	9	13815, 37068

Protein	Peptid- Ursprung	Peptid - code	Peptidsequenz	Position	nur IgE- positiv	nur IgG- positiv	IgE+IgG positiv	Total	IEDB Code
Gly m 6.01	Ph	DO12	--TEKYQQNSSG	194-197	1	3	b.t.	4	13815, 37068
Gly m 6.02	Ph	DL3	MIFPGCPS	101-104	b.t.	b.t.	b.t.	0	16009, 52031, 74358
Gly m 6.02	Pr	EM21	----EDDDDEEEQPQS	273-284	b.t.	b.t.	b.t.	0	
	Ph	DO3	GVYNSQVDDEEEQNQRD	276-281	3	b.t.	1	4	
Gly m 6.03	Pr	DN6	EPQQKGQSSRPQDRH	113-127	2	b.t.	1	3	538745
	Ph	DO22	-----DRHQKIYHF	128-131	2	b.t.	1	3	
Gly m 6.03	Pr	DN9	RRFYLAGNQEQEFL	177-190	3	1	1	5	13815, 53810
	Ph	EN6	-----HVVEQEFLD	186-190	17	3	8	28	13815, 53810
Gly m 6.05	Pr	DN10	-NPDIEHPETMQQ	192-203	b.t.	b.t.	b.t.	0	
	Ph	EN7	NQDDIEHPEPI	194-199	1	b.t.	b.t.	1	
Gly m 8	Pr	DO20	KCQCKALQKI	108-117	3	b.t.	b.t.	3	561750
Gly m agglut inin	Pr	DN11	HAAANVVGNKESQREAR	101-117	6	b.t.	b.t.	6	
	Ph	EN8	-----FDQKESQHFS	110-113	1	1	b.t.	2	
Gly m BD 28K	Ph	DO24	----QDQEEDEEDE	237-244	3	1	1	5	
	Ph	DO4	GYNPCRQEEDEELHHKC	239-244	4	4	3	11	
	Ph	DL7	GYVPCRQEEDEELHHKC	239-244	b.t.	b.t.	b.t.	0	
Gly m BD 30K	Pr	EN3	VYHNHEEEAK	54-63	11	1	3	15	74125
	Pr	DO15	YHNHEEEAKR	55-64	18	3	3	24	74125
	Ph	DO14	GENHEEQHVI	57-60	4	2	1	7	74125

Protein	Peptid- Ursprung	Peptid - code	Peptidsequenz	Position	nur IgE- positiv	nur IgG- positiv	IgE+IgG positiv	Total	IEDB Code
Gly m BD 30K	Ph	<u>DO7</u>	GEFS <u>KKYL</u> QAP	105-108	13	1	1	15	8803, 48988
	Ph	<u>DO6</u>	GPVQ <u>KKYL</u> FDS	105-108	9	5	b.t.	14	8803, 48988
Gly m BD 30K	Ph	<u>DO17</u>	-GEYSCDITYT	130-133	20	4	12	36	
	Ph	<u>DO18</u>	KEQYSCD <u>HPP</u>	132-135	1	b.t.	b.t.	1	
Gly m BD 30K	Ph	<u>DO16</u>	TCDHPVK <u>QQP</u>	132-135	4	3	b.t.	7	
Gly m BD 30K	Pr	<u>EL13</u>	GYGSADG <u>V</u> DYW	307-317	3	2	1	6	
	Ph	<u>DL6</u>	GVSSADG <u>V</u> GQINIDQN	310-314	b.t.	b.t.	b.t.	0	
Gly m KTI	Pr	<u>EN1</u>	--IRFIAEGH <u>PLSL</u>	88-89	1	2	1	4	
	Ph	<u>DO9</u>	GSVQVIAEGH <u>PHS</u>	91-96	5	3	b.t.	8	
	Ph	<u>DO8</u>	GCVQVIAEGH <u>PHC</u>	91-96	3	1	b.t.	4	
Gly m KTI	Pr	<u>DO10</u>	-----AEGH <u>PLSLK</u> F	92-101	2	b.t.	b.t.	2	
	Pr	<u>EO12</u>	-----Q <u>LDKESLAK</u>	198-207	b.t.	b.t.	b.t.	0	
	Ph	<u>DL9</u>	GATSSDR <u>QLDKES</u> PGYVI	200-204	b.t.	b.t.	b.t.	0	
Gly m KTI	Pr	<u>EM22</u>	-----S <u>LAKKNH</u> GLSR	204-214	2	1	b.t.	3	
	Ph	<u>DO5</u>	GTHFSKAVLG <u>KKNH</u> GDEF	207-211	7	2	1	10	
PR5	Pr	<u>EL12</u>	-GRTNST <u>FDNAG</u>	69-79	b.t.	1	b.t.	1	
	Ph	<u>DK3</u>	GVINCET <u>FDQYQ</u> PIEIGC	72-76	5	1	1	7	

### 3.6 Personenspezifische Bindungsmuster

Die meisten Epitope sind in den Proteinen Gly m BD 30K, Gly m 5, Gly m 6, Gly m BD28K und Gly m TI. Gly m BD 30K, Gly m 5 und Gly m 6 Epitope werden sowohl von IgE- als auch von IgG-Antikörpern etwas zu gleichen Teilen erkannt (Tabelle 15). Die bedeutendste Erkenntnis dieser Arbeit ist der Nachweis, dass jeder Mensch ein eigenes individuelles IgE und IgG Antikörpererkennungsmuster hat. Die Seren können in drei Gruppen einsortiert werden:

- IgE und IgG Antikörper
- nur IgG Antikörper
- nur IgE Antikörper

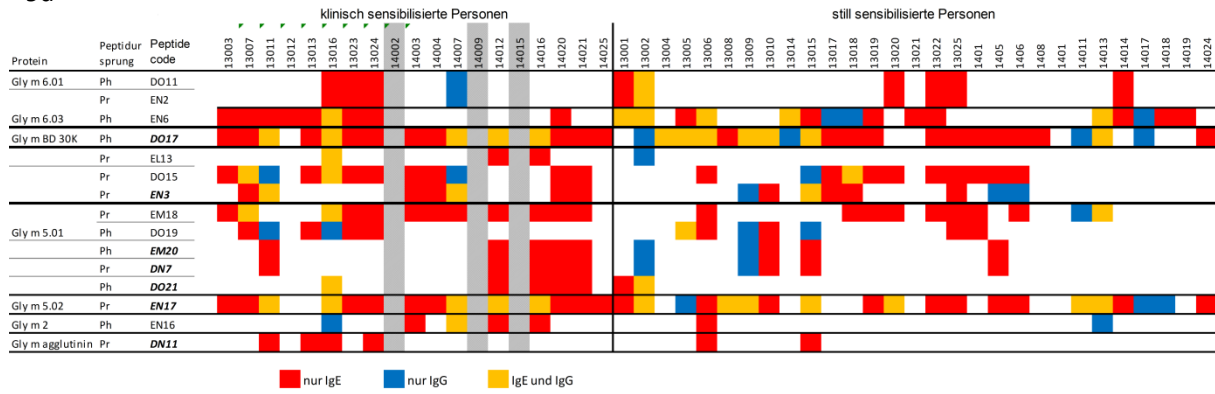
19 Personen berichteten über klinische Symptome nach dem Verzehr von Sojaprodukten. Die Bandbreite reichte von mild bis schwer, zu sehen in Tabelle 3. Diese Personen zeigten im Array kaum oder nur wenige IgG-bindende Peptide.

Drei Personen (14002; 14009 und 14015) wurden aus dieser Analyse herausgenommen. Diese Personen zeigten im Gegensatz zu allen anderen keine nachweisbaren Antikörper, obwohl sie über klinische Symptome nach dem Sojaverzehr berichteten. Es ist möglich, dass sie aufgrund einer Behandlung immunsuprimiert waren oder aber andere Effekte zu einer ungewöhnlich niedrigen Konzentration an Immunglobulinen im Serum unterhalb der Nachweisgrenze führten. Es wäre auch möglich, dass die klinischen Symptome ganz andere Ursachen haben, nicht vom Soja Verzehr verursacht und sie nur irrtümlich mit dem Sojaverzehr in Verbindung gebracht werden.

Abbildung 9 zeigt, dass einige Peptide von Seren mit klinischen Symptomen genauso wie Seren mit stiller Sensibilisierung im gleichen Ausmaß erkannt werden, z.B. EN1 und DN5. Andere Peptide werden hauptsächlich von Personen mit klinischen Symptomen erkannt, z.B. D017 und EM18.

Die IgE-Bindungsdaten nur für sich genommen zeigen, dass es mit nur drei Peptiden (EN17, D017, D021) möglich ist mehr als 80 % der Personen mit klinischen Symptomen zu identifizieren (Abbildung 9a). Für diese Anzahl der Peptide kann ein Unterschied zwischen den zwei Personengruppen (Personen mit klinischen Symptomen und Personen mit stiller Sensibilisierung) im IgE/IgG Verhältnis gesehen werden. Ohne weitere Optimierung der Peptide würde das IgE-Bindungsmuster von vier Peptiden (DN11, EM20, EM18, EN3) bereits ausreichen, um 90 % der Personen mit klinischen Symptomen gegenüber nur 50 % der sensibilisierten Personen abzugrenzen. Von denen zeigen wiederum einige ein Epitopmuster, das sich kaum von der ersten Gruppe unterscheidet. Retrospektivisch wäre hier eine genauere Untersuchung dieser Personen sicher noch sehr aufschlussreich.

9a



9b

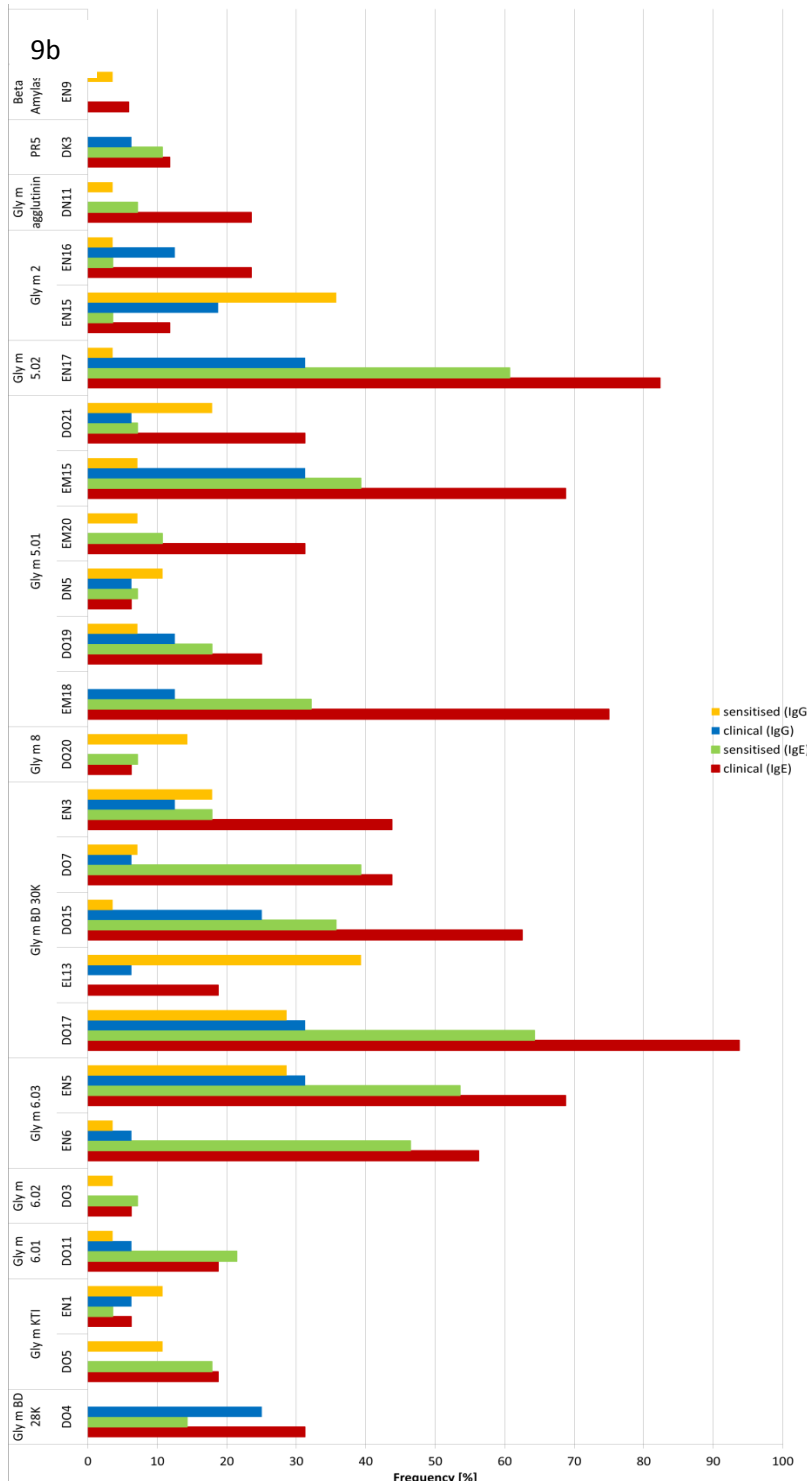


Abbildung 9a und b: Zusammenfassung des Antikörperbindungsmusters von Personen mit klinischen Symptomen versus Personen mit stiller Sensibilisierung im Mikroarray basierend auf den Sojaantikörper-Mimotopen. Serum IgEs von Personen mit klinisch relevanter Sensibilisierung binden häufiger an IgE positive Peptide als Immunglobuline von anderen Personen. Grau unterlegt sind die drei Personen, die aus der Analyse herausgenommen wurden (Abbildung 9a). Die Kombination von drei Peptiden (D017, D021 und EN17) erlaubt die Identifizierung von 90% der Personen mit klinisch relevanter Sensibilisierung vs. nur 50% Personen mit stiller Sensibilisierung. Die Kombination von 4 Peptiden (EN3, EM18, EM20 und DN11) erlaubt die Identifizierung von 80% aller Sojasensibilisierten Personen. In Abbildung 9b sind die Peptide abgebildet, die am meisten erkannt wurden. Serum IgE von Personen mit klinischen Symptomen erkennen häufiger Peptide als Immunglobuline der anderen Personen.

### 3.7 Zusammenfassung der Ergebnisse

Ziel des Projektes war die Identifizierung der sojaspezifischen Epitope. Dafür standen insgesamt 50 Seren Soja sensibilisierter Individuen zur Verfügung. Davon wurden 25 Seren ausgewählt und für die Peptid Phage Display Experimente verwendet. Aus der speziell designten Phagenbibliothek ENTE-1 wurden mehrere Millionen Phagemidklone selektiert. Mittels NGS konnte eine große Zahl der selektierten Peptidgene, was bis zu einer Million entsprach, sequenziert werden. Die Sequenzdaten wurden mittels trimmomatic und pear (Emboss Software) prozessiert und anschließend in eine JAVA™ basierte Webapplikation, genannt LibDB, eingelesen. Sequenzen, die nicht dem Design der Bibliothek entsprachen, werden von LibDB aussortiert. Zusätzlich werden die Sequenzdatensätze mithilfe des Programms in Motive aufgeteilt. Die angereicherten Motive wurden dann mit den Sojaproteinsequenzen abgeglichen und Übereinstimmungen ausführlicher untersucht. Es wurde angenommen, dass ein Motiv ein Teil eines potenziellen Epitops ist, wenn die Motive im Vergleich zur naiven Phagenbibliothek angereichert wurden und wenn die Motivsequenzen entsprechend zur Proteinsequenz auf 5mer oder 6mer Motive erweitert werden konnten. Mit diesem Vorgehen konnten insgesamt 405 potentielle Epitope in 14 Sojaproteinen identifiziert werden. Von diesen wurden 83 ausgewählt und im Microarray mit Seren getestet. Diese 83 Peptide repräsentierten 42 der am häufigsten gefundenen Epitopkandidaten aus elf Sojaproteinen. Insgesamt konnten 56 von 83 Epitopen bestätigt werden. 55 Peptide wurden trotz eines hohen Grenzwertes bei der Analyse der Arrays von IgE erkannt und 43 Peptide wurden von IgG erkannt. Sowohl Peptide basierend auf der Proteinsequenz als auch Mimotope konnten im Array Antikörper binden. Alle im Microarray validierten Peptide konnten IgE-Antikörper binden. Einige von diesen wurden auch von IgG erkannt. Teilweise besaß eine Person sowohl IgE- als auch IgG- Antikörper gegen das gleiche Peptid.

Zusätzlich wurde auch am Beispiel der in der Literatur beschriebenen Hauptallergene der Birke Bet v 1 und der Haselnuss Cor a 1 untersucht, ob Antikörper mit diesem Protein kreuzreagieren oder ob diese Person auch Antikörper gegen diese Proteine besitzen könnten. Mit Hilfe der Seren von Soja sensibilisierten Menschen konnten 12 Epitope des Bet v 1 und 14 Epitope des Cor a 1 Proteins identifiziert werden. Da jedes dieser Epitope bereits in der IEDB hinterlegt ist, kann man von validierten Epitopen ausgehen. Vermutlich haben diese Personen sowohl Antikörper gegen Soja als auch gegen Birke und Haselnuss. Diese Ergebnisse zeigen, dass die verwendete Peptide Phage Display Methode sehr gut geeignet ist, Antikörper aus Serum zu charakterisieren. Obwohl IgE-Antikörper in geringeren Mengen als IgG-Antikörper im Serum vorhanden sind, ist die eingesetzte Menge von weniger als 1 µl Serum ausreichend, um die IgE-Antikörper zu beschreiben und Epitope zu identifizieren. Die Ergebnisse zeigen auch, dass einige Personen sowohl IgG als auch IgE Antikörper gegen das gleiche Epitop besitzen. In jedem Fall würde auch die *in silico* Identifizierung der IgG Epitope ausreichen, da diese sich kaum von den IgE Epitopen unterscheiden würden. In einigen

Fällen wurde aber nur IgE Bindung gefunden, was darauf schließen lässt, dass tatsächlich Epitope auf IgE-Antikörpern selektiert wurden.



#### 4. Diskussion

Die vorliegende Arbeit beschäftigt sich mit der Identifizierung von Epitopen in Sojaproteinen. Die verwendete Methode des Phage Display stellte sich als viel leistungsfähiger heraus als bereits publizierte Arbeiten in diesem Bereich [26, 147, 148] zeigen. Daher konnte die ursprüngliche Aufgabe dieser Arbeit, die zuverlässige Identifizierung einiger Peptidepitope direkt aus Serum, noch erweitert werden. Es wurden erstmals überhaupt möglichst umfassend alle Epitope aller bekannten allergenen Sojabohnenproteine erfasst. Die große Anzahl der identifizierten Epitope erlaubt ein breites Screening von sensibilisierten Seren. Durch die hohe Auflösung der Epitope kann ein detaillierteres Abbild des individuellen Sensibilisierungsstatus erstellt werden, als bisher möglich. Zusätzlich beschreibt die Arbeit detailreich das IgE- bzw. IgG-Bindungsmuster auf Epitopebene. So wird gezeigt, dass bestimmte Epitope nur von IgE-Antikörpern, wogegen andere Epitope sowohl von IgE als auch von IgG-Antikörpern erkannt werden. Über die Ursache dieser Beobachtung kann nur spekuliert werden. Eine Erklärung lieferten Bøgh et al., 2014 [25]. Sie untersuchten das IgE- und IgG<sub>4</sub>-Epitoperkennungsmuster von drei Seren von erdnussallergischen Patienten mittels Phage Display. Insgesamt wurden 149 Ara h 1 Epitop-Mimotope selektiert. Davon waren 68 IgE – Epitop-Mimotope und 81 IgG<sub>4</sub> – Epitop-Mimotope. Es konnte gezeigt werden, dass die Ara h 1 spezifischen IgG<sub>4</sub> - Epitope in ihrer Sequenz heterogener sind. So scheint es, dass sich patientenspezifische IgG<sub>4</sub> - Epitopmuster von Patient zu Patient mehr unterscheiden, als IgE-Epitopmuster. Die IgG-Erkennungssequenz der Epitope, die in der vorliegenden Arbeit nur von IgE-Antikörpern erkannt wurden, unterscheidet sich möglicherweise so sehr in den einzelnen Individuen, dass die Konsensussequenz von keinem IgG-Antikörper erkannt wurde. Möglicherweise reifen auch mehr IgG-B-Zellen als IgE-Produzenten.

##### 4.1 Experimentelles Design

###### 4.1.1 Phagenbibliothek ENTE-1

Einer der wichtigsten Faktoren für eine erfolgreiche Selektion ist die Größe und die Aminosäureverteilung innerhalb der verwendeten kombinatorischen Bibliotheken. Je größer sie sind, desto wahrscheinlicher ist es, einen zum Zielmolekül passenden Binder zu finden. Dies ist darauf zurückzuführen, dass die Nachbarschaft des Epitops bei der Antikörperbindung wichtig sein könnte. Die Größe einer realisierbaren Bibliothek liegt etwa bei  $10^{11}$  individuellen Klonen. Größere Bibliotheken zu generieren wäre aufgrund von physikalischen und biologischen Hindernissen kaum realisierbar [158]. Das verwendete Phagemid enthält das Gen3 Fusionsprotein unter Kontrolle eines pL Promoters. Dieser gewährleistet die strikte Regulierung der Gene, obwohl dieser im Lambda-Lysogen nicht vollständig unterdrückt ist [159]. Die Expression ist ausreichend, dass wenigstens ein Fusionsprotein pro Phagemid-Partikel generiert wird [158, 160]. Die starke Repression durch den *ci*

Repressor des pL Promoter im verwendeten *E.coli* Wirtstamm gewährleistet minimalen Selektionsdruck und eine gleichbleibende Aminosäureverteilung der 16 Aminosäuren langer hypervariabler Region.

Ein weiterer zu beachtender Punkt ist, dass die Peptide auf der Phagemidoberfläche präsentiert werden und neben anderen Faltungsstrukturen auch Disulfidbrücken ausbilden. Die ENTE-1 Bibliothek wurde so konstruiert, dass biologisch sinnvolle Disulfidbrücken ausgebildet werden. Disulfidbrücken enthaltene Epitope können mit anderen Methoden, z. B. Immunoblot oder Proteinmicroarrays mit überlappenden Fragmenten nicht identifiziert werden. Beispielsweise wurde das strukturelle Epitop FIAEGHPLSLKFD in Gly m TI durch die Kombination von Phagemidklonen identifiziert (Abbildung 10). Eine Phagemidgruppe erkannte den ersten Teil des Epitops (QAGELLCVQVIAEGHPHCD) während ein anderer Abschnitt durch eine andere Phagemidgruppe identifiziert wurde (QAGHPLSHNGGLTYCEKFPD).

(A)

```

-----FIAEGHPLSLKFD
GELLCVQVIAEGHPHCD
-----GHPLSHNGLTYCEKFPD

```

(B)

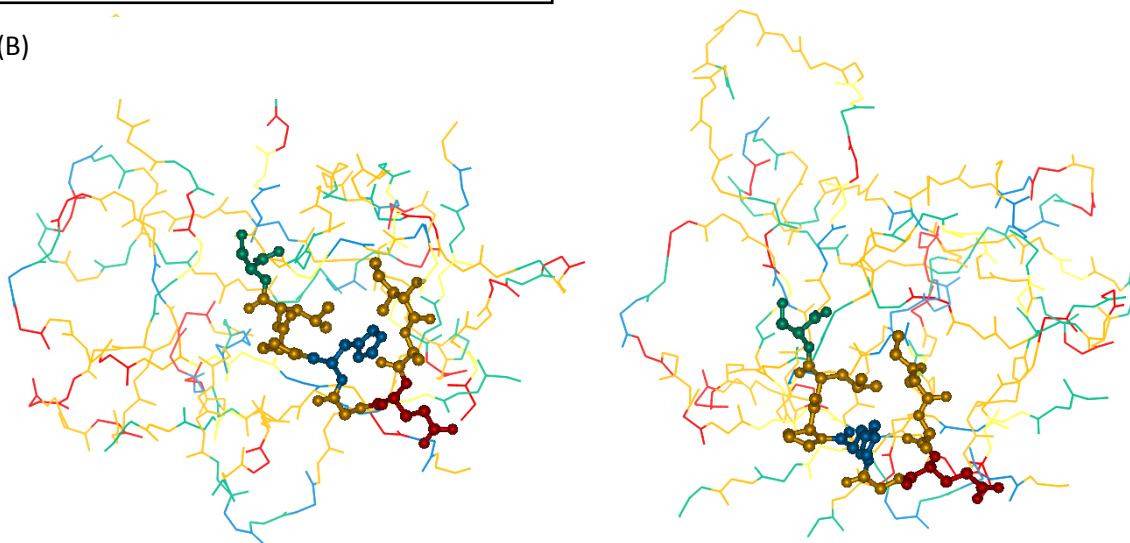


Abbildung 10: Strukturelles Epitop in Gly m TI. (A) zeigt die Sequenz, die mittels Phage Display identifiziert wurde. (B) zeigt die Struktur in zwei verschiedenen Orientierungen (PDB: 1ba7 [161]) Gly m TI. Das Epitop ist in Kugel-Stab-Modell dargestellt und befindet sich in einer Loop-Struktur.

Hauptsächlich werden mit dem Phage Display Mimotope erkannt und nicht das vollständige Epitop. Als Mimotop werden Strukturen bezeichnet, die das Paratop des Antikörpers binden können, die aber nur eine ähnliche Sequenz und/oder Struktur zu den Proteinantigenen haben [162]. Diese Mimotope haften an und werden so selektiert. Durch die Analyse des an den Antikörper bindendem Mimotop könnte man daher auf die räumliche Struktur des Epitops schließen [163].

## 4.2 Selektionen

In dieser Arbeit wurden Serumproteine ohne vorheriges Prozessieren des Serums an Immuntubes gebunden. Es konnte erstmalig gezeigt werden, dass viele Epitope mit nur zwei Selektionsrunden identifiziert werden konnten. Bisherige Arbeiten zeigten wenig Erfolg bei der Analyse der Serumantikörper mittels Peptid Phage Display. Üblicherweise wurden hohe Antikörpertiter für die erfolgreiche Charakterisierung einiger wenige Epitope benötigt [26, 147, 148]. Casey et al. [147], Folgari et al. [148] und Christiansen et al. [26] konnten in bis zu sechs Selektionsrunden nur ein bis zwei Epitope/Mimotope mittels einer arbeitsintensiven Methode identifizieren. In diesen Publikationen wurden Affinitätsselektionen benutzt um IgE oder IgG Antikörper aus Serum anzureichern.

Mit der detaillierten Identifizierung von Epitopen der Sojabohne befasste sich das MAVO-Projekt *LowAllergen*. Die in diesem Projekt angewandte Methode stellt eine neue Vorgehensweise für die *in silico* Auswertung von Phage Display-Daten dar, um diverse Peptidliganden zu identifizieren. Durch die in der Arbeitsgruppe entwickelte Selektion auf Seren [156], der Software LibDB und einer einzigartigen Phagenbibliothek wird die statistische Untersuchung von gebundenen Peptidsequenzen ermöglicht. Die Vervielfältigung der Phagen Bibliotheken in *E. coli* unterliegt häufig einem biologischen Selektionsdruck, weil einige Aminosäurekombinationen in den Bakterien langsamer amplifiziert und andere Sequenzen bevorzugt sekretiert werden. Toxizität, Sensitivität der Peptide gegen Proteasen und ungewünschte Bildung von Disulfidbrücken können für starke Variationen in der Amplifikation verantwortlich sein [142, 164]. Ein geringfügiger Unterschied in der Vermehrungsrate von *E. coli* kann einen erheblichen Einfluss auf die Zusammensetzung der Phagen haben, die unterschiedliche Peptide auf ihrer Oberfläche repräsentieren [164]. Beispielsweise kommt die Aminosäure Prolin einerseits häufig in Protein-Protein-Interaktionsstellen vor [165], andererseits kann sie aber auch durch die Eigenschaften des für die Phagenverpackung notwendigen *E.coli* Sekretionsapparats gehäuft vorkommen [166]. Unsere Daten zeigen allerdings keine Häufung von Prolin-enthaltenen Motiven. Bei einer größeren Anzahl von Amplifikationszyklen werden die benachteiligten Sequenzen, die ebenfalls Epitope enthalten können, noch stärker verringert. Die Häufigkeit der Aminosäuren in der Phagenbibliothek sind nicht mehr gleich verteilt und das Ergebnis wird verfälscht [144]. Es wurden nur zwei Selektionsrunden durchgeführt, um diese unspezifischen Phagenklone zu vermeiden und um möglichst alle Binder zu erfassen. Dies ist der beste Kompromiss um eine deutliche Anreicherung von den gewünschten Motiven zu erreichen und unspezifische niedrig zu halten. Teilweise konnte in den Daten der ersten Selektionsrunde schon eine Anreicherung zumindest von Motiven beobachten werden, die in der zweiten Selektionsrunde deutlich zunahm. Die Anreicherung in der ersten Selektionsrunde war jedoch nicht immer aussagekräftig genug, um auf potenzielle Epitope zu identifizieren.

#### 4.2.1 Next Generation Sequencing

Ohne Next Generation Sequencing (NGS) wäre eine Durchführung der vorgestellten Arbeiten nicht möglich gewesen. Diese Methode erlaubt es eine sehr große Anzahl von Sequenzen gleichzeitig zu sequenzieren. Ein Problem des NGS ist aber die hohe Fehlerrate beim Sequenzieren, darüber hinaus können Amplifizierungsfehler während der PCR auftreten. Um dies zu vermeiden wurde eine Vent® DNA Polymerase mit einer 3'-5' Korrekturlese- Exonukleaseaktivität für eine exakte Lesegenauigkeit verwendet. Die Vent® DNA Polymerase besitzt eine bis zu 15 mal höhere Lesegenauigkeit als die einer Taq DNA Polymerase [167, 168]. Zusätzlich wurden nur 15 Amplifikationszyklen pro PCR durchgeführt. Um die benötigte DNA-Menge für die Sequenzierung zu erzeugen, wurde die PCR drei bis viermal durchgeführt und die PCR-Produkte anschließend vereint.

Die Fehlerrate des MiSeq von illumina® soll im Vergleich zu anderen NGS-Systemen bei 1-2 % pro Base liegen. Dies wird durch die spezielle Brückenamplifikation gewährleistet, indem die DNA-Stränge in zwei Richtungen während der Sequenzierung abgelesen werden [169]. Durch den Vergleich der vor- und rückwärts gelesenen Sequenzen mit der Software Pear wurde die Fehlerrate der auszuwertenden Sequenzen möglichst gering gehalten. Bei Übereinstimmung der vor- und rückwärts gelesenen Sequenzen wird die Konsensussequenz ausgegeben. Gibt es Unterschiede zwischen beiden Leserastern, werden die Qualitätsscores, die für jede Base vorhanden sind, berücksichtigt. Die Sequenz mit den besseren Qualitätswerten wird schlussendlich als richtig ausgegeben. Die so bereinigten Datensätze wurden anschließend in die interaktive JAVA™ basierte Webapplikation, genannt LibDB, eingelesen. Ein Vorteil dieser Software ist, dass von den Motiven direkt auf die Sequenzen zugegriffen werden kann und umgekehrt. Dies ermöglicht ein sehr schnelles Analysieren der Daten. Sequenzen, die vom Design der Bibliothek abweichen, wurden von diesem Programm nicht in die Datenbank eingelesen. So wurden pro Selektionsrunde bis zu 50 % der Sequenzen aussortiert (Tabelle 11). Eine ähnlich hohe Ausschussrate von Sequenzen wurde auch in anderen Arbeiten gesehen [26].

Durch diese Prozedur ist eine Identifizierung potenzieller Sojabohnenepitope durch statistische Methoden möglich. Die NGS-basierte Untersuchung der selektierten Phagemidklone und dadurch möglichen Motiv-Analyse bietet eine kosteneffektive und schnelle Methode Serumproben zu analysieren.

#### 4.2.2 Identifizierung der Epitope über Vierermotive

Zuerst werden die Datensätze nach der statistischen Anreicherung der 4mer-Motive untersucht. Die Ente-1 Bibliothek besitzt  $5 \cdot 10^9$  Klone und eine Komplexität, die ausreichend ist, dass die 4mer-Motive in vielen verschiedenen Umgebungen existieren und Antikörperbindung möglich ist. Die Anzahl aller möglichen Vierermotive liegt bei 160.000 ( $20^4$ ). Für eine statistische Auswertung werden

mindestens ca. 1,2 Millionen Motive benötigt, dies entspricht ca.  $1-2 \cdot 10^5$  Peptidsequenzen. Die Selektion resultiert in ausreichend großen Datensätzen, speziell, wenn berücksichtigt wird, dass jede Sequenz mehrere 4mer-Motive enthält.

Eine statistische Analyse der 5mer-Motiven (oder höher) ist nicht möglich, weil mindestens, wie bei den Vierermotiven, zehnmal so viele Motive benötigt werden, wie theoretisch möglich sind. Wenn alle Fünfermotive berücksichtigt werden, ergibt das eine Anzahl von 3,2 Millionen ( $20^5$ ). Dies entspricht 32 Millionen Sequenzen. In den Experimenten wurden jedoch deutlich weniger Klone selektiert. Dies liegt nicht an der Größe der eingesetzten Bibliothek, sondern an der Anzahl der vorhandenen Zielmoleküle. In den verwendeten Immunotubes könnten nicht genügend Antikörpermoleküle binden, um ausreichend Sequenzen für die statistische Analyse der 5mer-Motiven zu selektieren. Zudem muss beachtet werden, dass sich, sobald die Anzahl der eingesetzten Phagenpartikel erhöht wird, auch die Anzahl der Hintergrundbinder zunehmen kann. Dies bestätigen auch Erfahrungen der Arbeitsgruppe beim Epitopmapping von monoklonalen Antikörpern.

Im nächsten Schritt der Epitopanalyse wurden die Epitope charakterisiert, indem die angereicherten 4mer-Motive auf erweiterbare Aminosäuren untersucht wurden. Die Größe der Datensätze ermöglichte den Vergleich von hinreichend vielen angereicherten Motiven mit der Proteinsequenz. Mit den Datensätzen, die im Durchschnitt 450.000 Sequenzen enthielten, war eine Erweiterung auf fünf Aminosäuren entsprechend der Proteinsequenz problemlos möglich (Abbildung 8). Beispielsweise konnte zum 4mer-Motiv ASQV die Sequenz 2013\_003 AGQ**FASQVV**WGPLEVYALC zugeordnet werden. In dieser Sequenz stimmten auch das C-terminale Phenylalanin und das N-terminale Valin des 4mer-Motivs mit der Proteinsequenz überein. Die Erkennung der gesamten potenziellen Helixstruktur „ASQV“ durch den Antikörper ist unwahrscheinlich, weil es von Antikörper nur einseitig gebunden werden kann. Der Antikörper erkennt wahrscheinlich „AXQV“, wobei X für jede beliebige AS steht. So werden alle Motive angereichert, die „AXQV“ enthalten, unter anderem auch „ASQV“. Zum anderen wird das Mimotop auch angereichert, weil verschiedene Antikörper im Serum vorhanden sind, von denen jeder ein etwas anderes Epitop erkennt. Durch den Konsensus aller Erkennungssequenzen kann das Epitop rekonstruiert werden.

Bei einigen Motiven wurde sogar eine Erweiterung auf sechs oder sieben Aminosäuren entsprechend der Proteinsequenz gefunden. Aufgrund der großen Anzahl der gebundenen Sequenzen war es möglich die Sequenzvarianz des Epitops abzulesen und eine Konsensussequenz zu bestimmen. Dies ist auch ein großer Vorteil gegenüber dem Peptid-Microarray auf Fragment Basis, der nur auf der nativen Sequenz beruht.

Für die durchgeführten Experimente eine Negativkontrolle zu finden, erwies sich als schwierig. Jeder Mensch konsumiert Soja bewusst, durch Verzehr von sojabasierten Lebensmitteln, oder unbewusst, durch den Verzehr von „versteckten“ sojabasierten Inhaltsstoffen in Lebensmitteln. Daher besitzt

auch jeder Mensch mit hoher Wahrscheinlichkeit Antikörper gegen Soja, weshalb keine Negativkontrolle definiert werden kann. Ersatzweise wurden viele verschiedene Serien sensibilisierter Personen und zwei nicht-sensibilisierter Personen miteinander verglichen. Jeder Mensch hat außerdem sein eigenes Erkennungsmuster an Antikörpern [170, 171]. Es zeigte sich, dass erst durch die Kombination von bestimmten Epitopen zwischen Patienten mit klinischen Symptomen und Soja sensibilisierten Personen unterschieden werden konnte.

Die Aminosäureverteilung innerhalb der Epitopkandidaten und der Proteinsequenzen wurde untersucht. Tabelle 13 zeigt die Ergebnisse dieser Untersuchungen. Die am häufigsten identifizierten Aminosäuren in den Epitopen der Sojabohne sind Glutaminsäure gefolgt von Glutamin. Prolin, Arginin und Cystein kommen in den Epitopen der Sojabohne häufiger vor, als man das anhand der Aminosäureverteilung innerhalb der Sojaproteinsequenzen erwarten würde. Die am seltensten vorkommenden Aminosäuren in den Epitopen der Sojabohne sind Methionin und Tryptophan. Die geringe Häufigkeit dieser beiden Aminosäuren liegt am Design der verwendeten ENTE-1 Bibliothek. Beide Aminosäuren sind hier stark unterrepräsentiert, so wie es auch in der Natur der Fall ist. So sind viele Methioninmotive in der Bibliothek nicht vorhanden. Die in den Epitopen häufig entdeckten Aminosäuren wurden teilweise von anderen Autoren bereits untersucht.

Auch Dall'Antonia beschrieb verschiedene Aminosäuren, die in Epitopen besonders häufig vorkommen. Diese waren: Tryptophan, Arginin, Glutamin, Asparagin und Asparaginsäure [20].

Asparagin, Arginin, Prolin und Lysin wurden von Rubinstein et al. als wichtige Aminosäuren in Epitopen identifiziert. Dagegen spielen Cystein, Alanin, Leucin, Valin und Phenylalain keine wichtige Rolle in Epitopen [172].

Die hier beschriebene Methode erlaubt die Identifizierung von einer großen Anzahl von Epitopen im Gegensatz zu der bereits zitierten Literatur [26, 147, 148], in der nur einige wenige Epitope identifiziert wurden. Die Identifizierung von vielen Epitopen wird neben dem verwendeten Promotor, der an derer Stelle diskutiert wird (siehe 4.1.1), unter anderem dadurch ermöglicht, dass die Anreicherung von Motiven untersucht wird und nicht wie berichtet [26, 147, 148] nach Anreicherung ganzer Sequenzen. Dieses Vorgehen würde auch mit der hier verwendeten Bibliothek in zwei Selektionsrunden nicht zum Erfolg führen (siehe Abbildung 11). In Abbildung 11 wird die Häufigkeit des Motivs „PRAV“ in den ersten 100, 200 und 1000 Sequenzen in zwei Selektionsexperimenten des Aktin-Antikörpers (1 und 2) gezeigt. Im ersten Selektionsexperiment enthalten fast die Hälfte (45/100) der Sequenzen in der zweiten Selektionsrunde (1b) das Motiv. Im Zweiten Experiment (2b) sind es aber nur 10 Sequenzen. Es ist fraglich, ob im zweiten Selektionsexperiment das richtige Mimotop gefunden werden würde. Betrachtet man allerdings die Anreicherung des Motivs über die Selektionsrunden, kann eine deutliche Anreicherung von 1. zur 2. Selektionsrunde beobachtet

werden. Nach zusätzlichen Selektionsrunden fände man aber nur noch wenige angereicherte Epitope für einzelne Antikörper.

Für die Auswertung werden nur Motive herangezogen, die in mehreren Sequenzen gefunden werden. Die Gefahr, dass Motive gefunden werden, die nur aufgrund von Sequenzierfehlern zufällig mit der Proteinsequenz übereinstimmen wird so minimiert. Die Suche nach der Anreicherung von Motiven, im Gegensatz zu der Anreicherung von Sequenzen, erlaubt die parallele Identifizierung einer großen Anzahl von möglichen Mimotopen.

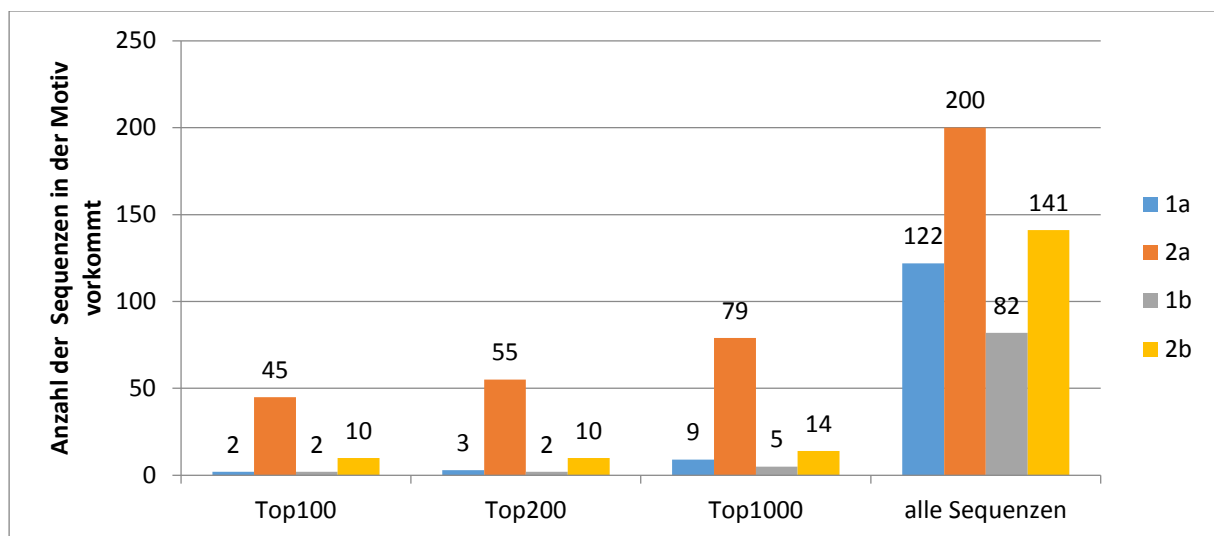


Abbildung 11: Anreicherung des Motivs „PRAV“ in zwei Selektionsexperimente des Aktins-Antikörper (1 und 2) über zwei Selektionsrunden (a und b). In der ersten Selektionsrunde enthalten nur 2 Sequenzen der 100 angereicherten Sequenzen, 79 Sequenzen der 1000 angereicherten Sequenzen und 200 verschiedenen Sequenzen insgesamt das Motiv. Die Größe der Datensätze variiert: 1a:499535 gültige Sequenzen; 1b: 598595 gültige Sequenzen; 2a: 556384 gültige Sequenzen und 2b: 682287 gültige Sequenzen.

#### 4.3 Array Datenauswertung

Für die Validierung der *in silico* identifizierten Peptide standen zum einen ein 15mer Peptid-Microarray mit 15 Sojaproteinen und zum anderen ein Epitop-Array mit den synthetisierten Peptid-Epitopen zur Verfügung. Es konnten 405 bzw. 232 im Phage Display und einem 15mer Peptid-Microarray mit überlappenden Peptiden erkannt werden. Die Differenz kann dadurch erklärt werden, dass mit dem Phage Display Ansatz zum einen mehr Proteine untersucht werden konnten und zum anderen enthielten überlappende Peptide des 15mer Peptid-Microarray teilweise mehrere Epitope. Die Arbeiten des 15mer Peptid-Microarray wurden im Rahmen einer anderen Doktorarbeit durchgeführt [173]. Allerdings wurden im 15mer Peptid-Microarray überlappenden Peptide nur IgE-Bindungsmotive untersucht, weil eine zusätzliche Untersuchung auf IgG einen viel größeren finanziellen Mehraufwand bedeutet hätte, der so im Projekt nicht darstellbar war. In der *in silico* Datenanalyse wurden hingegen alle Immunglobulin-Bindungsmotive identifiziert. Mittels des Phage

Displays konnten fünf bis vierzehn Aminosäuren lange Epitope gefunden werden, wogegen mit dem ersten Ansatz fünfzehn Aminosäuren lange Bindungsmotive bereitgestellt wurden. Durch die Verwendung von überlappenden Peptiden wurde im 15mer Peptid-Microarray eine Auflösung von vier Aminosäuren erreicht.

Einige Antigenstrukturen sind sehr wahrscheinlich strukturell oder chemisch in diesem 15mer Peptid-Microarray Format nicht darstellbar, dies spiegeln unsere Ergebnisse wider. Teilweise konnten nur Mimotope aus dem Phage Display Serumantikörper im Microarray binden. Mit dem Phage Display ist es möglich Mimotope zu identifizieren, welche Antikörper binden können, die nicht die native Struktur erkennen würden.

Zusätzlich scheinen 20 - 30 Ig-Moleküle in der Selektion für eine Bestimmung ausreichend zu sein, wodurch diese Technik eine sensitivere Methode als der obige Microarray darstellt. Ein Grund dafür könnte sein, dass die Peptide auf einem Objektträger in einer statischen Form den Antikörpern präsentiert werden. Im Phage Display werden die Peptide dagegen auf Phagemid-Partikeln präsentiert. Dies könnte bedeuten, dass die auf der Phagemid-Partikeloberfläche präsentierten Peptide unterschiedliche Strukturen mit unterschiedlichen Affinitäten haben. Im Fall des Phage Displays kann die Nachbarsequenz des Epitops zusätzlich für die Bindung hilfreiche Strukturen bilden, die mehr dem nativ gefalteten antigenen Protein entsprechen. Allerdings können mit der *in silico* Datenanalyse genauso wie im Peptidarray keine Epitope von Glyko- oder Lipidproteinen identifiziert werden.

Der Microarray ist für eine Messung im großen Maßstab und für Hochdurchsatzverfahren geeignet [174]. Diese Technik kann für die Krankheitsdiagnose, - Vorhersage, - Prävention und Arzneimittellentdeckung genutzt werden, indem DNA, RNA, Peptide, Proteine oder andere Biomoleküle getestet werden [175]. Es können aber auch drei dimensionale Strukturen, auch Scaffolds genannt, im Arrayformat getestet werden [176, 177]. Peptid-Microarrays mit überlappenden Fragmenten, welche eine typische Länge von 10 – 20 Aminosäuren haben, sind mittlerweile eine Standardmethode um Epitope zu kartieren [178–181]. Die Pionierarbeit leistete PepScan (Lelystad, The Netherlands) vor über 20 Jahren, indem sie die Epitop-Charakterisierung mithilfe von überlappenden Bibliotheken nicht nur linearer Peptide vorantrieben [182]. PepScan verwendet strukturell eingeschränkte Peptide. Der Vorteil einer starren Struktur ist in der Regel immer eine höhere Affinität, weil die Zahl der Freiheitsgrade sinkt. Viele Peptide sind zu flexibel, was eine schlechte Affinität zur Folge hat. Nachteile der Proteinarrays mit überlappenden Fragmenten sind die hohen Kosten und die Restriktion auf natürliche Aminosäuren des Antigens für die Epitoperkennung. Generell können die Epitope, die mittels Array identifiziert werden aufgrund der festgelegten Länge der aufgetragenen Peptide auch mehr als ein Epitop enthalten. Durch die festgelegte Länge der überlappenden Peptide, erkennt der Microarray nur die ungefähre Position des



Epitops in der nativen Sequenz. Die Phage Display-Methode basiert auf der Analyse der Motivanreicherung vieler Sequenzen und liefert viel detailliertere Informationen über potenzielle Epitope über eine variable Länge von fünf bis fünfzehn Aminosäuren (Tabelle 16). Jede Peptidsequenz wurde an einem anderen Antikörpermolekül angereichert. Alle Peptide haben eine ähnliche Struktur gemeinsam. Wenn die Anzahl der verschiedenen Sequenzen groß genug ist, kann ein Epitop-Fingerabdruck erstellt werden, der die akzeptierten Aminosäurevariationen innerhalb der Antikörperbindungsstelle anzeigt (Abbildung 8). Mithilfe dieser Daten kann eine Konsensussequenz für die Peptid-basierte Diagnostik für die spezielle Identifizierung von Kreuzreaktionen und/oder Multisensibilisierung ermittelt werden, beispielsweise wenn Seren Antikörper sowohl gegen Gly m 4 als auch Bet v 1 enthalten, sowie Antikörper, die beide Proteine erkennen.

Viele im Phage Display identifizierte Epitope zeigten auch im Array eine Antikörperbindung. Das ist insofern eine Bestätigung der Vorgehensweise bei der Auswahl der Peptide, weil diese ja auch aus dem Kontext der Phagemid-Partikeloberfläche herausgenommen wurden, aber bei der Auswahl ein Konsensus verschiedener Klone berücksichtigt wurde. Im Array wurden 83 Sojaepitope dargestellt, die von möglichst vielen Personen erkannt wurden, wovon 56 Peptide positiv getestet werden konnten. Sieben Peptide wurden sogar von 20 und mehr Seren entweder als IgE und/oder als IgG Binder identifiziert. Dazu gehörten Peptide aus Gly m 6.01, Gly m BD30K, Gly m 5.01 und Gly m 5.02. Die Signalintensität ist von vielen Faktoren abhängig. Erstens hat die Spotgröße und zweitens hat die Peptidreinheit ein Effekt auf die Intensität des Signals. Die Reinheit der genutzten Peptide variiert von 50 bis 90 % (ermittelt mittels Massenspektrometrie). Es wurden verschiedene Konzentrationen (9- 81 ng/ $\mu$ l) gespottet, um die verschiedenen Reinheitsgrade auszugleichen, wenn diese Variable einen Effekt auf die Intensität hat. Allerdings haben die meisten Peptide eine Reinheit von 90 %. Es stellte sich heraus, dass die unterschiedlichen gespotteten Konzentrationen keine Auswirkung auf die Bindung an sich haben. Der dritte Faktor, der die Signalintensität beeinflusst, ist, dass verschiedene Serumantikörper- Subklassen sich gegenseitig kompetieren können. Serum IgE und IgG Antikörper, die dasselbe Paratop haben, können die Signalintensität beeinflussen, indem sie um die Bindung kompetieren.

Es wurde nicht beobachtet, dass Peptide basierend auf den Mimotopen besser oder schlechter von den sensibilisierten Personen erkannt werden. Mal wird die Sequenz des Mimotopes und mal wird die Sequenz basierend aus der Sojabohne besser von den Antikörpern aus den Seren erkannt. Insgesamt wurden aber doppelt so viele IgE-Binder wie IgG-Binder identifiziert. Dies bedeutet nicht, dass ein Peptid nur IgE oder IgG bindet, es ist in der Tat eher so, dass das Peptid sowohl aus dem einem Serum IgE und aus einem anderem Serum IgG bindet. Vereinzelt enthält ein und dasselbe Serum sowohl IgE- als auch IgG-Binder gegen das gleiche Epitop. Eine Erklärung für diese Ergebnisse

ist, dass jeder Patient verschiedene Antikörpervariationen aufweist und Mimotope die Struktur von Fall zu Fall unterschiedlich besser wiedergeben.

Drei Patienten (14002; 14009 und 14015) zeigen keine Antikörperbindung im Epitoparray, obwohl der Pricktest positiv war. Dies könnte auf ein falsch positives Testergebnis des Pricktest hindeuten. In nur 60 % der Fälle weist der Haut-Pricktest auf eine IgE-basierte Reaktion hin [78], die durch Nachweis des IgE bestätigt werden kann. IgE in Konzentrationen unterhalb der Nachweisgrenze kann aber trotzdem im Serum oder vor allem gebunden an MAST-Zellen vorhanden sein. Abbildung 13

zeigt exemplarisch die zusammengefassten Ergebnisse für das Protein Gly m 2. In dieser Abbildung sind die alle vorhandenen Daten zusammengetragen worden. Es zeigt die mit dem Peptide Phage Display und die mit dem Peptid Microarray mit überlappenden Peptiden identifizierten Epitope, sowie die im Microarray bestätigten Antikörperbindungsmotive. Anhand dieser Grafik sieht man zum einen, dass die Epitope, die mit der Peptide Phage Display Methode beschrieben worden sind, auch mit den anderen Methoden identifiziert wurden und umgekehrt. Ein Vorteil der im Peptide Phage Display identifizierten Epitope ist, dass sie wesentlich kürzer sind. Die Abbildung zeigt außerdem, dass es bestimmte Epitopcluster gibt, d. h. bestimmte Bereiche innerhalb des Proteins werden häufiger als andere Stellen von Antikörpern erkannt. Allerdings gibt es keine Region innerhalb des Proteins, die von jedem getesteten Serum erkannt wird. Auch andere Wissenschaftler beobachteten schon, dass jeder Mensch ein individuelles Ig-Genrepertoire besitzt und zudem ein individuelles IgE-Repertoire ausbildet [11, 25, 170, 171]. So zeigten Awazuhara et al. zum Beispiel in einer Studie mit 30 Patienten, dass die spezifische Proteinreaktivität zwischen den verschiedenen Patienten sehr unterschiedlich ist. Dies würde sich ggf. in unterschiedlichen Signalthöhen im Microarray widerspiegeln. Seren einiger Personen reagierten nur schwach gegen ein oder zwei Sojaproteine, wogegen andere Patienten auf kein Protein reagierten. Viele Proteinbanden waren in dieser Studie positiv, aber keine war dominant und wurde von allen Patientenseren erkannt. Dies ist auch unwahrscheinlich aufgrund des schon erwähnten individuellen Antikörperrepertoires. Es muss allerdings berücksichtigt werden, dass im Immunoblot strukturelle Epitope nicht gefunden werden können [170].

Zusammenfassend zeigen diese Ergebnisse, dass die Peptid Phage Display Methode sehr gut geeignet ist, Antikörper aus Seren zu charakterisieren. Obwohl IgE-Antikörper in geringerer Menge als IgG-Antikörper im Serum vorhanden sind, ist die eingesetzte Menge von weniger als 1µl Serum ausreichend, um IgE-Antikörper zu beschreiben und ihre Epitope zu identifizieren. Die Ergebnisse zeigen außerdem, dass einige Personen sowohl IgG als auch IgE Antikörper gegen das gleiche Epitop besitzen.

#### 4.4 Vergleich mit den in der Literatur beschriebenen Soja-Epitopen

Es sind schon sehr viele Lebensmittelallergene und damit die allergenen Bestandteile bekannt. Andere Studien berichteten bereits über stark allergene Sojaproteine, sie untersuchten allerdings nur ganze Proteine und nicht einzelne Epitope. Ogawa et al. [102] benutzten 69 Seren von sojasensitiven Patienten, um IgE-bindende Epitope an Sojaproteine zu beschreiben. Drei Hauptallergene, Gly m BD 28K, Gly m BD 30K und die  $\alpha$ -Untereinheit von  $\beta$ -Conglycinin (Gly m 5.01) wurden identifiziert. Andere Hauptallergene sind die Speicherproteine (Gly m 6) und Profilin (Gly m 3) [183]. Die saure Untereinheit von Glycinin G1 (Gly m 6.01) und die basische Untereinheit von Glycinin G2 [184] gelten als signifikante Allergene in Patienten mit Sojaallergie. Der wichtigste Unterschied, zu der hier vorliegenden Arbeit ist, dass in den publizierten Studien die Antikörperbindestellen meist durch Immunoblot oder ELISpot identifiziert wurden [115, 184]. Potenzielle Epitope von 15 und mehr Aminosäuren sind das Ergebnis der publizierten Arbeiten. Die Charakterisierung der Antikörperbindungsstelle auf Aminosäureebene ist nur möglich durch den zeit- und kostenintensiven Ala-Scan [185]. Die im LowAllergen Projekt beschriebenen Epitope haben eine viel höhere Auflösung. Der Epitop-Fingerabdruck in Abbildung 8 ist abgeleitet von allen Motivsequenzen und zeigt die wichtigsten Aminosäuren.

In dieser Arbeit werden 405 Epitope der Sojabohne beschrieben. Die Länge der Epitope variiert von fünf bis sechzehn Aminosäuren. Berücksichtigt man die Länge der Proteine, wurden die meisten Epitope in den Speicherproteinen Gly m 5 und Gly m 6, die über 50 % des Proteingehalts der Sojabohne entsprechen [184], identifiziert. Diese Proteine beinhalten mehr als 20 Epitope pro 100 Aminosäuren. In der Immune Epitop Database ([www.IEDB.org](http://www.IEDB.org)) sind zurzeit 195 Epitope registriert. Die Länge der Epitope variiert von 10 bis 53 Aminosäuren, im Durchschnitt 16 Aminosäuren. Die meisten Epitope wurden in Gly m 5, Gly m 6 und Gly m BD30K gefunden. Neben Gly m 28K wurden auch Gly m 5 und Gly m BD30K von Ogawa et al. als die Hauptallergene der Sojabohne identifiziert [186, 187]

##### 4.4.1 Gly m 5 Epitope

In dieser Promotionsarbeit wurden 107 Epitope in Gly m 5 gefunden. 66 % davon wurden schon publiziert [8–10]. 66 Epitope des Protein Gly m 5 ( $\beta$ -Conglycinin) sind in der Datenbank IEDB vermerkt (19. November 2018) welche auf drei Publikationen zurückzuführen sind [8–10]. 20 Epitope werden Gly m 5.01 und 46 Epitope Gly m 5.03 zugeordnet. Es sind keine Epitope von Gly m 5.02 identifiziert worden. Die publizierten Epitope erstrecken sich über die gesamte Proteinsequenz, sowohl in Gly m 5.01 als auch in Gly m 5.03. In Gly m 5.01 können die Epitope in sechs Cluster unterteilt werden. Im Bereich zwischen der 151.-252. Aminosäure befinden sich allein zehn Epitope, welche sich in großen Bereichen überlappen. Von 66 publizierten Epitopen [8–10] konnten zu sieben

Epitopen keine Daten aus dem Phage Display zugeordnet werden. Alle anderen Epitope konnten als kurze Peptide von fünf bis fünfzehn Aminosäuren identifiziert werden.

Von 20 Epitopen, die von Fu et al. [8] und Sun et al. [9] publiziert wurden, konnten auch 18 mit der Phage Display Methode gefunden werden. Beide Autoren berichteten, dass die Epitope nur am N-Terminus des Proteins befanden, da diese flexible, hydrophobe und aus polaren Aminosäuren bestehende Region sehr immunogen ist. Die in dieser Arbeit identifizierten Epitope befinden sich über das gesamte Protein verteilt. Allerdings befinden sich viele Epitope und solche, die von einer großen Anzahl von Personen erkannt werden, häufig am Anfang der Proteinsequenz.

Fu et al. berichteten, dass die Seren trotz der hohen Sequenzidentität zwischen den verschiedenen Untereinheiten des Gly m 5 Proteins spezifisch nur an Gly m 5.01 banden [8]. Dies kann in dieser Promotionsarbeit nicht bestätigt werden, da keine ganzen Proteine verwendet wurden. Sowohl im Peptid Microarray mit überlappenden Fragmenten als auch mit den Phage Display Daten konnten die gefundenen Epitop-Motive zu allen Untereinheiten passen.

Die einzige Arbeit, in der Epitope in Gly m 5.03 beschrieben sind, ist von Taliercio et al. [10]. Von den 46 IgG und IgE Epitopen, die in seiner Publikation beschrieben werden, konnten 32 Epitope mittels der Phage Display Methode identifiziert werden. Dies ist vermutlich darauf zurückzuführen, dass Taliercio et al. [10] das Epitopmuster von verschiedenen Tierspezies, z.B. Schwein und Hund, untersuchte und nicht das von Menschen.

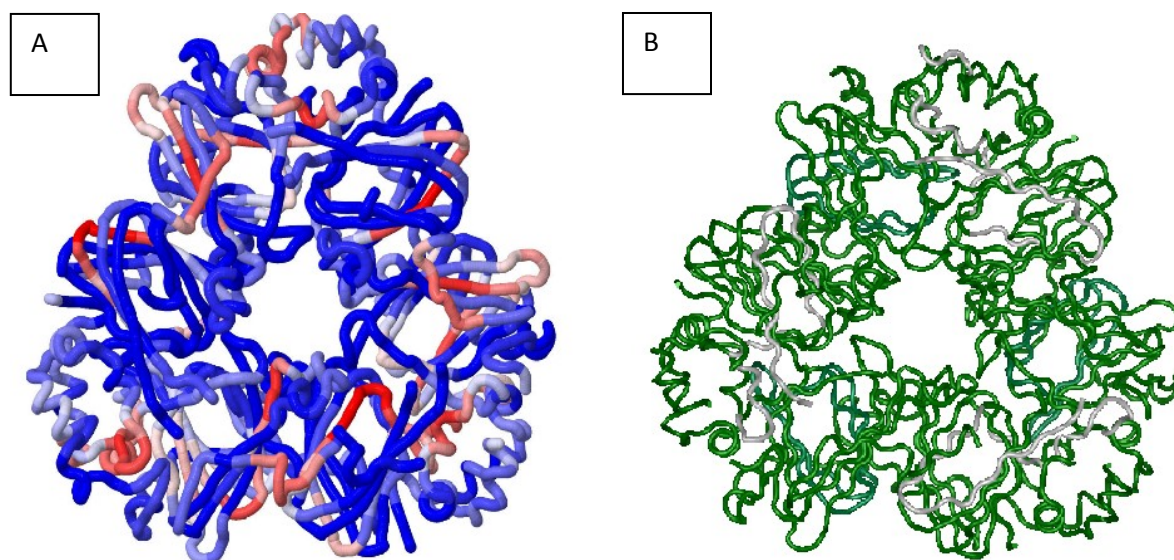


Abbildung 12: Epitopkarte von Gly m 5 (PDB 1IPK [188]). A zeigt die Epitope, die mittels Phage Display identifiziert worden sind. In rot sind die Abschnitte markiert, die von besonders vielen Menschen erkannt werden, blau zeigt die Regionen, die kaum oder nicht erkannt werden. B zeigt die Epitope, die in der IEDB gelistet sind. Nimmt man alle Epitope zusammen wird fast das ganze Molekül von Antikörpern erkannt.

#### 4.4.2 Gly m 6 Epitope

In der vorliegenden Arbeit werden 129 Epitope des Gly m 6 beschrieben. 32 % der identifizierten Epitope wurden bereits in anderen Arbeiten publiziert [115, 179, 184, 189–191]. Es sind 54 Epitope für Gly m 6 in der IEDB beschrieben, die in sechs Publikationen publiziert [115, 179, 184, 189–191] worden sind. Die Gly m 6.01 Epitope sind in drei Cluster unterteilt, wobei das Cluster von Position 216-235 am häufigsten erkannt wurde.

Beardslee et al. [115] untersuchten Gly m 6 und bestimmten eine Hauptbinderegion für IgE: F192-A232. In dieser Region wurden zwei Hauptepitope von Gly m 6.01 identifiziert: GGSILSGFTLEFLEHAFSV und GAIVTVKGGLSVI. Die Epitope, die mittels Phage Display erkannt wurden, sind unterstrichen dargestellt.

In dieser Arbeit wurden 27 von 30 Epitopen identifiziert, die bereits Helm et al. [184] publizierten. Diese Wissenschaftler lokalisierten Epitope von Gly m 6.02 mithilfe der Genosys Spot Methode. Mit diesem Ansatz wurden sechs große IgE-Binderegionen identifiziert: 1–23, 57–111, 169–215, 249–271, 329–383 und 449–471.

Es konnten alle neun Epitope in Gly m 6.02 erkannt werden, die bereits von Saeed et al [190] im Western Blot charakterisiert wurden. Diese Gruppe untersuchte auch Gly m 6.03, sie benutzte allerdings eine andere Proteinsequenz, daher konnten keine Epitope für dieses Protein bestätigt werden.

Xiang et al [189] ermittelten ein dominantes IgE-bindendes Epitop in Gly m 6.02 in der Region S219–N233: SGFAPEFLKEAFGVN. Das unterstrichen Epitop wurde im Phage Display erkannt.

Es werden hier auch acht von elf Epitopen beschrieben, die zuvor von Taliencio et al. [179] publiziert wurden. Diese Wissenschaftler identifizierten drei Epitopregionen, die von mehreren Seren erkannt wurden: LIQTWNSQHPELKCAGVTVSK, LIETWNPNNKPFQCAGVALSR, IVTVEGGLSVISPKWQE. Die markierten Aminosäuren stellen die für die Immunglobulinbindung wichtigen Aminosäuren dar.

#### 4.4.3 Epitope anderer Sojaproteine

Es wurden 21 Epitope in Gly m BD30K in dieser Arbeit identifiziert. In der IEDB-Datenbank sind 29 Epitope basierend auf drei Publikationen [192–194] registriert. Zehn von 16 Epitopen, die von Helm et al. [193] beschrieben werden, wurden mittels Phage Display erkannt.

Hiroshi et al. [194] benutzten zwei monoklonale Mausantikörper, um zwei Epitope in Gly m BD30K zu identifizieren: QGGCGRGWAFSATGAIEA und DKVTIDGYETLIMSDEST. Die Epitope, die mittels Phage Display identifiziert wurden, sind unterstrichen dargestellt. Dreizehn Epitope in Gly m 8 wurden auch in dieser Arbeit identifiziert. Für Gly m 8 sind 12 Epitope in der IEDB-Datenbank gelistet. Diese wurden von Han et al. [195] mittels überlappender Peptide im Microarray identifiziert. Acht von

diesen elf Epitopen konnten auch im Phage Display Ansatz identifiziert werden. Teilweise enthält ein beschriebenes Proteinfragment mehrere, potenziell unabhängige Epitope.

Es konnten sieben Epitope in Gly m 3 identifiziert werden. Dies deckt sich nicht mit den Beobachtungen von Rihs et al. [97], der nur ein Epitop mit der Länge von 53 Aminosäuren charakterisierte. Diese Wissenschaftler hatten ein rekombinantes Gly m 3 hergestellt und in drei Fragmente unterteilt. Sowohl das gesamte rGly m 3 als auch die einzelnen Fragmente wurden auf ihrer IgE-Reaktivität untersucht. In dieser Publikation wurde nur das komplette rGly m 3 von acht humanen Seren erkannt. Rihs et al. schlossen daraus, dass Gly m 3 ein diskontinuierliches Epitop besitzt [97].

Mittels der Phage Display Methode konnten 14 Epitope in Gly m 4 identifiziert werden. In der Datenbank sind 35 Epitope basierend auf zwei Publikationen hinterlegt. Ein 15 Aminosäuren langes Peptid wurde von Jahn-Schmid et al. [196] als T-Zell Epitop identifiziert, indem sie 10 Bet v 1 sensitive Patienten auf Kreuzreaktion untersuchten. Dies konnte mithilfe von Phage Display nicht als Epitop gefunden werden. Dies ist wenig überraschend, da das 15mer Epitop als T-Zell Epitop beschrieben wird und es nicht zusätzlich auch ein B-Zell Epitope sein muss. In einer Studie von Schulten et al. [24] wurden T-Zell Antigene aus Wiesen-Lieschgras Extrakt identifiziert, von denen einige weder IgE- noch IgG Epitope waren.

Andere Epitope, die in der Datenbank registriert sind, basieren auf einem Peptid Microarray mit überlappenden Fragmenten und Phage Display. Diese Ergebnisse wurden bereits von unserem Projektpartner publiziert [197].

In dieser Arbeit wurden auch zahlreiche noch nicht publizierte Epitope identifiziert. Mehr als 60 % der Epitope sind bisher noch nicht publiziert. Es konnten mehrere Epitope in Proteinen charakterisiert werden, die vorher noch nicht als immunogen beschrieben worden sind, zum Beispiel 18kD Seed Maturation Protein oder Basic 7S Globulin. Inwiefern diese Epitope auch eine allergische Reaktion verursachen, müsste noch durch Versuche mit Patienten nachgewiesen werden.

#### 4.5 Kreuzreaktionen

Häufig reagieren Allergiker auf mehr als ein Lebensmittel. Dies kann viele Gründe haben. So kann es innerhalb des Allergikers zur Multisensibilisierung, zur Spezies-spezifischen Co-Sensibilisierung oder zur Kreuzreaktion gekommen sein [198].

Mit der Phage Display Methode ist es möglich alle Serumantikörper gleichzeitig zu erfassen, deshalb wurde anhand der Beispiele Bet v 1 und Cor a 1 überprüft, ob potenzielle Epitope anderer Proteine außer Soja zu identifiziert werden können. Immer wieder wird zum Beispiel von einer Kreuzreaktion zwischen Bet v 1 und anderen PR10-Proteinen berichtet [63, 199].

Eine andere Möglichkeit Kreuzreaktionen zu untersuchen wäre die Vorselektion des Serums, indem beispielsweise alle Anti-Erdnuss-Antikörper vorher mit einem Erdnuss-spezifischen Peptid entfernt werden. Anschließend wird das Serum mit Sojapeptiden getestet. Alle positiv getesteten Peptide wären dann sojaspezifisch und nicht mit Erdnuss kreuzreaktiv. Zusätzlich müssen weitere Mitglieder aus den Proteinfamilien, die Vertreter in vielen Pflanzen besitzen, untersucht werden. Deren Epitope und Epitope der Sojabohne könnten dann als Peptide im Microarray getestet und so die kreuzreaktiven Epitope bestimmt werden. Zusätzlich könnte man die Seren auf weitere Allergenquellen untersuchen, um festzustellen, ob die Person multisensibilisiert sind. Die Untersuchung auf multiple Sensibilisierungen wäre aber sehr aufwendig und aufgrund der Verschiedenheit der Seren nicht immer aufschlussreich. Im anschließenden FoodAllergen Projekt werden allerdings Seren auf eine Vielzahl von Sensibilisierungen untersucht.

In dieser Arbeit wurden viele Epitope unter anderem aus Gly m 4 und Bet v 1, dem homologen PR10-Protein aus der Birke, sowie Gly m 5 und Gly m 6 identifiziert. Diese Proteine teilen sich Sequenz- oder Strukturmerkmale mit vielen anderen Proteinen. Diese Proteine wurden bisher 27 Familien pflanzlicher Lebensmittelallergene zugeordnet [39, 199]. Einige dieser Proteinfamilien sind sehr groß: die Cupin Familie, die Bet v 1 (PR10) Familie und die Profilin Familie. Alleine die Bet v 1 - und Profilin-Familien beinhalten mehr als 65 % aller beschriebenen Pflanzenallergene [199], was aber auch an der leichten Gewinnung dieser Proteine in reiner Form liegen könnte. Proteine, welche einer dieser Proteinfamilien zugeordnet werden können, werden auch Panallergene genannt und sind für zahlreiche Kreuzreaktionen zwischen inhalierten und essbaren Allergenquellen verantwortlich [200–202]. Aufgrund dieser gemeinsamen Sequenz- und Strukturmerkmale kann es zu zahlreichen Kreuzreaktionen kommen. So wurde für Bet v 1 ein Strukturepitop publiziert, welches in allen Proteinen dieser Familie zu finden ist [203]. Dies könnte die Kreuzreaktion der Bet v 1 Antikörper mit PR10 Proteine aus vielen anderen Pflanzen erklären. James et al. meinten allerdings, dass einzelne Antikörpermoleküle verschiedene Paratopkonformationen annehmen und deshalb auch an unverwandten Antigenen binden könnten [204]. Man könnte daraus aber auch schließen, dass sie schlicht unspezifisch sind, weil während der Maturierung das zweite Antigen nicht anwesend oder eine spezifische Selektion nicht möglich war.

Bet v 1 verursacht allergische Symptome in mehr als 90 % der Allergiker, die auf Birkenpollen allergisch reagieren [205]. Proteine dieser Familie sind 154-160 AS lang und besitzen eine hohe Sequenzähnlichkeit. Bet v 1 verwandte Proteine sind bekannt als Inhalationsallerge aus Baumpollen der Familie *Fagale*, die im Frühjahr blühen. Nach der primären Sensibilisierung mit dem Pollenallergen entwickeln die Mehrzahl der Patienten allergische Symptome nach der Aufnahme von Bet v 1-verwandten Proteinen. Diese sind in vielen Früchten (z.B. Apfel, Pfirsich, Kirsche), in Nüsse (z.B. Haselnüsse) und Gemüse (in Karotten, Sellerie und Petersilie) vorhanden [39]. Das Gly m 4

Protein der Sojabohne gehört auch zu den Bet v 1-verwandten Proteinen [3, 31], sowie das Protein Cor a 1, welches bis zu 80 % Sequenzähnlichkeit mit dem Bet v 1 Protein besitzt [206]. Abbildung 12 zeigt, dass sowohl Epitope des Proteins Gly m 4 als auch Epitope des Proteins Bet v 1 und Cor a 1 in den Soja sensitiven Seren gefunden wurden. Fast alle Seren, die Antikörper gegen Gly m 4 enthalten, scheinen auch Antikörper gegen Bet v 1 und Cor a 1 zu enthalten. 7 von 22 Seren mit Antikörpern gegen Gly m 4 und/oder Bet v 1 bzw. Cor a 1 enthalten Antikörper, die ähnliche Positionen innerhalb der verschiedenen Proteine erkennen. Allerdings wird anscheinend nicht dieselbe Position von allen Seren erkannt, es werden mehrere Positionen innerhalb der Proteine von einem Serum erkannt. Anhand des Erkennungsmusters kann man keine Hauptbinderegion innerhalb der Proteine erkennen. Es wurde allerdings berichtet, dass das Hauptepitop des Birkenpollenallergens um den Aminosäurerest 45, in der sogenannten P-Schleife, sei [207]. Der Bereich um den Aminosäurerest 45 ist in allen drei Proteinen (Gly m 4, Bet v 1 und Cor a 1) sehr konserviert, siehe Abbildung 12. Dieser Bereich wird von vielen Seren in allen drei Proteinen erkannt. In dieser Region scheint aber kein wichtiges Epitop des Sojabohnenproteins zu liegen. Ein sehr großer Bereich beider Proteine wird von vielen Seren erkannt. Alle mit der Phage Display Methode identifizierten Epitope des Bet v 1 Proteins wurden bereits als Teile längerer Epitope publiziert (siehe Tabelle 14). So kann man von bereits validierten Epitopen ausgehen. Anhand des Erkennungsmusters innerhalb der Seren kann man davon ausgehen, dass die Soja sensitiven Personen sowohl kreuzreaktive Antikörper, die alle drei Proteine erkennen, als auch Spezies-spezifische Antikörper, die entweder Gly m 4, Bet v 1 oder Cor a 1 erkennen, enthalten.

Von den (bisher bekannten) 16 allergenen Proteinen der Sojabohne gehören einige Proteine (Gly m 3, Gly m 4, Gly m 8, Gly m BD30K) zu Proteinfamilien, die als Panallergen bekannt sind [33, 201, 208, 209]. Gly m 5, Gly m 6 und Gly m BD28K gehören zu der Cupin-Familie, von der ebenfalls bekannt ist, dass sie mit Proteinen aus anderen Pflanzen, z.B. Erdnuss, Lupine, Kichererbse oder andere Leguminosen kreuzreagieren [95, 121, 210, 211]. Viele Proteine aus Lebensmittelallergenen gehören großen Proteinfamilien an. Da diese Proteine in vielen verschiedenen Pflanzen vorkommen, ist die Gefahr einer Kreuzreaktion allgegenwärtig. Obwohl Kreuzreaktionen oftmals nicht zu klinischen Reaktionen führen [39, 212], gibt es doch einige allergen wirkende Proteine, die eine klinisch relevante Kreuzreaktion auslösen, beispielsweise Bet v 1 spezifische - und Profilin spezifische IgEs [213]. In einer zukünftigen Allergiediagnostik sollten Kreuzreaktionen innerhalb verschiedener Lebensmittel berücksichtigt werden. Dies zeigen auch bereits durchgeführte Studien zum Thema. Melioli et al. [63] testeten 321 Patienten mit dem ImmunoCap ISAC. Hier wurde bei 52 % der Patienten eine Kreuzreaktion zwischen Panallergenen gefunden. In 37 % der Fälle wurden IgE-Antikörper gegen Bet v 1, das Hauptbirkenallergen, gefunden und 93 % von diesen Individuen hatten



IgE-Antikörper gegen andere PR10-Proteine. Nur 58 % der Patienten zeigten IgE-Antikörper gegen Lebensmittel-spezifische Proteine.

Ob Patienten gegen ein individuelles Allergen reagieren oder ob es nicht eine Kreuzreaktion auf ein Panallergen ist, kann derzeit noch nicht umfassend getestet werden. Mit dem speziellen Verfahren zur Identifizierung von Epitopen wird zum einen versucht die Aminosäuren der Epitope genau zu charakterisieren und zum anderen individuelle Erkennungsmuster für jeden Patienten zu erstellen, um so Informationen über die Epitope zu liefern. Dieses Wissen kann helfen zukünftig die Patienten besser zu charakterisieren, die unter einer Allergie leiden.

#### 4.6 Zusammenfassung und Ausblick

Mit der detaillierten Identifizierung von Epitopen der Sojabohne befasste sich das MAVO-Projekt *LowAllergen*. Die hier vorgestellte Methode wird das Verständnis der Allergie an sich und die Nahrungsmittelallergie-Diagnostik zu verbessern. Mithilfe der Peptid Phage Display Technik und dem NGS im Anschluss wurden mehr als 400 Epitope in 14 Sojaproteinen identifiziert. Für die Charakterisierung der Epitope und dessen spätere Validierung im Microarray wurden lediglich ca. 10 µl Serum benötigt. So konnte in dieser Arbeit gezeigt werden, dass das Peptid Phage Display in Kombination mit NGS und *in silico* Datenauswertung eine geeignete Methode ist, zahlreiche Epitope zu beschreiben. Ein Epitop- Microarray ist für die Detektion von individuellen Patientenmustern, sowie das Sensibilisierungsmuster für beispielsweise geografische Varianten bei den Patienten geeignet. Wie am Beispiel von Bet v 1 und Cor a 1 gezeigt werden konnte, ist die Identifizierung von kreuzreaktiven Antikörpern und Multisensibilisierung ein großer Vorteil. Auch kann der exakte Nachweis einer Sensibilisierung gegen bestimmte Proteine und/oder Epitope auf das verstärkte Risiko schwerer Reaktionen hinweisen [214]. Ein diagnostischer Test sollte Epitope von drei bis fünf verschiedenen Sojaproteinen, z.B. Gly m 5, Gly m 6 und Gly m BD30K enthalten, da unsere Ergebnisse zeigen, dass kein Individuum gegen alle Proteine sensibilisiert ist.

Inwieweit die identifizierten Epitope auch für die klinische Diagnostik relevant sind, werden zukünftige Analysen klären. So könnte man eine Kombination von Epitopen, die sich als zuverlässig für die Diagnostik herausgestellt haben, als kurze Peptide synthetisieren und im Prick-Test testen. Ein einzelnes Epitop wäre hier nicht ausreichend, da jeder Mensch ein individuelles Erkennungsmuster besitzt und es unwahrscheinlich ist, dass ein Epitop von der Mehrheit der Patienten erkannt wird. Durch die Kombination von Allergie-relevanten Peptiden kann aber die Mehrheit der Patienten besser als mit herkömmlichen Methoden erfasst werden. Bisher wird von den Klinikern allerdings nur der Provokationstest als sicherer Nachweis einer Allergie akzeptiert. Nach entsprechender weiterer Validierung könnte die Epitop-basierte Allergiediagnostik allerdings sowohl den Pricktest als auch den Provokationstest mit all seinen Risiken ersetzen.

## 5. Literaturverzeichnis

- 1 Zuberbier T, Edenharter G, Worm M, Ehlers I, Reimann S, Hantke T, Roehr CC, Bergmann KE, Niggemann B. Prevalence of adverse reactions to food in Germany - a population study. *Allergy* 2004;**59**(3):338–45. eng.
- 2 COMMISSION DIRECTIVE 2006/142/EC. European Parliament, 2006 [cited 2018 May 9]. 2 p.
- 3 Ballmer-Weber BK, Vieths S. Soy allergy in perspective. *Current opinion in allergy and clinical immunology* 2008;**8**(3):270–5. eng.
- 4 Belitz H-D, Grosch W, Schieberle P. Lehrbuch der Lebensmittelchemie. 6th ed. Berlin, Heidelberg: Springer, 2008. (Springer-Lehrbuch). ISBN: 9783540732013. Available from: <http://dx.doi.org/10.1007/978-3-540-73202-0>. ger.
- 5 FAO. Technical Consultation on Food Allergies: Food and Agriculture Organization of the United Nations, 1995. Available from: <https://books.google.de/books?id=EA2RYgEACAAJ>.
- 6 Caubet J-C, Sampson HA. Beyond skin testing: state of the art and new horizons in food allergy diagnostic testing. *Immunology and allergy clinics of North America* 2012;**32**(1):97–109. eng.
- 7 Duke WW. Soy bean as a possible important source of allergy. *Journal of Allergy* 1934;**5**(3):300–2.
- 8 Fu CJ, Jez JM, Kerley MS, Allee GL, Krishnan HB. Identification, characterization, epitope mapping, and three-dimensional modeling of the alpha-subunit of beta-conglycinin of soybean, a potential allergen for young pigs. *Journal of agricultural and food chemistry* 2007;**55**(10):4014–20. eng.
- 9 Sun X, Shan X, Yan Z, Zhang Y, Guan L. Prediction and characterization of the linear IgE epitopes for the major soybean allergen  $\beta$ -conglycinin using immunoinformatics tools. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association* 2013;**56**:254–60. eng.
- 10 Taliencio E, Loveless TM, Turano MJ, Kim SW. Identification of epitopes of the  $\beta$  subunit of soybean  $\beta$ -conglycinin that are antigenic in pigs, dogs, rabbits and fish. *Journal of the science of food and agriculture* 2014;**94**(11):2289–94. eng.
- 11 Cordle CT. Soy protein allergy: incidence and relative severity. *The Journal of nutrition* 2004;**134**(5):1213S-1219S. eng.
- 12 Pirquet C v. Allergy [Münchener Medizinische Wochenschrift/24 July 1906: Allergy by Clemens v. Pirquet, Vienna]. *Münchener medizinische Wochenschrift* 1978;**120**(14):474. ger.
- 13 Gell PGH, Coombs RRA. Clinical aspects of immunology. *Clinical Aspects of Immunology* 1963.
- 14 Schroeder HW, Cavacini L. Structure and function of immunoglobulins. *The Journal of allergy and clinical immunology* 2010;**125**(2 Suppl 2):S41-52. eng.
- 15 Williams A. The Immunoglobulin Superfamily: Domains For Cell Surface Recognition. *Annual Review of Immunology* 1988;**6**(1):381–405.
- 16 Kirkham PM, Schroeder HW. Antibody structure and the evolution of immunoglobulin V gene segments. *Seminars in immunology* 1994;**6**(6):347–60. eng.
- 17 Goldsby R, Kindt TJ, Osborne BA, Janis Kuby. Immunology. 5th ed. New York: Freeman, 2003. ISBN: 0716749475. eng.

- 18 Chatchatee P, Jarvinen K-M, Bardina L, Vila L, Beyer K, Sampson HA. Identification of IgE and IgG binding epitopes on beta- and kappa-casein in cow's milk allergic patients. *Clin Exp Allergy* 2001;**31**(8):1256–62.
- 19 Opinion of the Scientific Panel on Dietetic products, nutrition and allergies [NDA] on a request from the Commission relating to the evaluation of allergenic foods for labelling purposes. *EFS2* 2004;**2**(3):478.
- 20 Dall'antonia F, Pavkov-Keller T, Zangger K, Keller W. Structure of allergens and structure based epitope predictions. *Methods (San Diego, Calif.)* 2014;**66**(1):3–21. eng.
- 21 Ichikawa S, Takai T, Yashiki T, Takahashi S, Okumura K, Ogawa H, Kohda D, Hatanaka H. Lipopolysaccharide binding of the mite allergen Der f 2. *Genes to cells : devoted to molecular & cellular mechanisms* 2009;**14**(9):1055–65. eng.
- 22 Niemi M, Jänis J, Jylhä S, Kallio JM, Hakulinen N, Laukkanen M-L, Takkinen K, Rouvinen J. Characterization and crystallization of a recombinant IgE Fab fragment in complex with the bovine beta-lactoglobulin allergen. *Acta crystallographica. Section F, Structural biology and crystallization communications* 2008;**64**(Pt 1):25–8. eng.
- 23 Dall'antonia F, Gieras A, Devanaboyina SC, Valenta R, Keller W. Prediction of IgE-binding epitopes by means of allergen surface comparison and correlation to cross-reactivity. *The Journal of allergy and clinical immunology* 2011;**128**(4):872-879.e8. eng.
- 24 Schulten V, Greenbaum JA, Hauser M, McKinney DM, Sidney J, Kolla R, Lindestam Arlehamn CS, Oseroff C, Alam R, Broide DH, Ferreira F, Ferreira-Briza F, Grey HM, Sette A, Peters B. Previously undescribed grass pollen antigens are the major inducers of T helper 2 cytokine-producing T cells in allergic individuals. *Proceedings of the National Academy of Sciences of the United States of America* 2013;**110**(9):3459–64. eng.
- 25 Bøgh KL, Nielsen H, Eiwegger T, Madsen CB, Mills ENC, Rigby NM, Szépfalusi Z, Roggen EL. IgE versus IgG4 epitopes of the peanut allergen Ara h 1 in patients with severe allergy. *Molecular immunology* 2014;**58**(2):169–76. eng.
- 26 Christiansen A, Kringelum JV, Hansen CS, Bøgh KL, Sullivan E, Patel J, Rigby NM, Eiwegger T, Szépfalusi Z, Masi F de, Nielsen M, Lund O, Dufva M. High-throughput sequencing enhanced phage display enables the identification of patient-specific epitope motifs in serum. *Scientific reports* 2015;**5**:12913. eng.
- 27 Yang WW, Mejia EG de, Zheng H, Lee Y. Soybean allergens: presence, detection and methods for mitigation. *Soybean and Health* 2011.
- 28 Matsuo H, Yokooji T, Taogoshi T. Common food allergens and their IgE-binding epitopes. *Allergology International* 2015;**64**(4):332–43.
- 29 Bonds RS, Midoro-Horiuti T, Goldblum R. A structural basis for food allergy: the role of cross-reactivity. *Current opinion in allergy and clinical immunology* 2008;**8**(1):82–6. eng.
- 30 Bohle B, Radakovics A, Jahn-Schmid B, Hoffmann-Sommergruber K, Fischer GF, Ebner C. Bet v 1, the major birch pollen allergen, initiates sensitization to Api g 1, the major allergen in celery: evidence at the T cell level. *European journal of immunology* 2003;**33**(12):3303–10. eng.
- 31 Hauser M, Egger M, Wallner M, Wopfner N, Schmidt G, Ferreira F. Molecular Properties of Plant Food Allergens: A Current Classification into Protein Families. *TOIJ* 2008;**1**(1):1–12.
- 32 Vieths S, SCHEURER S, BALLMER-WEBER B. Current Understanding of Cross-Reactivity of Food Allergens and Pollen. *Annals of the New York Academy of Sciences* 2002;**964**(1):47–68.

- 33 Hauser M, Roulias A, Ferreira F, Egger M. Panallergens and their impact on the allergic patient. *Allergy, asthma, and clinical immunology : official journal of the Canadian Society of Allergy and Clinical Immunology* 2010;**6**(1):1. eng.
- 34 Dhyani A, Arora N, Jain VK, Sridhara S, Singh BP. Immunoglobulin E (IgE)-mediated cross-reactivity between mesquite pollen proteins and lima bean, an edible legume. *Clinical and experimental immunology* 2007;**149**(3):517–24. eng.
- 35 Herrera-Mozo I, Ferrer B, Luís Rodriguez-Sanchez J, Juarez C. Description of a novel panallergen of cross-reactivity between moulds and foods. *Immunological investigations* 2006;**35**(2):181–97. eng.
- 36 Hugues B, Didierlaurent A, Charpin D. Cross-reactivity between cypress pollen and peach: a report of seven cases. *Allergy* 2006;**61**(10):1241–3. eng.
- 37 Bublin M, Radauer C, Wilson IBH, Kraft D, Scheiner O, Breiteneder H, Hoffmann-Sommergruber K. Cross-reactive N-glycans of Api g 5, a high molecular weight glycoprotein allergen from celery, are required for immunoglobulin E binding and activation of effector cells from allergic patients. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2003;**17**(12):1697–9. eng.
- 38 Aalberse RC. Food allergens. *Environmental Toxicology and Pharmacology* 1997;**4**(1-2):55–60.
- 39 Hoffmann-Sommergruber K, Mills ENC. Food allergen protein families and their structural characteristics and application in component-resolved diagnosis: new data from the EuroPrevall project. *Analytical and bioanalytical chemistry* 2009;**395**(1):25–35. eng.
- 40 Breiteneder H, Radauer C. A classification of plant food allergens. *The Journal of allergy and clinical immunology* 2004;**113**(5):821-30; quiz 831. eng.
- 41 Pastorello EA, Pravettoni V, Calamari AM, Banfi E, Robino AM. New plant-origin food allergens. *Allergy* 2002;**57**(s72):106–10.
- 42 Scientific Opinion on the evaluation of allergenic foods and food ingredients for labelling purposes. *EFS2* 2014;**12**(11):54.
- 43 Breiteneder H, Ebner C. Molecular and biochemical classification of plant-derived food allergens. *The Journal of allergy and clinical immunology* 2000;**106**(1 Pt 1):27–36. eng.
- 44 Mills ENC, Jenkins JA, Alcocer MJC, Shewry PR. Structural, biological, and evolutionary relationships of plant food allergens sensitizing via the gastrointestinal tract. *Critical reviews in food science and nutrition* 2004;**44**(5):379–407. eng.
- 45 Witke W. The role of profilin complexes in cell motility and other cellular processes. *Trends in cell biology* 2004;**14**(8):461–9. eng.
- 46 Radauer C, Hoffmann-Sommergruber K. Profilins. In: Mills ENC, Shewry PR, eds. Plant food allergens. Oxford, Malden, MA: Blackwell Science, 2004, 105–24. ISBN: 9780470995174.
- 47 Saraste M, Sibbald PR, Wittinghofer A. The P-loop — a common motif in ATP- and GTP-binding proteins. *Trends in Biochemical Sciences* 1990;**15**(11):430–4.
- 48 Marković-Housley Z, Degano M, Lamba D, Roepenack-Lahaye E v., Clemens S, Susani M, Ferreira F, Scheiner O, Breiteneder H. Crystal Structure of a Hypoallergenic Isoform of the Major Birch Pollen Allergen Bet v 1 and its Likely Biological Function as a Plant Steroid Carrier. *Journal of Molecular Biology* 2003;**325**(1):123–33.
- 49 Jenkins JA, Breiteneder H, Mills ENC. Evolutionary distance from human homologs reflects allergenicity of animal food proteins. *The Journal of allergy and clinical immunology* 2007;**120**(6):1399–405. eng.

- 50 Brown JH, Cohen C. Regulation of Muscle Contraction by Tropomyosin and Troponin: How Structure Illuminates Function. In: Squire JM, ed. *Fibrous proteins: muscle and molecular motors*. Amsterdam: Elsevier, 2005, 121–59. (Advances in Protein Chemistry; vol. 71). ISBN: 9780120342716.
- 51 Yong Jeong K, Hong C-S, Yong T-S. Allergenic Tropomyosins and Their Cross-Reactivities. *PPL* 2006;**13**(8):835–45.
- 52 Lewit-Bentley A, Réty S. EF-hand calcium-binding proteins. *Current Opinion in Structural Biology* 2000;**10**(6):637–43.
- 53 Pearce N, Douwes J. The global epidemiology of asthma in children. *The International Journal of Tuberculosis and Lung Disease* 2006. (10):125–32.
- 54 Pawankar R, Canonica GW, Holgate ST, Lockey RF. World Allergy Organization (WAO) white book on allergy. United Kingdom: WAO, 2011. 210 p. ISBN: 978-0-615-46182-3.
- 55 Wood RA. The natural history of food allergy. *Pediatrics* 2003;**111**(6 Pt 3):1631–7. eng.
- 56 Sampson HA. Food allergy. Part 1: Immunopathogenesis and clinical disorders. *Journal of Allergy and Clinical Immunology* 1999;**103**(5):717–28.
- 57 Turnbull JL, Adams HN, Gorard DA. Review article: the diagnosis and management of food allergy and food intolerances. *Alimentary pharmacology & therapeutics* 2015;**41**(1):3–25. eng.
- 58 Boyce JA, Assa'ad A, Burks AW, Jones SM, Sampson HA, Wood RA, Plaut M, Cooper SF, Fenton MJ, Arshad SH, Bahna SL, Beck LA, Byrd-Bredbenner C, Camargo CA, Eichenfield L, Furuta GT, Hanifin JM, Jones C, Kraft M, Levy BD, Lieberman P, Luccioli S, McCall KM, Schneider LC, Simon RA, Simons FER, Teach SJ, Yawn BP, Schwaninger JM. Guidelines for the diagnosis and management of food allergy in the United States: summary of the NIAID-Sponsored Expert Panel report. *Journal of the American Academy of Dermatology* 2011;**64**(1):175–92. eng.
- 59 Burks AW, James JM, Hiegel A, Wilson G, Wheeler JG, Jones SM, Zuerlein N. Atopic dermatitis and food hypersensitivity reactions. *The Journal of Pediatrics* 1998;**132**(1):132–6.
- 60 Sherwood E, Boyd A. Food Allergy in Children and Young People. *InnovAiT* 2012;**5**(2):76–82.
- 61 AMLOT PL, KEMENY DM, ZACHARY C, PARKES P, LESSOF MH. Oral allergy syndrome (OAS): symptoms of IgE-mediated hypersensitivity to foods. *Clin Exp Allergy* 1987;**17**(1):33–42.
- 62 Sicherer SH, Sampson HA, Burks AW. Peanut and soy allergy: a clinical and therapeutic dilemma. *Allergy* 2000;**55**(6):515–21.
- 63 Melioli G, Bonifazi F, Bonini S, Maggi E, Mussap M, Passalacqua G, Rossi ER, Vacca A, Canonica GW. The ImmunoCAP ISAC molecular allergology approach in adult multi-sensitized Italian patients with respiratory symptoms. *Clinical biochemistry* 2011;**44**(12):1005–11. eng.
- 64 Skypala IJ, Calderon MA, Leeds AR, Emery P, Till SJ, Durham SR. Development and validation of a structured questionnaire for the diagnosis of oral allergy syndrome in subjects with seasonal allergic rhinitis during the UK birch pollen season. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2011;**41**(7):1001–11. eng.
- 65 Sampson HA, Muñoz-Furlong A, Campbell RL, Adkinson NF, Bock SA, Branum A, Brown SGA, Camargo CA, Cydulka R, Galli SJ, Gidudu J, Gruchalla RS, Harlor AD, Hepner DL, Lewis

- LM, Lieberman PL, Metcalfe DD, O'Connor R, Muraro A, Rudman A, Schmitt C, Scherrer D, Simons FER, Thomas S, Wood JP, Decker WW. Second symposium on the definition and management of anaphylaxis: summary report--Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. *The Journal of allergy and clinical immunology* 2006;**117**(2):391–7. eng.
- 66 Niggemann B, Reibel S, Roehr CC, Felger D, Ziegert M, Sommerfeld C, Wahn U. Predictors of positive food challenge outcome in non-IgE-mediated reactions to food in children with atopic dermatitis. *The Journal of allergy and clinical immunology* 2001;**108**(6):1053–8. eng.
- 67 Arbes SJ, Gergen PJ, Elliott L, Zeldin DC. Prevalences of positive skin test responses to 10 common allergens in the US population: results from the third National Health and Nutrition Examination Survey. *The Journal of allergy and clinical immunology* 2005;**116**(2):377–83. eng.
- 68 Niggemann B, Rolinck-Werninghaus C, Mehl A, Binder C, Ziegert M, Beyer K. Controlled oral food challenges in children--when indicated, when superfluous? *Allergy* 2005;**60**(7):865–70. eng.
- 69 Soares-Weiser K, Takwoingi Y, Panesar SS, Muraro A, Werfel T, Hoffmann-Sommergruber K, Roberts G, Halken S, Poulsen L, van Ree R, Vlieg-Boerstra BJ, Sheikh A. The diagnosis of food allergy: a systematic review and meta-analysis. *Allergy* 2014;**69**(1):76–86. eng.
- 70 Wood RA, Segall N, Ahlstedt S, Williams PB. Accuracy of IgE antibody laboratory results. *Annals of Allergy, Asthma & Immunology* 2007;**99**(1):34–41.
- 71 Wang J, Godbold JH, Sampson HA. Correlation of serum allergy (IgE) tests performed by different assay systems. *The Journal of allergy and clinical immunology* 2008;**121**(5):1219–24. eng.
- 72 Commins SP, Satinover SM, Hosen J, Mozena J, Borish L, Lewis BD, Woodfolk JA, Platts-Mills TAE. Delayed anaphylaxis, angioedema, or urticaria after consumption of red meat in patients with IgE antibodies specific for galactose-alpha-1,3-galactose. *The Journal of allergy and clinical immunology* 2009;**123**(2):426–33. eng.
- 73 Cox L, Williams B, Sicherer S, Oppenheimer J, Sher L, Hamilton R, Golden D. Pearls and pitfalls of allergy diagnostic testing: report from the American College of Allergy, Asthma and Immunology/American Academy of Allergy, Asthma and Immunology Specific IgE Test Task Force. *Annals of Allergy, Asthma & Immunology* 2008;**101**(6):580–92.
- 74 Fontaine C, Mayorga C, Bousquet PJ, Arnoux B, Torres M-J, Blanca M, Demoly P. Relevance of the determination of serum-specific IgE antibodies in the diagnosis of immediate beta-lactam allergy. *Allergy* 2007;**62**(1):47–52. eng.
- 75 Eigenmann PA, Sampson HA. Interpreting skin prick tests in the evaluation of food allergy in children. *Pediatr Allergy Immunol* 1998;**9**(4):186–91.
- 76 Sampson HA, Ho DG. Relationship between food-specific IgE concentrations and the risk of positive food challenges in children and adolescents☆☆☆★. *Journal of Allergy and Clinical Immunology* 1997;**100**(4):444–51.
- 77 Verstege A, Mehl A, Rolinck-Werninghaus C, Staden U, Nocon M, Beyer K, Niggemann B. The predictive value of the skin prick test weal size for the outcome of oral food challenges. *Clin Exp Allergy* 2005;**35**(9):1220–6. eng.
- 78 Costa AJF, Sarinho ESC, Motta MEFA, Gomes PN, Oliveira Melo SM de, da Silva GAP. Allergy to cow's milk proteins: what contribution does hypersensitivity in skin tests have to

- this diagnosis? *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* 2011;**22**(1 Pt 2):e133-8. eng.
- 79 Hamilton RG. Clinical laboratory assessment of immediate-type hypersensitivity. *The Journal of allergy and clinical immunology* 2010;**125**(2 Suppl 2):S284-96. eng.
- 80 Burks AW, Jones SM, Wood RA, Fleischer DM, Sicherer SH, Lindblad RW, Stablein D, Henning AK, Vickery BP, Liu AH, Scurlock AM, Shreffler WG, Plaut M, Sampson HA. Oral immunotherapy for treatment of egg allergy in children. *The New England journal of medicine* 2012;**367**(3):233-43. eng.
- 81 Larenas-Linnemann D. Sublingual immunotherapy in children: complete and updated review supporting evidence of effect. *Current opinion in allergy and clinical immunology* 2009;**9**(2):168-76. eng.
- 82 Boissieu D de, Dupont C. Sublingual immunotherapy for cow's milk protein allergy: a preliminary report. *Allergy* 2006;**61**(10):1238-9. eng.
- 83 Nadeau KC, Kohli A, Iyengar S, DeKruyff RH, Umetsu DT. Oral immunotherapy and anti-IgE antibody-adjunctive treatment for food allergy. *Immunology and allergy clinics of North America* 2012;**32**(1):111-33. eng.
- 84 Anagnostou K, Islam S, King Y, Foley L, Pasea L, Bond S, Palmer C, Deighton J, Ewan P, Clark A. Assessing the efficacy of oral immunotherapy for the desensitisation of peanut allergy in children (STOP II): a phase 2 randomised controlled trial. *The Lancet* 2014;**383**(9925):1297-304.
- 85 Nadeau KC, Schneider LC, Hoyte L, Borrás I, Umetsu DT. Rapid oral desensitization in combination with omalizumab therapy in patients with cow's milk allergy. *The Journal of allergy and clinical immunology* 2011;**127**(6):1622-4. eng.
- 86 Lemon-Mulé H, Sampson HA, Sicherer SH, Shreffler WG, Noone S, Nowak-Wegrzyn A. Immunologic changes in children with egg allergy ingesting extensively heated egg. *The Journal of allergy and clinical immunology* 2008;**122**(5):977-983.e1. eng.
- 87 Codina R, Arduoso L, Lockey RF, Crisci C, Medina I. Allergenicity of varieties of soybean. *Allergy* 2003;**58**(12):1293-8. eng.
- 88 Schäfer T, Böhler E, Ruhdorfer S, Weigl L, Wessner D, Heinrich J, Filipiak B, Wichmann H-E, Ring J. Epidemiology of food allergy/food intolerance in adults: associations with other manifestations of atopy. *Allergy* 2001;**56**(12):1172-9.
- 89 Bindslev-Jensen C, Briggs D, Osterballe M. Can we determine a threshold level for allergenic foods by statistical analysis of published data in the literature? *Allergy* 2002;**57**(8):741-6.
- 90 Matricardi PM, Bockelbrink A, Beyer K, Keil T, Niggemann B, Grüber C, Wahn U, Lau S. Primary versus secondary immunoglobulin E sensitization to soy and wheat in the Multi-Centre Allergy Study cohort. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2008;**38**(3):493-500. eng.
- 91 Kattan JD, Cocco RR, Järvinen KM. Milk and soy allergy. *Pediatric clinics of North America* 2011;**58**(2):407-26. x. eng.
- 92 Savage JH, Kaeding AJ, Matsui EC, Wood RA. The natural history of soy allergy. *The Journal of allergy and clinical immunology* 2010;**125**(3):683-6. eng.
- 93 Foucard T, Malmheden Yman I. A study on severe food reactions in Sweden - is soy protein an underestimated cause of food anaphylaxis? *Allergy* 1999;**54**(3):261-5.

- 94 Moroz LA, Yang WH. Kunitz soybean trypsin inhibitor: a specific allergen in food anaphylaxis. *The New England journal of medicine* 1980;**302**(20):1126–8. eng.
- 95 Xiang P, Haas EJ, Zeece MG, Markwell J, Sarath G. C-Terminal 23 kDa polypeptide of soybean Gly m Bd 28 K is a potential allergen. *Planta* 2004;**220**(1):56–63. eng.
- 96 WHO/IUIS. WHO/IUIS Allergen Nomenclature Home Page [Internet] [cited 2019 Apr 17]. Available from: <http://www.allergen.org/index.php>.
- 97 Rihs H-P, Chen Z, Ruëff F, Petersen A, Rozynek P, Heimann H, Baur X. IgE binding of the recombinant allergen soybean profilin (rGly m 3) is mediated by conformational epitopes. *Journal of Allergy and Clinical Immunology* 1999;**104**(6):1293–301.
- 98 Thanh VH, Shibasaki K. Major proteins of soybean seeds. Subunit structure of .beta.-conglycinin. *Journal of agricultural and food chemistry* 1978;**26**(3):692–5.
- 99 Staswick PE, Hermodson MA, Nielsen NC. Identification of the acidic and basic subunit complexes of glycinin. *The Journal of biological chemistry* 1981;**256**(16):8752–5. eng.
- 100 Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB. Characterization of the glycinin gene family in soybean. *The Plant cell* 1989;**1**(3):313–28. eng.
- 101 Schiller D, Hellmuth M, Gubesch M, BALLMER-WEBER B, Bindslev-Jensen C, Niggemann B, Scibillia J, Hanschmann K-M, Kühne Y, Reuter A, Vieths S, Wangorsch A, Holzhauser T. Soybean allergy: IgE epitopes of glycinin (Gly m 6), an important soybean allergen. *Clin Transl Allergy* 2014;**4**(Suppl 2):O2.
- 102 OGAWA T, BANDO N, TSUJI H, OKAJIMA H, NISHIKAWA K, SASAOKA K. Investigation of the IgE-Binding Proteins in Soybeans by Immunoblotting with the Sera of the Soybean-Sensitive Patients with Atopic Dermatitis. *Journal of Nutritional Science and Vitaminology, J Nutr Sci Vitaminol* 1991;**37**(6):555–65.
- 103 Djurtoft R, Pedersen HS, Aabin B, Barkholt V. Studies of Food Allergens: Soybean and Egg Proteins. In: Friedman M, ed. Nutritional and toxicological consequences of food processing. [Place of publication not identified]: Springer-Verlag New York, 2013, 281–93. (Advances in Experimental Medicine and Biology; vol. 289). ISBN: 978-1-4899-2628-9.
- 104 Baur X, Pau M, Czuppon A, Fruhmant G. Characterization of soybean allergens causing sensitization of occupationally exposed bakers. *Allergy* 1996;**51**(5):326–30.
- 105 Holzhauser T, Wackermann O, Ballmer-Weber BK, Bindslev-Jensen C, Scibilia J, Perono-Garoffo L, Utsumi S, Poulsen LK, Vieths S. Soybean (*Glycine max*) allergy in Europe: Gly m 5 (beta-conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for severe allergic reactions to soy. *The Journal of allergy and clinical immunology* 2009;**123**(2):452–8. eng.
- 106 Fukutomi Y, Sjölander S, Nakazawa T, Borres MP, Ishii T, Nakayama S, Tanaka A, Taniguchi M, Saito A, Yasueda H, Nakamura H, Akiyama K. Clinical relevance of IgE to recombinant Gly m 4 in the diagnosis of adult soybean allergy. *The Journal of allergy and clinical immunology* 2012;**129**(3):860-863.e3. eng.
- 107 Klemans RJB, Knol EF, Michelsen-Huisman A, Pasmans SGMA, Kruijf-Broekman W de, Bruijnzeel-Koomen CAFM, van Hoffen E, Knulst AC. Components in soy allergy diagnostics: Gly m 2S albumin has the best diagnostic value in adults. *Allergy* 2013;**68**(11):1396–402. eng.



- 108 Codina R, Lockey RF, Fernández-Caldas E, Rama R. Purification and characterization of a soybean hull allergen responsible for the Barcelona asthma outbreaks. II. Purification and sequencing of the Gly m 2 allergen. *Clin Exp Allergy* 1997;**27**(4):424–30. eng.
- 109 Kosma P, Sjölander S, Landgren E, Borres MP, Hedlin G. Severe reactions after the intake of soy drink in birch pollen-allergic children sensitized to Gly m 4. *Acta Paediatrica* 2011;**100**(2):305–6.
- 110 Hong X, Caruso D, Kumar R, Liu R, Liu X, Wang G, Pongracic JA, Wang X. IgE, but not IgG4, antibodies to Ara h 2 distinguish peanut allergy from asymptomatic peanut sensitization. *Allergy* 2012;**109**(Suppl 2):n/a-n/a.
- 111 Csáky I, Fekete S. Soybean: feed quality and safety. Part 1: Biologically active components. A review. *Acta Veterinaria Hungarica* 2004;**52**(3):299–313.
- 112 Kern K, Havenith H, Delaroque N, Rautenberger P, Lehmann J, Fischer M, Spiegel H, Schillberg S, Ehrentreich-Foerster E, Aurich S, Treudler R, Szardenings M. The immunome of soy bean allergy: Comprehensive identification and characterization of epitopes. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2019;**49**(2):239–51. eng.
- 113 Matricardi PM, Kleine-Tebbe J, Hoffmann HJ, Valenta R, Hilger C, Hofmaier S, Aalberse RC, Agache I, Asero R, Ballmer-Weber B, Barber D, Beyer K, Biedermann T, Bilò MB, Blank S, Bohle B, Bosshard PP, Breiteneder H, Brough HA, Caraballo L, Caubet JC, Cramer R, Davies JM, Douladiris N, Ebisawa M, Eigenmann PA, Fernandez-Rivas M, Ferreira F, Gadermaier G, Glatz M, Hamilton RG, Hawranek T, Hellings P, Hoffmann-Sommergruber K, Jakob T, Jappe U, Jutel M, Kamath SD, Knol EF, Korosec P, Kuehn A, Lack G, Lopata AL, Mäkelä M, Morisset M, Niederberger V, Nowak-Węgrzyn AH, Papadopoulos NG, Pastorello EA, Pauli G, Platts-Mills T, Posa D, Poulsen LK, Raulf M, Sastre J, Scala E, Schmid JM, Schmid-Grendelmeier P, van Hage M, van Ree R, Vieths S, Weber R, Wickman M, Muraro A, Ollert M. EAACI Molecular Allergology User's Guide. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* 2016;**27** Suppl 23:1–250. eng.
- 114 Treudler R, Werner M, Thierry J, Kramer S, Gebhardt C, Averbek M, Suess A, Simon JC. High risk of immediate-type reactions to soy drinks in 50 patients with birch pollinosis. *Journal of investigational allergology & clinical immunology* 2008;**18**(6):483–4. eng.
- 115 Beardslee TA, Zeece MG, Sarath G, Markwell JP. Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3. *International archives of allergy and immunology* 2000;**123**(4):299–307. eng.
- 116 Sicherer SH. Clinical implications of cross-reactive food allergens. *The Journal of allergy and clinical immunology* 2001;**108**(6):881–90. eng.
- 117 Burks W, Sampson HA, Bannon GA. Peanut allergens. *Allergy* 1998;**53**(8):725–30.
- 118 Green TD, LaBelle VS, Steele PH, Kim EH, Lee LA, Mankad VS, Williams LW, Anstrom KJ, Burks AW. Clinical characteristics of peanut-allergic children: recent changes. *Pediatrics* 2007;**120**(6):1304–10. eng.
- 119 Aalberse RC. Assessment of allergen cross-reactivity. *Clinical and molecular allergy : CMA* 2007;**5**:2. eng.
- 120 Ivanciuc O, Garcia T, Torres M, Schein CH, Braun W. Characteristic motifs for families of allergenic proteins. *Molecular immunology* 2009;**46**(4):559–68. eng.

- 121 Mills ENC, Jenkins J, Marigheto N, Belton PS, Gunning AP, Morris VJ. Allergens of the cupin superfamily. *Biochim. Soc. Trans.* 2002;**30**(6):925–9.
- 122 Smith G. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 1985;**228**(4705):1315–7.
- 123 Kehoe JW, Kay BK. Filamentous Phage Display in the New Millennium. *Chem. Rev.* 2005;**105**(11):4056–72.
- 124 Gray BP, Brown KC. Combinatorial peptide libraries: mining for cell-binding peptides. *Chemical reviews* 2014;**114**(2):1020–81. eng.
- 125 Vodnik M, Molek P, Strukelj B, Lunder M. Peptides binding to the hunger hormone ghrelin. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et métabolisme* 2013;**45**(5):372–7. eng.
- 126 Devlin J, Panganiban L, Devlin P. Random peptide libraries: a source of specific protein binding molecules. *Science* 1990;**249**(4967):404–6.
- 127 Pali-Schöll I, Jensen-Jarolim E. Biopanning for the characterization of allergen mimotopes. *Methods in molecular medicine* 2008;**138**:271–83. eng.
- 128 Szardenings M. Phage Display of Random Peptide Libraries: Applications, Limits, and Potential. *Journal of Receptors and Signal Transduction* 2003;**23**(4):307–49.
- 129 Smith J, Kontermann RE, Embleton J, Kumar S. Antibody phage display technologies with special reference to angiogenesis. *The FASEB Journal* 2005;**19**(3):331–41.
- 130 Houghten RA, Pinilla C, Blondelle SE, Appel JR, Dooley CT, Cuervo JH. Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* 1991;**354**(6348):84–6.
- 131 Wells JA, Lowman HB. Rapid evolution of peptide and protein binding properties in vitro. *Current Opinion in Biotechnology* 1992;**3**(4):355–62.
- 132 Rowley MJ, O'Connor K, Wijeyewickrema L. Phage display for epitope determination: a paradigm for identifying receptor-ligand interactions. *Biotechnology annual review* 2004;**10**:151–88. eng.
- 133 Felici F, Galfrè G, Luzzago A, Monaci P, Nicosia A, Cortese R. Phage-displayed peptides as tools for characterization of human sera. *Methods in enzymology* 1996;**267**:116–29. eng.
- 134 Hust M, Frenzel A, Meyer T, Schirrmann T, Dübel S. Construction of human naive antibody gene libraries. *Methods in molecular biology (Clifton, N.J.)* 2012;**907**:85–107. eng.
- 135 Parhami-Seren B, Viswanathan M, Margolies MN. Selection of high affinity p-azophenyarsonate Fabs from heavy-chain CDR2 insertion libraries. *Journal of Immunological Methods* 2002;**259**(1-2):43–53.
- 136 Hertveldt K, Robben J, Volckaert G. Whole genome phage display selects for proline-rich Boi polypeptides against Bem1p. *Biotechnology letters* 2006;**28**(16):1233–9. eng.
- 137 Cramer R, Suter M. Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. *Gene* 1993;**137**(1):69–75.
- 138 Noren KA, Noren CJ. Construction of high-complexity combinatorial phage display peptide libraries. *Methods (San Diego, Calif.)* 2001;**23**(2):169–78. eng.

- 139 Brunet E. A novel strategy for the functional cloning of enzymes using filamentous phage display: the case of nucleotidyl transferases. *Nucleic acids research* 2002;**30**(9):40e-40.
- 140 Pande J, Szewczyk MM, Grover AK. Phage display: Concept, innovations, applications and future. *Biotechnology Advances* 2010;**28**(6):849–58.
- 141 Messing J. Cloning single-stranded DNA. *Molecular biotechnology* 1996;**5**(1):39–47. eng.
- 142 Lowman HB; Clackson T, eds. Phage display, A practical approach. Oxford, New York: Oxford University Press, 2004. 332 p. (Practical approach series; vol. no. 266). ISBN: 9780199638741. Available from: <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=264953>. eng.
- 143 Bábíčková J, Tóthová Ľ, Boor P, Celec P. In vivo phage display--a discovery tool in molecular biomedicine. *Biotechnology Advances* 2013;**31**(8):1247–59. eng.
- 144 't Hoen PAC, Jirka SMG, Broeke BR ten, Schultes EA, Aguilera B, Pang KH, Heemskerk H, Aartsma-Rus A, van Ommen GJ, den Dunnen JT. Phage display screening without repetitious selection rounds. *Analytical Biochemistry* 2012;**421**(2):622–31.
- 145 Aghebati-Maleki L, Bakhshinejad B, Baradaran B, Motallebnezhad M, Aghebati-Maleki A, Nickho H, Yousefi M, Majidi J. Phage display as a promising approach for vaccine development. *Journal of biomedical science* 2016;**23**(1):66. eng.
- 146 La Cruz VF de, Lal AA, McCutchan TF. Immunogenicity and epitope mapping of foreign sequences via genetically engineered filamentous phage. *The Journal of biological chemistry* 1988;**263**(9):4318–22. eng.
- 147 Casey JL, Coley AM, Parisi K, Foley M. Peptide mimics selected from immune sera using phage display technology can replace native antigens in the diagnosis of Epstein-Barr virus infection. *Protein engineering, design & selection : PEDS* 2009;**22**(2):85–91. eng.
- 148 Folgori A, Tafi R, Meola A, Felici F, Galfré G, Cortese R, Monaci P, Nicosia A. A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *The EMBO Journal* 1994;**13**(9):2236–43.
- 149 Sambrook J, Russell DW. Molecular cloning, A laboratory manual. 3rd ed. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press, 2001. 3 p. ISBN: 978-0-879-69577-4. eng.
- 150 Vieira J, Messing J. [1] Production of single-stranded plasmid DNA. In: Wu R, ed. Recombinant DNA. San Diego, Calif.: Academic Press, 1987, 3–11. (Methods in Enzymology; vol. 153). ISBN: 9780121820541.
- 151 Szardenings M, Delaroque N, inventors. Phage Display Library.
- 152 Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 1976;**72**(1-2):248–54.
- 153 Russell B, Levitt J, Polson A. Particle-size classification by electrophoresis through graded agarose gels. *Biochimica et Biophysica Acta (BBA) - Specialized Section on Biophysical Subjects* 1964;**79**(3):622–4.
- 154 Metzker ML. Sequencing technologies - the next generation. *Nature reviews. Genetics* 2010;**11**(1):31–46. eng.

- 155 NCBI [Internet] [cited 2018 May 9]. Available from:  
<https://www.ncbi.nlm.nih.gov/gene>.
- 156 Kulpa K. Screening for Allergy Markers in Serum using Peptide Phage Display. 2011 Jun 27.
- 157 Heinze K. Versuche zur Identifizierung von Serummarkern in Allergiker-Seren mittels Peptid-Phagen-Display. 2012.
- 158 Collins J, Horn N, Wadenbäck J, Szardenings M. Cosmix-plexing®: a novel recombinatorial approach for evolutionary selection from combinatorial libraries. *Reviews in Molecular Biotechnology* 2001;**74**(4):317–38.
- 159 Remaut E, Tsao H, Fiers W. Improved plasmid vectors with a thermoinducible expression and temperature-regulated runaway replication. *Gene* 1983;**22**(1):103–13.
- 160 Szardenings M, Collins J. A phasmid optimised for protein design projects: pMAMPF. *Gene* 1990;**94**(1):1–7.
- 161 Meester P de, Brick P, Lloyd LF, Blow DM, Onesti S. SOYBEAN TRYPSIN INHIBITOR, 1998.
- 162 Huang J, Ru B, Zhu P, Nie F, Yang J, Wang X, Dai P, Lin H, Guo F-B, Rao N. MimoDB 2.0: a mimotope database and beyond. *Nucleic acids research* 2012;**40**(Database issue):D271-7. eng.
- 163 Aalberse RC, Cramer R. IgE-binding epitopes: a reappraisal. *Allergy* 2011;**66**(10):1261–74.
- 164 Derda R, Tang S, Li SC, Ng S, Matochko W, Jafari M. Diversity of Phage-Displayed Libraries of Peptides during Panning and Amplification. *Molecules* 2011;**16**(2):1776–803.
- 165 Kini RM, Evans HJ. A Hypothetical Structural Role for Proline Residues in the Flanking Segments of Protein-Protein Interaction Sites. *Biochemical and Biophysical Research Communications* 1995;**212**(3):1115–24.
- 166 Rodi DJ, Soares AS, Makowski L. Quantitative Assessment of Peptide Sequence Diversity in M13 Combinatorial Peptide Phage Display Libraries. *Journal of Molecular Biology* 2002;**322**(5):1039–52.
- 167 Eckert KA, Kunkel TA. DNA polymerase fidelity and the polymerase chain reaction. *Genome Research* 1991;**1**(1):17–24.
- 168 Kong H, Kucera RB, Jack WE. Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *The Journal of biological chemistry* 1993;**268**(3):1965–75. eng.
- 169 Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology* 2012;**2012**:251364. eng.
- 170 AWAZUHARA H, KAWAI H, MARUCHI N. Major allergens in soybean and clinical significance of IgG4 antibodies investigated by IgE- and IgG4-immunoblotting with sera from soybean-sensitive patients. *Clinical & Experimental Allergy* 1997;**27**(3):325–32.
- 171 Vickery BP, Lin J, Kulis M, Fu Z, Steele PH, Jones SM, Scurlock AM, Gimenez G, Bardina L, Sampson HA, Burks AW. Peanut oral immunotherapy modifies IgE and IgG4 responses to major peanut allergens. *Journal of Allergy and Clinical Immunology* 2013;**131**(1):128-134.e3.

- 172 Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, Gershoni JM, Pupko T. Computational characterization of B-cell epitopes. *Molecular immunology* 2008;**45**(12):3477–89.
- 173 Havenith H. Expression und Charakterisierung von Sojaallergen-Varianten für Lebensmittelanalytik, Allergiediagnostik und Therapeutik. 2017.
- 174 Zhu H, Snyder M. Protein chip technology. *Current Opinion in Chemical Biology* 2003;**7**(1):55–63.
- 175 Barbulovic-Nad I, Lucente M, Sun Y, Zhang M, Wheeler AR, Busmann M. Bio-microarray fabrication techniques--a review. *Critical reviews in biotechnology* 2006;**26**(4):237–59. eng.
- 176 Noble JE. Quantification of protein concentration using UV absorbance and Coomassie dyes. *Methods in enzymology* 2014;**536**:17–26. eng.
- 177 Ock J, Li W. Fabrication of a three-dimensional tissue model microarray using laser foaming of a gas-impregnated biodegradable polymer. *Biofabrication* 2014;**6**(2):24110. eng.
- 178 Lin J, Bardina L, Shreffler WG, Andrae DA, Ge Y, Wang J, Bruni FM, Fu Z, Han Y, Sampson HA. Development of a novel peptide microarray for large-scale epitope mapping of food allergens. *Journal of Allergy and Clinical Immunology* 2009;**124**(2):315–322.e3.
- 179 Taliercio E, Kim SW. Epitopes from two soybean glycinin subunits are antigenic in pigs. *J. Sci. Food Agric.* 2013;**93**(12):2927–32.
- 180 Martínez-Botas J, Cerecedo I, Zamora J, Vlaicu C, Dieguez MC, Gómez-Coronado D, Dios V de, Terrados S, La Hoz B de. Mapping of the IgE and IgG4 Sequential Epitopes of Ovomuroid with a Peptide Microarray Immunoassay. *International archives of allergy and immunology* 2013;**161**(1):11–20.
- 181 Vereda A, Andrae DA, Lin J, Shreffler WG, Ibañez MD, Cuesta-Herranz J, Bardina L, Sampson HA. Identification of IgE sequential epitopes of lentil (Len c 1) by means of peptide microarray immunoassay. *Journal of Allergy and Clinical Immunology* 2010;**126**(3):596–601.e1.
- 182 <https://www.facebook.com/pages/Pepscan-Therapeutics-BV/177773366121185>. Epitope Mapping Services | Pepscan [Internet] [cited 2018 Oct 30]. Available from: <https://www.pepscan.com/epitope-mapping-services>.
- 183 Kleine-Tebbe J, Wangorsch A, Vogel L, Crowell DN, Hausteiner U-F, Vieths S. Severe oral allergy syndrome and anaphylactic reactions caused by a Bet v 1– related PR-10 protein in soybean, SAM22. *Journal of Allergy and Clinical Immunology* 2002;**110**(5):797–804.
- 184 Helm RM, COCKRELL G, CONNAUGHTON C, Sampson HA, Bannon GA, Beilinson V, Nielsen NC, Burks AW. A soybean G2 glycinin allergen. 2. Epitope mapping and three-dimensional modeling. *International archives of allergy and immunology* 2000;**123**(3):213–9. eng.
- 185 Morrison KL, Weiss GA. Combinatorial alanine-scanning. *Current Opinion in Chemical Biology* 2001;**5**(3):302–7.
- 186 OGAWA T, TSUJI H, BANDO N, Kitamura K, Zhu Y-L, Hirano H, NISHIKAWA K. Identification of the Soybean Allergenic Protein, Gly m Bd 30K, with the Soybean Seed 34-kDa Oil-body-associated Protein. *Bioscience, Biotechnology, and Biochemistry* 1993;**57**(6):1030–3.

- 187 OGAWA T, SAMOTO M, TAKAHASHI K. Soybean Allergens and Hypoallergenic Soybean Products. *Journal of Nutritional Science and Vitaminology, J Nutr Sci Vitaminol* 2000;**46**(6):271–9.
- 188 Maruyama N, Adachi M, Takahashi K, Yagasaki K, Kohno M, Takenaka Y, Okuda E, Nakagawa S, Mikami B, Utsumi S. CRYSTAL STRUCTURES OF RECOMBINANT AND NATIVE SOYBEAN BETA-CONGLYCININ BETA HOMOTRIMERS, 2002.
- 189 Xiang P, Beardslee TA, Zeece MG, Markwell J, Sarath G. Identification and analysis of a conserved immunoglobulin E-binding epitope in soybean G1a and G2a and peanut Ara h 3 glycinins. *Archives of biochemistry and biophysics* 2002;**408**(1):51–7. eng.
- 190 Saeed H, Gagnon C, Cober E, Gleddie S. Using patient serum to epitope map soybean glycinins reveals common epitopes shared with many legumes and tree nuts. *Molecular immunology* 2016;**70**:125–33.
- 191 Saeed H, Gagnon C, Cober E, Gleddie S. Data on the epitope mapping of soybean A2 and A3 glycinin. *Data in brief* 2016;**8**:123–31. eng.
- 192 Helm RM, COCKRELL G, CONNAUGHTON C, West CM, Herman E, Sampson HA, Bannon GA, Burks AW. Mutational analysis of the IgE-binding epitopes of P34/Gly m Bd 30K. *The Journal of allergy and clinical immunology* 2000;**105**(2 Pt 1):378–84. eng.
- 193 HELM RM, Cockrell G, Herman E, Burks AW, Sampson HA, Bannon GA. Cellular and Molecular Characterization of a Major Soybean Allergen. *International archives of allergy and immunology* 1998;**117**(1):29–37.
- 194 Hiroshi H, Akio O, Noriko B, Hideaki T, Tadashi O. Epitope Analysis of Soybean Major Allergen Gly m Bd 30K Recognized by the Mouse Monoclonal Antibody Using Overlapping Peptides. *Bioscience, Biotechnology, and Biochemistry* 1996;**60**(7):1181–2.
- 195 Han Y, Lin J, Bardina L, Grishina G, Lee C, Seo W, Sampson H. What Characteristics Confer Proteins the Ability to Induce Allergic Responses? IgE Epitope Mapping and Comparison of the Structure of Soybean 2S Albumins and Ara h 2. *Molecules* 2016;**21**(5):622.
- 196 Jahn-Schmid B, Radakovics A, Lüttkopf D, SCHEURER S, Vieths S, Ebner C, Bohle B. Bet v 1142-156 is the dominant T-cell epitope of the major birch pollen allergen and important for cross-reactivity with Bet v 1-related food allergens. *Journal of Allergy and Clinical Immunology* 2005;**116**(1):213–9.
- 197 Havenith H, Kern K, Rautenberger P, Spiegel H, Szardenings M, Ueberham E, Lehmann J, Buntru M, Vogel S, Treudler R, Fischer R, Schillberg S. Combination of two epitope identification techniques enables the rational design of soy allergen Gly m 4 mutants. *Biotechnol. J.* 2017;**12**(2):1600441.
- 198 Weber RW. Patterns of pollen cross-allergenicity. *Journal of Allergy and Clinical Immunology* 2003;**112**(2):229–39.
- 199 Jenkins JA, Griffiths-Jones S, Shewry PR, Breiteneder H, Mills EC. 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**(1):163–70.
- 200 Canonica GW, Ansoategui IJ, Pawankar R, Schmid-Grendelmeier P, van Hage M, Baena-Cagnani CE, Melioli G, Nunes C, Passalacqua G, Rosenwasser L, Sampson H, Sastre J, Bousquet J, Zuberbier T. A WAO - ARIA - GA<sup>2</sup>LEN consensus document on molecular-based allergy diagnostics. *The World Allergy Organization journal* 2013;**6**(1):17. eng.

- 201 Valenta R. Profilins constitute a novel family of functional plant pan-allergens. *Journal of Experimental Medicine* 1992;**175**(2):377–85.
- 202 Valenta R, Duchene M, Pettenburger K, Sillaber C, Valent P, Bettelheim P, Breitenbach M, Rumpold H, Kraft D, Scheiner O. Identification of profilin as a novel pollen allergen; IgE autoreactivity in sensitized individuals. *Science* 1991;**253**(5019):557–60.
- 203 Aalberse RC, Akkerdaas J, van Ree R. Cross-reactivity of IgE antibodies to allergens. *Allergy* 2001;**56**(6):478–90.
- 204 James LC. Antibody Multispecificity Mediated by Conformational Diversity. *Science* 2003;**299**(5611):1362–7.
- 205 JAROLIM E, Rumpold H, ENDLER AT, EBNER H, Breitenbach M, Scheiner O, Kraft D. IgE and IgG antibodies of patients with allergy to birch pollen as tools to define the allergen profile of *Betula verrucosa*\*. *Allergy* 1989;**44**(6):385–95.
- 206 Breiteneder H, Ferreira F, Hoffmann-Sommergruber K, Ebner C, BREITENBACH M, RUMPOLD H, Kraft D, Scheiner O. Four recombinant isoforms of Cor a I, the major allergen of hazel pollen, show different IgE-binding properties. *Eur J Biochem* 1993;**212**(2):355–62.
- 207 Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, van Joost Neerven RJ, Schou C, Løwenstein H, Spangfort MD. X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nature structural biology* 1996;**3**(12):1040–5. eng.
- 208 Moreno FJ, Clemente A. 2S Albumin Storage Proteins: What Makes them Food Allergens? *The open biochemistry journal* 2008;**2**:16–28. eng.
- 209 Kalinski A, Weisemann JM, Matthews BF, Herman EM. Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. *The Journal of biological chemistry* 1990;**265**(23):13843–8. eng.
- 210 Eigenmann PA, Burks AW, Bannon GA, Sampson HA. Identification of unique peanut and soy allergens in sera adsorbed with cross-reacting antibodies. *Journal of Allergy and Clinical Immunology* 1996;**98**(5 Pt 1):969–78. eng.
- 211 BERNHISELBROADBENT J, SAMPSON H. Cross-allergenicity in the legume botanical family in children with food hypersensitivity. *Journal of Allergy and Clinical Immunology* 1989;**83**(2):435–40.
- 212 Crespo JF, Rodriguez J, James JM, Daroca P, Reano M, Vives R. Reactivity to potential cross-reactive foods in fruit-allergic patients: implications for prescribing food avoidance. *Allergy* 2002;**57**(10):946–9.
- 213 Asero R, Monsalve R, Barber D. Profilin sensitization detected in the office by skin prick test: a study of prevalence and clinical relevance of profilin as a plant food allergen. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2008;**38**(6):1033–7. eng.
- 214 Sanz ML, Blázquez AB, Garcia BE. Microarray of allergenic component-based diagnosis in food allergy. *Current opinion in allergy and clinical immunology* 2011;**11**(3):204–9. eng.

## 6. Anhang

Tabelle 16: Liste der identifizierten, potentiellen Epitopkandidaten. Solche, die in mehreren Untereinheiten vorkommen sind nur einmal gelistet.

Nr.	Epitope	Position	Homologie in anderen Proteinen	IEDB ID publizierter Epitope
<b>GenBank ID / Protein: A29385/ Gly m 1</b>				
1	RPSCPDL	05-11		
2	LGGSLGTV	18-25		
3	VDDCC	25-29		
4	CCALIG	28-33		
5	IEAIVCLC	38-45		
6	LCIQLRA	44-50		
7	CGRSY	67-71		
8	RSYPSNA	69-75		
9	NATCPR	74-79		
<b>GenBank ID / Protein: Z13956/ Gly m 2</b>				
10	RKSCGF	04-09		
11	FASQVV	18-23		
12	SQVVVQTEGRVCESQS	20-35		
13	SQSHGFH	33-39		
14	HGLCNRDH	39-46		
15	RDHNCALVC	44-52		
16	LVCRNE	50-55		
17	EGFSGG	55-60		
18	SRRCF	65-69		
<b>GenBank ID / Protein: AJ223981/ Gly m 3.01</b>				
19	LLCGI	11-15		
20	AAIIGQD	23-29	AAIIGQD/ Gly m 3.02	



21	DGSVWLQS	29-36	DGSVW / Gly m 3.02	
22	TPGQCN	111-116	TPGQCN / Gly m 3.02	
23	GDYLID	123-128	GDYLID / Gly m 3.02	
24	IDQGY	127-131	IDQGY / Gly m 3.02	

**GenBank ID / Protein: AJ223982/ Gly m 3.02**

25	DGSVWAQS	29-36	DGSVW / Gly m 3.01	78563
26	VERPGDYL	119-126		

**GenBank ID / Protein: x60043/ Gly m 4**

27	EINSP	09-13		
28	NSPVA	11-15		
29	PVAPA	13-17		
30	KSVENV	39-44		
31	ENVEG	42-46		
32	FLEDGET	58-64		
33	GETKVV	62-67		
34	LGYSY	79-83		
35	SYSVV	82-86		
36	VVGGAA	85-90		
37	NGGSAGK	109-115		
38	YETKG	120-124		
39	GDAEP	124-128		
40	AEPNQDE	126-132		
41	QDELK	130-134		

**GenBank ID / Protein: X17698/ Gly m 5.01**

42	FLASVSVSFGI	15-25	FLASVSVSFGI / Gly m 5.02, Gly m 5.03	
43	YWEKE	27-31	YWEKE / Gly m 5.03	

44	CLQSCN	38-43	CLQSCN / Gly m 5.03m, CLQSC / sucrose binding protein	
45	SERDSY	44-49	SERDSY / Gly m 5.02	
46	SYRNQACHC	48-56	SYRNQACHC / Gly m 5.02, Gly m 5.03	
47	ARCNLL	56-61	ARCNL / Gly m 5.02, Gly m 5.03	
48	EKEECEEG	64-71	EKEECEEG / Gly m 5.03; EECEE / Gly m 5.02	181299
49	EIPRPR	72-77	EIPRPR / Gly m 5.03, IPRPR / Gly m 5.02	181299
50	RPRPQHPE	77-84	RPRPQHPE / Gly m 5.02, Gly m 5.03	181299
51	HPEREPQQPG	82-91	HPEREPQQPG / Gly m 5.03; HPERE / Gly m 5.02; EPQQP / Gly m 6.01	181299
52	QQPGEKE	88-94	QQPGEKE / Gly m 5.03	181299
53	EEDEDE	94-99	EEDEDE / Gly m 5.03, EDEDE / Gly m 6.04, Gly m 6.05	181432
54	EDEQPRP	97-103	EDEQPRP / Gly m 5.03	
55	PRPIPFPR	101-108		181447
56	PFPRPQ	105-110	PFPRPQ / Gly m 5.03; PFPRP / Gly m 5.02	181432, 181447
57	PRQEEEH	111- 117	PRQEEEH/ Gly m 5.02, Gly m 5.03; PRQEE / Gly m 6.04	181447
58	EEEHEQ	114-119	EEEHEQ / Gly m 5.02, Gly m 5.03	181447
59	EHEQREEQE	116-124		181334, 181447
60	SEEEDEDE	137-144	EEEDE/ Gly m 6.01, sucrose binding protein; EEEDEDE / Gly m 5.03; EDEDE / Gly m 6.04, Gly m 6.05	181334, 181432
61	EDEEQDERQFP	144-154	EDEEQDERQFP / Gly m 5.03, EDEEQDER / Gly m 5.02, EEQDE / sucrose binding protein	
62	RPPHQKEER	157-165	RPPHQKEER / Gly m 5.03; PHQKEE / Gly m 5.02	181432
63	ERNEEE	164-169		181334, 181432, 11732
64	EEEDEDEE	167-174	EEEDEDE / Gly m 5.03; EEEDE / Gly m 6.01, sucrose binding protein; EDEDE / Gly m 6.04, Gly m 6.05	181288, 181334, 181432, 11305, 11732
65	EDEEQQ	171-176	EDEEQ / Gly m 5.02, Gly m 5.03	181288, 181432, 11305

66	EQQRESE	174-180		181288, 181432, 11305
67	ESESEDS	178-185	ESESEDS / Gly m 5.03	181288, 11305
68	DSELRR	184-189	DSELRR / Gly m 5.03	181288, 57535
69	LFGSNR	197-202	LFGSNR / Gly m 5.03	32585, 181406
70	RFETLF	202-207	RFETLF / Gly m 5.03	32585, 45674, 181406
71	KNQYGRI	208-214	KNQYG / Gly m 5.02; KNQYGRI / Gly m 5.03	45674, 181406, 45646
72	QRFNQ	218-222	QRFNQ / Gly m 5.03	38973, 45646
73	SPQLQN	224-229	SPQLQN / Gly m 5.03	38973
74	YRILEFN	233-239	YRILEFN / Gly m 5.02, Gly m 5.03	181386
75	NSKPN	239-243	NSKPN / Gly m 5.02, Gly m 5.03	181386
76	SYRLQSGDA	276-284	OSGDA / Gly m 5.02; SYRLQSGDA / Gly m 5.03	181474, 9963
77	VPSGTT	287-292	VPSGTT / Gly m 5.03	181474
78	TTYVV	291-296	TTYVV / Gly m 5.02, Gly m 5.03	
79	FLSSTEAQQSYLQGFSRNIL	323-342	FLSSTEAQQSYLQGFSRNIL / Gly m 5.03; FLSST, AQQSYLQGFS / Gly m 5.02	181316, 17831
80	LEASYDT	342-348	LEASYDT / Gly m 5.02, Gly m 5.03	17831
81	FEEIN	350-354	FEEIN / Gly m 5.02, Gly m 5.03	27677, 227839
82	VLFSREEGQQQGEQR	356-370	VLFSREEGQQQGEQR / Gly m 5.03; GQQQGE / Gly m 5.02	11655, 27677, 227839,
83	QRLQESVIVE	369-378	RLQESVIVE / Gly m 5.02, Gly m 5.03; SVIVE / $\beta$ -Amylase	11655
84	ISKEQIRA	379-386	ISKEQIRA / Gly m 5.03	227835, 227858, 181307
85	AKSSSRKT	391-398	AKSSSRKT / Gly m 5.02; Gly m 5.03	181307; 227829, 227834
86	NLRSRDP	407-413	NLRSRDP / Gly m 5.02, Gly m 5.03	181455, 227814, 227847
87	PQLRD	431-435	PQLRD / Gly m 5.02, Gly m 5.03	181292, 227769, 227787, 227812
88	IFLSIVD	438-444	IFLSIVD / Gly m 5.03	
89	HFNSK	454-458	HFNSK / Gly m 5.02, Gly m 5.03	2716, 227784
90	LKEQQQE	477-483	LKEQQQE / Gly m 5.03, KEQQQ / Gly m 5.02	13307
91	QEEQPL	486-491	QEEQPL / Gly m 5.02, Gly m 5.03	

92	EVRKYRAELSE	491-502	EVRKYRAELSE / Gly m 5.02, Gly m 5.03	
93	QDIFVIPAGYPVVVN	503-517	QDIFVIPAGYPVVVNA / Gly m 5.02, Gly m 5.03	26180
94	ATSNLN	518-523	ATSNLN / Gly m 5.03	227740
95	RNFLA	536-540	RNFLA / Gly m 5.02, Gly m 5,03	181403, 27590
96	GSQDNVISQIPS	541-552	GSQDNVISQIPS / Gly m 5.03; ISQIPS / Gly m 5.02	
97	SQVQELAF	552-559	SQVQELA / Gly m 5.02, Gly m 5.03	227813, 227819
98	FPGSAQAVE	559-567	FPGSAQVE/ Gly m 5.03, FPGSA/ Gly m 5.02,	227819, 22286
99	NQRESYFVDAQPKK	572-585	NQRESYFVDAQPKK / Gly m 5.03; ESYFVDAQP / Gly m 5.02	181434, 227756, 227797
100	PLSSILR	596-602	PLSSILR / Gly m 5.02, Gly m 5.03	181434

**GenBank ID / Protein: M13759/ Gly m 5.02**

101	YWEKQ	27-31		
102	QNPSHNK	31-37		
103	RSCNSE	40-45	SCNSE / Gly m 5.01, Gly m 5.03	
104	KVEEEEECEEGQ	62-73	VEEEE / sucrose binding protein; EEECEEGQ / Gly m 5.0101; EEECEEGQ / Gly m 5.03	11655, 27677
105	RPRPQHPERERQQHGEK	76-92	RPRPQHPERE / Gly m 5.01, Gly m 5.03	181299
106	EDEGEQPR	94-101		
107	PRPRQ	106-110	PRPRQ / Gly m 5.0101, Gly m 5.03, Gly m 6.04	
108	QPHQEE	110-115		
109	EEEHEQ	114-119	EEEHEQ / Gly m 5.0101, Gly m 5.03	181447
110	EQKEEHE	118-124	EQKEE / Gly m 5.0101, Gly m 5.03	181432, 227821
111	KEEKHG	128-133		
112	SEEEQDERE	137-145	SEEEQDER / Gly m 5.0101, Gly m 5.03; EEEQDE / sucrose binding protein	
113	EHPRPH	145-150	EHPRPH / Gly m 5.0101	
114	QPHQKEEE	151-158	PHQKE / Gly m 5.01, Gly m 5.03	181334, 181432, 11732, 17285
115	HEWQHKQEKH	160-169		
116	ESEEE	173-177	ESEEE / Gly m 5.0101, Gly m 5.03	

117	EEEDQDEDEEQK	177-188	DQDEDE / Gly m 6.04; DEDEE / Gly m 8, Gly m 5.0101, Gly m 5.03, Gly m 6.05	181432
118	ESQESEGSESQREP	190-203	ESEGS / Gly m BD 30K; SESQRE / Gly m 5.0101, ESQRE / Seed biotinylated protein	
119	HFNSKR	213-218	HFNSK / Gly m 5.01, Gly m 5.03	227784
120	RFQTLF	218-223	FQTLF / Gly m 5.03	
121	NKRSQQLQ	237-244		
122	LPHHAD	263-268	LPHHAD / Gly m 5.03	227782
123	SYNLQ	292-296		
124	LRVPAG	301-306		
125	FLSSTQAQQSYLQGF	358-373	FLSST / Gly m 5.01; FLSSTQAQQSYLQGF / Gly m 5.03; AQQSYLQGF / Gly m 5.01	181316
126	GREEGQ	394-399	REEGQ / Gly m 5.01 , Gly m 5.03	
127	EERLQESVIVE	403-413	RLQESVIVE / Gly m 5.01, Gly m 5.03	11655
128	KQIREL	417-422		
129	TQRNPQLRD	461-469	NPQLRD / Gly m 5.01, Gly m 5.03	181292, 227769, 227787, 227812
130	VFLSVVD	472-478		
131	EGALF	481-485		
132	INEGEAN	499-505		
133	IKEQQQ	511-516	IKEQQQ / Gly m 5.03; KEQQQ / Gly m 5.01	
134	QEEQPL	520-525	QEEQPL / Gly m 5.0101, Gly m 5.03	
135	NATSD	551-555		
136	FPRSAK	593-598		
137	KSQSES	605-610		
138	AQPQQ	615-619	AQPQQ / Gly m 5.03	227797

**GenBank ID / Protein: S44893/ Gly m 5.03**

139	FLASVCVS	16-23	FLASV / Gly m 5.01	
140	YFRSSNSFQTLFEN	35-48	FQTLF / Gly m 5.02	227751, 227798
141	QNVRIPLLQ	49-57		227822

142	SPQLEN	63-68		227764, 227832
143	ENLRD	67-71		227764
144	YRIVQFQSK	72-80		
145	FVLSGRAIL	97-105		227775
146	SYNLHP	115-120		227783
147	DAQRIP	122-127		227776
148	YLVNPHDHQN	133-142		227772, 227805
149	KPGRYD	154-159		227744, 227790
150	FLSSTQAQQSYLQGFS	162-177	FLSST, AQQSYLQGFS / Gly m 5.01; FLSSTQAQQSYLQGFS / Gly m 5.02	181316
151	LETSFHSEFEEINR	181-194	FEEIN / Gly m 5.0101, Gly m 5.02	227820, 227839
152	NRVLFG	193-198		227839
153	LFGEEEE	196-202		
154	QRQQEGVI	203-210		227835
155	VELSKEQI	211-218	SKEQI / Gly m 5.01	227835, 227858
156	IFLSSVDIN	272-280		227812
157	VILVINE	295-301	VILVINE / Gly m 5.0101; LVINE / Gly m 5.02	227765
158	AYPFVV	345-350		227740

**GenBank ID / Protein: AB113349/ Gly m 6.01**

159	KPFQCAGVA	60-68	KPFQCAGVA / Gly m 6.02, Gly m 6.03; QCAGV / Gly m 6.05	187829, 32708
160	RPSYT	81-85	RPSYT / Gly m 6.02, Gly m 6.03	54971
161	NGPQE	86-90	NGPQE / Gly m 6.02	44051, 54971, 76131
162	EIYIQQG	90-96	EIYIQQG / Gly m 6.02, Gly m 6.03	44051, 76131, 74358
163	IYPGCPSTFEE	103-113	PGCPSTFE / Gly m 6.02, Gly m 6.03; PGCP / Gly m 6.04, Gly m 6.05, Gly m TI; PSTF / Lectin; TFEE / Gly m 6.04	16009, 52031
164	EPQQPQQRG	113-121	EPQQP / Gly m 5.0101, Gly m 5.03	
165	GQSSRPQD	121-128	GQSSRPQD / Gly m 6.02, Gly m 6.03	

166	PQDRHQ	126-131	PQDRHQ / Gly m 6.02, Gly m 6.03	
167	IAVPTG	142-147	IAVPTG/ Gly m 6.02, Gly m 6.03	
168	ENQLDQ	172-177	ENQLDQ / Gly m 6.02; NQLDQ / Gly m 6.03, Gly m 6.04, Gly m 6.05	45574
169	PRRFY	179-183	PRRFY / Gly m 6.02, Gly m 6.03	45574
170	GNQEQ	186-190	GNQEQ / Gly m 6.02, Gly m 6.03	53180
171	FLKYQQE	192-198	FLKYQQ / Gly m 6.02	13815, 37068
172	QQEQGGHQ	196-203		
173	HQSQKG	202-207	QSQKG / Gly m 6.02, Gly m 6.03	52128
174	KHQQEE	208-213	HQQEE / Gly m 6.04	
175	GFTLEF	223-228		19632, 20021, 58037, 187850
176	FLEHA	228-232	FLEHA / Gly m 6.02	19632, 58037, 187850
177	EHAFSVD	230-236		58037
178	NLQGE	242-246	NLQGE / Gly m 6.02	
179	QGENE	244-248	QGENE / Gly m 6.02, Gly m 6.03	
180	LSVIKP	262-267		
181	PTDEQQQ	268-274		
182	QRPQEE	274-279		
183	EEEEDEK	281-287	EEEEED / Gly m Gly m 5.02, sucrose binding protein, EEEEDE / Gly m 5.0101, Gly m 5.03, sucrose binding protein	
184	EKPQC	286-290		
185	KDKHC	293-297	KDKHC / Gly m 6.02, Gly m 6.03	
186	ETICT	314-318	ETICT / Gly m 6.02, Gly m 6.03	
187	FVPHYNL	370-374	FVPHYNL / Gly m 6.02, Gly m 6.03	41548, 59438
188	FLVPPQESQK	482-491	FLVPP / Gly m 6.02, Gly m 6.03, VPPQESQ / Gly m 6.02	
<b>GenBank ID / Protein: D00216/ Gly m 6.02</b>				
189	QQEQGGHQ	12-18	LFSGC / Gly m 6.01, Gly m 6.03	6238, 32233

190	HQSQKG	19-31	QQNECQIQ / Gly m 6.01, Gly m 6.03	6238
191	KHQQEE	38-43	PDNRIE / Gly m 6.01, Gly m 6.03	
192	GFTLEF	41-46	RIESEG / Gly m 6.01, Gly m 6.03	187845
193	FLEHA	44-49	SEGGFI / Gly m 6.03	
194	EHAFSVD	100-108	IFPGCPS / Gly m 6.03; FPGCP / Gly m 6.05; PGCPST / Gly m 6.01; GCPST / Gly m 6.03	16009, 52031
195	NLQGE	110-116	PQESQ / Gly m 6.01	
196	QGENE	112-118	QQRG / Gly m 6.01	
197	LSVIKP	118-128	RPQDRHQ / Gly m 6.01, Gly m 6.03	
198	PTDEQQQ	128-133		
199	QRPQEE	142-147	PTGVAV / Gly m 6.01	
200	EEEEDEK	189-195	FLKYQQ / Gly m 6.01	13815, 37068
201	EKPQC	195-201		13815, 37068, 52128
202	KDKHC	199-205	QSQKG / Gly m 6.01, Gly m 6.03	37068, 52128
203	ETICT	209-214	EEENEG / Gly m 6.01, Gly m 6.03	60470
204	FVPHYNL	219-223	SGFAP / Gly m 6.03	58026
205	FLVPPQESQK	238-244	NLQGEN / Gly m 6.01; LQGEN / Gly m 6.03	
206	LFSGCFA	257-262		19919, 57973
207	LREQAQQNECQIQ	275-282		
208	PDNRIE	287-292		538733
209	RIESEG	313-317		
210	SEGGFI	317-322		
211	IFPGCPSTY	323-328	IYNPQA / Gly m 6.01	
212	EPQESQQ	345-352	LKLSAQY / Gly m 6.03; LSAQY / Gly m 6.04, Gly m 6.05	36948, 35160, 187962
213	QESQQRG	350-357	GSLRKN / Gly m 6.01; SLRKN / Gly m 6.03	36948, 59438
214	GRSRPQDRHQ	369-374	NSIYY / Gly m 6.01, Gly m 6.03	41548, 43262, 45850, 70332,
215	QKVHRF	455-460	NLKSQQ / Gly m 6.01	51031, 69084
216	PTGVAV	460-464	QARQV / Gly m 6.03	33488, 51031
217	FLKYQQQ	384-390	VNCNGER / Gly m 6.01, Gly m 6.03	



218	QQQGSQ	387-392	NGERVF / Gly m 6.01, Gly m 6.03	
219	GSQSQKG	394-400	GELQEG / Gly m 6.01, Gly m 6.03	

**GenBank ID / Protein: X15123/ Gly m 6.03**

220	SGCCFAF	14-20	SGCCFAF / Gly m 6.01, Gly m 6.02	
221	CFAFSF	17-22	CFAFS / Gly m 6.01; CFAFSF, Gly m 6.02	
222	FSFREQPQ	20-27	FSFREQPQ / Gly m 6.02; REQPQ / Gly m 6.01	
223	QPQQNECQIQ	25-34	QPQQNECQIQ / Gly m 6.01, Gly m 6.02	
224	NRIESEG	43-49	NRIESEG / Gly m 6.01, Gly m 6.02	
225	YIQQSGI	92-99	YIQQG / Gly m 6.01; YIQQSGI / Gly m 6.02	44051, 74358, 76131
226	EPQQKG	113-118	EPQQK / Gly m 6.02	
227	QKKGQSSR	115-122	QKKGQSSR / Gly m 6.02; GQSSR / Gly m 6.01	
228	IYHFRE	130-135	IYHFRE / Gly m 6.02	
229	IAVPTGF	139-145	IAVPVG / Gly m 6.01; IAVPTGF / Gly m 6.02	
230	FQNQLDQ	168-174	FQNQLDQ / Gly m 6.02; NQLDQ / Gly m 6,0101, Gly m 6,0301, Gly m 6,0401, Gly m 6,0501	45574
231	EQEFL	186-190	EQEFL / Gly m 6.01, Gly m 6.02	45544, 13815, 53810
232	FLQYQPQ	189-195	FLQYQPQ / Gly m 6.02	
233	GLSVISP	260-266	GLSVISP / , Gly m 6.02, Gly m 6.04, Gly m 6.05; LSVIS / Gly m 6.01	18633, 187789, 187855
234	PPTEEQQ	266-272	PPTEEQQ / Gly m 6.02	
235	RPEEEKP	274-281	RFEEKP / Gly m 6.02	
236	EKPCDCDE	278-285	EKPCD / Gly m 6.02, Lea Protein, EKPCDCDE / Gly m 6.02	
237	QSQSRN	291-296	QSQSRN / Gly m 6.02	
238	TSSPDIF	314-320	TSSPDIF / Gly m 6.01, Gly m 6.02, TSSPDI / Gly m 6.01	

**GenBank ID / Protein: X02626/ Gly m 6.04**

239	LEPDH	40-44	LEPDH / Gly m 6.05	
240	SQHPEL	58-63	SQHPEL / Gly m 6.05	187903

241	GVTVSK	67-72	GVTVSK / Gly m 6.05	187903
242	NGLHSP	78-83		
243	KQQLQD	128-133	KQQLQD / Gly m 6.05	
244	NEGDVLV	142-148	NEGDVLV / Gly m 6.05	
245	QGQHQQ	215-220	QGQHQQ / Gly m 6.05	
246	QDEDE	279-284	QDEDE / Gly m 5.02	
247	GQDEDE	335-340	GQDEDE / Gly m 5.02	
248	PRQEEPR	363-369	PRQEE / Gly m 5.0101, Gly m 5.02, Gly m 5.50301	
249	ARPSRA	394-399	ARPSRA / Gly m 6.05	187752
250	QFQLSAQY	422-429	LSAQY / Gly m 6.02, Gly m 6.05	187962, 35160, 36948
251	AVTSY	507-511		
252	GPLVN	545-549	GPLVN / Gly m 6.05	

**GenBank ID / Protein: M10962/ Gly m 6.05**

253	ECQLNN	32-37	ECQLNN / Gly m 6.04	
254	VESEGG	47-52	VESEGG / Gly m 6.04; ESEGG / Gly m 6.01, Gly m 6.02, Gly m 6.03, Seed biotinylated protein	
255	NRNGSH	77-82		
256	NGSHLPS	79-85		
257	LPSYLPY	83-89		
258	YLPYPQ	86-91		
259	AFPGCP	105-110	FPGCP / Gly m 6.02, Gly m 6.03	16009, 52031
260	GCPETFE	108-114	GCPETFE / Gly m 6.04	
261	HFNEG	140-144	HFNEG / Gly m 6.04	
262	LVIPLG	147-152		
263	EPVVAI	163-168	EPVVAI / Gly m 6.04	
264	DIEHPET	194-200		
265	QEEEGG	221-226	EEEGG / Gly m 6.04	
266	EGGSV	224-228	EGGSV / Gly m 6.04	

267	KLRSPDD	250-256		
268	GGLSVISP	266-273	GGLSVI / Gly m 6.0101; GGLSVISP / Gly m 6.02, Gly m 6.03, Gly m 6.04	18633, 187789, 187855
269	SPKWQ	272-276	SPKWQ / Gly m 6.04	187855
270	WQEQEDED	275-282		
271	DEDEEYG	284-290	DEDEE / Gly m 8, Gly m 5.0101, Gly m 5.02, Gly m 5.03	
272	TPSYPP	292-297	TPSYP / $\beta$ - Amylase	
273	RPSHGKH	299-305	RPSHGK / Gly m 6.04	187775
274	EDEDE	309-313	EDEDE / Gly m 5.0101, Gly m 5.03, Gly m 6.04	
275	EDEEEDQPRP	311-320	EDEEE / Gly m 8, Gly m BD28K; EEEDQ / Gly m 5.02; EDQPR / Gly m 6.04	
276	QPRPDH	317-322		
277	PDHPPQ	320-325		
278	QRPSRPEQQE	325-334	RPSRP / Gly m 6.04	
279	QQEPRG	332-337		
280	RGRGCQ	336-341		
281	GCQTRN	339-344		
282	AVFDGE	433-438	AVFDGE / Gly m 6.04; VFDGE / Gly m 6.01, Gly m 6.02; Gly m 6.03	
283	THHNA	469-473	THHNA / Gly m 6.04	
284	VSSYIK	474-479		
285	LSNSYN	490-495		
286	NSYNLG	492-497		
287	NSGPLVN	509-515	GPLVN / Gly m 6.04	

**GenBank ID / Protein: U59626/ Gly m 7**

288	KEIHVEKHR	16-24		
289	KHRVPK	22-27		
290	HAGKA	67-71		

291	KESQREA	110-116	KESQR / Gly m 6.02, Gly m 6.03; ESQRE / Gly m 5.01, Gly m 5.02
292	GGQVVAEK	163-170	
293	HVVEGA	388-393	
294	EEAQREL	432-438	
295	QPQEAERP	445-453	
296	EERPSEGIG	450-458	
297	GGESE	508-512	
298	VTGRA	630-634	
299	AKHEEG	634-639	

**GenBank ID / Protein: AAB71140/ Gly m 8**

300	LFCIAHTCS	12-20	
301	SASKWQH	20-26	
302	QQDSCR	27-32	
303	SCRKQL	30-35	
304	KQLQGVN	33-39	
305	NLTPCEK	39-45	
306	QGRGD	52-56	
307	EDEEEEG	86-92	EDEEE /Gly m BD 28K, Gly m 6.05 ; EEEEG / Gly m 6.04
308	QKCCT	95-99	
309	TEMSEL	99-104	
310	CKALQK	111-116	
311	NQSEELEEK	120-128	
312	MCRFGP	143-148	
313	IQCDLS	150-155	

**GenBank ID / Protein: S45092/ Gly m TI**

314	RAAPTG	54-59	
-----	--------	-------	--

315	CVGIPT	110-115	
316	TEWSV	115-120	

**GenBank ID / Protein: K00821/ Gly m agglutinin**

317	QTHAGY	134-139	
318	FRNSWD	160-165	
319	SWDPP	163-167	
320	FASNLPH	260-266	
321	HASSNID	266-272	
322	DLTSF	275-279	

**GenBank ID / Protein: J05560/ Gly m BD30K**

323	SSSSIST	16-22		37300
324	HRSILD	23-28		55861
325	QKQVSSLFQL	37-46	SLFQL / $\beta$ - Amylase	39985, 61155
326	HEEEA	58-62		74125
327	ITPQEFSK	98-105		8803, 48987, 48988
328	KYLQAP	106-111		8803, 74873
329	DVSQQIK	113-119		50328, 60532
330	QYSCDH	129-134		
331	GVITQV	145-150		31559
332	EESEGCYNG	192-200	ESEGS / Gly m 5.02	14141
333	GWQYQS	200-205		
334	HYQSFQV	202-209		14141, 44118
335	KYQGGCG	151-157		78253
336	QGGCGR	153-158		78253
337	LEHGGI	210-215		12334
338	DDYPYR	219-224		12334, 63127
339	EGRCKAN	227-233		75579, 53282

340	IDGYE	241-245		78118
341	IMSDES	248-253		78118
342	SVSIDAK	272-278		28777
343	GYGSAD	307-312		23875
344	GEDGYI	328-333		11472
345	NLLGVC	341-346		

**GenBank ID / Protein: AAA33984/ 18kD Seed Maturation Protein**

346	IGASA	18-22	IGASA/ Lea protein	
347	ATVQEKAER	31-39	ATVQEKAER/ Lea protein	
348	ARDPVQ	42-47	ARDPVQ/ Lea protein	
349	ELATQK	49-54	ELATQK/ Lea protein	
350	QHNTA	71-75	QHNTA/ Lea protein	
351	GHGHHT	87-92	GHGHHT/ Lea protein	
352	GEYGQP	106-111	GEYGQP/ Lea protein	
353	HQTSA	115-119	HQTSA/ Lea protein	
354	HGTGQPTGH	123-131	HGTGQPTGH/Lea protein	
355	HVTEGV	131-136	HVTEGV / Lea protein	
356	VGSHPIG	137-143	VGSHPIG / Lea protein	

**GenBank ID / Protein: AJ011935/ Albumin**

357	NGACSPFEV	23-31		
358	PPCRSRD	32-38		
359	RDCRC	37-41		
360	RCVPIGLF	70-47		
361	GLSSVA	56-61		
362	VDEHP	64-68		
363	HPNLCQSD	67-74		
364	QSDDEC	72-77		

365	GNFCARYP	83-90	
366	SDSEA	102-106	

**GenBank ID / Protein: AB084260/ Basic 7S Globulin**

367	ASILHY	02-06	
368	LSFSFL	13-18	
369	SDSVPIP	22-28	
370	VPIPQH	25-30	
371	VLPVQN	42-47	
372	QVPVLVD	66-72	
373	HLWVNCE	77-83	
374	NCEQHYS	81-87	
375	YSSKTYQA	86-93	
376	TYQAPFCH	90-97	
377	PFCHSTQCSR	94-103	
378	QCLSCPAASRPGCHKN	108-123	
379	NPITQ	131-135	
380	GELGQDVL	140-147	
381	IHATQGSTQQL	149-159	
382	TQQLGPL	156-162	
383	VTVPQF	163-168	
384	RNIQGVA	184-190	
385	LASHFGLQ	203-210	
386	GLQHQFT	208-214	
387	TSKGAL	222-227	
388	QQFHNQ	237-242	
389	VTPQGEYNVR	255-264	
390	VRVSSIR	263-269	
391	NQHSV	271-275	

392	ISTSTPH	294-300		
393	SLYQAFTQ	306-313		
394	QVFAQQLEKQ	312-322		
395	LEKQAQVKS	319-328		
396	KSVAPF	326-331		
397	GLCFNSN	332-338		
398	NAYPSVDL	341-348		
399	VQAQPG	367-372		
400	VTCLGV	373-378		
401	QPRAEV	384-389		
402	GTRQLE	392-397		
403	ARSRVG	406-411		
404	FSTSSLHSH	412-420		



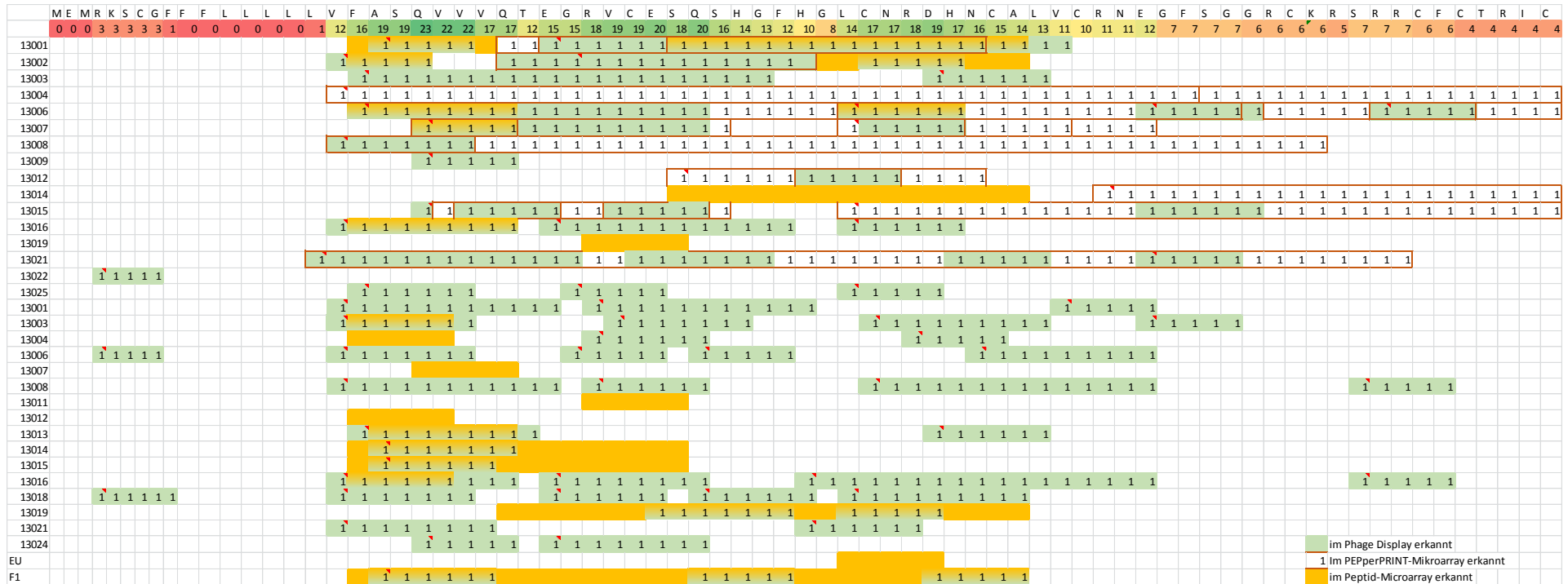


Abbildung 13: Epitopkarte des Proteins Gly m 2. Dargestellt sind die Epitope, die mit unterschiedlichen Methoden identifiziert wurden. Mimotope, die im Peptide Phage Display identifiziert wurden, sind als grüne Balken, Epitope, die im PEPperPRINT Microarray identifiziert worden, sind mit orangem Rahmen und Peptide die im Peptid-Microarray gefunden wurden, sind mit orangem Balken dargestellt.



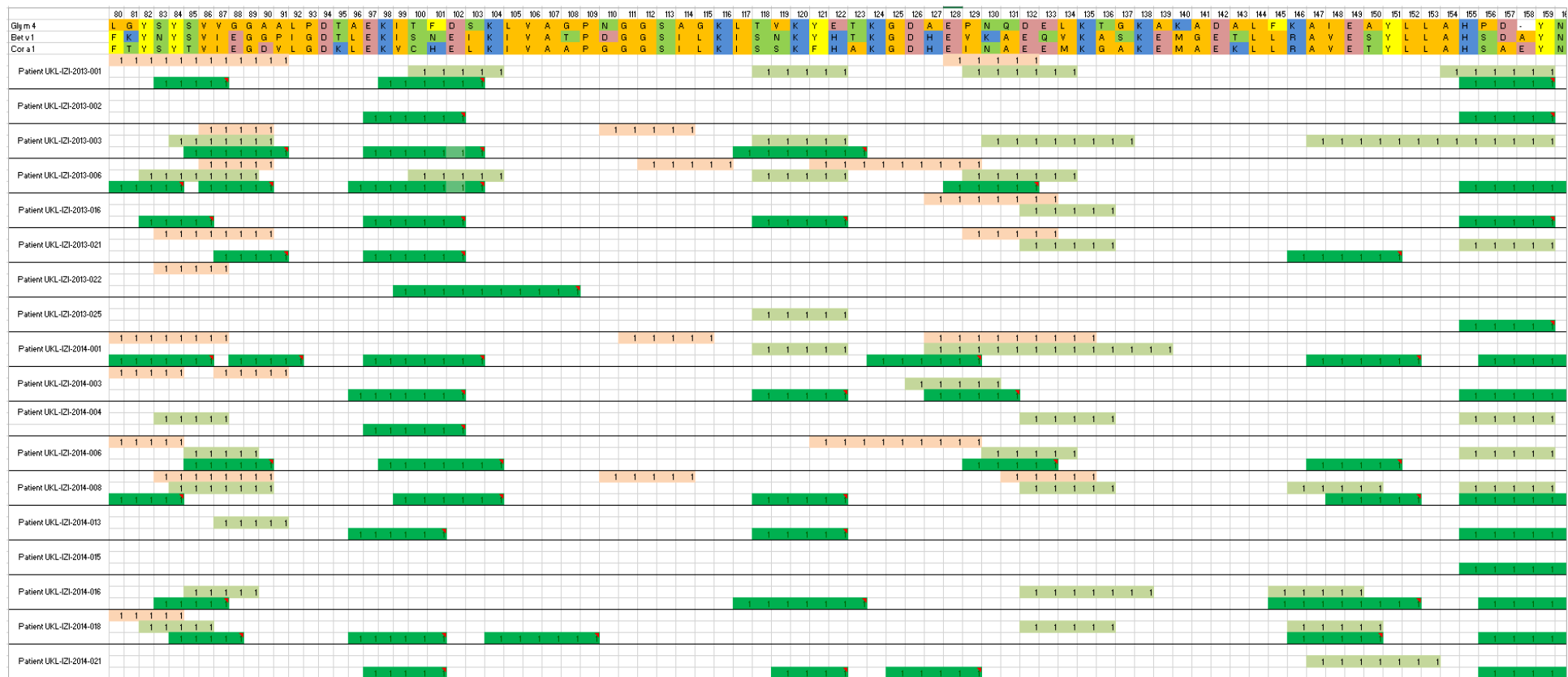


Abbildung 14: Epitopkarte von Gly m 4 (oben), Bet v 1 (mitte) und Cor a 1 (unten) von Position 80 -160. Die Epitope der Proteine sind in braun für Gly m 4, hellgrün für Bet v 1 und dunkelgrün für Cor a 1 dargestellt.

## **Danksagung**

Meinem Betreuer, Herrn Dr. Szardenings, danke ich für die Überlassung des aktuellen Themas und die Möglichkeit der Anfertigung einer Dissertation in der Arbeitsgruppe Liganden-Entwicklung des Fraunhofer Instituts für Zelltherapie und Immunologie. Bedanken möchte ich mich auch für seine Hilfs- und Diskussionsbereitschaft, sowie an seinem Interesse am Fortschritt dieser Arbeit und der hilfreichen Kommentare bezüglich dieses Manuskriptes.

Herrn Prof. Dr. Lorkowski danke ich für die Unterstützung bei der Anfertigung der außeruniversitären Arbeit sowie für das Gutachten.

Bei Frau Prof. Dr. med. habil. Regina Treudler von der Universitätsklinikum Leipzig bedanke ich mich für die Bereitstellung der Seren, die ich für diese Arbeit verwendet habe.

Besonderer Dank gilt zudem den Mitarbeitern der Arbeitsgruppe, sowie den zahlreichen Studenten, die mich in dieser Arbeit unterstützt haben.

Für ihre Anregung und konstruktive Kritik an der Arbeit danke ich René Kern, sowie Dorothee Wehrmann.

Bedanken möchte ich mich auch bei meiner Familie für die Unterstützung während dieser Arbeit.

Betreuer:

Herr Prof. Dr. Lorkowski, Institut für Ernährungswissenschaften, Friedrich-Schiller-Universität  
Jena

Herr Dr. Szardenings, Fraunhofer Institut für Zelltherapie und Immunologie

### **Eidesstattliche Erklärung**

Teilergebnisse der vorliegenden Arbeit wurden veröffentlicht in:

Kern, Karolin, et al. "The immunome of soy bean allergy: Comprehensive identification and characterization of epitopes." *Clinical & Experimental Allergy* 49.2 (2019): 239-251.

Havenith, Heide, et al. "Combination of two epitope identification techniques enables the rational design of soy allergen Gly m 4 mutants." *Biotechnology journal* 12.2 (2017): 1600441.

Ich versichere, dass ich die vorliegende Arbeit selbständig und ohne unerlaubte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die inhaltlich oder wörtlich aus Veröffentlichungen stammen, sind kenntlich gemacht. Diese Arbeit lag in der gleichen oder ähnlichen Weise noch keiner Prüfungsbehörde vor und wurde bisher noch nicht veröffentlicht.

Datum/ Unterschrift

## Publikationen

## Poster

Kern, Karolin (2013): Identification of allergy-related epitopes of soybean. Poster presented at Research Festival .Leipzig, 19.12.2013.

Kern, Karolin (2014): Identification of Allergy-Related Epitopes of Soybean Proteins. Poster presented at PEGS, Boston, 06.05.2014.

Kern, Karolin (2014): Characterization of Allergy-Related Epitopes of Soybean Proteins. Poster presented at Science Day, Leipzig, 10.07.2014.

Kern, Karolin (2014): Identification of Allergy-Related Epitopes of Soybean Proteins. Poster presented at Annual meeting of German Society of Immunology (dgfi) Bonn, 19.09.2014.

Kern, Karolin (2014): Identifizierung potentiell allergener Sojaprotein Epitope mittels Peptid Phage Display. Poster presented at Research Festival .Leipzig, 18.12.2014.

Kern, Karolin (2015): Identification of Soy-Related Epitopes. Poster presented at Science Day, Leipzig, 03.07.2015.

Kern, Karolin (2016): Mapping of Immunogenic and Allergenic Soybean Protein Epitopes. Poster presented at PEGS, Boston, 27.04.2016.

Kern, Karolin (2018): Clinically relevant, allergy related soy peptide epitopes from phage display. Poster presented at Science Day, Leipzig, 19.06.2018.

## Paper

Kern, Karolin, et al. "The immunome of soybean allergy: Comprehensive identification and characterization of epitopes." *Clinical & Experimental Allergy* (2018).

Havenith, Heide, et al. "Combination of two epitope identification techniques enables the rational design of soy allergen Gly m 4 mutants." *Biotechnology journal* 12.2 (2017).