

MS. ANNA M EHLERS (Orcid ID : 0000-0001-7833-2900)

Article type : Original Article: Basic and Translational Allergy Immunology

Distinction between peanut allergy and tolerance by characterization of B-cell receptor repertoires

Manuscript Acceptance Date: 25-Apr-2021

Running title: Specific B-cell repertoires in peanut allergy

*AnnaM. Ehlers, PhD^{1,2}

Constance F. den Hartog Jager, BSc^{1,2}

André C. Knulst^{1,2}, MD PhD

Henny G. Otten¹, PhD

¹Center for Translational Immunology, University Medical Center Utrecht, Utrecht University, The Netherlands

²Department of Dermatology/Allergology, University Medical Center Utrecht, Utrecht University, The Netherlands

***Corresponding author:**

Anna M. Ehlers, MSc

University Medical Center Utrecht

Center for Translational Immunology

Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands

Tel: +31 88 7559024, Fax: +31 88 7555404

Email: a.m.ehlers@umcutrecht.nl

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/ALL.14897](#)

This article is protected by copyright. All rights reserved

Acknowledgments

We would like to thank all participants for their efforts and dedication in this study. Additionally, we would like to thank Anouska Michelsen-Huisman, Anne van Dijk, AnsLebens and Mehran Alizadeh Aghdam for their support in the patient inclusion and Michael Drefahl, Rowena Melchers and Miriam Katsburg for their technical support. For critical discussion and reading of the manuscript, we would like to thank Melanie Abram, Edward Knol and Femke van Wijk. For financial support, we want to acknowledge EUROIMMUN AG, Lübeck, Germany.

Author contributions

AE, CHJ, HO: experimental design, AE, AC: patient recruitment, AE, CHJ: experimental performance, AE, AC, HO: data analysis and interpretation, AE: drafting manuscript, CHJ, AC, HO: critical revision of the manuscript, all authors approved the final manuscript

Funding

This research was financially supported by EUROIMMUN AG, Lübeck, Germany. They were not involved in the overall design of the study.

Conflict of Interest

The research position of A. Ehlers is partially funded by EUROIMMUN AG, Lübeck, Germany. Other authors have no conflicts of interest to declare.

Abstract

Background

Specific IgE against a peanut 2S albumin (Ara h2 or 6) is the best predictor of clinically relevant peanut sensitization. However, sIgE levels of peanut allergic and those of peanut sensitized but tolerant patients partly overlap, highlighting the need for improved diagnostics to prevent incorrect diagnosis and consequently unnecessary food restrictions. Thus, we sought to explore differences in V(D)J gene transcripts coding for peanut 2S albumin-specific monoclonal antibodies (mAbs) from allergic and sensitized but tolerant donors

Methods

2S albumin-binding B-cells were single-cell sorted from peripheral blood of peanut allergic (n=6) and tolerant (n=6) donors sensitized to Ara h2 and/or 6 (≥ 0.1 kU/l) and non-atopic controls (n=5). Corresponding heavy and light chain gene transcripts were heterologously expressed as mAbs and tested for specificity to native Ara h2 and 6. HCDR3 sequence motifs were identified by Levenshtein distances and hierarchically clustering.

Results

The frequency of 2S albumin-binding B-cells was increased in allergic (median: 0.01%) compared to tolerant (median: 0.006%) and non-atopic donors (median: 0.0015%, $p=0.008$). The majority of mAbs (74%, 29/39) bound specifically to Ara h2 and/or 6. Non-specific mAbs (9/10) were mainly derived from non-atopic controls. In allergic donors, 89% of heavy chain gene transcripts consisted of VH3-family genes, compared with only 54% in sensitized but tolerant and 63% of non-atopic donors. Additionally, certain HCDR3 sequence motifs were associated with allergy (n=4) or tolerance (n=3) upon hierarchical clustering of their Levenshtein distances.

Conclusions

Peanut allergy is associated with dominant VH3 family gene usage and certain public antibody sequences (HCDR3 motifs).

Keywords: clinically irrelevant sensitization, food allergy diagnostics, monoclonal antibodies, peanut allergy, VH family gene usage

Word count: 3500

Figure/Table count: 3 Figures and 3 Tables

Abbreviations

BAT	Basophil activation test
DBPCFC	double-blind placebo-controlled food challenge
HCDR3	complementary-determining region 3 of the heavy chain
mAbs	monoclonal antibodies
OD	optical density
PBMCs	peripheral blood mononuclear cells
RBL	rat basophil leukemia cells
sIgE	specific IgE
V(D)J	rearranged variable (V), diversity (D) and joining (J) gene segments
VH	V (variable) gene of the heavy chain

Introduction

Current food allergy diagnostics comprise careful history, skin prick tests, measuring specific IgE (sIgE) and double-blind placebo-controlled food challenges (DBPCFC) as the gold standard. However, DBPCFCs are burdensome for the patient, costly and require dedicated hospital facilities^{1, 2}. On the other hand, current technologies to measure sIgE detect both clinically relevant and irrelevant sensitization, potentially leading to incorrect diagnosis and consequently unnecessary food restrictions^{3, 4}. Clinically relevant peanut sensitization is associated with sIgE against the major peanut allergens belonging to the 2S albumin family, Ara h2 and 6⁵. In previous studies in adults, 100% positive predictive values for sIgE against Ara h2 and 6 were found using sIgE positivity thresholds of respectively 1.75 kU/l and 1.8 kU/l. Specific IgE levels below these thresholds, however, overlapped between allergic and tolerant subjects, preventing precise diagnosis^{6, 7}.

The occurrence of clinically irrelevant sensitization to Ara h2 and 6 might be explained by differences in peanut (Ara h2 and 6) specific antibody repertoires comprising IgE and non-IgE antibodies. These differences may include the presence of non-IgE antibodies blocking sIgE binding to clinically relevant epitopes^{8, 9}, whose recognition induces degranulation and leads to the presentation of allergic symptoms. Moreover, differences may also be based on sIgE antibody affinity and recognition patterns of clinically relevant and irrelevant epitopes¹⁰. However, no clear differences between allergic and tolerant subjects were observed by current epitope mapping approaches¹¹. We hypothesize that these evaluations may have been hampered by the use of patient sera consisting of polyclonal specific IgE and non-IgE antibodies. Sera of allergic subjects may usually contain mixtures of antibodies recognizing both clinically relevant and irrelevant epitopes. On the other hand, sera from tolerant subjects may contain IgE antibodies recognizing clinically relevant epitopes with insufficient affinity for successful FcεRI receptor crosslinking, sIgE antibodies recognizing clinically irrelevant epitopes, and non-IgE blocking antibodies¹². Hence, deep analysis of monoclonal antibodies (mAbs) from specific B-cells may provide more insights into differences in specific IgE and non-IgE antibody repertoires between allergic and tolerant subjects.

To this end, we analyzed gene sequences encoding the variable region of peanut 2S albumin-specific mAbs from 6 allergic and 6 sensitized but tolerant adults. In particular, sequence motifs of the HCDR3 region, the most important region for recognizing antigens¹³, were hierarchically clustered. Clustering of HCDR3 regions resulted in four motifs exclusively present in allergic donors and three motifs associated with tolerance.

Methods

Blood samples (100 ml) were drawn from allergic (n=6) and tolerant (n=6) adults (≥ 18 years) sensitized to Ara h2 and/or 6 (≥ 0.1 kU/l¹⁴). Classification was achieved by DBPCFC or convincing history (n=2) of tolerated peanut ingestion. Non-atopic donors (n=5) served as reference group. The present study was approved by the ethical committee of the University Medical Center Utrecht (No. 17-945) and informed consent was given by all participants.

2S albumins isolated from roasted peanuts were used for tetramer formation with Streptavidin-APC or -PE^{15, 16}. B-cells double positive for tetramer-PE/APC staining were single cell sorted and the V(D)J gene transcript of the heavy and corresponding light chain were amplified as described previously^{15, 17, 18}. Quality checked sequences of successfully amplified gene transcripts were evaluated with the IgBlast web interface¹⁹. pFUSEss vectors (Invivogen), containing the IgE or IgG1 backbone, were used for antibody cloning and expression in HEK293 cells. Generated mAbs were tested for their specificity to nAra h2 and/or 6²⁰ in comparison with their binding to transferrin. The ability of generated mAbs to induce degranulation was determined by CD63 upregulation of re-sensitized and stimulated human basophils.

Descriptive gene lineage analysis consisted of isotype distribution, mutational status and VH-family gene usage^{21, 22}. Sequence motifs were identified by calculating Levenshtein distances of HCDR3 regions combined with hierarchically clustering. HCDR3 sequences with five or less differences were defined as one motif. Detailed descriptions are shown in Supplementary File 1. These sequence data have been submitted to the GenBank database under submission number 2395667.

Results

Specific IgE levels overlapped between allergic and tolerant donors

In order to study differences in antibody repertoires between adult peanut allergic and peanut sensitized (IgE) but tolerant patients, blood was drawn from peanut allergic (n=6, age: 27-41) and tolerant (n=6, age: 27-63) donors sensitized to Ara h2 and/or 6 (≥ 0.1 kU/l). Specific IgE levels to Ara h2 ranged from 1.0-72 kU/l in the allergic and from 0-1.7 kU/l in the tolerant group. Specific IgE levels to Ara h 6 were lower compared with IgE levels to Ara h2 and ranged from 0-13 kU/l in the allergic and from 0-0.9 kU/l in the tolerant group. Mono-sensitization to either Ara h2 or 6 was detected in both groups, although it occurred more often within the tolerant group (allergic: Pt-01, tolerant: Pt-02, 03, 07) as shown in Table 1.

Frequency of peanut 2S albumin-binding B-cells was significantly higher in allergic donors

As a first step, the frequencies of 2S albumin-binding B-cells were compared between peanut allergic patients, peanut sensitized but tolerant patients, and non-atopic controls. 2S albumin-binding B-cells, double-positive for allergen-tetramer staining, were defined as putatively specific, and their frequency was expressed as percentage of total B-cells acquired from the respective sample. While the frequency in allergic patients (median: 0.01%, 95% CI: 0.005-0.164) was only slightly increased compared to the tolerant patients (median: 0.006%, 95% CI: 0.0016-0.014), the frequency was significantly elevated compared with non-atopic controls (median: 0.002%, 95% CI: 0.0004-0.004, $p=0.008$). Those 2S albumin-binding B-cells were single-cell sorted and served as source for heavy and corresponding light chain gene transcript amplification. No correlation was found between the frequency of 2S albumin-binding B-cells and the number of 2S albumin-binding B-cells from which the heavy and the corresponding light chain gene transcripts were successfully amplified and sequenced ($R^2 = 0.1825$, $p = 0.0872$). A high number of V(D)J gene transcripts were successfully amplified from the B-cells of patients 4 ($n = 63$) and 6 ($n = 61$) – both allergic – while none and only one V(D)J gene transcript was successfully amplified from B-cells of patient 2 (tolerant) and 9 (allergic), respectively. For both of these patients, only a small number of 2S albumin-binding B-cells were sorted (Pt-2: 8, Pt-9: 12). Frequencies, number of sorted 2S albumin-binding B-cells and successfully amplified corresponding V(D)J gene transcripts are shown in Figure 1¹. Taken together, those successfully sequenced gene transcripts - heavy chain: 280, light chain: 221 - provide a good basis for further analyses.

¹Sorted B-cells/amplified corresponding V(D)J gene transcripts - allergic: Pt-1 (32/49), Pt-4 (63/192), Pt-5 (8/96), Pt-6 (61/192), Pt-9 (1/12), Pt-10 (19/96); tolerant: Pt-2 (0/8), Pt-3 (7/108), Pt-7 (13/40), Pt-8 (7/36), Pt-11 (36/144), Pt-12 (17/72); non-atopic: NA-1 (21/50), NA-2 (14/32), NA-3 (7/36), NA-4 (8/48), NA-5 (12/24)

2S albumin-binding B-cells show specific binding to Ara h 2 and 6

In order to verify the specificity of identified 2S albumin-binding B-cells, 42 different sets of heavy and corresponding light chain gene transcripts were randomly heterologously expressed as IgE or IgG1 mAbs to enable the examination of specificity (IgE and IgG1), ability to induce degranulation (IgE) and inhibitory potencies (IgG1, Supplementary Figure S3 and 5). These mAbs derived from allergic (except Pt 9), tolerant (except Pt 2 and 8) and non-atopic donors (except NA-3). The majority of these mAbs (81%, 34/42), mostly derived from allergic and tolerant patients, showed specific binding to a mixture of Ara h 2 and 6 at 10 µg/ml and lower in a concentration-dependent manner (1-10 µg/ml, Figure 2A). These heterologously expressed mAbs originated from 2S albumin-binding IgM+ (41%, 14/34), IgG1+ (3%, 1/34), IgG2+ (23%, 8/34), IgE+ (3%, 1/34), IgA1+ (18%, 6/34) and IgA2+ (9%, 3/34) B-cells. While the two mAbs with relatively high optical density (OD) values originated from IgG1+ and IgA1+ B-cells, mAbs with relatively low OD values originated predominately from IgM+ B-cells. About two-thirds of the specific mAbs (70%, 19/27) tested for separate binding to Ara h 2 or 6 recognized both allergens with comparable OD values. Eight mAbs exclusively recognized either Ara h 2 or 6 (15%, 4/27 Ara h 2 and 15%, 4/27 Ara h 6) (Figure 2B). Notably, the OD values measured for separate binding for either Ara h 2 or 6 tended to be higher compared with the OD values measured for binding to the mixture of Ara h 2 and 6. The remaining antibodies without specific binding to Ara h 2 and 6 at 10 µg/ml originated predominately from IgM+ B-cells of non-atopic controls (88%, 7/8). Additionally, specific binding to the peanut allergens Ara h 1 and 3 was assessed as it has been described that sIgE to Ara h 2 and 6 can cross-react with those peanut allergens²³⁻²⁵. Of our mAbs, 67% (22/33) bound additionally to Ara h 3 while only 9% (3/33) of them bound additionally to Ara h 1 (Figure 2D). These data confirm the specificity of most selected 2S albumin-binding B-cells, which supports the relevance of the gene analyses described in the following paragraphs.

Selected 2S albumin-specific mAbs can induce degranulation

As IgE B-cell memory is potentially stored in IgG1+ memory B-cells and to a lesser extent in B-cells belonging to another isotype, the functionality of mAbs originated from non-IgE 2S albumin-binding B-cells was examined²⁶⁻²⁸. To determine mAb functionality, stripped human basophils were loaded with mixtures of randomly selected mAbs (IgE) and subsequently stimulated with Ara h 2 and 6 (Figure 2C and Supplementary, Figure S4). Stimulated basophils showed individual CD63 upregulation upon stimulation when loaded with mixtures containing the mAb Pt-6-2 (Mix-1, 2, 3 and 6). Such responses were already achieved at low concentrations (5 ng/ml Ara h 2/6). Moreover, the overall CD63 upregulation induced by Mix 3 and 6 was even higher than the CD63 upregulation induced by the anti-IgE control at all allergen concentrations, pointing to an additive effect by using mAb Pt-6-2 (Mix-1) in combination with additional mAbs. The functionality of mAb Pt-6-2 and the observed additive effect confirm the selection of specific and functional B-cells.

Allergic and tolerant donors possess more class-switched 2S albumin-binding B-cells than non-atopic controls

All successfully sequenced heavy chain gene transcripts were initially analyzed for isotype distribution, as class-switching is mostly accompanied by somatic hypermutation maturation upon antigen challenge. Overall, IgM was the most dominant isotype across all groups. Notably, its prevalence was much lower in allergic (47%) and tolerant (53%) donors than in the non-atopic reference group (94%), indicating more antigen-challenged and matured specific B-cells originated from allergic and tolerant patients. Specific binding assessment in Figure 2A and B showed that IgM⁺ B cells derived from the non-atopic control group reflected a certain degree of non-specific selection of IgM⁺ B cells. However, IgM⁺ B cells from allergic and tolerant donors were nearly exclusively specific to Ara h 2 and 6 (93%, 14/15). Moreover, the prevalence of IgA2-expressing B-cells was increased in the tolerant group (20%), while a comparable distribution was observed in allergic and non-atopic donors (2%), as shown in Figure 3A. This finding suggests potential protection by specific IgA in tolerant patients, as IgA is generally able to prevent mucosal antigen crossing²⁹. Regarding IgE, only a small number of V(D)J gene transcripts derived from IgE⁺ B-cells were successfully amplified (allergic: 1; tolerant: 2), possibly due to the extreme low abundance (0.002-1%) of IgE⁺ B-cells within the circulation³⁰. Even though slightly higher proportions are known for IgG3⁺ (~1%) and IgG4⁺ B-cells (~0.75%)^{31, 32}, no V(D)J gene transcripts derived from IgG3⁺ and IgG4⁺ B-cells were amplified. Overall, the individual isotype distribution varied for each patient, e.g. 2S albumin-binding B-cells of the allergic patient 4 were predominantly IgM⁺ B-cells, while those of patient 1 were predominantly IgG⁺ B-cells with a shift to class-switched B-cells in peanut allergic and tolerant patients. This indicates that the 2S albumin-binding B-cells from allergic and tolerant patients were more often matured and antigen-challenged B-cells compared with the non-atopic control group.

IgM⁺ B-cells from allergic donors partly show strong maturation by somatic hypermutations

The introduction of non-silent (mutations on amino acid level) somatic hypermutations can increase target specificity and affinity during B-cell maturation. As shown in Figure 3B, the median number of non-silent mutations was increased in VDJ heavy chain and VJ light chain gene transcripts of class-switched B-cells and was the highest for IgA1⁺ and IgA2⁺ B-cells (heavy chain IgA1: 0-24 mutations, median: 11; heavy chain IgA2: 0-16 mutations, median: 16). The number of non-silent mutations in VDJ heavy chain gene transcripts of IgM-expressing B-cells, on the other hand, ranged from 0 to more than 10 in all study groups. Of note, a small number of IgM⁺ B-cells, mostly originating from the allergic group, consisted of highly mutated heavy chain gene transcripts with mutation numbers comparable to the other isotypes, pointing to potential maturation of 2S albumin-binding IgM⁺ B-cells by antigen challenge.

VH3-family genes are predominately used in gene transcripts of allergic donors

The variable region of the heavy chain gene transcript consists of recombined V, D and J genes. The V gene accounts for most nucleotides of the variable region and contributes, besides D and J genes, to the CDR3 region. Hence, heavy chain V gene (VH) lineage can greatly affect B-cell receptor specificity and affinity. In line with B-cell repertoires described in literature³³, VH3-family genes were observed to dominate across all groups, with a shift to higher proportions (89% of sequences, range: 80-100%, mainly IGHV3-30, VH3-23 and VH3-72) in allergic donors ($\chi^2(2, n=183)=23.67, p<.0001$). This significant difference was still present upon neglecting redundant sequences from the same donor, excluding a bias by proliferation of certain 2S albumin-binding B-cells. In turn, IGHV4 genes accounted for a larger part of incorporated VH-genes in tolerant donors (21% of sequences, range: 7-50%) than in allergic ones (7% of sequences, range: 0-12%). In contrast, no significant differences were detected for light chain V (VL) family gene usage ($\chi^2(2, n=248)=22.90, p=.1165$, Table 2). Furthermore, pairs of VH genes and corresponding VL genes are shown in Figure 3C. Although certain VH-VL pairs (IGHV3-30 – IGKV1 family, IGHV3-72 – IGKV1-9, IGHV3-23 – IGHV3-11) emerged repeatedly in allergic and partly in non-atopic donors (IGHV3-72 – IGKV1-9), most VH-VL pairs appeared to be highly individual; particularly within tolerant donors. Overall, VH-family gene usage, especially the dominant VH3 family gene usage, differs between allergic and tolerant patients sensitized to Ara h2 and/or 6, supporting the hypothesis of differences in allergen-specific antibody repertoires between allergic and tolerant patients.

An IgE+ B-cell is largely clonally related to an IgG1+ B-cell in allergic patients

In total, three VDJ gene transcripts were successfully amplified from IgE+ B-cells, representing the smallest isotype group when disregarding IgG3 and IgG4. Where two VDJ gene transcripts of IgE+ B-cells derived from one tolerant donor were not closely related to another amplified VDJ gene transcript derived from the same donor, the VDJ gene transcript of one IgE+ B-cell originating from an allergic donor (Pt-10) was nearly identical to the gene transcript of an IgG1+ B-cell derived from an unrelated allergic donor (Pt-04). These two sequences differed only in one silent mutation within the FR4 region and one non-silent mutation within the FR2 region, pointing towards conserved clones between unrelated donors ($\cong$ public antibody sequence, Figure 3C). Although it has to be emphasized that only a limited number of IgE+ B cells were selected, this finding may support that public antibody sequences, attributed to peanut allergy, are shared between individuals and isotypes^{27, 28}.

HCDR3 sequence motifs are related to allergy or tolerance

As HCDR3 regions have a great impact on antigen binding, their characteristics were more deeply analyzed. Detailed analysis of physiochemical properties, such as amino acid distribution, aliphatic index, Boman index and Kidera factors 1-10, are shown in the Supplementary, Figure S6 and 7²¹. The Kidera factor 5, expressing the preference to form double-bended structures, was significantly higher for the HCDR3 region of allergic donors compared with that of tolerant donors ($p = 0.0226$, Bonferroni correction: non-significant). Nevertheless, the mean Boman index, describing the theoretical ability to bind proteins, was comparable between HCDR3 regions of allergic, tolerant, and non-atopic donors. For deeper analysis, HCDR3 sequence motifs associated with peanut allergy and/or tolerance, were evaluated by Levenshtein distances (≤ 5 replacements, deletions or insertions) and hierarchical clustering. Levenshtein distances were favored over clonal relationship analysis because of individual genetic variation regarding haplotype diversity, single nucleotide polymorphisms, gene copy number and preferred gene lineage³⁴. For this analysis, all successfully amplified and sequenced heavy chain gene transcripts were used irrespectively of successfully sequenced corresponding light chain gene transcripts. Overall, four unique HCDR3 sequence motifs were associated with peanut allergy and three with peanut tolerance (Figure 3E, Table 3). The most dominant motif was 'CARDSSALEIYNRFDPW' (motif 1), which was derived from 36 different B-cells belonging to three different allergic donors (Pt-1 (n=32), Pt-5 (n=3), Pt-6 (n=1)) including 32 nearly identical B-cells of Pt-1 (monoclonal proliferation). This motif was formed by VH3-30, DH3-3 and JH5 genes. Besides motifs exclusively related to either allergy or tolerance, five motifs were shared between allergic, tolerant, and non-atopic donors. It must be mentioned that one motif (CARNVFDGYWL VYW) associated with tolerance was only found in Pt-11 and no motif was shared between all allergic or tolerant donors. In comparison, LCDR3 motifs, defined as Levenshtein distances ≤ 2 based on reduced heterogeneity of LCDR3 sequences compared to HCDR3 sequences, were detected in allergic (n=7) and tolerant donors (n=1). However, more dominant motifs (n=10) containing up to 24 LCDR3 sequences were shared between allergic, tolerant and non-atopic donors (Figure 3F, detailed description Supplementary Table S2). Heterologously expressed mAbs corresponding to these motifs showed specific binding to Ara h2 and/or 6, supporting the identification of public antibody sequences potentially attributed to peanut allergy or tolerance despite sIgE sensitization.

Discussion

Specific IgE levels to Ara h2 and 6 between 0.1 and 1.8 kU/l overlap between peanut allergic and tolerant subjects^{6, 7}, risking inaccurate diagnosis, and therefore indicating the need for new diagnostic strategies. In the present study, we observed a preferential usage of VH3-family genes in peanut 2S albumin-specific B-cells from peanut allergic patients. Additionally, we identified 2S albumin-binding B-cells carrying HCDR3 sequence motifs either related to peanut allergy or tolerance. These findings indicate the association of certain public antibody sequences to allergy or tolerance in sensitized patients with suspected peanut allergy.

Despite a large number of amplified heavy chain VDJ gene transcripts (n=280), only three of them belonged to IgE class-switched B-cells. This low abundance may be potentially explained by dominant class-switching to IgE in the tissue²⁶, low BCR surface expression of plasmablast-like IgE⁺ B-cells^{23, 35, 36} and the extreme low abundance of IgE⁺ B-cells within the circulation³⁰. Nevertheless, the relevance of the B-cells sorted within the present study is supported by the close relationship between one heavy chain gene transcript from an IgE⁺ B-cell and a gene transcript from an IgG1⁺ B-cell derived from an unrelated peanut allergic donor. Additionally, some generated mAbs were able to slightly inhibit patient-derived serum IgE-binding to Ara h2 and 6 and to partly inhibit activation of basophils loaded with serum IgE upon stimulation with Ara h 2 and 6 (Supplementary, Figure S3 and S5). Correspondingly, there is evidence that human IgE⁺ B-cells are predominately plasmablasts or plasma cells generated by sequential class-switching from IgG1⁺ B-cells (γ switch region remnants) and that the humoral IgE memory is contained in IgG⁺ memory B-cells²⁸. Further evidence for this theory was obtained by clonal relationship analyses, since IgE⁺ B-cells were dominantly clonally related to IgG1⁺ B-cells, but also, to a lesser extent, to B-cells of other isotypes^{23, 26, 27}.

The majority of mAbs generated in the present study showed specific binding to peanut 2S albumins and two-thirds of these mAbs showed specific binding to both allergens Ara h 2 and 6, pointing towards a high degree of cross-reactivity. Strikingly, 67% of the mAbs were even able to bind, partly to a lesser extent, to the non-homologous peanut allergen Ara h 3. In current literature, a certain degree of serum IgE cross-reactivity has been reported between Ara h 2 and 6^{37, 38} and even serum IgE cross-reactivity between Ara h 2 and the non-homologous allergens Ara h 1 and 3 has been described by Bublin and colleagues²⁴. Auxiliary, mAbs derived from peanut allergic patients showed cross-reactivity between Ara h 2 and Ara h 3 but not to Ara h 1²³.

Regarding isotype distribution of the selected 2S albumin-binding B-cells, IgM⁺ B-cells represented the largest isotype group, as detected in respectively 47, 53 and 94% of allergic, tolerant and non-atopic donors. Additionally, IgM⁺ B-cells from non-atopic donors, all specific for Ara h2 and/or 6, shared HCDR3 motifs (motif 4, 8, 10) with IgM⁺, IgG2⁺ and IgA1⁺ B-cells from allergic donors. Clonal relationships between IgA/G/M⁺ and IgE⁺ B-cells have been described for B-cells derived from gut tissues

by Hoh and colleagues^{26, 39}. Combining the findings from the present study with the finding of Hoh and colleagues leads to the suggestion that non-atopic donors can potentially possess non-IgE antibodies with required specificity or affinity to theoretically induce an allergic reaction.

Moreover, tolerant donors showed a higher proportion of IgA⁺ B-cells and tended to have higher specific IgA serum levels than allergic donors (Supplementary, Figure S3). Allergen-specific IgA serum levels have been shown to be increased in peanut allergic subjects undergoing oral compared with their baseline levels²⁹. Moreover, salivary IgA levels have been closely associated with the degree of tolerance induction by sublingual immunotherapy confirmed by DBPCFCs, pointing to a protective role of specific IgA against mucosal allergen absorption⁴⁰. Such a protective role is supported by increased intestinal permeability in the absence of IgA in mice⁴¹. Taken together, these findings suggest a potential protective effect of allergen-specific IgA in the mucosa of sensitized but tolerant patients.

Regarding gene lineage, VH3-family genes were significantly more often used ($p < .0001$) in heavy chain gene transcripts of peanut 2S albumin-specific B-cells from allergic donors (89%) than in those from tolerant (54%) and non-atopic donors (63%). Previous studies on VH-gene usage of heavy chain gene transcripts in healthy donors showed VH3-family gene usage in 40 to 65% of the B-cells, which was comparable to the usage in the tolerant and non-atopic groups in our study (Table 2)^{42, 43}. These findings suggest a shift towards VH3-family gene usage in 2S albumin specific B-cells from allergic donors. Contrary to our findings, other datasets of Ara h2 specific B-cells did not observe a similar shift in VH-gene usage^{15, 18}. These datasets, however, were generated from patients undergoing peanut oral immunotherapy with a focus on different time points during immunotherapy. This fact hampers the comparison between the present and previous Ara h2-related datasets. Moreover, conflicting results regarding preferred VH-gene usage were described for different food and respiratory allergies, with a shift to VH3 usage for anti- α Gal antibodies and anti-grass pollen Phl p6 and 11 antibodies^{44, 45}, indicating that the preferred VH-family gene usage may be allergen dependent. Moreover, the preferred usage might depend on the status of allergy or tolerance and can potentially be used for diagnostic purposes upon validation in a larger patient cohort. A validation study will also provide information about the number of detected 2S albumin-binding B-cells is sufficient for a diagnostic workflow.

Besides differences in VH-family gene usage, certain HCDR3 sequence motifs were associated with either peanut allergy or tolerance and appear to have the ability to discriminate between those two groups. The main HCDR3 sequence motif 'CARDSSALEIYNRFDPW' was associated with peanut allergy and derived from recombined VH3-30, DH3-3 and JH5 genes. Comparably, VH3-30*18 was incorporated in the VDJ gene transcript of clonally related IgE⁺ B-cells specific for Ara h2 in the study of Croote and colleagues²³. Additionally, a highly similar HCDR3 region (CAREGYESSGFDYW) to motif 6 (CAREGHYSSQFDYW), associated with tolerance, has been described for peanut allergic subjects undergoing oral immunotherapy¹⁸. Oral immunotherapy may shape the antibody repertoire towards

Accepted Article

repertoires present in tolerant subjects. These comparisons support the identification of public antibody sequences potentially attributed to peanut allergy or tolerance.

In conclusion, peanut allergy is associated with dominant usage of VH3-family genes and certain public antibody sequences, expressed as HCDR3 sequence motifs, are attributed to either peanut allergy or tolerance despite sIgE sensitization. Validation of these findings in a larger patient cohort with higher ethnical diversity, leading to different gene allele usage, may be achieved using next-generated sequencing approaches, potentially combined with the sorting strategy of 2S albumin-binding B-cells presented in this study..

References

1. Muraro A, Werfel T, Hoffmann-Sommergruber K, Roberts G, Beyer K, Bindslev-Jensen C, et al. EAACI Food Allergy and Anaphylaxis Guidelines: diagnosis and management of food allergy. *Allergy* 2014; 69; doi: 10.1111/all.12429.
2. Yanagida N, Sato S, Asaumi T, Ogura K, Ebisawa M. Risk Factors for Severe Reactions during Double-Blind Placebo-Controlled Food Challenges. *Int Arch Allergy Immunol* 2017; 172:173-82; doi: 10.1159/000458724.
3. Portnoy J. Appropriate Allergy Testing and Interpretation. *Mo Med* 2011; 108:339-43.
4. Yu W, Freeland D, Nadeau K. Food allergy: immune mechanisms, diagnosis and immunotherapy. *Nat Rev Immunology* 2016; 16:751-65; doi: 10.1038/nri.2016.111.
5. Hemmings O, Du Toit G, Radulovic S, Lack G, Santos A. Ara h 2 is the dominant peanut allergen despite similarities with Ara h 6. *J Allergy Clin Immunol* 2020; 146:621-30.e5.
6. Klemans R, Broekman H, Knol E, Bruijnzeel-Koomen C, Otten H, Pasmans S, et al. Ara h 2 is the best predictor for peanut allergy in adults. *J Allergy Clin Immunol Pract* 2013; 1:632-8.e1; doi: 10.1016/j.jaip.2013.07.014.
7. Kansen H, van Erp F, Knulst A, Ehlers A, Lyons S, Knol E, et al. Peanut allergy is accurately predicted in one third of adults with suspected peanut allergy using specific IgE to Ara h 2 and Ara h 6. EAACI. London, 2020.
8. Santos A, James L, Bahnson H, Shamji M, Couto-Francisco N, Islam S, et al. IgG4 inhibits peanut-induced basophil and mast cell activation in peanut-tolerant children sensitized to peanut major allergens. *J Allergy Clin Immunol* 2015; 135:1249-56.
9. Strait R, Morris S, Finkelman F. IgG-blocking antibodies inhibit IgE-mediated anaphylaxis in vivo through both antigen interception and FcγRIIb cross-linking. *J Clin Invest* 2006; 116:833-41.
10. Knol E. Requirements for effective IgE cross-linking on mast cells and basophils. *Mol Nutr Food Res* 2006; 50:620-4.
11. Ehlers A, Blankestijn M, Knulst A, Klinge M, Otten H. Can alternative epitope mapping approaches increase the impact of B-cell epitopes in food allergy diagnostics? *Clin Exp Allergy* 2019; 49:17-26; doi: 10.1111/cea.13291.
12. Bucaite G, Kang-Pettinger T, Moreira J, Gould H, James L, Sutton B, et al. Interplay between Affinity and Valency in Effector Cell Degranulation: A Model System with Polcalcin Allergens and Human Patient-Derived IgE Antibodies. *J Immunol* 2019; 203:1693-700; doi: 10.4049/jimmunol.1900509.

- Accepted Article
13. Xu J, Davis M. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 2000; 13:37-45; doi: 10.1016/s1074-7613(00)00006-6.
 14. Martinet J, Couderc L, Renosi F, Bobée V, Marguet C, Boyer O. Diagnostic Value of Antigen-Specific Immunoglobulin E Immunoassays against Ara h 2 and Ara h 8 Peanut Components in Child Food Allergy. *Int Arch Allergy Immunol* 2016; 169:216-22; doi: 10.1159/000446181.
 15. Patil S, Ogunniyi A, Calatroni A, Tadigotla V, Ruiter B, Ma A, et al. Peanut oral immunotherapy transiently expands circulating Ara h 2-specific B cells with a homologous repertoire in unrelated individuals. *J Allergy Clin Immunol* 2015; 136:125-34; doi: 10.1016/j.jaci.2015.03.026.
 16. Franz B, May K, Dranoff G, Wucherpfennig K. Ex vivo characterization and isolation of rare memory B cells with antigen tetramers. *Blood* 2011; 118:348-57; doi: 10.1182/blood-2011-03-341917.
 17. Tiller T, Meffre E, Yurasov S, Tsuiji M, Nussenzweig M, Wardemann H. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J Immunol Methods* 2008; 329:112-24; doi: 10.1016/j.jim.2007.09.017.
 18. Hoh R, Joshi S, Liu Y, Wang C, Roskin K, Lee J, et al. Single B cell Deconvolution of Peanut-specific Antibody Responses in Allergic Patients. *J Allergy Clin Immunol* 2016; 137:157-67; doi: 10.1016/j.jaci.2015.05.029.
 19. Ye J, Ma N, Madden T, Ostell J. IgBLAST: An Immunoglobulin Variable Domain Sequence Analysis Tool *Nuc Acid Res* 2013; 41:W34-40; doi: 10.1093/nar/gkt382.
 20. Koppelman S, De Jong G, Laaper-Ertmann M, Peeters K, Knulst A, Hefle S, et al. Purification and immunoglobulin E-binding properties of peanut allergen Ara h 6: evidence for cross-reactivity with Ara h 2. *Clin Exp Allergy* 2005; 35:490-7; doi: 10.1111/j.1365-2222.2005.02204.x.
 21. Margreitter C, Lu H-C, Townsend C, Stewart A, Dunn-Walters D, Fraternali F. BRepertoire: a user-friendly web server for analysing antibody repertoire data. *Nuc Acid Res* 2018; 46:W264-70; doi: 10.1093/nar/gky276.
 22. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, et al. Clustal W and Clustal X version 2.0 *Bioinformatics* 2007; 23:2947-8; doi: 10.1093/bioinformatics/btm404.
 23. Croote D, Darmanis S, Nadeau K, Quake S. High-affinity allergen-specific human antibodies cloned from single IgE B cell transcriptomes. *Science* 2018; 362:1306-9; doi: 10.1126/science.aau2599.
 24. Bublin M, Kostadinova M, Radauer C, Hafner C, Szépfalusi Z, Varga EM, et al. IgE cross-reactivity between the major peanut allergen Ara h 2 and the nonhomologous allergens Ara h 1 and Ara h 3 *J Allergy Clin Immunol* 2013; 132:118-24.

25. Smit JJ, Pennings MT, Willemsen K, van Roest M, van Hoffen E, Pieters RH. Heterogeneous responses and cross reactivity between the major peanut allergens Ara h 1, 2,3 and 6 in a mouse model for peanut allergy Clin Transl Allergy 2015; 5:13.
26. Hoh R, Joshi S, Lee J, Martin B, Varma S, Kwok S, et al. Origins and clonal convergence of gastrointestinal IgE+ B cells in human peanut allergy. Sci Immunol 2020; doi: 10.1126/sciimmunol.aay4209.
27. Looney T, Lee J, Roskin K, Hoh R, King J, Glanville J, et al. Human B-cell isotype switching origins of IgE. J Allergy Clin Immunol 2016; 137:579-86; doi: 10.1016/j.jaci.2015.07.014.
28. Saunders S, Ma E, Aranda C, Curotto de Lafaille M. Non-classical B Cell Memory of Allergic IgE Responses. Front immunol 2019; 10:715; doi: 10.3389/fimmu.2019.00715.
29. LaHood N, Patil S, Shreffler W, Virkud Y, Ruiter B, Tang H, et al. Ara h 2 Specific IgA B Cell Repertoire Matures During Peanut Oral Immunotherapy. J Allergy Clin Immunol 2020; 145.
30. Jiménez-Saiz R, Ellenbogen Y, Bruton K, Spill P, Sommer D, Lima H, et al. BCR analysis of single-sorted, putative IgE+ memory B cells in food allergy: an ephemeral existence? J Allergy Clin Immunol 2019; 144:336-9; doi: 10.1016/j.jaci.2019.04.001.
31. Karagiannis P, Villanova F, Josephs D, Correa I, Van Hemelrijck M, Hobbs C, et al. Elevated IgG4 in patient circulation is associated with the risk of disease progression in melanoma. Oncoimmunology 2015; 4; doi: 10.1080/2162402X.2015.1032492.
32. Marsh-Wakefield F, Ashhurst T, Trend S, McGuire H, Juillard P, Zinger A, et al. IgG3 + B cells are associated with the development of multiple sclerosis. Clin Trans Immunology 2020; 9:e01133; doi: 10.1002/cti2.1133.
33. Brezinschek H, Foster S, Brezinschek R, Dörner T, Domiati-Saad R, Lipsky P. Analysis of the human VH gene repertoire. Differential effects of selection and somatic hypermutation on human peripheral CD5(+)/IgM+ and CD5(-)/IgM+ B cells. J Clin Invest 1997; 99:2488-501; doi: 10.1172/JCI119433.
34. Watson C, Glanville J, Marasco W. The Individual and Population Genetics of Antibody Immunity. Trends Immunol 2017; 38:459-70; doi: 10.1016/j.it.2017.04.003.
35. Karnowski A, Acatz-Straussberger G, Klockenbusch C, Achatz G, Lamers M. Inefficient processing of mRNA for the membrane form of IgE is a genetic mechanism to limit recruitment of IgE-secreting cells. Eur J Immunol 2006; 36:1917-25; doi: 10.1002/eji.200535495.
36. Yang Z, Sullivan B, Allen C. Fluorescent in vivo detection reveals that IgE(+) B cells are restrained by an intrinsic cell fate predisposition. Immunity 2012; 36:857-72; doi: 10.1016/j.immuni.2012.02.009.

37. Blankestijn M, Otten H, Suer W, Weimann A, Knol E, Knulst A. Specific IgE to peanut 2S albumin Ara h 7 has a discriminative ability comparable to Ara h 2 and 6. *Clin Exp Allergy* 2018; 48:60-5.
38. Hazebrouck S, Guillon B, Paty E, Dreskin SC, Adel-Patient K, Bernard H. Variable IgE cross-reactivity between peanut 2S-albumins: The case for measuring IgE to both Ara h 2 and Ara h 6. *Clin Exp Allergy* 2019; 49:1107-15.
39. Iweala O, Burks A. IgE producers in the gut expand the gut's role in food allergy. *Nat Rev Gastroenterol Hepatol* 2020; 17:384-6; doi: 10.1038/s41575-020-0309-5.
40. Kulis M, Saba K, Kim E, Bird J, Kamilaris N, Vickery B, et al. Increased peanut-specific IgA levels in saliva correlate with food challenge outcomes after peanut sublingual immunotherapy. *J Allergy Clin Immunol* 2012; 129:1159-62; doi: 10.1016/j.jaci.2011.11.045.
41. Strait R, Mahler A, Hogan S, Khodoun M, Shibuya A, Finkelman F. Ingested allergens must be absorbed systemically to induce systemic anaphylaxis. *J Allergy Clin Immunol* 2012; 127:982-9.e1; doi: 10.1016/j.jaci.2011.01.034.
42. Graichy M, Galson J, Kovaltsuk A, von Niederhäusern V, Pachlopnik Schmid J, Recher M, et al. Maturation of the Human B-Cell Receptor Repertoire with Age SSRN Electronic Journal 2020; doi: 10.2139/ssrn.3518535.
43. Goldstein L, Chen Y-J, Wu J, Chaudhuri S, Hsiao Y-C, Schneider K, et al. Massively Parallel Single-Cell B-cell Receptor Sequencing Enables Rapid Discovery of Diverse Antigen-Reactive Antibodies *Commun Biol* 2019; 2:304; doi: 10.1038/s42003-019-0551-y.
44. Wang L, Radic M, Galili U. Human anti-Gal Heavy Chain Genes. Preferential Use of VH3 and the Presence of Somatic Mutations. *J Immunol* 1995; 155:1276-85.
45. Persson H, Karbalaeei Sadegh M, Greiff L, Ohlin M. Delineating the specificity of an IgE-encoding transcriptome. *J Allergy Clin Immunol* 2007; 120:1186-92; doi: 10.1016/j.jaci.2007.06.041.

Tables

Table 1: Patient characteristics

	Sex	Age	Last reaction	Müller score†	Subj. ED [mg]	Obj. ED [mg]	Ara h2 [kU/l]	Ara h 6 [kU/l]	Total IgE [kU/l]
Allergic group									
Pt-05	f	34	3 years	3	10	3000	1.0	1.4	625
Pt-09	f	29	1 year	3	4	4000	2.4	5.3	807
Pt-01	f	39	1.5 years	1	100	3000	3.3	0	14.5
Pt-04	m	27	7 years	1	10	-	8.2	1.7	417
Pt-10	f	27	3 years	4	100	100	23	7.9	252
Pt-06	m	41	2 years	3	100	100	72	13.1	83.3
	Sex	Age	Last consumption	Amount	Processing/ Manufacturing		Ara h2 [kU/l]	Ara h 6 [kU/l]	Total IgE [kU/l]
Tolerant group									
Pt-02	m	45	< 2 weeks	< 1 g	contained in products		0	0.9	14.2
Pt-07	f	29	< 2 weeks	< 1 g	contained in products		0.2	0	250
Pt-08	m	36	< 2 weeks	> 20 g	roasted		0.6	0.2	98.3
Pt-11	m	27	1-3 days	> 20 g	roasted, peanut butter		1.3	0.8	>1000
Pt-12	f	31	< 2 weeks	< 1 g	contained in products		2.3	0.3	>1000
Pt-03	m	63	2 days	> 20 g	peanut butter		1.7	0	208

† Müller score 0: oral cavity, 1: skin and mucous membranes, 2: gastrointestinal, 3: respiratory, 4: cardiovascular

Table 2: Comparison of the VH and VL-gene family usage between the present study, healthy controls and Ara h2-related datasets (OIT trials); (transcripts from allergic, n = 120; tolerant, n = 63; non-atopic, n = 52)

VH/L family	Allergic group	Tolerant group	Non-atopic group	Hoh 2016 ¹⁸	Patil 2015 ¹⁵	Goldstein 2019 ⁴³	Graichy [†] 2020 ⁴²
VH1 [%]	2	21	3.8	18	23	15-18	5-18
VH2 [%]	-	-	-	-	-	4	2
VH3 [%]	89 [‡]	54	63	60	62	46-50	40-65
VH4 [%]	6.3	21	29	15	15	22-25	25-35
VH5 [%]	1.3	2.9	3.8	-	-	5-7	1-2
VH7 [%]	1.3	-	-	-	-	0.05-1	0.5-1
VK1 [%]	54.5	52.6	46.2	-	-	-	-
VK2 [%]	3.3	10.5	2.6				
VK3 [%]	42.3	34.2	52.6				
VK4 [%]	-	2.6	-				
VL1 [%]	31	17	33	-	-	-	-
VL2 [%]	45	67	44				
VL3 [%]	3.5	8	22				
VL4 [%]	3.5	-	-				
VL7 [%]	17	-	-				
VL8 [%]	-	8	-				

[†] Age group between 25-40 years used for comparison, most closely related to the average age of the study population; data for different subsets were combined

[‡]dominant usage VH3-family genes in all individuals

Table 3: HCDR3 sequence motifs associated with peanut allergy or tolerance

bold: differences of HCDR3 sequences; italic: tolerant donors; bold patient numbers: indicate heterologous expression mAb, n. e. = not expressed

Motif	HCDR3 sequence	Antibodies	Isotype	Specificity
Allergy				
1	CARDSSALEIYNRFDPW	Pt-1-1 – Pt-1-32 [†] , Pt-5-3, Pt-5-4, Pt-5-8 [†] , Pt-6-43 [†]	IgG2	Ara h2 + 6
2	CVKDRQQYSSRWLDSW	Pt-5-2, Pt-5-5, Pt-6-32 [†] , Pt-6-42, Pt-6-45, Pt-6-47, Pt-6-48	IgA1	Ara h2 + 6
3	CASMDILAANTHFGMDVW	Pt-6-17 [†] , Pt-6-37, Pt-6-52 [†] , Pt-10-19 [†]	IgG2	Ara h2 + 6
4	CARGLVGANFYYYMDVW	Pt-4-10 [†] , Pt-4-41, Pt-4-61	IgM	Ara h2
4	CARGRSGATYYYYMDVW	NA-4-5 [†]	IgM	Ara h2 + 6
4	CARGRAGPSYYYYMDVW	Pt-4-44	IgM	
Tolerance				
5	CARNVFDGYWLVYW	<i>Pt-11-3, Pt-11-4, Pt-11-14[†], Pt-11-15, Pt-11-23, Pt-11-24, Pt-11-25, Pt-11-26</i>	IgA2	Ara h2 + 6
6	CAREGHYSSQFDYW	<i>Pt-8-3, Pt-8-4, Pt-8-5, NA-2-3</i>	IgM	n. e.
7	CARDYGGYPHAAFDIW	<i>Pt-11-11</i>	IgM	n. e.
7	CTRDTGTYPHAAFNW	<i>Pt-3-1, Pt-3-5</i>	IgM, IgG2	Ara h2 + 6
Shared				
8	CTRPYRAFNAIGHW	NA-1-15	IgM	n. e.
8	CTRPYRAFNWATGHW	NA-1-17, NA-5-2 [†] (only Ara h2), NA-5-5 [†] , NA-5-11, Pt-6-19	IgM, IgA1	Ara h2 + 6
9	CARVSSSWHTEYW	<i>Pt-3-7, Pt-6-33, Pt-6-46[†], Pt-6-53, Pt-6-60, Pt-6-61</i>	IgA1	Ara h2 + 6
10	CARGIIDKYGMDVW	Pt-6-38, <i>Pt-7-6</i> , Pt-10-13, Pt-10-14	IgG2	n. e.
10	CAREYYYGMDVW	NA-4-1	IgM	n. e.
10	CARTLGYGMDVW	<i>Pt-11-16</i>	IgM	n. e.
11	CTRGAVSYTRHFQFW	Pt-4-20 [†]	IgM	Ara h2
11	CVRGALAYTRHFQYW	Pt-4-34 [†]	IgM	
11	CVRGAMSNTRHFQYW	Pt-4-28 [†]	IgM	Ara h2/6 [‡]
11	CARGAMSYTRHFQYW	Pt-4-8, Pt-4-40, Pt-4-63 [†] , <i>Pt-8-6</i>	IgM	Ara h2/6 [‡]

12	CAKAYGSGSYLFDYW	<i>Pt-7-4</i> , NA-1-10 †	IgM	Ara h2 + 6
12	CARGGSGSYTFDYW	NA-1-4, NA-1-6, NA-1-13	IgM	
12	CARSGGSGSYTFDYW	Pt-10-6 †	IgM	Ara h2 + 6

†specificity to Ara h2/6 proven

‡only specificity against mixture of Ara h2 and 6

Figure Legends

Figure 1 Frequency of 2S albumin-binding B-cells

Frequency of 2S albumin-binding B-cells, double-positive for allergen-tetramer (PE and APC) staining, expressed as percentage of the total B-cell number acquired from the respective sample, indicated with the median. Each frequency is labelled with the respective patient number and the number of successfully amplified corresponding heavy and light V(D)J gene transcripts in relation to sorted B-cells. Statistical significance was proven with the Kruskal-Wallis test and corrected for multiple testing using the Dunn's multiple comparison test (adjusted p-value); all = allergic, tol = tolerant, n-a = non-atopic

Figure 2 Specificity and functionality of heterologously expressed mAbs

A Concentration-dependent (1 to 10 $\mu\text{g/ml}$) binding of heterologously expressed mAbs ($n = 42$) to a mixture of Ara h 2 and 6 defined as OD values corrected for their binding to transferrin; filled dots indicate mAbs expressed as IgE and open dots mAbs expressed as IgG1; bold dash line is the cut-off for IgE mAbs and non-bold dash line the cut-off for IgG1 antibodies; dot = allergic (Pt 1 = red, Pt 4 = black, Pt 5 = orange, Pt 6 = green, Pt 10 = blue, square = tolerant (Pt 3 = red, Pt 7 = black, Pt 11 = green, Pt 12 = blue), triangle = non-atopic (NA 1 = red, NA 2 = black, NA 4 = green, NA 5 = blue); A = allergen, T= transferrin

B Binding of heterologously expressed mAbs to Ara h2 and 6 ($n = 33$, 10 $\mu\text{g/ml}$) defined as OD values corrected for their binding to transferrin; filled dots indicate mAbs expressed as IgE and open dots mAbs expressed as IgG1, bold dash line is the cut-off for IgE mAbs and non-bold dash line the cut-off for IgG1 mAbs; dot = allergic (Pt 1 = red, Pt 4 = black, Pt 5 = orange, Pt 6 = green, Pt 10 = blue, square = tolerant (Pt 3 = red, Pt 7 = black, Pt 11 = green, Pt 12 = blue), triangle = non-atopic (NA 1 = red, NA 2 = black, NA 4 = green, NA 5 = blue); A = allergen, T= transferrin

C Binding of heterologously expressed mAbs to Ara h 1, 2, 3 and 6 ($n = 33$, 5 $\mu\text{g/ml}$) defined as OD values corrected for their binding to transferrin and the respective blank value (HEK cell supernatant); heat map at the top: binding of mAbs expressed as IgE, heat map at the bottom: binding of mAbs expressed as IgG1, red star = corrected OD value > 1.0

D Upregulation of CD63 upon loading of human basophils with mAb mixture and subsequent stimulation with Ara h2 and 6 in a concentration-dependent manner (5 to 1000 ng/ml); the dots indicate the mean of duplicate measurements and their range; Mix 1 = Pt-6-2, Mix 2 = Pt-6-2 + Pt-6-17, Mix 3 = Pt-6-2 + Pt-3-2 + Pt-6-11, Mix 4 = Pt-6-17, Pt-3-2, Pt-6-11, Mix 5 = Pt-1-32, Pt-6-11, Pt-6-6, Pt-6-8, Mix 6 = NA-1-10 +

Pt-4-28 + Pt-4-34 + Pt-6-2, Mix 7 = Pt NA-1-10 + Pt-4-28 + Pt-4-34, Mix 8 = Pt-1-32 + Pt-6-15 + Pt-6-21, stripped basophils before loading, Rituximab expressed as IgE; Supplementary, Figure S2: comparison to stimulation with cow's milk extract, dash line: threshold level set to 5%

Figure 3: Descriptive gene analysis and motif analysis of HCDR3 regions

Gene lineage analysis of successfully sequenced heavy chain gene transcripts defined as productive (allergic n = 151; tolerant n = 68; non-atopic, n = 51), irrespectively of successfully sequenced corresponding light chain gene transcript

A Isotype distribution split for allergic (all), tolerant (tol) and non-atopic (n-a) donors

B Numbers of non-silent mutations located on the variable region of the heavy and light chain expressed with the median (95% CI) and split for allergic (all), tolerant (tol) and non-atopic (n-a) donors; filled dots: variable region heavy chain, open dots: variable region light chain

CV gene lineage of the heavy and corresponding light chain was visualized with circled diagrams showing the heavy chain at the bottom connected to the light chain (grey) at the top. VH1 family genes are shown in red, VH3 family genes in blue, VH4 family genes in green and VH5/7 family genes in orange.

DC clonotypes were defined as identical V gene, J gene and HCDR3 length and related clones were aligned using ClustalW2. The IgE mAb of Pt-10 was highly related to an IgG1 antibody of Pt-4.

EH CDR3 sequence motifs were analyzed by Levenshtein distances and hierarchical clustering (detailed description in the method section) and sequences with distances ≤ 5 were grouped into one motif. More detailed description of the motifs is presented in Table 3.

FL CDR3 sequence motifs were analyzed by Levenshtein distances and hierarchical clustering (detailed description in the method section) and sequences with distances ≤ 2 were grouped into one motif. More detailed description of the motifs is presented in Table S2.

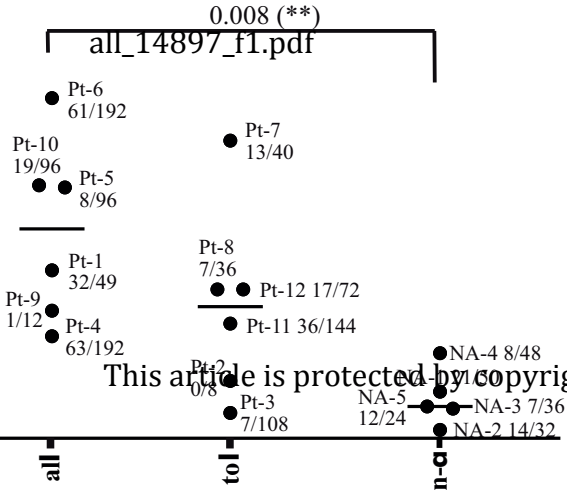
2S albumin-binding B cells
(% CD19+ B cells)

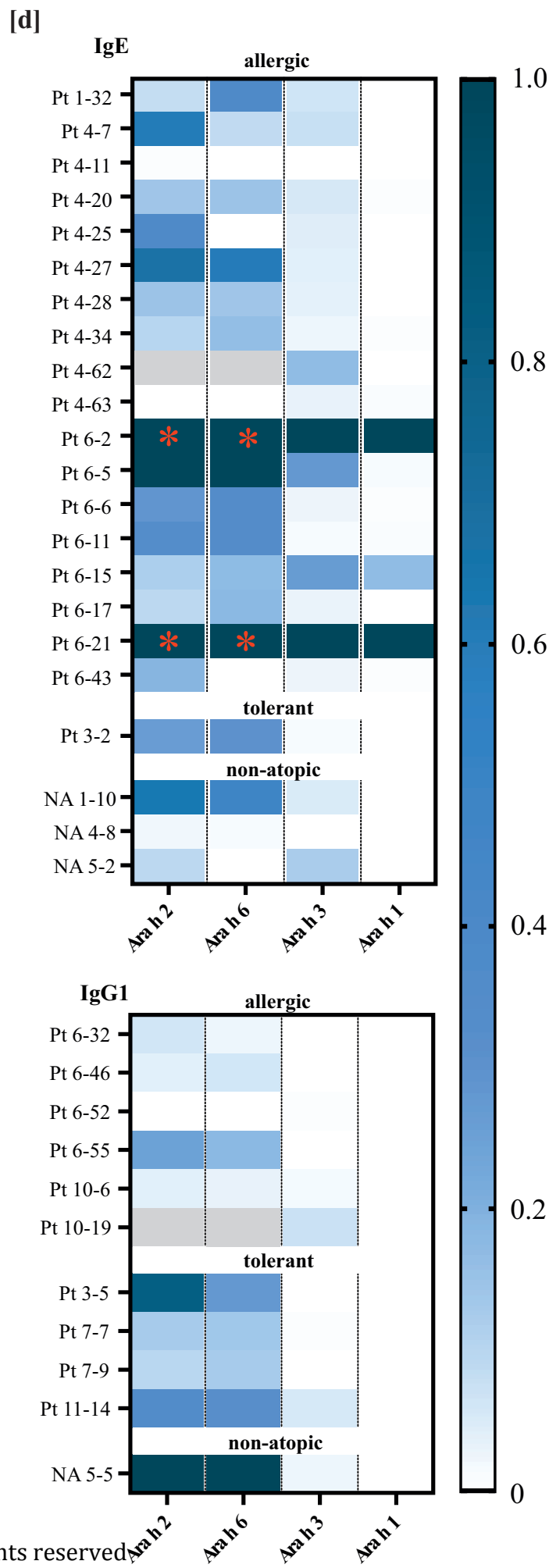
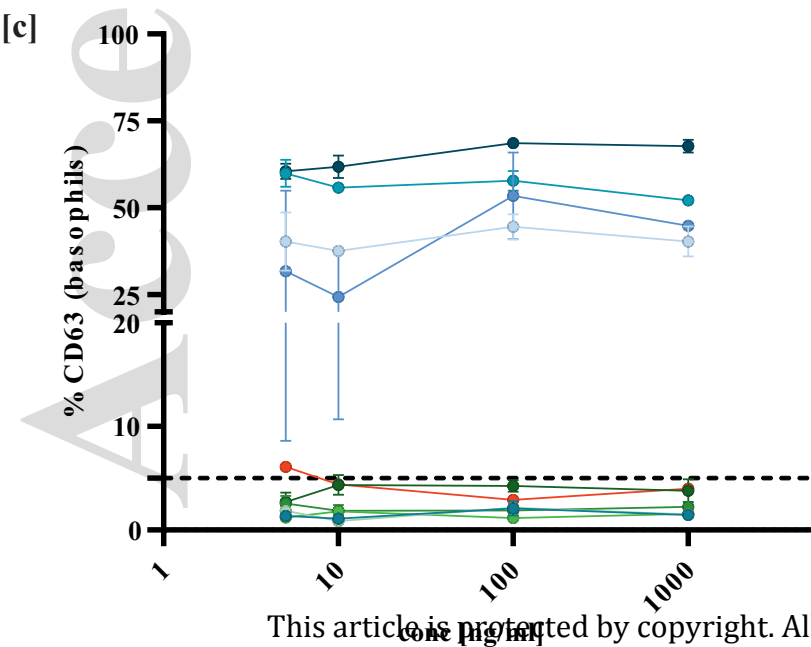
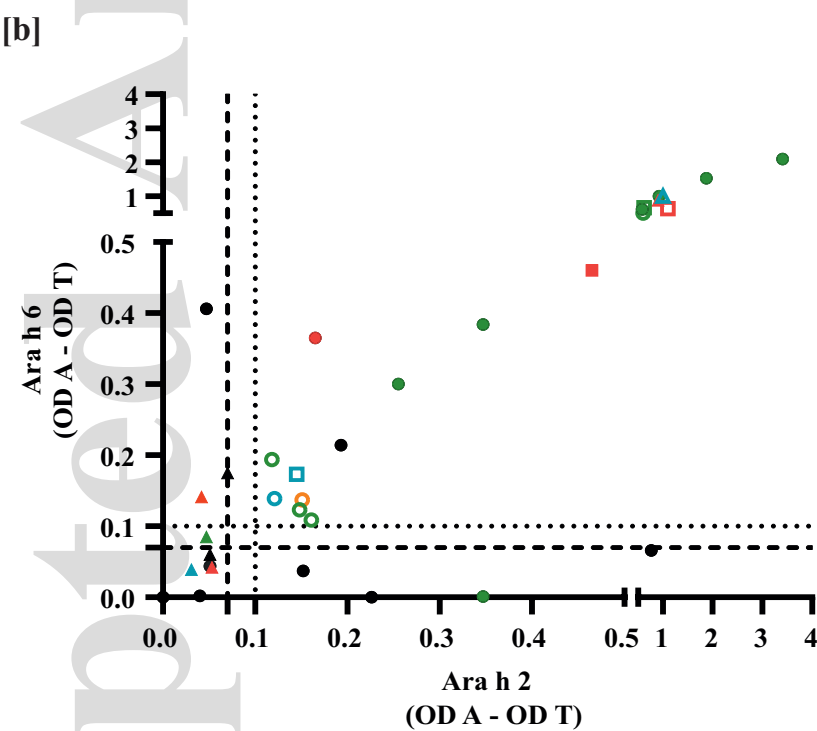
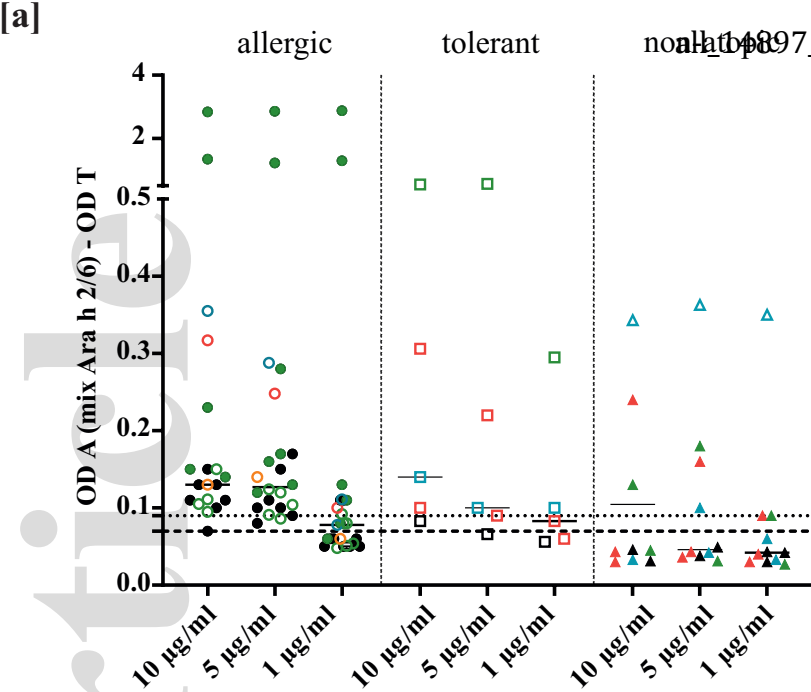
0.020
0.015
0.010
0.005
0.000

all

tol

n-a





Supplementary File 1: Materials and methods

1 Approach

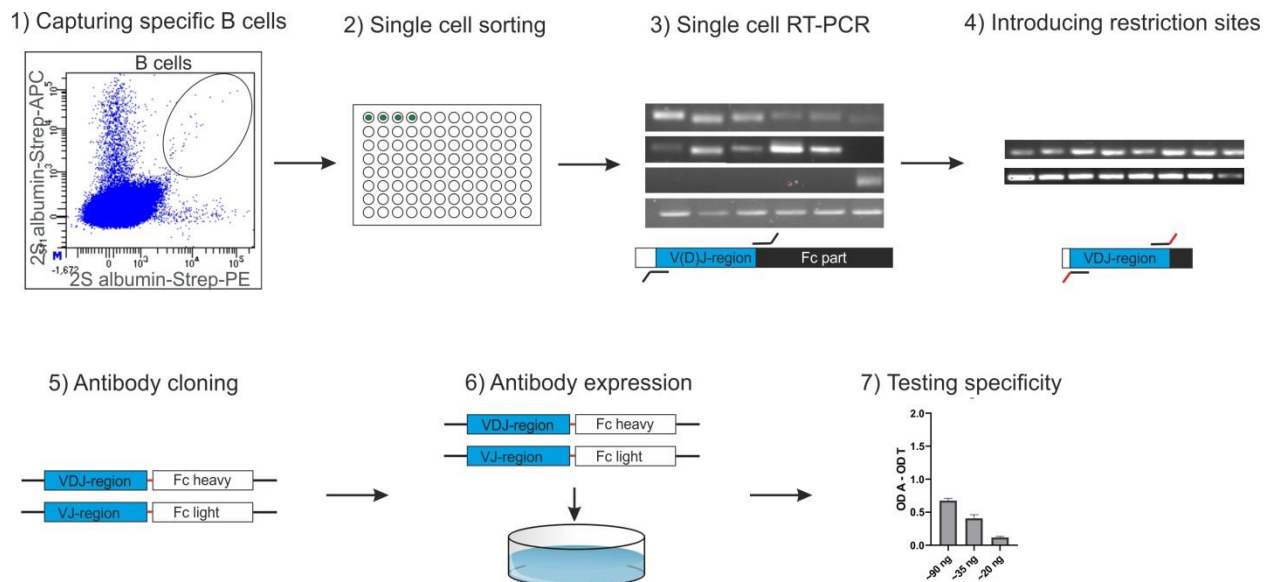


Figure 1: Peanut 2S albumin-binding B cells, double-positive for allergen tetramer-staining, were single cell sorted. The variable gene transcript of the heavy and the corresponding light chain were amplified and subsequently sequenced by Sanger Sequencing. Successfully amplified and sequenced variable regions were prolonged with restriction sites for cloning them into a vector containing the backbone of IgE, IgG1 or IgG4. Plasmids checked for sequence correctness were heterologously expressed in HEK293 cells and the expressed monoclonal antibodies were screened for their binding to native Ara h 2 and 6.

Patient selection

Blood samples (100 ml) were drawn once from allergic (n=6) and tolerant (n=6) adults sensitized to Ara h 2 and/or Ara h 6 (≥ 0.1 kU/l¹). Their diagnoses were confirmed by DBPCFCs or convincing history (n=2) of tolerated peanut ingestion evaluated by a trained physician. Tolerant patients consumed at least small amounts of peanuts ≤ 2 weeks before study inclusion and five non-atopic donors (total IgE ≤ 100 kU/l, sIgE Ara h 2/6 ≤ 0.1 kU/l) served as reference. The study, conducted in accordance with the declaration of Helsinki, was ethically approved (number 17-945) by the medical ethical committee of the University Medical Center (UMC) Utrecht and informed consent was given by all participants.

Patient selection

Blood samples (100 ml) were drawn from allergic (n=6) and tolerant (n=6) adults sensitized to Ara h 2 and/or 6 (ImmunoCAP ≥ 0.1 kU/l¹). Their diagnoses were confirmed by DBPCFCs or in the case of two

tolerant subjects, by convincing history of tolerated peanut ingestion, as evaluated by a trained physician. Tolerant patients consumed at least small amounts of peanuts ≤ 2 weeks before study inclusion. Five non-atopic donors (total IgE ≤ 100 IU/ml) served as reference. The study, conducted in accordance with the declaration of Helsinki, was ethically approved (number 17-945) by the medical ethical committee of the University Medical Center Utrecht and informed consent was given by all participants.

Isolation of peanut 2S albumins

Roasted, salted, and ground peanuts were soaked in 50 mM sodium citrate buffer (pH 5.5) and incubated overnight at 4°C. The filtrate was applied onto a cation ion exchange column (Capto S, GE Healthcare) and the pooled fraction, containing proteins with molecular masses between 10 and 25 kDa, was separated by size exclusion (Superdex 75 pg, GE Healthcare). Enclosed proteins were analyzed by mass spectrometry, as described previously², and by inhibition assays.

Inhibition assays were performed with a concentration range of 0-100 μ g/ml peanut 2S albumin fraction and the commercially available EUROLINE (EL) DPA-Dx Peanut 1 strip (EUROIMMUN AG) including the seed storage proteins Ara h 2, 6 and 7. Briefly, 1:11 in washing buffer diluted sera were pre-incubated with the respective concentration of the peanut 2S albumin fraction for 30 min at room temperature. Pre-incubated sera were applied onto EL strips and incubated overnight at room temperature. Bound IgE antibodies were detected by applying anti-human IgE conjugate coupled with alkaline phosphatase for 1 hour at room temperature. Visualization was achieved by applying nitro-blue tetrazolium/5-bromo-4-chloro-3'-indolylphosphate substrate for ten minutes and specific IgE levels were evaluated as EUROLINE (EL)-intensities. Inhibition was calculated as the percentage of remaining EL-intensities compared to the EL-intensities without pre-incubation with the peanut 2S albumin fraction. Results of the MS analysis and inhibition experiments are shown in the Supplementary, Table S1 and Figure S1.

Allergen-tetramer formation

An excess of biotin was added to the isolated peanut 2S albumin fraction resulting on average in four biotin molecules per one molecule protein (EZ-Link™ Sulfo-NHS-Biotin, ThermoFisher Scientific). Allergen-tetramers were formed by separately adding streptavidin- R-phycoerythrin or streptavidin-Allophycocyanin (ThermoFisher Scientific) to the biotinylated 2S albumin fraction in a molecular ratio of 1:1 (streptavidin: protein)^{3,4}.

Single-cell sorting of allergen specific B cells

Blood drawn into heparin-coated tubes was freshly processed within 24 hours and peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation. B cells were negatively enriched from the PBMCs fraction by magnetic beads (B cell isolation kit II, Miltenyi Biotec) and they were stained for allergen specificity with CD45-Pacific Orange (1:40, ThermoFisher Scientific), CD3-Pacific Blue (1:160, BD Biosciences), CD19- Fluorescein isothiocyanate (1:20, BD Biosciences), CD14-R-phycoerythrin/cyanine dye 7 (1:800, Sony Biotechnology), CD16- R-phycoerythrin/cyanine dye 7 (1:1000, BD Biosciences) and allergen-tetramers (75 pg $\pm$ 30*10⁶ antigen-tetramers sufficient for approximately 150-300 resting B cells⁵). Cells gated for CD45+, CD3-, CD19+, allergen tetramers (double positive) + and CD14/16- were single cell sorted into 96 well plates supplemented with 14 mM DTT and 11.2 U RNase inhibitor (N8080119, ThermoFisher Scientific) in 0.5x PBS. Sorted cells were immediately put on dry ice and stored at -80 °C until further processing. PBMCs from a donor who showed specific B cells in previous experiments served as positive control and staining with only biotin served as negative control. The gating was based on the negative control (example FACS plot is shown in the Supplementary, Figure S2).

Gene amplification of B cell receptors

B cell receptors were amplified as previously described with minor modifications^{3, 6, 7}. Briefly, frozen B cells were put on dry ice, supplemented with 1.4% NP-40, 3 U RNase inhibitor (N8080119, ThermoFisher Scientific) and 7 μ M hexamer primers and subsequently incubated for one minute at 68 °C for primer annealing. cDNA transcription was performed accordingly to manufacture's instruction (SuperScript III, ThermoFisher Scientific) and cDNA was used as a template for amplifying the variable gene transcript of the heavy and the corresponding light chain. Amplification was performed with 1.25 U hot-start DNA polymerase (AmpliTaq Gold, ThermoFisher Scientific) and 40 nmol multiplex primers for 50 cycles and the annealing temperature was set to 62, 60 and 58 °C for heavy, λ and κ light chain, respectively.

Sequence analysis of B cell receptors

Enzyme purified heavy chain amplification products were sequenced (Sanger sequencing, MacroGen) with either multiplex reverse primers or framework (FR) 1 forward primers (isotype determination) while the light chain amplification products were all sequenced with the respective reverse primers. All sequences were checked for their quality using Chromas Lite 2.6.5, and double peaks, potentially resulting from amplification errors in the beginning of the amplification reaction, were aligned and corrected for the

germline if plausible. Quality-checked gene sequences, saved as FASTA files, were used for automatic germline alignment using the IgBLAST web interface (reference: IMGT database) ⁸ and the output was automatically written into a SQLite database using R, version 3.6.3. For further evaluation, sequences were considered if the sequence was valid and declared as 'productive'.

Cloning of gene products from single cells

Restriction sites were introduced with specific V gene forward primers and J gene reverse primers, ensuring that the entire variable region is cloned without any additional amino acids, using 0.5 U Phusion high-fidelity DNA polymerase (NEB). The amplification was performed for 30 cycles with an annealing temperature of 62, 60 and 58 °C for heavy, κ and λ , respectively, and purified gene products were digested with the respective restriction enzymes Eco RI and Nhe I (heavy chain), BspWI (κ light chain) or Avr II (λ light chain). Before cloning these digests into human IgH (IgE, IgG1, IgG4) and IgL (κ , λ) expression vectors (pFUSEss vector series, Invivogen), they were purified and phosphorylated (T4-PNK, NEB). All vectors contained an hEF1-HTLV promotor, an IL-2 signal sequence, a multiple cloning site upstream the constant regions and zeocin (IgH) or blasticidin (IgL) resistance genes for selection. To prevent self-ligation, the digested parent vector was dephosphorylated (parent vector, FAST-AP, ThermoFisher Scientific) before the digests were added in a molecular ratio of 5:1 (T4 ligase, NEB). Heat-shock transformation was performed by adding the ligated vector to competent *E. coli* Top10 cells and incubating them for 45 seconds at 42 °C. Grown colonies were screened for correctly incorporated variable regions by PCR, using forward primers binding to the respective FR2 region and reverse primers suitable for the constant part of the vector. Positive clones were Sanger sequenced and the correctness was checked by aligning the vector sequence with the first sequence result.

Heterologous expression of monoclonal antibodies

Human embryonic kidney (HEK) 293F cells were cultured in FreeStyle 293 expression medium (ThermoFisher Scientific) using 125 ml shaking culture flasks. Exponentially growing cells (80% confluence, 90% viability) were transiently transfected with VH and VL expression vectors in a ratio of 2:3 (total 0.5 μ g plasmid DNA per 1×10^6 cells) in 150 mm culture plates using 293fectin (2 μ l/ μ g plasmid DNA, ThermoFisher Scientific). Three days upon transfection, the supernatant was harvested and stored at -20 °C for further analyses.

Determining allergen specificity

Specificity of heterologously expressed mAbs was tested using a direct ELISA. Coating was performed overnight at room temperature with either 0.3 µg/well Ara h 2/6⁹ or transferrin (negative control). Upon blocking (PBS + 1% BSA + 0.1% Tween-20), serial-diluted thawed (1x) supernatant was applied for one hour at room temperature and bound mAbs were detected by goat α-human IgE (1:5000, KPL) or α-human IgG (1:2000, Jackson ImmunoResearch) antibodies coupled with horse-radish peroxidase. Visualization was provided by adding tetramethylbenzidine (TMB) for 15 minutes in the dark and the OD was measured at 450 nm. The cut-off levels were determined for IgE and IgG1 mAbs individually and were defined as the blank value (HEK293F cells supernatant) plus three times its standard deviation. A mAb was defined as specific when the sample had an OD greater than the set cut-off levels upon subtraction of the negative control (transferrin) OD value. The negative control OD value was defined as the OD value of the mAb upon subtraction of its blank value.

Specificity of patient serum and inhibition experiments

ELISA to measure serum sIgA and inhibition experiments were performed as described in the previous paragraph with the following modifications:

[1] Patients' serum samples were 1:10 diluted in blocking buffer before application (duplicates, same sample)

[2] Detection of sIgA binding by goat α-human IgA (1:2000, Southern Biotech) coupled with horse-radish peroxidase and incubation for 1 hour at room temperature

[3] Determination of optimal dilution for sIgE inhibition experiments by applying serum samples in a dilution series (1:10, 1:20, 1:50, 1:100); optimal serum dilution for inhibition experiments was defined as the first dilution showing lowered OD values in relation to the dilution yielding the maximum value or maintaining a plateau phase

[4] Additional incubation step prior to serum application with the respective mAb in a concentration-dependent manner (1 to 100 ng/ml) and incubation for one hour at room temperature

Indirect basophil activation test

Functionality of 2S albumin-specific mAbs were evaluated by indirect basophil activation test (BAT)¹⁰. Briefly, isolated PBMCs were treated with 13.4 mM lactic acid buffer to strip off receptor-bound IgE. Receptors were reloaded with human mAbs expressed as IgE, supplemented with 10 µg/ml heparin and 8 mM EDTA and incubated for 90 min at 37 °C. After extensive washing, re-loaded cells were cultured overnight in RPMI supplemented with 10% fetal calf serum and 1 ng/ml IL-3 at 37 °C. Dose-dependent stimulation was performed with a mixture of native Ara h 2/6 and 1 ng/ml IL-3 (1 µg/ml to 5 ng/ml). RPMI + 1 ng/ml IL-3, 100 ng/ml C5a (R&D Systems, 2037-C5), 1 µg/ml anti-human IgE (Vector Laboratories, BA-3040-.5) and 10 µM fMLP (Sigma Aldrich, F3506) served as negative and positive controls, respectively. Assessment of mAb blocking capacity of serum IgE and allergen interaction was performed with 10 ng/ml Ara h 2 or 6 as this concentration was the last concentration before the plateau was reached (Pt 6 and 10). Pre-incubation of these allergens was performed for 30 min at room temperature with mAbs in a concentration-dependent manner (5 to 100 ng/ml $\triangleq$ around 2 molecules mAb per 1 molecule allergen). Stimulated cells were stained for IL-3 activated basophils with CD123-Fluorescein isothiocyanate (1:400, BioLegend), CD63-R-phycoerythrin (1:800, Sanbio B.V.), CD203c-Allophycocyanin (1:160, Sony Biotechnology) and CD193-eridin-Chlorophyll-Protein/Cyanine5.5 (1:160, Sony Biotechnology) and basophils were defined as CD203c+ CD123+ CD193+. Basophil activation was calculated as the percentage of CD63+ basophils and the threshold was set to 5% CD63+ cells. All measurements were performed in duplicates and the gating strategy is shown in the Supplementary, Figure S2.

Data analyses

Comparisons of continuous data between allergic and tolerant patients were statistically analyzed with the two-sided non-parametric Mann-Whitney-U-test. Correction for multiple testing was obtained by Bonferroni correction. Comparison of continuous data between allergic, tolerant and non-atopic donors was statistically evaluated using the non-parametric Kruskal-Wallis test combined with Dunn's multiple testing correction. Categorical comparisons were performed with the Chi Square test. Descriptive analysis consisted of isotype distribution, mutational status and VH-VL usage. VH-VL usage was visualized with circular diagrams created with the R package 'Circos'¹¹. Moreover, related clones were defined as sequences resulted from identical VH and JH genes and HCDR3 length and related clones were aligned using ClustalW2¹². Sequence motifs were identified by calculating Levenshtein distances (R package 'factoextra') of HCDR3 or LCDR3 regions combined with hierarchical clustering (default unsupervised clustering in R). HCDR3 sequences with five or fewer differences and LCDR3 region with two or fewer

171 differences were defined as one motif. Physical properties of the HCDR3 regions were calculated using the
172 BRepertoire interface ¹³. For the latter analysis, non-redundant HCDR3 amino acid sequences were used.
173
174

Primer	Sequence
Heavy chain primer forward	
<i>Single cell RT-PCR</i>	
5' L-Vh1	ACAGGTGCCCCACTCCCAGGTGCAG
5' L-Vh3	AAGGTGTCCAGTGTGARGTGCAG
5' L-Vh4/6	CCCAGATGGGTCCTGTCCCAGGTGCAG
5' L-Vh5	CAAGGAGTCTGTTCCGAGGTGCAG
<i>Introduction restriction sites</i>	
5' EcoRI_VH1	ATATTGAATTCGCAGGTGCAGCTGGTGCAG
5' EcoRI_VH1/5	ATATTGAATTCGGAGGTGCAGCTGGTGCAG
5' EcoRI_VH1-18	ATATTGAATTCGCAGGTTCAGCTGGTGCAG
5' EcoRI_VH1-24	ATATTGAATTCGCAGGTCCAGCTGGTACAG
5' EcoRI_VH3	ATATTGAATTCGGAGGTGCAGCTGGTGGAG
5' EcoRI_VH3-23	ATATTGAATTCGGAGGTGCAGCTGTTGGAG
5' EcoRI_VH3-33	ATATTGAATTCGCAGGTGCAGCTGGTGGAG
5' EcoRI_VH3-9	ATATTGAATTCGGAAGTGCAGCTGGTGGAG
5' EcoRI_VH4	ATATTGAATTCGCAGGTGCAGCTGCAGGAG
5' EcoRI_VH4-34	ATATTGAATTCGCAGGTGCAGCTACAGCAGTG
5' EcoRI_VH4-39	ATATTGAATTCGCAGCTGCAGCTGCAGGAG
5' EcoRI_VH6-1	ATATTGAATTCGCAGGTACAGCTGCAGCAG
<i>Sanger Sequencing</i>	
5' Vh1-FR1_(1-2)	GGCCTCAGTGAAGGTCTCCTGCAAG
5' Vh2-FR1_(2-5)	GTCTGGTCCTACGCTGGTGAAACCC
5' Vh3-FR1_(3-7)	CTGGGGGGTCCCTGAGACTCTCCTG
5' Vh4-FR1_(4-4)	CTTCGGAGACCCTGTCCCTCACCTG
5' Vh5-FR1_(5-51)	CGGGGAGTCTCTGAAGATCTCCTGT
5' Vh6-FR1_(6-1)	TCGCAGACCCTCTCACTCACCTGTG
<i>Colony screening</i>	
5' Vh1-FR2_(1-2)	CTGGGTGCGACAGGCCCCCTGGACAA
5' Vh2-FR2_(2-5)	TGGATCCGTCAGCCCCCAGGGAAGG
5' Vh3-FR2_(3-7)	GGTCCGCCAGGCTCCAGGGAA
5' Vh4-FR2_(4-4)	TGGATCCGCCAGCCCCCAGGGAAGG
5' Vh5-FR2_(5-51)	GGGTGCGCCAGATGCCCCGGGAAAGG
5' Vh6-FR2_(6-1)	TGGATCAGGCAGTCCCCATCGAGAG
5' Vh7-FR2_(7)	TTGGGTGCGACAGGCCCCCTGGACAA
Heavy chain primer reverse	
<i>Single cell RT-PCR</i>	
3' CH1_IgA*	AGCCCTGGACCAGGCA
3' CH1_IgE*,**	GAAGACGGATGGGCTCTGT
3' CH1_IgG*,**	GGAAGGTGTGCACGCCGCTG
3' CH1_IgM*	GGAATTCTCACAGGAGACG
<i>Introduction restriction sites</i>	
3' NheI_JH1/2/4/5	ATGCTAGCTGAGGAGACGGTGACCAG
3' NheI_JH3	ATGCTAGCTGAAGAGACGGTGACCATTG
3' NheI_JH6	ATGCTAGCTGAGGAGACGGTGACCGTG

Kappa light chain primer forward

Single cell RT-PCR

5' L-V κ 1/2	ATGAGGSTCCCYGCTCAGCTGGTGG
5' L-V κ 3	CTCTTCCTCCTGCTACTCTGGCTCCCAG
5' L-V κ 4	ATTTCTCTGTTGCTCTGGATCTCTG

Introduction restriction sites

5' EcoRI_V κ 1-5	ATATTGAATTCAGACATCCAGATGACCCAGTC
5' EcoRI_V κ 1-9	ATATTGAATTCAGACATCCAGTTGACCCAGTCT
5' EcoRI_V κ 1D-43	ATATTGAATTCAGCCATCCGGATGACCCAGTC
5' EcoRI_V κ 2-24	ATATTGAATTCAGATATTGTGATGACCCAGAC
5' EcoRI_V κ 2-28	ATATTGAATTCAGATATTGTGATGACTCAGTC
5' EcoRI_V κ 2-30	ATATTGAATTCAGATGTTGTGATGACTCAGTC
5' EcoRI_V κ 3-11	ATATTGAATTCAGAAATTGTGTTGACACAGTC
5' EcoRI_V κ 3-15	ATATTGAATTCAGAAATAGTGATGACGCAGTC
5' EcoRI_V κ 3-20	ATATTGAATTCAGAAATTGTGTTGACGCAGTCT
5' EcoRI_V κ 4-1	ATATTGAATTCAGACATCGTGATGACCCAGTC

Colony screening

5' V κ 1f/6	TCAAGGTTTCAGCGGCAGTGGATCTG
5' V κ 2f	GGCCTCCATCTCCTGCAGGTCTAGTC
5' V κ 3f	CCCAGGCTCCTCATCTATGATGCATCC
5' V κ 4_int	CAACTGCAAGTCCAGCCAGAGTGTTTT
5' V κ 5_int	CCTGCAAAGCCAGCCAAGACATTGAT
5' V κ 7_int	GACCGATTTACCCCTCACAAATTAATCC

Kappa light chain primer reverse

Single cell RT-PCR

3' C κ 494*,**	GTGCTGTCCTTGCTGTCCTGCT
-----------------------	------------------------

Introduction restriction sites

3' BsiWI_J κ 1/4	ATCGTACGTTTGATYTCCACCTTGGTC
3' BsiWI_J κ 2	ATCGTACGTTTGATCTCCAGCTTGGTC
3' BsiWI_J κ 3	ATCGTACGTTTGATATCCACTTTGGTC
3' BsiWI_J κ 5	ATCGTACGTTTAATCTCCAGTCGTGTC

Lambda light chain primer forward

Single cell RT-PCR

5' L-V λ 1	GGTCCTGGGCCCAGTCTGTGCTG
5' L-V λ 2	GGTCCTGGGCCCAGTCTGCCCTG
5' L-V λ 3	GCTCTGTGACCTCCTATGAGCTG
5' L-V λ 4/5	GGTCTCTCTCSCAGCYGTTGCTG
5' L-V λ 6	GTTCTTGGGCCAATTTTATGCTG
5' L-V λ 7	GGTCCAATTCYCAGGCTGTGGTG
5' L-V λ 8	GAGTGGATTCTCAGACTGTGGTG

Introduction restriction sites

5' EcoRI_Vλ1	ATATTGAATTCGCAGTCTGTGCTGACKCAG
5' EcoRI_Vλ2	ATATTGAATTCGCAGTCTGCCCTGACTCAG
5' EcoRI_Vλ3	ATATTGAATTCGTCCTATGAGCTGACWCAG
5' EcoRI_Vλ4/5	ATATTGAATTCGCAGCYTGTGCTGACTCA
5' EcoRI_Vλ6	ATATTGAATTCGAATTTTATGCTGACTCAG
5' EcoRI_Vλ7/8	ATATTGAATTCGCAGRCTGTGGTGACYCAG

Colony screening

5' Vλ1/2_int	ATTCTCTGGCTCCAAGTCTGGC
5' Vλ3_int	GGATCCCTGAGCGATTCTCTGG

Lambda light chain primer reverse

Single cell RT-PCR

3' Cλ*,**	CACCAGTGTGGCCTTGTTGGCTTG
-----------	--------------------------

Introduction restriction sites

3' AvrII_Jλ1	ATTCCTAGGACGGTGACCTTGGT
3' AvrII_Jλ2/3	ATTCCTAGGACGGTCAGCTTGGT
3' AvrII_Jλ6	ATTCCTAGGACGGTCACCTTGGT
3' AvrII_Jλ7-1	ATTCCTAGGACGGTCAGCTGGGT
3' AvrII_Jλ7-2	ATTCCTAGGCGGTCAGCTGGGT

176 *also used for Sanger Sequencing; **also used for colony screening

177

178

179 **References**

- 180 1. Martinet J, Couderc L, Renosi F, Bobée V, Marguet C, Boyer O. Diagnostic Value of
181 Antigen-Specific Immunoglobulin E Immunoassays against Ara h 2 and Ara h 8 Peanut
182 Components in Child Food Allergy. *Int Arch Allergy Immunol* 2016; 169:216-22; doi:
183 10.1159/000446181.
- 184 2. Scharf M, Miske R, Kade S, Hahn S, Denno Y, Begemann N, et al. A Spectrum of Neural
185 Autoantigens, Newly Identified by Histo-Immunoprecipitation, Mass Spectrometry, and
186 Recombinant Cell-Based Indirect Immunofluorescence. *Front Immunol* 2018; 9:1447; doi:
187 10.3389/fimmu.2018.01447.
- 188 3. Patil S, Ogunniyi A, Calatroni A, Tadigotla V, Ruiter B, Ma A, et al. Peanut oral
189 immunotherapy transiently expands circulating Ara h 2-specific B cells with a
190 homologous repertoire in unrelated individuals. *J Allergy Clin Immunol* 2015; 136:125-34;
191 doi: 10.1016/j.jaci.2015.03.026.
- 192 4. Franz B, May K, Dranoff G, Wucherpfennig K. Ex vivo characterization and isolation of
193 rare memory B cells with antigen tetramers. *Blood* 2011; 118:348-57; doi: 10.1182/blood-
194 2011-03-341917.
- 195 5. Yang J, Reth M. Oligomeric organization of the B-cell antigen receptor on resting cells.
196 *Nature* 2010; 467:465-9; doi: 10.1038/nature09357.
- 197 6. Tiller T, Meffre E, Yurasov S, Tsuiji M, Nussenzweig M, Wardemann H. Efficient generation
198 of monoclonal antibodies from single human B cells by single cell RT-PCR and
199 expression vector cloning. *J Immunol Methods* 2008; 329:112-24; doi:
200 10.1016/j.jim.2007.09.017.
- 201 7. Hoh R, Joshi S, Liu Y, Wang C, Roskin K, Lee J, et al. Single B cell Deconvolution of
202 Peanut-specific Antibody Responses in Allergic Patients. *J Allergy Clin Immunol* 2016;
203 137:157-67; doi: 10.1016/j.jaci.2015.05.029.
- 204 8. Ye J, Ma N, Madden T, Ostell J. IgBLAST: An Immunoglobulin Variable Domain Sequence
205 Analysis Tool *Nuc Acid Res* 2013; 41:W34-40; doi: 10.1093/nar/gkt382.
- 206 9. Koppelman S, De Jong G, Laaper-Ertmann M, Peeters K, Knulst A, Hefle S, et al.
207 Purification and immunoglobulin E-binding properties of peanut allergen Ara h 6:
208 evidence for cross-reactivity with Ara h 2. *Clin Exp Allergy* 2005; 35:490-7; doi:
209 10.1111/j.1365-2222.2005.02204.x.
- 210 10. Santos A, James L, Bahnson H, Shamji M, Couto-Francisco N, Islam S, et al. IgG4 inhibits
211 peanut-induced basophil and mast cell activation in peanut-tolerant children sensitized
212 to peanut major allergens. *J Allergy Clin Immunol* 2015; 135:1249-56; doi:
213 10.1016/j.jaci.2015.01.012.
- 214 11. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. CIRCOS: an
215 information aesthetic for comparative genomics. *Genome Research* 2009; 19:1639-45;
216 doi: 10.1101/gr.092759.109.
- 217 12. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, et al. Clustal W
218 and Clustal X version 2.0 *Bioinformatics* 2007; 23:2947-8; doi:
219 10.1093/bioinformatics/btm404.

220 13. Margreitter C, Lu H-C, Townsend C, Stewart A, Dunn-Walters D, Fraternali F. BRepertoire:
221 a user-friendly web server for analysing antibody repertoire data. Nuc Acid Res 2018;
222 46:W264-70; doi: 10.1093/nar/gky276.

223
224

Supplementary File

1. Analysis peanut 2S albumin fraction

Table S1: Mass spectrometry analysis of the 2S albumin fraction from roasted peanut

Protein	Protein (Accession number)	Cut- Off	Protein Score	Sequence coverage [%]	Molecular mass [kDa]
1	Ara h 2.01 allergen (ACN62248.1)	80/51	49/39	32/16	19.0
2	Ara h 2.01 allergen (ACN62248.1)	80/51	93/56	51/16	19.0
3	Ara h 6 (Q647G9.1)	80/48	103/70	71/17	17.5
	Ara h 7.0301 (AAU21496.1)	80	64	37	18.8
4	Ara h 6 (Q647G9.1)	80/48	54/43	39/8	17.5
5	Ara h 6 (Q647G9.1)	49	14	8	17.5

MS/MS analyses are highlighted in blue.

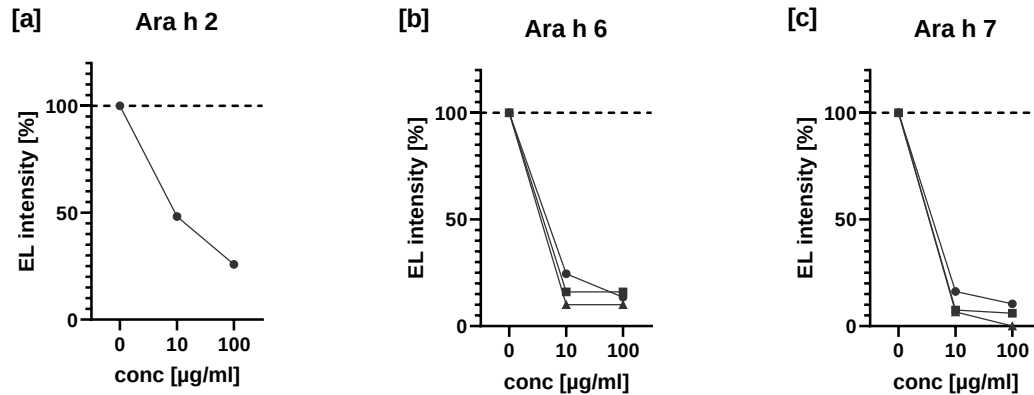


Figure S1: Inhibition of sIgE against Ara h 2, 6 and 7 by pre-incubation with peanut 2S albumin fraction
Presence of peanut 2S albumins in the isolated fraction confirmed by inhibition experiments
Inhibition of serum binding to (a) Ara h 2, (b) Ara h 6 and (c) Ara h 7 in concentration-dependent manner (0-100 µg/ml peanut 2S albumin fraction). dots: serum with increased sIgE titers for Ara h 2, 6 and 7, Squares: serum with increased sIgE titers to Ara h 6 and 7, Triangle: serum with increased sIgE titers to Ara h 6 and 7

2. FACS gating strategies

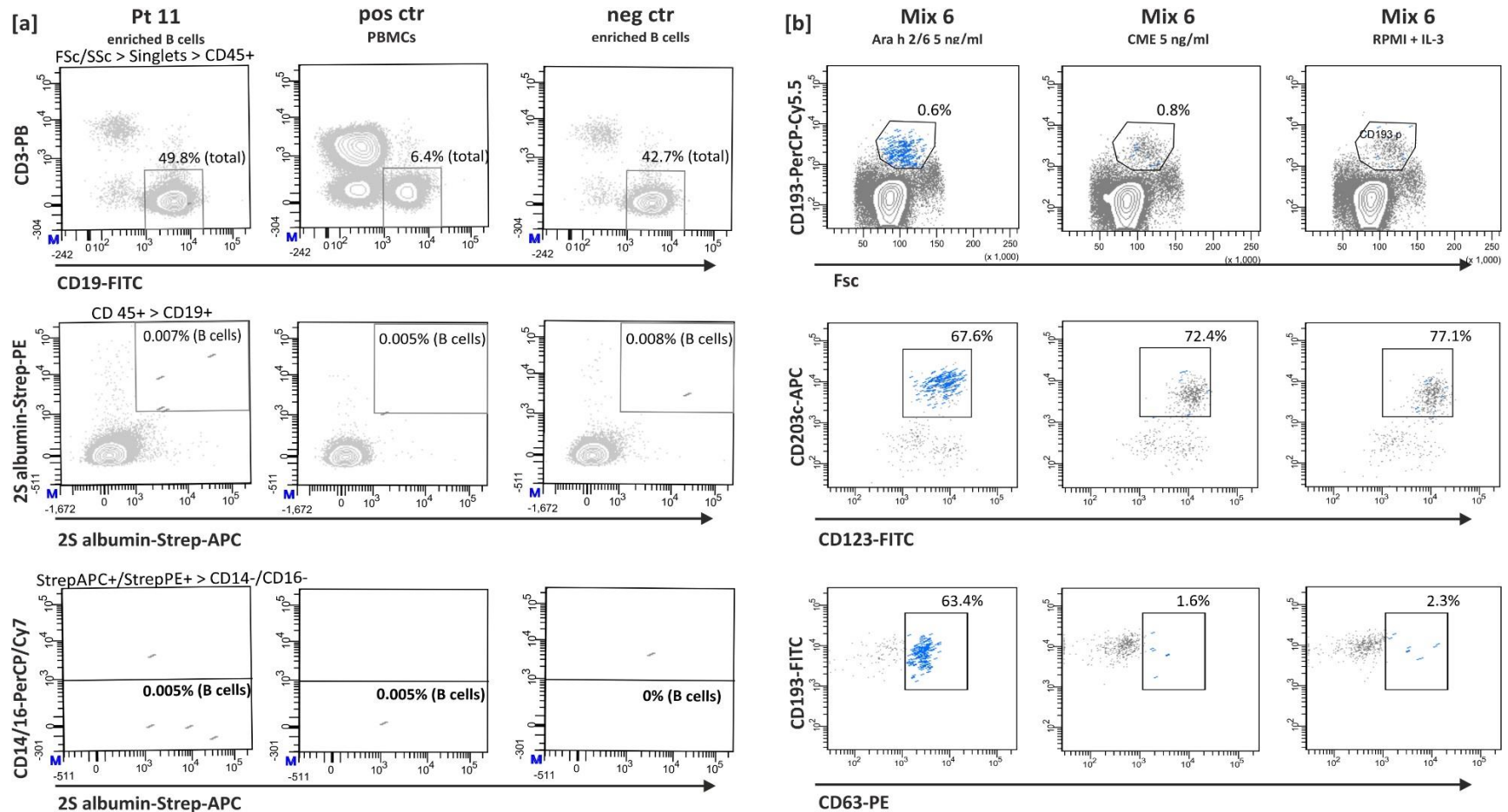


Figure S2 FACS gating strategies for the selection of 2S albumin-specific B cells and indirect basophils activation test

[a] selection of 2S albumin-binding B cells – the lowest panel of gates was used for sorting, [b] indirect basophil activation test; CME = cow's milk extract (control)

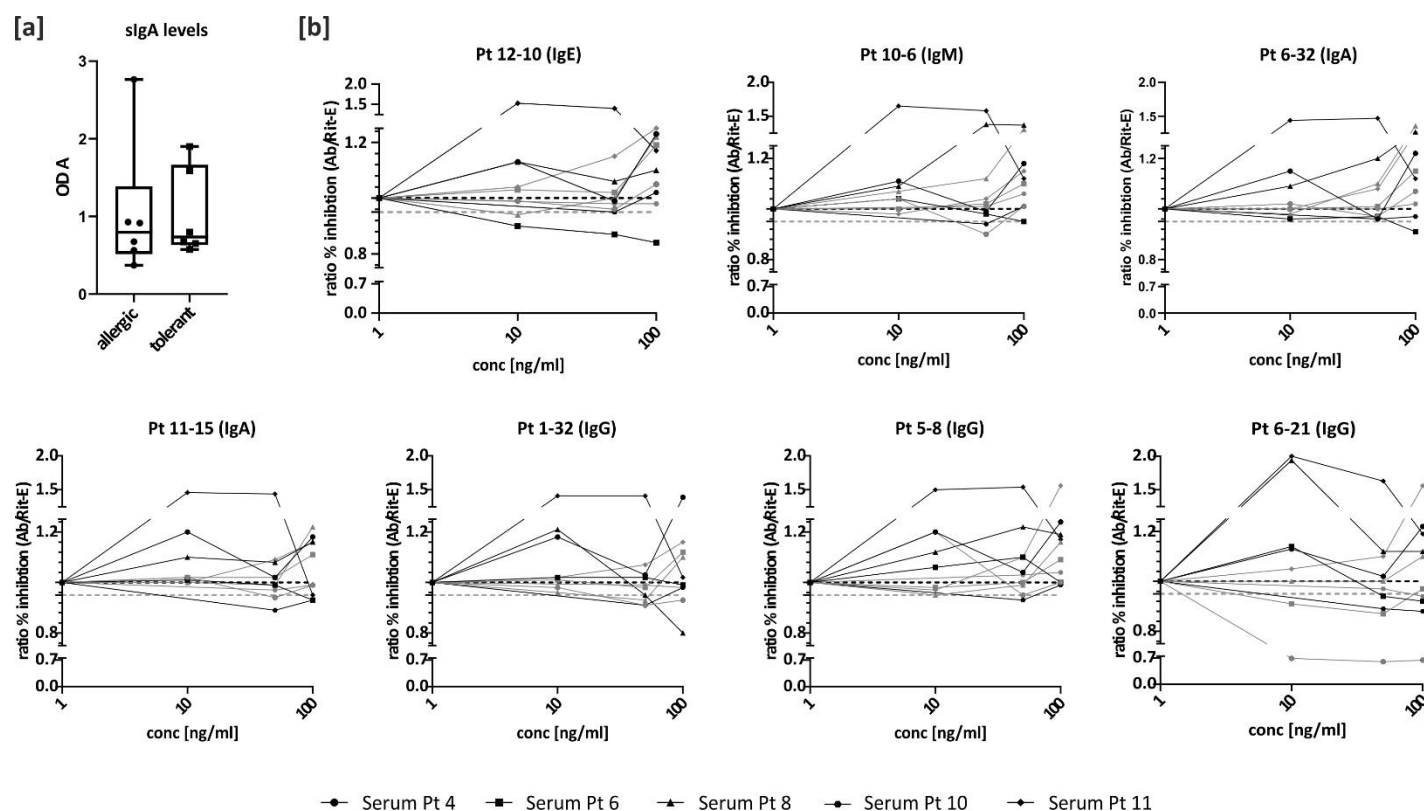
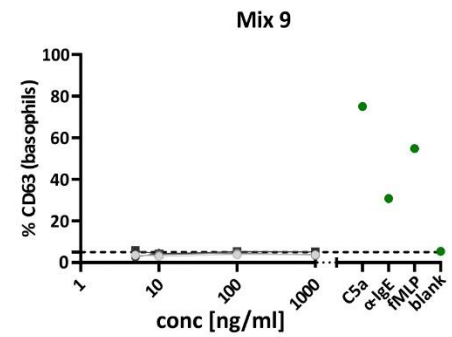
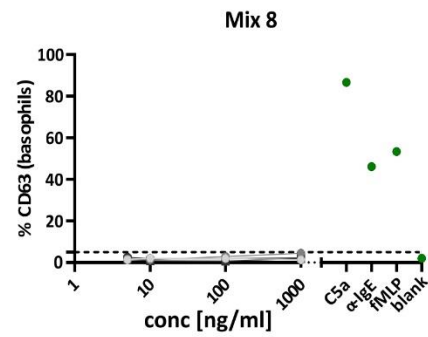
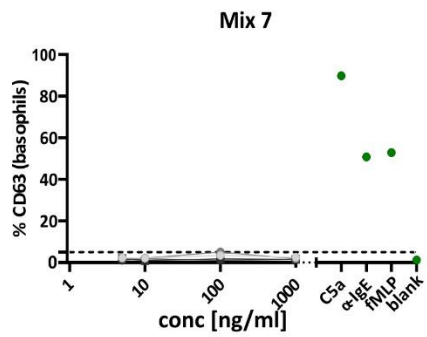
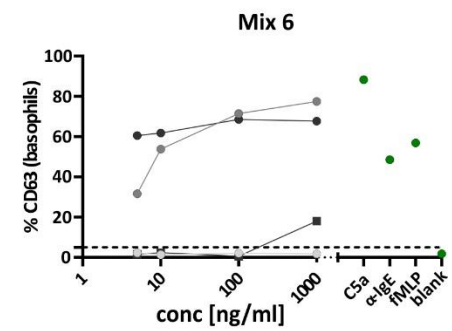
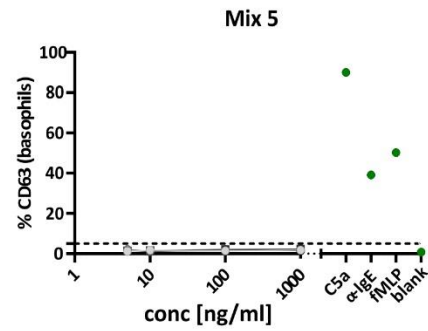
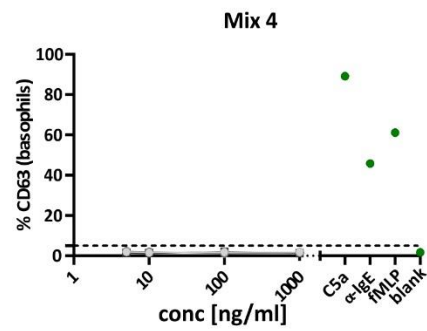
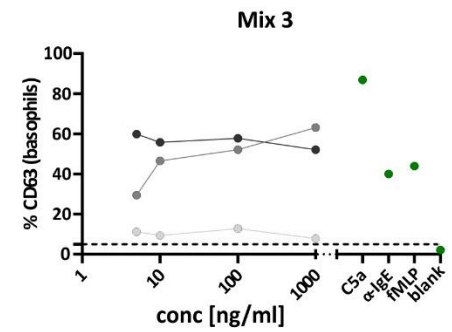
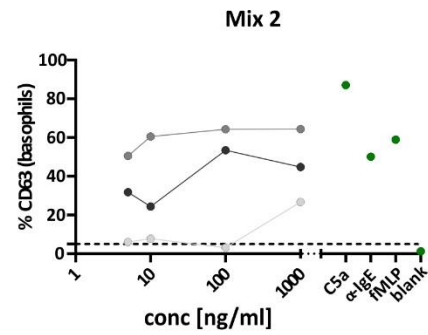
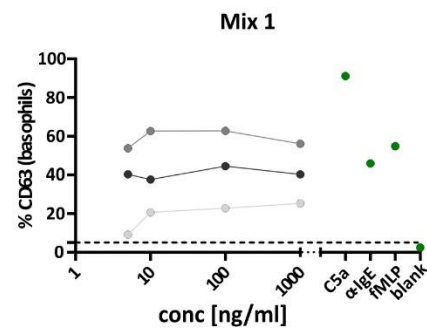


Figure S3: Binding of specific IgA to peanut 2S albumins and inhibition of sIgE binding to 2S albumins by 2S albumin-specific mAbs

[a] Specific IgA levels against Ara h 2 and 6 were measured in patients' serum samples, each dot is the mean of duplicate measurements, box plots shows the interquartile range of measured sIgA levels

[b] Inhibition of sIgE binding from patients' serum samples by pre-incubation with 7 distinct 2S albumin-specific mAbs expressed as IgG1 in a dose-dependent manner (1 to 100 ng/ml, mean of duplicates, the original isotype is displayed in brackets); inhibition is the remaining OD signal of the serum without pre-incubation and is displayed as the ratio of the inhibition by the respective mAb and Rituximab (negative control) expressed as IgG1; the critical range of 0 to 5% inhibition is highlighted by two dash lines



● Ara h 2
 ● Ara h 6
 ● Ara h 2/6
 ■ CME

Figure S4: Indirect basophil activation tests with 2S albumin-specific mAbs including all controls

Upregulation of CD63 upon loading of human basophils with mAb mixture and subsequent stimulation with Ara h 2 and 6 or CME in a concentration-dependent manner (5 to 1000 ng/ml); the dots indicate the mean of duplicate measurements; Mix 1 = Pt 6-2, Mix 2 = Pt 6-2 + Pt 6-17, Mix 3 = Pt 6-2 + Pt 3-2 + Pt 6-11, Mix 4 = Pt 6-17, Pt 3-2, Pt 6-11, Mix 5 = Pt 1-32, Pt 6-11, Pt 6-6, Pt 6-8, Mix 6 = NA 1-10 + Pt 4-28 + Pt 4-34 + Pt 6-2, Mix 7 = Pt NA 1-10 + Pt 4-28 + Pt 4-34, Mix 8 = Pt 1-32 + Pt 6-15 + Pt 6-21, stripped basophils before loading, Rituximab expressed as IgE; dash line: threshold level set to 5%

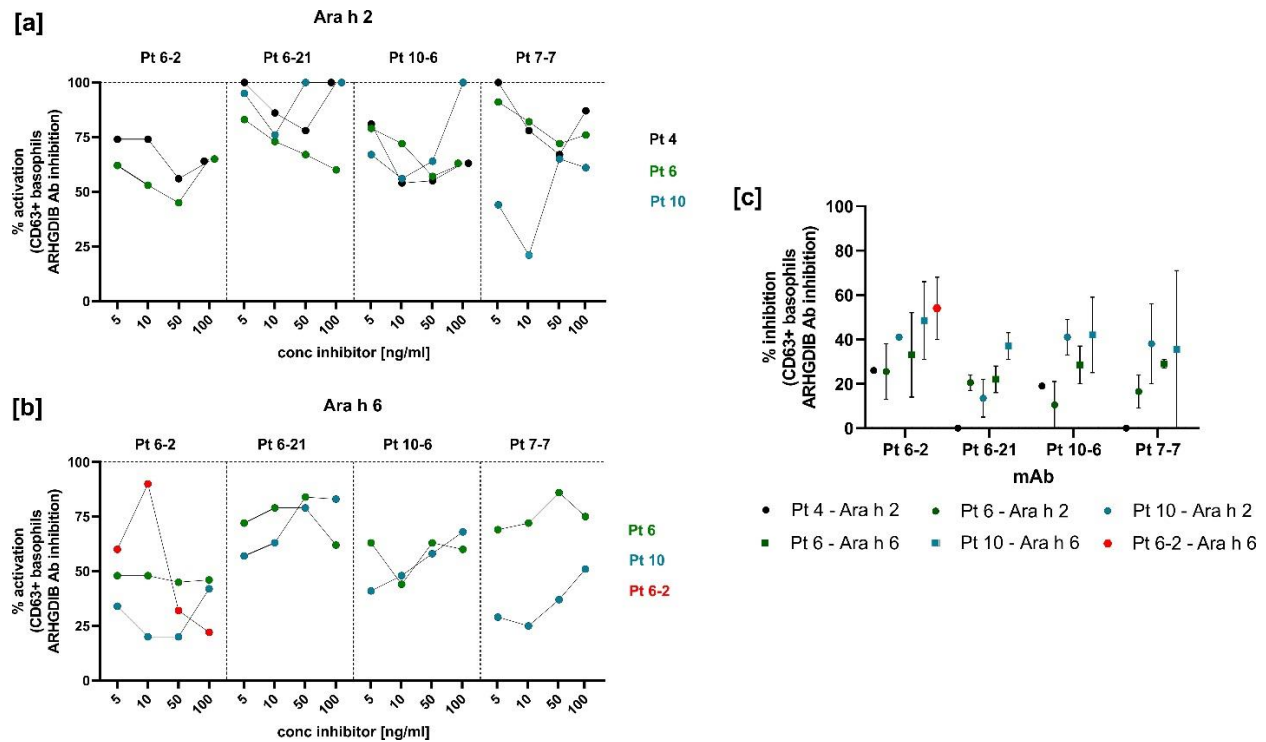


Figure S5: Blocking ability of mAbs to inhibit the interaction between serum IgE and Ara h 2 or 6

[a] Inhibitor (mAb: Pt 6-2, Pt 6-21, Pt 10-6, Pt 7-7) concentration-dependent blocking (5-100 ng/ml) of the interaction between serum IgE (Pt 4, 6, 10) and **Ara h 2** (10 ng/ml) to induce basophil activation (mean, duplicate measurements); ability expressed as remaining activation after pre-incubation of 10 ng/ml Ara h 2 with different concentrations of mAbs in comparison to the inhibition with an irrelevant mAb directed against ARHGDI B

[b] Inhibitor (mAb: Pt 6-2, Pt 6-21, Pt 10-6, Pt 7-7) concentration-dependent blocking (5-100 ng/ml) of the interaction between serum IgE (Pt 6, 10) and **Ara h 6** (10 ng/ml) to induce basophil activation (mean, duplicate measurements); positive control: self-inhibition with mAb Pt 6-2 upon loading basophils with mAb Pt 6-2; ability expressed as remaining activation after pre-incubation of 10 ng/ml Ara h 2 with different concentrations of mAbs in comparison to the inhibition with an irrelevant mAb directed against ARHGDI B

[c] Inhibition of the interaction between serum IgE (Pt 4, 6, 10) and Ara h 2 or 6 (10 ng/ml) with 5 ng/ml inhibitor (Pt 6-2, Pt 6-21, Pt 10-6, Pt 7-7) in two independent experiments (duplicate measurements each, mean with range); ability expressed as inhibition of CD63 upregulation in comparison to the inhibition with an irrelevant mAb directed against ARHGDI B

Table S2: LCDR3 sequence motifs associated with peanut allergy or tolerance

bold: differences of LCDR3 sequences; italic: tolerant donors; bold patient numbers: indicate heterologous expression mAb, n. e. = not expressed

Motif	LCDR3 sequence	Antibodies	Isotype	Specificity
Allergy				
1	CQQSYST P RTF	Pt 1-16, Pt 4-43	IgG2, IgM	n. e.
	CQQSYST P YTF	Pt 1-10, Pt 1-32 [†] , Pt 4-26	IgG2, IgG1	Ara h 2+6
	CQQSYST P P TF	Pt 6-54	IgM	n. e.
	CQQSYST P ITF	Pt 1-9	IgG2	n. e.
2	CQQRST W PLTF	Pt 4-8, Pt 4-20 [†] , Pt 4-28 [†] , Pt 4-63 [‡] , Pt 6-51	IgM	Ara h 2+6
	CQQRSS W PLTF	NA 4-8 [†]	IgM	Ara h 2+6
	CQQRS N YPLTF	Pt 4-11 [‡]	IgM	Ara h 2+6
3	CQQ H KSYPLTF	Pt 4-10 [†] , Pt 4-41	IgM	Ara h 2
	CQQ L KSYPLTF	Pt 4-60	IgM	n. e.
	CQQR S SYP LTF	Pt 4-48	IgM	n. e.
	CQQ L SYP LTF	Pt 4-7 [†]	IgM	Ara h 2+6
4	CQQ A NSFPLTF	Pt 4-35, NA 1-11	IgM	n. e.
	CQQ A YSFPLTF	Pt 4-29	IgM	n. e.
	CQQ L NSFPLTF	Pt 4-62 [†]	IgM	Ara h 2
	CQQ Y NSFPLTF	Pt 5-7	IgM	n. e.
5	CQQYYS P PLTF	Pt 1-25, Pt 1-30	IgG2	n. e.
	CQQYYSYP L TF	Pt 1-29, Pt 4-6	IgG2, IgM	n. e.
	CQQYYSYP V TF	Pt 4-3	IgM	n. e.
6	CQQYNN W PLTF	Pt 6-42	IgA1	n. e.
	CQQYNN W PRTF	NA 5-3	IgM	n. e.
	CQQYNN W S F	Pt 4-46, NA 1-6	IgM	n. e.
7	CQQR T TWPLTF	Pt 4-34 [†] , Pt 4-45, Pt 4-47, Pt 4-56	IgM	Ara h 2+6
Tolerance				
8	CSSYTSS S SVVF	<i>Pt 3-1, Pt 11-11</i>	IgM	n. e.
	CSSYTSS S IVVF	NA 4-2	IgM	n. e.

Motif	LCDR3 sequence	Antibodies	Isotype	Specificity
Shared				
9	CQQLNSYPLTF	Pt 4-1, Pt 4-22, Pt 4-33, Pt 4-36, Pt 4-39, Pt 4-42, Pt 6-4, Pt 6-23, <i>Pt 8-7, Pt 11-16</i> , NA 1-1	IgM, IgA2	n. e.
	CQQVNSYPLTF	NA 1-2, NA 1-13, NA 1-20, NA 1-21	IgM	n. e.
	CQQLNSYPITF	Pt 6-60, <i>Pt 7-5</i>	IgA1, IgM	n. e.
	CQQFNSYPLTF	Pt 4-27[†]	IgM	Ara h 2 + 6
	CQQHNSYPLTF	Pt 4-44	IgM	n. e.
	CQQLNSYPYTF	Pt 1-3	IgG2	n. e.
	CQQLNSYPFTF	NA 5-10	IgM	n. e.
	CQQNNSYPRTF	Pt 6-18	IgA2	n. e.
10	CQQYGSSP RTF	Pt 4-51, Pt 6-15[†] , NA 2-6, NA 3-1	IgM, IgA1	Ara h 2 + 6
	CQQYGSSP ITF	Pt 1-5, Pt 4-55, NA 2-12	IgG2, IgM	n. e.
	CQQYGSSPPITF	Pt 1-2, Pt 1-18	IgG2	n. e.
	CQQYGSSPPWTF	<i>Pt 3-3</i> , NA 4-1	IgM	n. e.
	CQQYGSSPPYTF	Pt 4-16, Pt 4-38	IgM	n. e.
	CQQYGSSPPLTF	Pt 1-6	IgG2	n. e.
	CQQYGSSPLITF	<i>Pt 7-1</i>	IgM	n. e.
	CQQYGSSP VTF	NA 1-3	IgM	n. e.
	CQQYGSSP WTF	Pt 1-19	IgG2	n. e.
11	CQQRSNWP LTF	<i>Pt 11-22</i> , NA 1-7, NA 1-8, NA 2-8	IgM	n. e.
	CQQRSNWPALTF	Pt 1-31	IgG2	n. e.
	CQQRSNWPPHTF	Pt 6-52[‡]	IgG2	Mix Ara h 2 + 6
	CQQRSNWPPFTF	NA 1-10[†]	IgM	Ara h 2 + 6
12	CQQRSKWPYTF	Pt 4-9[†] , Pt 4-21, Pt 6-8[†]	IgM, IgA1	Ara h 6, Ara h 2+6
	CQQRSNWPYTF	Pt 5-5	IgA1	n. e.
	CQQRSNWPITF	Pt 4-40	IgM	n. e.
	CQQRSNWPRTF	<i>Pt 11-24</i>	IgM	n. e.
13	CLQHNSYPL TF	<i>Pt 11-23</i> , NA 2-2	IgA2, IgM	n. e.
	CLQHNSYPY TF	Pt 1-27	IgG2	n. e.

14	CQQVNSFPITF	NA 5-2[†], NA 5-5[†], NA 5-11	IgM	Ara h 2, Ara h 2+6
	CQQVNSYPITF	Pt 6-24	IgM	n. e.
15	CQKYNSAPLTF	Pt 6-22, NA 3-3	IgM	n. e.
	CQKYDSAPLTF	Pt 6-5[†]	IgM	Ara h 2+6
	CQQYGSAPLTF	<i>Pt 11-7</i>	IgM	n. e.
16	CQQYDSLPTYF	Pt 1-14, Pt 1-23	IgG2	n. e.
	CQQYDSTPTYF	NA 3-3	IgM	n. e.
	CQQYDSSPTYF	<i>Pt 11-5</i>	IgM	n. e.
	CQQYDSYPITF	<i>Pt 3-6</i>	IgA1	n. e.
17	CQQYNNGPPVTF	<i>Pt 11-13</i>	IgM	n. e.
	CQQYNNWPPLTF	<i>Pt 11-4</i>	IgA2	n. e.
	CQQYNNWPPWTF	<i>Pt 7-3</i>	IgM	n. e.
	CQQYNNWPPYTF	Pt 6-16	IgA1	n. e.
18	CSSYTSSSTL VF	Pt 6-39, NA 4-7	IgM	n. e.
	CSSYTSSSTLVVF	Pt 1-11	IgG2	n. e.
	CSSYTSSNTLVVF	<i>Pt 8-1</i>	IgA2	n. e.
A	CATWDDSLSG YVF	Pt 4-58	IgA1	n. e.
	CATWDDSLSG VVF	Pt 6-30		n. e.
	CAAWDDSLNGHVVF	Pt 5-3	IgG2	n. e.
	CAAWDDSLSGHWVF	Pt 10-4	IgA1	n. e.
	CAAWDDSLSGPWVF	Pt 1-1	IgG2	n. e.

† specificity to Ara h2/6 proven

‡ only specificity against mixture of Ara h2 and 6

3. Physical properties of the HCDR3 regions

Physical properties of the HCDR3 regions were calculated using the BRepertoire web server ¹. For this analysis, only non-redundant HCDR3 amino acid sequences (allergic n = 88; tolerant n = 55; non-atopic, n = 48) were used.

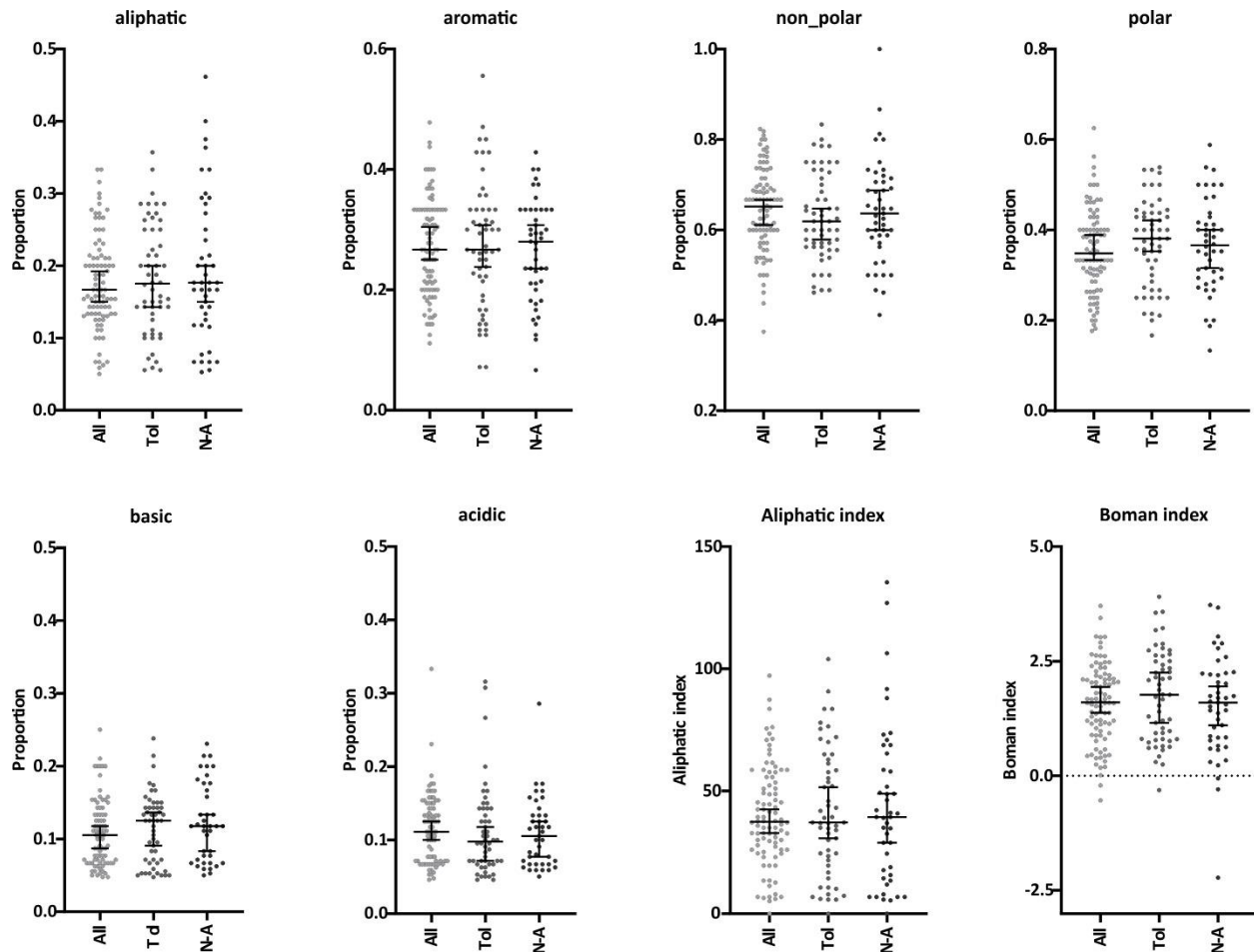


Figure S6: Proportion of amino acids within the HCDR3 region

Proportion of amino acids are categorized into different groups (aliphatic, aromatic, non-polar, polar, basic and acidic); calculation of the aliphatic index and the Boman index; Boman indices ≥ 2.48 indicate high probability for protein-protein interactions, all sample sets show the median with the 95% confidence interval; all = allergic, tol = tolerant, n-a = non-atopic

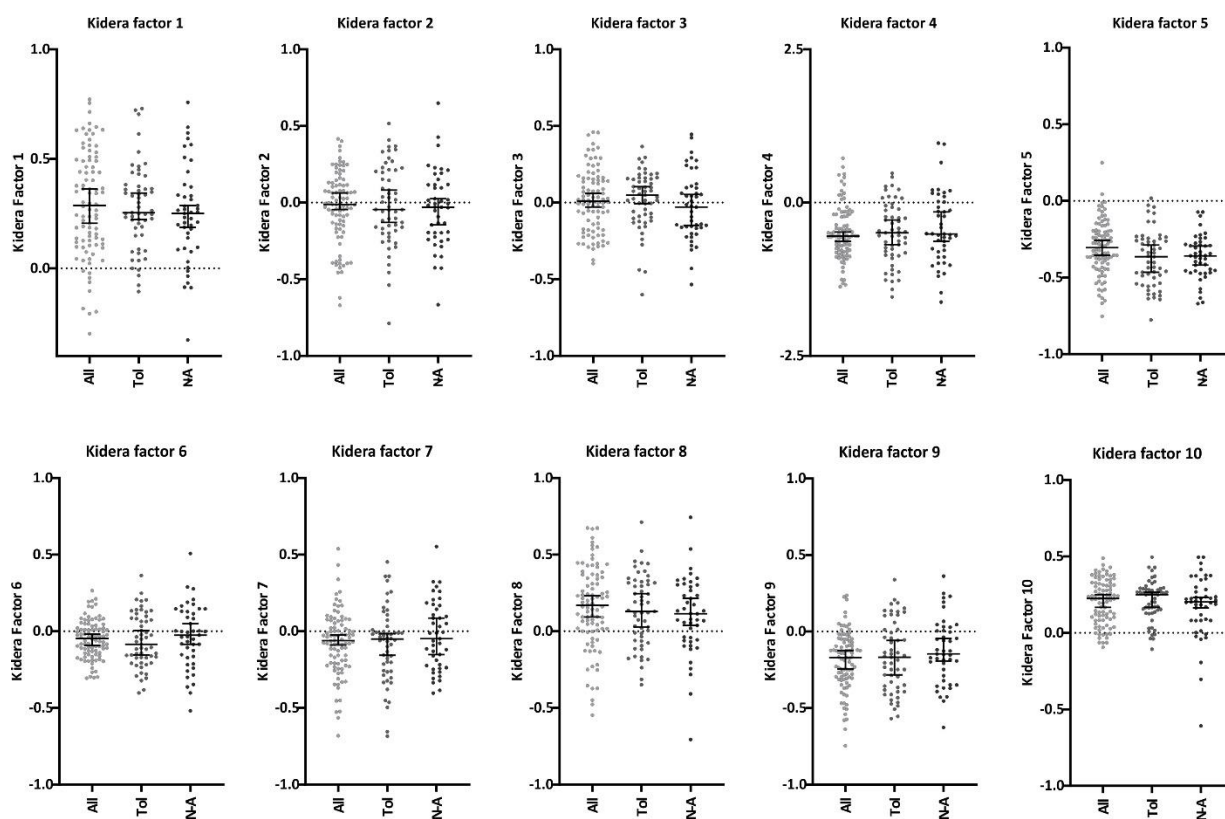


Figure S7: Properties of HCDR3 region expressed as Kidera factors (KF) 1-10 ²

KF1: Helix/bend preference, KF2: side chain size, KF3: extended structure preference, KF4:

Hydrophobicity, KF5: double-bend preference (*, $p=0.0242$), KF6: partial specific volume, KF7: flat extended preference, KF8: occurrence in alpha region, KF9: pK-C, KF10: surrounding hydrophobicity, all sample sets show the median with the 95% confidence interval; all = allergic, tol = tolerant, n-a = non-atopic

References

1. Margreitter, C. *et al.* BRepertoire: a user-friendly web server for analysing antibody repertoire data. *Nuc Acid Res* **46**, W264-270 (2018).
2. Kidera, A., Konishi, Y., Oka, M., Ooi, T. & Scheraga, H. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *J Protein Chem* **4**, 23-55 (1985).

```

install.packages("httr", dependencies = TRUE);
install.packages("RSQLite", dependencies = TRUE);
install.packages("XML", dependencies = TRUE);
install.packages("xml2", dependencies = TRUE);
install.packages("seqinr", dependencies = TRUE);
install.packages("stringr", dependencies = TRUE);
library(RSQLite);
library(httr);
library(XML);
library(xml2);
library(seqinr);
library(stringr);

database = "/Users/anna/Desktop/Cloud/Database/database_iallergen.db"
data_dir = "/Users/anna/Desktop/Cloud/Database/edited_data_sequencing"
con <- dbConnect(RSQLite::SQLite(), database);
for(dir in dir(data_dir)){
  pat_dir = paste(data_dir, dir, sep = "/");
  fasta_files <- list.files(path = pat_dir, pattern = "*.fasta");
  if(length(fasta_files) > 0){
    for(f in fasta_files){
      array_filename <- strsplit(f, "-");
      date <-
paste(array_filename[[1]][2], array_filename[[1]][3], array_filename[[1]][4],
sep= "-");
      fasta <- read.fasta(file = paste(pat_dir, f, sep = "/"), as.string =
TRUE);
      head_fasta <- getAnnot(fasta[[1]]);
      array_head_fasta <- strsplit(head_fasta, "-");
      p_acronym <- array_head_fasta[[1]][2];
      b_acronym <- array_head_fasta[[1]][3];
      array_head_fasta_end <- strsplit(array_head_fasta[[1]][4], "_");
      primer_use <- array_head_fasta_end[[1]][1];
      well_patient <- array_head_fasta_end[[1]][2];
      barcode <- array_head_fasta_end[[1]][4];
      well_sequencing <- array_head_fasta_end[[1]][5];
      seq_data <- fasta[[1]][1];
      post_data <- list(queryseq = seq_data, queryfile = "", organism =
"human", germline_db_V = "IG_DB/imgt.Homo_sapiens.V.f.orf.p", custom_db_V =
"", germline_db_D = "IG_DB/imgt.Homo_sapiens.D.f.orf", custom_db_D
="", germline_db_J = "IG_DB/imgt.Homo_sapiens.J.f.orf", custom_db_J
="", program = "blastn", min_D_match = "5", D_penalty = "-2", min_V_length =
"9", min_J_length = "0", J_penalty = "-2", num_alignments_V =
"3", num_alignments_D = "3", num_alignments_J = "3", translation =
"true", domain = "imgt", num_clonotype = "100", outfmt = "19", eval =
"1", num_alignments_additional = "10", SEARCH_TYPE = "IG", igsource =
"new", analyze = "on", CMD = "request");
      igblast_post <-
POST("https://www.ncbi.nlm.nih.gov/igblast/igblast.cgi", body = post_data,
encode = 'form');
      igblast_xml <- htmlParse(igblast_post, encoding = "UTF-8", asText =
TRUE);
      igblast_csv <-
xpathSApply(igblast_xml, "//div[@id='content']/pre", xmlValue);
      igblast_csv_data <- read.csv2(text = igblast_csv, sep = "\t",
encoding = "UTF-8");
      igblast_data <- head(igblast_csv_data, 1);
      sequence_alignment_aa <-
strsplit(as.character(igblast_data[1, "sequence_alignment_aa"]), "");
      germline_alignment_aa <-
strsplit(as.character(igblast_data[1, "germline_alignment_aa"]), "");

```

```

mutations_aa_count <- 0;
mutations_aa_fwr1_count <- 0;
mutations_aa_fwr2_count <- 0;
mutations_aa_fwr3_count <- 0;
mutations_aa_cdr1_count <- 0;
mutations_aa_cdr2_count <- 0;
mutations_aa_cdr3_count <- 0;
mutations_aa_positions <- NULL;
mutations_aa_differences_sequence <- NULL;
mutations_aa_differences_germline <- NULL;
mutations_aa_valid <- FALSE;
cdr1_start <- as.integer(igblast_data[1,"cdr1_start"]);
cdr1_end <- as.integer(igblast_data[1,"cdr1_end"]);
cdr2_start <- as.integer(igblast_data[1,"cdr2_start"]);
cdr2_end <- as.integer(igblast_data[1,"cdr2_end"]);
cdr3_start <- as.integer(igblast_data[1,"cdr3_start"]);
cdr3_end <- as.integer(igblast_data[1,"cdr3_end"]);
fwr1_start <- as.integer(igblast_data[1,"fwr1_start"]);
fwr1_end <- as.integer(igblast_data[1,"fwr1_end"]);
fwr2_start <- as.integer(igblast_data[1,"fwr2_start"]);
fwr2_end <- as.integer(igblast_data[1,"fwr2_end"]);
fwr3_start <- as.integer(igblast_data[1,"fwr3_start"]);
fwr3_end <- as.integer(igblast_data[1,"fwr3_end"]);
if(lengths(sequence_alignment_aa) == lengths(germline_alignment_aa)){
  mutations_aa_valid <- TRUE;
  for(i in 1:lengths(sequence_alignment_aa)){
    if(germline_alignment_aa[[1]][i] != 'X'){
      if(sequence_alignment_aa[[1]][i] !=
germline_alignment_aa[[1]][i]){
        mutations_aa_count <- mutations_aa_count+1;
        mutations_aa_positions <-
paste(mutations_aa_positions,paste(i,"",sep=""),sep="");
        mutations_aa_differences_sequence <-
paste(mutations_aa_differences_sequence,paste(sequence_alignment_aa[[1]][i]
,"",sep=""),sep="");
        mutations_aa_differences_germline <-
paste(mutations_aa_differences_germline,paste(germline_alignment_aa[[1]][i]
,"",sep=""),sep="");
        if(!is.na(cdr1_start) && !is.na(cdr1_end) && 3*i >=
cdr1_start && 3*i <= cdr1_end){
          mutations_aa_cdr1_count <- mutations_aa_cdr1_count+1;
        }
        if(!is.na(cdr2_start) && !is.na(cdr2_end) && 3*i >=
cdr2_start && 3*i <= cdr2_end){
          mutations_aa_cdr2_count <- mutations_aa_cdr2_count+1;
        }
        if(!is.na(cdr3_start) && !is.na(cdr3_end) && 3*i >=
cdr3_start && 3*i <= cdr3_end){
          mutations_aa_cdr3_count <- mutations_aa_cdr3_count+1;
        }
        if(!is.na(fwr1_start) && !is.na(fwr1_end) && 3*i >=
fwr1_start && 3*i <= fwr1_end){
          mutations_aa_fwr1_count <- mutations_aa_fwr1_count+1;
        }
        if(!is.na(fwr2_start) && !is.na(fwr2_end) && 3*i >=
fwr2_start && 3*i <= fwr2_end){
          mutations_aa_fwr2_count <- mutations_aa_fwr2_count+1;
        }
        if(!is.na(fwr3_start) && !is.na(fwr3_end) && 3*i >=
fwr3_start && 3*i <= fwr3_end){
          mutations_aa_fwr3_count <- mutations_aa_fwr3_count+1;
        }
      }
    }
  }
}

```

```

    }
  }
}

}

sql <- paste("INSERT OR IGNORE INTO b_cells
(p_acronym,b_acronym,date,primer_use,well_patient,barcode,well_sequencing,s
equence,locus,stop_codon,vj_in_frame,productive,rev_comp,v_call,d_call,j_ca
ll,sequence_alignment,germline_alignment,sequence_alignment_aa,germline_ali
gnment_aa,v_alignment_start,v_alignment_end,d_alignment_start,d_alignment_e
nd,j_alignment_start,j_alignment_end,junction,junction_length,junction_aa,j
unction_aa_length, v_identity, d_identity, j_identity,
v_sequence_start,v_sequence_end,v_germline_start,v_germline_end,d_sequence_
start,d_sequence_end,d_germline_start,d_germline_end,j_sequence_start,j_seq
uence_end,j_germline_start,j_germline_end,fwr1_start,fwr1_end,cdrl_start,cd
r1_end,fwr2_start,fwr2_end,cdr2_start,cdr2_end,fwr3_start,fwr3_end,cdr3_sta
rt,cdr3_end,valid,mutations_aa_valid,mutations_aa_count,mutations_aa_positi
ons,mutations_aa_differences_sequence,mutations_aa_differences_germline,mu
tations_aa_fwr1_count,mutations_aa_fwr2_count,mutations_aa_fwr3_count,mutati
ons_aa_cdr1_count,mutations_aa_cdr2_count,mutations_aa_cdr3_count) VALUES
('",paste(p_acronym,b_acronym,date,primer_use,well_patient,barcode,well_seq
uencing,igblast_data[1,"sequence"],igblast_data[1,"locus"],igblast_data[1,"
stop_codon"],igblast_data[1,"vj_in_frame"],igblast_data[1,"productive"],igb
last_data[1,"rev_comp"],igblast_data[1,"v_call"],igblast_data[1,"d_call"],i
gblast_data[1,"j_call"],igblast_data[1,"sequence_alignment"],igblast_data[1
,"germline_alignment"],igblast_data[1,"sequence_alignment_aa"],igblast_data
[1,"germline_alignment_aa"],igblast_data[1,"v_alignment_start"],igblast_dat
a[1,"v_alignment_end"],igblast_data[1,"d_alignment_start"],igblast_data[1,"
d_alignment_end"],igblast_data[1,"j_alignment_start"],igblast_data[1,"j_ali
gnment_end"],igblast_data[1,"junction"],igblast_data[1,"junction_length"],i
gblast_data[1,"junction_aa"],igblast_data[1,"junction_aa_length"],igblast_d
ata[1,"v_identity"],igblast_data[1,
"d_identity"],igblast_data[1,"j_identity"],igblast_data[1,"v_sequence_start
"],igblast_data[1,"v_sequence_end"],igblast_data[1,"v_germline_start"],igbl
ast_data[1,"v_germline_end"],igblast_data[1,"d_sequence_start"],igblast_dat
a[1,"d_sequence_end"],igblast_data[1,"d_germline_start"],igblast_data[1,"d_
germline_end"],igblast_data[1,"j_sequence_start"],igblast_data[1,"j_sequenc
e_end"],igblast_data[1,"j_germline_start"],igblast_data[1,"j_germline_end"]
,igblast_data[1,"fwr1_start"],igblast_data[1,"fwr1_end"],igblast_data[1,"cd
r1_start"],igblast_data[1,"cdr1_end"],igblast_data[1,"fwr2_start"],igblast_
data[1,"fwr2_end"],igblast_data[1,"cdr2_start"],igblast_data[1,"cdr2_end"],
igblast_data[1,"fwr3_start"],igblast_data[1,"fwr3_end"],igblast_data[1,"cdr
3_start"],igblast_data[1,"cdr3_end"],"TRUE",mutations_aa_valid,mutations_aa
_count,mutations_aa_positions,mutations_aa_differences_sequence,mutations_a
a_differences_germline,mutations_aa_fwr1_count,mutations_aa_fwr2_count,muta
tions_aa_fwr3_count,mutations_aa_cdr1_count,mutations_aa_cdr2_count,mutati
ons_aa_cdr3_count,sep = "','"),('')",sep="");
sql <- str_replace_all(sql,"'NA'",'');
dbExecute(con, sql);
sql <- paste(paste("update b_cells set date =
'",date,"'",,"well_patient = '",well_patient,"'",,"barcode =
'",barcode,"'",,"well_sequencing = '",well_sequencing,"'",,"sequence =
'",igblast_data[1,"sequence"],"'",,"locus =
'",igblast_data[1,"locus"],"'",,"stop_codon =
'",igblast_data[1,"stop_codon"],"'",,"vj_in_frame =
'",igblast_data[1,"vj_in_frame"],"'",,"productive =
'",igblast_data[1,"productive"],"'",,"rev_comp =
'",igblast_data[1,"rev_comp"],"'",,"v_call =
'",igblast_data[1,"v_call"],"'",,"d_call =
'",igblast_data[1,"d_call"],"'",,"j_call =
'",igblast_data[1,"j_call"],"'",,"sequence_alignment =

```

```

'",igblast_data[1,"sequence_alignment"],"',,"germline_alignment =
'",igblast_data[1,"germline_alignment"],"',,"sequence_alignment_aa =
'",igblast_data[1,"sequence_alignment_aa"],"',,"germline_alignment_aa =
'",igblast_data[1,"germline_alignment_aa"],"',,"v_alignment_start =
'",igblast_data[1,"v_alignment_start"],"',,"v_alignment_end =
'",igblast_data[1,"v_alignment_end"],"',,"d_alignment_start =
'",igblast_data[1,"d_alignment_start"],"',,"d_alignment_end =
'",igblast_data[1,"d_alignment_end"],"',,"j_alignment_start =
'",igblast_data[1,"j_alignment_start"],"',,"j_alignment_end =
'",igblast_data[1,"j_alignment_end"],"',,"junction =
'",igblast_data[1,"junction"],"',,"junction_length =
'",igblast_data[1,"junction_length"],"',,"junction_aa =
'",igblast_data[1,"junction_aa"],"',,"junction_aa_length =
'",igblast_data[1,"junction_aa_length"],"',,"v_identity =
'",igblast_data[1,"v_identity"],"',,"d_identity =
'",igblast_data[1,"d_identity"],"',,"j_identity =
'",igblast_data[1,"j_identity"],"',,"v_sequence_start =
'",igblast_data[1,"v_sequence_start"],"',,"v_sequence_end =
'",igblast_data[1,"v_sequence_end"],"',,"v_germline_start =
'",igblast_data[1,"v_germline_start"],"',,"v_germline_end =
'",igblast_data[1,"v_germline_end"],"',,"d_sequence_start =
'",igblast_data[1,"d_sequence_start"],"',,"d_sequence_end =
'",igblast_data[1,"d_sequence_end"],"',,"d_germline_start =
'",igblast_data[1,"d_germline_start"],"',,"d_germline_end =
'",igblast_data[1,"d_germline_end"],"',,"j_sequence_start =
'",igblast_data[1,"j_sequence_start"],"',,"j_sequence_end =
'",igblast_data[1,"j_sequence_end"],"',,"j_germline_start =
'",igblast_data[1,"j_germline_start"],"',,"j_germline_end =
'",igblast_data[1,"j_germline_end"],"',,"fwr1_start =
'",igblast_data[1,"fwr1_start"],"',,"fwr1_end =
'",igblast_data[1,"fwr1_end"],"',,"cdr1_start =
'",igblast_data[1,"cdr1_start"],"',,"cdr1_end =
'",igblast_data[1,"cdr1_end"],"',,"fwr2_start =
'",igblast_data[1,"fwr2_start"],"',,"fwr2_end =
'",igblast_data[1,"fwr2_end"],"',,"cdr2_start =
'",igblast_data[1,"cdr2_start"],"',,"cdr2_end =
'",igblast_data[1,"cdr2_end"],"',,"fwr3_start =
'",igblast_data[1,"fwr3_start"],"',,"fwr3_end =
'",igblast_data[1,"fwr3_end"],"',,"cdr3_start =
'",igblast_data[1,"cdr3_start"],"',,"cdr3_end =
'",igblast_data[1,"cdr3_end"],"',,"mutations_aa_valid =
'",mutations_aa_valid,""',,"mutations_aa_count =
'",mutations_aa_count,""',,"mutations_aa_positions =
'",mutations_aa_positions,""',,"mutations_aa_differences_sequence =
'",mutations_aa_differences_sequence,""',,"mutations_aa_differences_germlin
e = '",mutations_aa_differences_germline,""',,"mutations_aa_fwrl_count =
'",mutations_aa_fwrl_count,""',,"mutations_aa_fwrl2_count =
'",mutations_aa_fwrl2_count,""',,"mutations_aa_fwrl3_count =
'",mutations_aa_fwrl3_count,""',,"mutations_aa_cdr1_count =
'",mutations_aa_cdr1_count,""',,"mutations_aa_cdr2_count =
'",mutations_aa_cdr2_count,""',,"mutations_aa_cdr3_count =
'",mutations_aa_cdr3_count,"' where p_acronym = "',p_acronym,"' and
b_acronym = "',b_acronym,"' and primer_use = "',primer_use,"'", sep="");
sep="");
    sql <- str_replace_all(sql,"'NA'",'');
    #befehle fehlerhafte zeilen ausgeben
    dbExecute(con, sql);
  }
}
}

```

```

sql <- "UPDATE b_cells SET
p_acronym=NULLIF(p_acronym,''),b_acronym=NULLIF(b_acronym,''),date=NULLIF(d
ate,''),primer_use=NULLIF(primer_use,''),well_patient=NULLIF(well_patient,'
'),barcode=NULLIF(barcode,''),well_sequencing=NULLIF(well_sequencing,''),se
quence=NULLIF(sequence,''),locus=NULLIF(locus,''),stop_codon=NULLIF(stop_co
don,''),vj_in_frame=NULLIF(vj_in_frame,''),productive=NULLIF(productive,'')
,rev_comp=NULLIF(rev_comp,''),v_call=NULLIF(v_call,''),d_call=NULLIF(d_call
,''),j_call=NULLIF(j_call,''),sequence_alignment=NULLIF(sequence_alignment,
''),germline_alignment=NULLIF(germline_alignment,''),sequence_alignment_aa=
NULLIF(sequence_alignment_aa,''),germline_alignment_aa=NULLIF(germline_alig
nment_aa,''),v_alignment_start=NULLIF(v_alignment_start,''),v_alignment_end
=NULLIF(v_alignment_end,''),d_alignment_start=NULLIF(d_alignment_start,''),
d_alignment_end=NULLIF(d_alignment_end,''),j_alignment_start=NULLIF(j_align
ment_start,''),j_alignment_end=NULLIF(j_alignment_end,''),junction=NULLIF(j
unction,''),junction_length=NULLIF(junction_length,''),junction_aa=NULLIF(j
unction_aa,''),junction_aa_length=NULLIF(junction_aa_length,''),v_identity=
NULLIF(v_identity,''),d_identity=NULLIF(d_identity,''),j_identity=NULLIF(j_
identity,''),v_sequence_start=NULLIF(v_sequence_start,''),v_sequence_end=NU
LLIF(v_sequence_end,''),v_germline_start=NULLIF(v_germline_start,''),v_germ
line_end=NULLIF(v_germline_end,''),d_sequence_start=NULLIF(d_sequence_start
,''),d_sequence_end=NULLIF(d_sequence_end,''),d_germline_start=NULLIF(d_ger
mline_start,''),d_germline_end=NULLIF(d_germline_end,''),j_sequence_start=N
ULLIF(j_sequence_start,''),j_sequence_end=NULLIF(j_sequence_end,''),j_germl
ine_start=NULLIF(j_germline_start,''),j_germline_end=NULLIF(j_germline_end,
''),fwr1_start=NULLIF(fwr1_start,''),fwr1_end=NULLIF(fwr1_end,''),cdr1_star
t=NULLIF(cdr1_start,''),cdr1_end=NULLIF(cdr1_end,''),fwr2_start=NULLIF(fwr2
_start,''),fwr2_end=NULLIF(fwr2_end,''),cdr2_start=NULLIF(cdr2_start,''),cd
r2_end=NULLIF(cdr2_end,''),fwr3_start=NULLIF(fwr3_start,''),fwr3_end=NULLIF
(fwr3_end,''),cdr3_start=NULLIF(cdr3_start,''),cdr3_end=NULLIF(cdr3_end,'')
,valid=NULLIF(valid,''),vector_sequence=NULLIF(vector_sequence,''),isotype=
NULLIF(isotype,''),Specificity=NULLIF(Specificity,''),note=NULLIF(note,''),
PCR2=NULLIF(PCR2,'')";
dbExecute(con, sql);
dbDisconnect(con);
rm(list=ls());
detach("package:stringr", unload=TRUE);
detach("package:seqinr", unload=TRUE);
detach("package:xml2", unload = TRUE);
detach("package:XML", unload = TRUE);
detach("package:httr", unload=TRUE);
detach("package:RSQLite", unload=TRUE)

```