HHS Public Access

Author manuscript

Clin Exp Allergy. Author manuscript; available in PMC 2018 April 01.

Published in final edited form as:

Clin Exp Allergy. 2017 April ; 47(4): 577–592. doi:10.1111/cea.12829.

Immunoproteomic analysis of house dust mite antigens reveals distinct classes of dominant T cell antigens according to function and serological reactivity

Carla Oseroff[#1], **Lars H. Christensen**[#2], **Luise Westernberg**[1], **John Pham**[1], **Jerome Lane**[1], **Sinu Paul**[1], **Jason Greenbaum**[1], **Thomas Stranzl**[2], **Gitte Lund**[2], **Ilka Hoof**[2], **Jens Holm**[2], **Peter A Würtzen**[2], **Kåre H. Meno**[2], **April Frazier**[1], **Veronique Schulten**[1], **Peter S. Andersen**[2], **Bjoern Peters**[1], and **Alessandro Sette**[1]

[1]La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA

[2]ALK-Abelló A/S, Global Research, Hørsholm, Denmark

[#] These authors contributed equally to this work.

Abstract

BACKGROUND—House dust mite (HDM) allergens are a common cause of allergy and allergic asthma. A comprehensive analysis of proteins targeted by T cells, which are implicated in the development and regulation of allergic disease independent of their antibody reactivity, is still lacking.

OBJECTIVE—To comprehensively analyze the HDM-derived protein targets of T cell responses in HDM-allergic individuals, and investigate their correlation with IgE/IgG responses and protein function.

METHODS—Proteomic analysis (liquid chromatography tandem mass spectrometry) of HDM extracts identified 90 distinct protein clusters, corresponding to 29 known allergens and 61 novel proteins. Peripheral blood mononuclear cells (PBMC) from 20 HDM-allergic individuals were stimulated with HDM extracts and assayed with a set of ~2500 peptides derived from these 90 protein clusters and predicted to bind the most common HLA class II types. 2D immunoblots were made in parallel to elucidate IgE and IgG reactivity and putative function analyses were performed *in silico* according to gene ontology (GO) annotations.

RESULTS—Analysis of T cell reactivity revealed a large number of T cell epitopes. Overall response magnitude and frequency was comparable for known and novel proteins, with 15 antigens (nine of which were novel) dominating the total T cell response. Most of the known allergens that were dominant at the T cell level were also IgE-reactive, as expected, while few novel dominant T cell antigens were IgE reactive. Among known allergens, hydrolase activity and

Corresponding Author: Alessandro Sette, Dr. Biol. Sci., Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037. Telephone: 858-752-6916, Fax: 858-752-6987, alex@lji.org.

Conflict of interest

Alessandro Sette and Bjoern Peters are consultants for ALK-Abelló A/S Bøge Allé 6 DK-2970 Hørsholm, Denmark. Lars H. Christensen, Thomas Stranzl, Gitte Lund, Ilka Hoof, Jens Holm, Peter A Würtzen' Kåre H. Meno, Peter S. Andersen are employed by ALK-Abelló A/S Bøge Allé 6 DK-2970 Hørsholm, Denmark

detectable IgE/IgG reactivity are strongly correlated, while no protein function correlates with immunogenicity of novel proteins. A total of 106 epitopes accounted for half of the total T-cell response, underlining the heterogeneity of T cell responses to HDM allergens.

CONCLUSIONS AND CLINICAL RELEVANCE—Herein, we define the T cell targets for both known allergens and novel proteins, which may inform future diagnostics and immunotherapeutics for allergy to HDM.

Keywords

house dust mite; allergens; proteomics; T cell epitopes

Introduction

Atopic allergies are associated with highly complex immune responses, in terms of both antibodies (IgE and IgG) and T cells [1]. Allergen extracts containing tens or hundreds of distinct molecular and immunogenic species are commonly used to characterize immune reactivity. In the vast majority of cases, allergic T cell responses are complex and target multiple antigens, but previous studies of extract-specific responses have typically been limited to major allergens that by definition trigger significant IgE responses. Furthermore, the biochemical and molecular protein characteristics determining allergenicity and immunological dominance in the context of IgE and T cell responses are not well understood.

In house dust mite (HDM) allergy, the major allergens so far reported to induce IgE responses are associated with proteolytic/enzymatic activity [2-4], and also trigger innate immune responses [5-7]. Sequence conservation or homology also seems to be an important factor in allergen-specific responses, which are often highly cross-reactive [8]. Furthermore, the immunodominance of T cell and B cell response does not consistently correlate [9]. As has previously been shown in HDM allergy, overall T cell and IgE reactivity do not generally correlate [10], but several studies have pointed out a lack of correlation at the level of individual proteins and individual patients in other allergenic systems [11, 12].

HDMs, which belong to the genus *Dermatophagoides* [13], are one of the most frequent indoor allergen sources worldwide and are potent inducers of perennial asthma and rhinitis [14-16]. Several groups of allergens from *D. pteronyssinus* (Der p) and *D. farina* (Der f) with diverse biological functions have been described (www.allergen.org) [17, 18]. Der p/f 1 and Der p/f 2 are the major allergens, as defined by the prevalence of specific IgE in HDM-allergic patients [19, 20]. In addition, Der p 23 has recently been shown to have an IgE prevalence comparable to those against Der p 1 and 2 and it hence represents another major HDM allergen [21, 22]. A previous study, analyzing T cell reactivity to Der p/f 1, Der p/f 2 and Der p 23, showed that while Der 1 and 2 are dominant T cell antigens, Der 23 is associated with marginal T cell reactivity [10].

Several other known allergens from the genus *Dermatophagoides* associated with variable levels of IgE have been described. Investigations of IgE responses to these other allergens have focused on Der p, revealing that allergens that account for most of the remaining IgE

reactivity apart from Der p 1, 2, and 23 are Der p 4, 5, 7, and 21 [23, 24]. IgE reactivity against the remaining allergens (Der p/f 3, 6, 8-10, 11, 13-18, 20-22, 24-33) is infrequent and of low titer [21, 23-25].

At the level of T cell responses, several studies have described epitopes [26-50]. Most epitopes are, however, derived from Der p/f 1, Der p/f 2, Der p 23 and Der p 4; very little information is available for other allergens. A comprehensive characterization of HDM T cell epitopes, associated antigens, and patterns of T helper cell responses is required to better understand pathogenic immune responses.

Using an immunoproteomics approach to define and characterize novel T cell targets in Timothy grass and cockroach allergic patients, we have previously demonstrated that the allergic T cell response extends beyond IgE-reactive allergens [10, 51, 52]. We have also improved the feasibility of high-throughput screening of a large number of potential T cell-reactive peptides with limited cell availability using a sequential lyophilization approach that generates megapools of 100 or more peptides [53]. The sequential lyophilization approach is based on the simple principle that one peptide can act as an excipient for another different one, thus facilitating reaching higher solubility of a peptide pool (e.g. the solubility of negatively charged peptides is increased in presence of positively charged ones that act as counter-ions).

In this study, we use this approach to comprehensively analyze novel T cell targets within the HDM proteome, including known allergens and novel proteins. We identify several dominant T cell targets and map the associated epitopes for each. We further address the underlying immunodominance in allergen-specific T cell responses by examining the correlation of T cell reactivity to serological reactivity, biological function and evolutionary conservation.

Methods

Identification of novel HDM proteins by liquid chromatography tandem mass spectrometry (LC-MS/MS)

Preparation of Der f and Der p body and feces extracts—Two separate fractions consisting of mite bodies and feces, respectively, from each of two HDM cultures, Der p and Der f, were gently extracted in 10% (w/v) PBS, pH 7.2 for 10 minutes at RT (mimicking the events taking place at the respiratory mucosal surface when HDM particles are inhaled). 50 μg of each dried extract was digested with trypsin and analyzed by LC-MS/MS as follows.

LC-MS/MS—Reverse-phase liquid chromatography (Ultimate 3000 RSLC nano, Thermo Scientific, Waltham, MA, USA) was performed using C18 pre and analytical columns at a flow rate of 300 nL/min. Peptides eluting from the LC were sprayed directly into an ESI-QTOF mass spectrometer (maXis, Bruker, Billerica, MA, USA) and spectra were acquired in the mass range of 50-2200 m/z at 2 Hz and MS/MS sequencing at a spectral rate of 4-16 Hz.

Protein identification—Proteins were identified by searching (via MASCOT 2.2 and X! Tandem search engines) the MS/MS spectral data against a database compiled of: in-house transcriptomes (Illumina HiSeq 2000, Trinity assemblies [54, 55]) of the two HDM species Der f and Der p, in-house transcriptomes of four other mite/storage mite species; *Blomia tropicalis* (Blo t), *Glycyphagus domesticus* (Gly d), *Lepidoglyphus destructor* (Lep d) and *Tyrophagus putrescentiae* (Tyr p), Swissprot and Trembl sequences from the *Acari* subclass, and previously identified allergens from Der f and Der p (extracted from allergen.org and allergome.org as of November 2014), adding up to a database of a total of 409,187 protein sequences. In total, we identified 438 potential proteins, isoforms included, given a false discovery rate <2%.

Clustering of identified proteins

The mass-spectrometry studies described in the section above identified a large number of sequences derived from Der p and Der f. These corresponded to a total of 438 different protein sequences that included isoforms and overlapping sequences. To avoid over-representing essentially identical protein sequences, the next analysis generated a set of non-redundant sequences by removing pairs of sequences that shared 80% or more identity [56]. As a result of this analysis, the 438 sequences identified in mass spectrometry plus additional Der p and Der f allergen sequences from public databases were classified into 96 groups (also referred to as clusters). All sequences contained in each cluster were aligned to each other using the MEGA software tool (using ClustalW) [57].

Six of the 96 clusters corresponded to isoforms of Der p/f 1, 2 and 23. As epitopes in these antigens were previously mapped [10], these six clusters were removed from consideration, leaving a set of 90 clusters for epitope prediction. Of those, a minority corresponded to known Der p/f proteins that were previously described as associated with some IgE reactivity (while not necessarily considered major allergens), and 61 were novel, to the best of our knowledge previously undescribed protein sequences, at least in terms of IgE/T cell reactivity. For each protein/cluster, **Supplementary Table 1** lists the protein name, accession number and molecular function.

Molecular function annotation

To assign putative functions to the identified protein sequences, we compared their sequences to protein with known functions annotated with Gene Ontology (GO) [58, 59] terms using BLAST2GO [60] GOslim reduced the number of GO terms [61]. Terms with low representation were regrouped into common parent concepts (“Category” column in **Supplementary Table 1**). ORFs/proteins with no classification were left blank. To investigate possible correlations between each protein's molecular function and IgE, IgG and T cell reactivity, ORFs/proteins were grouped according to function and the association of each with IgE, IgG or T cell reactivity was determined by the Fisher exact or chi-square test, for comparisons of two or three categories, respectively. Statistical tests were done on the full protein set, as well as separately for previously described allergens other proteins.

Promiscuous HLA class II binding predictions and pool generation

Each sequence cluster was aligned separately using the MEGA software tool with ClustalW. Stretches of fifteen amino acid residues overlapping by 10 amino acids were generated overlapping the sequence alignment. Peptides from each sequence in the alignment were generated. In case of gaps in the alignment, the gaps were removed in the peptide generation, and additional residues were added at the end. For example, peptide "---PPQ----PKMAD" from region "---PPQ----PKMADQLTEEQI" becomes "PPQPKMADQLTEEQI". Further, peptides with unknown residues in the alignment (indicated by "X") and duplicate peptides were removed, leaving a total of 14,783 unique peptides.

HLA class II binding predictions optimized for global coverage were performed for seven class II alleles (HLA-DRB1*03:01, HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*01:01) as previously described using the standalone version of the IEDB class II binding prediction tool [62]. Briefly, the binding of a given peptide was predicted to each of the seven alleles listed above. Binding affinities were quantified as percentile ranks by comparing the predicted affinity to a large set of random peptide sequences. The medium rank of the seven alleles was taken as an overall single metric of binding affinity to the allele panel, which measures both breadth and affinity of binding. Using this seven-allele binding predictions of each peptide, peptides were ranked from highest to lowest affinity binder. If two peptides shared ten or more overlapping 10 amino acids (which mostly occurred due to repetitive protein sequences), the lower ranking one was removed. Similarly, peptide variants that originated from the same sequence alignment position were also removed, retaining the better peptide based on the median consensus percentile rank and conservation among the sequences within its respective cluster. Peptides with median consensus percentile rank ≤10.0 and conserved in ≥35% of sequences in the same cluster were finally selected, also including additional selected peptides chosen to maximize DRB1 allele coverage, for a grand total of 2,589 peptides.

Peptide synthesis

Peptides were purchased from Mimotopes (Clayton, Victoria, Australia) and/or A and A (San Diego, CA, USA) as crude material on a small (1 mg) scale. Individual peptides were resuspended in DMSO at a final concentration of 40 mg/mL. Peptide "megapools" of 30-65 peptides/pool were generated as described [53]. According to this procedure, each individual lyophilized peptide is dissolved in 100% DMSO at 20 mg/ml (for 1 mg of peptide, this corresponds to a volume of 50 μl). Then equal amounts of each peptide are mixed; for 100 peptides, this corresponds to 50 μl × 100 = 5 ml; the total peptide concentration is still 20 mg/ml, but the concentration of each peptide is now 0.2 mg/ml, all in 100% DMSO. The resulting 5 ml are then lyophilized again, adding water if required. The resulting megapool is carefully resuspended in as small amounts of DMSO as feasible. Usually a megapool easily dissolves in 500 μl of 100% DMSO. This corresponds to 2 mg/ml of each peptide in 100% DMSO, and a final concentration of 5 μg/ml in an assay corresponds to 0.25% DMSO. This approach has been used to develop megapools specific for Timothy grass, TB, DENV, pertussis and tetanus [53, 63, 64].

In the end, each pool (regular, meso or mega) was reconstituted in DMSO so that each peptide was present at a concentration of 4 mg/mL. To facilitate deconvolution of positive megapools, each megapool was further broken down into 2-6 “mesopools” (259 mesopools in total), each containing 8-14 peptides (**Supplementary Table 2)**. Each mesopool was then deconvoluted to identify individual positive peptides.

Study population

PBMCs from 20 European HDM-allergic individuals (defined by clinical evaluation and Der p and Der f extract IgE titers greater than or equal to 0.35 kU_A/L). For these European donors, the clinical evaluation was based on patient history and verbal reporting of previous allergy testing. Positive mite responses were based on year round/out of season symptoms. In addition, PBMCs from 10 American HDM-allergic individuals (defined by Der p extract IgE titers greater than or equal to 0.35 kU_A/L) were obtained at LJI. Donors were recruited in the Copenhagen region (ALK) or in San Diego (LJI) following informed consent (IRB approved protocols include ALK Project ID: H-3-2014-129 and LJI Project ID: VD-112-0315). PBMCs were isolated from whole blood by density gradient centrifugation according to manufacturers’ instructions (Ficoll-Hypaque, Amersham Biosciences, Uppsala, Sweden). Der p- and Der f-specific extract IgE titers were determined using the ImmunoCAP system (Thermo Fisher, Uppsala, Sweden). Age, gender and IgE information are provided in **Supplementary Table 3**. In a separate series of experiments, pooled plasma from 10 European and 10 American HDM atopic individuals from the San Diego region, respectively, was utilized to run 2D immunoblots to elucidate IgE and IgG reactivity towards the novel HDM proteins and allergens (within HDM extracts), covered by the synthetic peptides (as described in a later section). The donors used in the 2D immunoblots are identified in **Supplementary Table 3**.

Stimulation and expansion of HDM specific T cells and dual ELISPOT assays

HDM-specific T cells were expanded in vitro as previously described [10]. Briefly, PBMCs from HDM-allergic individuals were stimulated with HDM extract (5 μg/mL) and expanded over 14-17 days with IL-2 (added every 3 days). Cells were harvested on day 14, restimulated with HDM extract (5 μg/mL), individual peptides (10 μg/mL) or peptide pools (5 μg/mL) and screened for IFNγ/IL-5-production by ELISPOT as described previously [11]. Criteria for positivity were 100 or 20 spot forming cells (SFCs) per 10^6 PBMCs for peptide pools or single peptides, respectively, $p < 0.05$, and a stimulation index > 2 [65-68].

Determination of IgE and IgG reactivity

Methods for determining IgE and IgG reactivity have previously been described for cockroach [52] and Timothy grass [51] allergens. Briefly, extracts of Der p and Der f were mixed 1:1 and 300 μg of protein was run on 2D gels (3 –10 pH range, 12% 138 (vol/vol) acrylamide) at Applied Biomics (Hayward, CA). The 2D-immunoblots of the labeled extracts were incubated with either (1) pooled plasma (diluted 1:20) from 10 HDM allergic donors recruited in San Diego (**Supplementary Table 3**), or (2) pooled sera from 10 HDM allergic donors recruited in Europe (diluted 1:33) (**Supplementary Table 3**).

Blots were incubated with goat anti-human IgE and mouse anti-human IgG (Sigma-Aldrich, St. Louis, MO, USA), and HDM donor antibody reactivity visualized using Cy2-conjugated donkey anti-goat IgG and Cy5-conjugated donkey anti-mouse IgG antibodies (Biotium, Fremont, CA, USA). We then determined the antibody reactivity of each spot by visual inspection of the 2D gel images. Reactivities of both the San Diego and European serum pools were included in the analysis (**Supplementary Table 1**). In total 237 IgE and/or IgG-reactive protein spots were picked and analyzed by mass spectrometry (**Supplementary Figure 1)**. The MS spectra were compared to the MS database previously described in the LC-MS/MS method section to identify the most likely protein within each spot. Each protein was linked to the assigned protein cluster identified by the proteomic approach described in a previous section.

Homology analysis

To determine if the HDM proteins identified in the proteomics analysis were conserved within other *arachnids*, their sequences were compared to three *arachnid* proteomes (Ixodes scapularis, Metaseiulus occidentalis, Stegodyphus mimosarum) derived from *de novo* sequence assembly. Each HDM protein was aligned against each proteome to identify proteins that had > 70% sequence identity for > 50% of the length of the proteome transcript, and these proteins were considered as conserved in the corresponding species. Similar analyses were performed for each of the HDM protein sequences against 1,130 proteins of the aero, bacteria, contact and venom or salivary categories from the AllergenOnline Database version 15 [69].

RESULTS

Identification of a comprehensive set of known and novel HDM proteins and elucidation of antibody reactivity patterns

To comprehensively map T cell reactivity to known and novel HDM proteins, and to broadly examine the relationship between IgE and T cell responses, we initially identified a large number of known and novel HDM proteins by LC-MS/MS analysis of four HDM extracts. The mass spectra were searched against a comprehensive MS database assembled from both known HDM sequences including, among others, the IUIS allergen database, and from novel ORFs/proteins (two HDM transcriptomes plus four storage mite transcriptomes).

To remove overlapping sequences and isoforms, the identified HDM protein sequences were grouped together into as described in the methods. Clusters corresponding to the antigens Der p 1, 2, and 23 (which were previously studied [10]) were removed, leaving 90 clusters of protein sequences. Out of these 90 clusters, 29 corresponded to known Der p/f allergens previously associated with IgE reactivity. Most of the remaining 61 protein clusters corresponded to protein sequences not previously described. Identifiers (IDs) for each protein/cluster are listed in **Supplementary Table 1**. To determine putative functions of the protein sequences in each cluster in order to later correlate them with immune reactivity, we examined if there were homologs of these genes with known functions. We relied on gene ontology (GO) molecular function annotations freely available online as described in the Methods section, which are displayed in **Supplementary Table 1**.

To elucidate the antibody reactivity (IgE and IgG) to these HDM protein clusters, 2D immunoblots were prepared utilizing a mixture of Der p/f extracts and pooled plasma from 10 European or 10 American HDM atopic individuals, respectively (**Supplementary Figures 1a-b**). For each spot, IgE and IgG reactivity was determined and used to identify IgE^+ and IgG^+ protein clusters, as tabulated in **Supplementary Table 1**. The two different plasma pools differing in terms of geographical locations were used to generalize our finding to different patient populations; while in general the reactivity was similar, spots were selected on the basis of being positive with any of the two plasma pools.

Prediction of HLA class II promiscuous binders and generation of mega- and mesopools

Protein sequences within each cluster were aligned and their binding affinity to a panel of HLA class II molecules was predicted as described in the methods. This resulted in the selection of 2,589 15-mer peptide sequences. To allow for effective screening of this large number of peptides, we adopted a strategy in which larger pools (megapools) were initially screened, and the peptides encompassing any positive megapool were screened again in smaller pools (mesopools). Positive mesopools were eventually deconvoluted to identify individual positive peptides. Megapools of 30-65 peptides/pool were prepared for screening by sequential lyophilization as previously described [53]. Each megapool was further broken down into 2-6 “mesopools” (259 mesopools in total), each containing 8-14 peptides, as described in the Methods. Each mesopool yielding positive responses was deconvoluted to identify individual positive peptides.

T cell reactivity against novel and known HDM proteins

T cell reactivity was measured by ELISPOT in PBMCs from 20 HDM-allergic subjects in response to peptide pool restimulation after 14-17 day stimulation with Der p/f extract. The donor population was 21 to 76 years in age (average 49 ± 15 years), and with an M/F ratio of 50%; IgE titers for Der p and Der f ranged between 0.77 kU_A/L and 116.44 kU_A/L with a median of 13 kU_A/L (**Supplementary Table 3**).

The reactivity detected against peptides derived from each protein cluster, both in terms of magnitude (sum of SFC/donor) and frequency of positive responses is listed in **Supplementary Table 1.** A detailed list of all reactive epitopes and associated magnitude and frequency of responses is shown in **Supplementary Table 4**. The overall reactivity, defined as the sum of all reactivity against any individual positive peptide, is illustrated in **Figure 1A**, and the corresponding overall frequency of donor recognition of each allergen is shown in **Figure 1B**. In general, the hierarchy of immunodominance of response frequency was similar to that observed for response magnitude.

Overall, seven known HDM allergens elicited a response magnitude > 500 $SFC/10^6$ PBMC/donor (**Table 1**). For the known allergens, reactivity to peptides derived from Der p/f 3, Der p/f 6, Der p/f 9 and Der p/f 20 were particularly prominent, ranging from 1,000 to 4,000 $SFC/10^6$/donor. Peptides from nine of the novel HDM proteins were also associated with a similarly strong response (**Table 1**). T cell recognition of peptides from novel antigens 10, 75 and 102 were prominent (> 100 $SFC/10^6$ PBMC/donor for several individual peptides from antigen 10 and 75 and > 300 for antigen 102). The average sum of responses to

peptides from the 29 known HDM allergens in all 20 donors was 13,263 SFC, while the average sum for peptides from the 61 novel HDM proteins was 15,659 **(Supplementary Table 1)**.

Table 1 lists the protein clusters for which T cell responses against derived peptides were above an arbitrary cutoff of > 500 SFC /donor, including seven of the 29 known allergens and nine of 61 novel antigens. These antigens accounted for 86% and 62% of the total known and novel antigen reactivity, respectively. As mentioned above, the reactivity of Der p/f 1, 2, and 23 was not investigated here because the T cell reactivity of these antigens was investigated in a previous study [53]. In that study, the reactivity was as follows: Der p 1, 3046 SFC; Der f 1, 3450; Der p 2, 1461; Der f 2, 2245 and Der p 23, 344. While it should be emphasized that these data result from testing complete sets of overlapping peptides, and the present study utilized an approach based on predictive HLA binding that is expected to capture about 50% of the total response [53], by these criteria, Der p/f 1, 2 would be classified as above the 500 SFC threshold and Der p 23 below the same threshold.

The data was also assessed for polarization of Th responses based on IFNγ vs. IL-5-production in ELISPOT assays. Consistent with the allergic status of the donors tested, overall Th2 (IL-5) responses dominated Th1 (IFNγ) responses by a factor of approximately 25:1 (**Table 1**); a similar pattern was observed regardless of the antigen.

Association between T cell immunodominance and IgE/IgG reactivity

We next investigated in more detail potential correlates of immunodominance. We found a strong correlation between T cell reactivity and IgE/IgG reactivity for the known Der p/f allergens (**Table 2a**). Specifically, six of the 7 (86%) known allergens that were dominant at the T cell level were also positive for IgE reactivity in the cohort of the present study. Six of the 14 intermediate T cell antigens (43%) were IgE-reactive, while only one of the 8 (13%) of the T cell unreactive proteins were IgE-positive among the known allergens.

By contrast, among the novel proteins, only one of the 9 (11%) dominant T cell antigens was IgE-reactive, which is significantly less than the proportion of IgE-reactive novel HDM proteins that were negative for T cell activity (5/19; 26%). Indeed, most of the IgE reactivity in the novel HDM proteins was associated with intermediate T cell reactivity (15/33; 46%). A very similar pattern was observed in the case of IgG responses (**Table 2b**). Thus, while IgE reactivity appears to be associated with T cell reactivity for the known HDM proteins, a different class of novel HDM antigens induces potent T cell responses, associated in general with only marginal IgE and IgG reactivity.

HDM allergens are not preferentially conserved in other Arachnida genomes

Conservation amongst taxonomically related allergen sources is important in determining allergenicity of pollen allergens [70]. In the case of HDM, no other well-described allergenic species are closely related and could contribute to modulation of allergenicity by this mechanism. Nevertheless, we assessed conservation of the sequences corresponding to each of the various HDM allergens and novel antigens relative to the proteomes of each of three species relatively close to HDM from an evolutionary point of view, namely the two *acari,*

Ixodes scapularis and *Metaseiulus occidentalis* and the other *arachnida, Stegodyphus mimosarum*. The results of this analysis are presented in **Supplementary Table 1**, listing the number of species for which each HDM protein had at least > 70% sequence identity for > 50% of the length of the transcript. No significant association was noted between protein conservation and either IgE, IgG or T cell reactivity. Similar analyses were performed comparing each of the sequences against the AllergenOnline Database [71], which similarly did not reveal any significant association (data not shown).

Association of HDM allergens' molecular function with IgE/IgG reactivity and T cell immunodominance

Next, we investigated whether the broad range of molecular functions among the investigated proteins correlated with immunodominance. Among the functions listed in **Supplementary Table 1,** six classes (transferase, structural molecule, oxidoreductase, isomerase, hydrolase and binding-related) included five or more occurrences. The association of each with IgE, IgG or T cell reactivity was determined by Fisher exact or chi-square tests. Significant associations were detected only for the hydrolase class (data not shown). As shown in **Figure 2a**, T cell reactivity correlated with hydrolase activity when all antigens were considered together ($p = 0.01$).

When known and novel HDM proteins were considered separately, the association remained significant for the known allergens (**Figure 2B**, $p = 0.04$), but the same visual trend did not achieve statistical significance (**Figure 2C,** $p = 0.11$) among novel proteins. Hydrolase functionality was also significantly associated with IgE/IgG reactivity among known HDM allergens (**Table 2c;** $p = 0.005$ and 8.0×10^{-5}, respectively), but not for novel HDM proteins ($p = 0.066$ and 0.25, respectively). Overall, the association of T cell reactivity, hydrolase activity and serological reactivity among known HDM allergens does not appear to hold true for novel T cell antigens.

Frequency and magnitude of responses and identification of dominant T cell epitopes

Finally, individual dominant T cell epitopes were identified by analysis similar to that for the dominant HDM antigens. Overall, (**Supplementary Table 4)** we detected responses against 674 different peptides, 241 of which were recognized in two or more donors (corresponding to recognition frequency of ≥ 10%).

These 241 peptides accounted for 21,339 SFC/donor (74% of the total 28,922 response). Of those, 106 epitopes were recognized with an average magnitude of 50 SFC/donor or higher and accounted for 16,658 SFC/donor (or 58% of the total response against all 674 epitopes). **Table 3** lists these 106 main epitopes with the associated antigen, the percentage of donors responding, and the percentage of the average SFC response accounted for by each epitope.

As mentioned above, the reactivity of Der p/f 1, 2, and 23 was investigated in a previous study [10]. The list of epitopes that would fulfill the same criteria of an average magnitude of 50 SFC/donor or higher, as previously published [10] is listed for reference purpose in **Supplementary Table 5**. These correspond to 12 Der p 1, 11 Der f 1, 6 Der p 2, 6 Der f 2 and 1 Der p 23 epitopes, encompassing 90%, 90%, 94%, 94% and 98% of the total response,

respectively. Overall, these data demonstrate the extreme degree of heterogeneity of T cell responses specific for HDM antigens, both among donors and in the sense of a large number of antigens being recognized by T-cells with different frequency/magnitude.

Discussion

Here, we present the most detailed analysis of the allergic immune response to HDM to date. This study demonstrates the general applicability of an immunoproteomic approach to study allergen-specific T cell responses, previously applied to Timothy grass and cockroach allergy, leading to the discovery of dozens of novel T cell targets [51, 52]. It further validates the applicability of the use of epitope megapools as a powerful tool to screen and define T cell epitopes in a high-throughput fashion [10, 53].

The data presented in this study contribute to the understanding of allergic immune responses in HDM allergy on three levels. First, we identify the protein targets of T cell responses and general features associated with T cell immunogenicity for HDM proteins. For the known HDM allergens, we find that allergens that elicit strong T cell reactivity are mostly IgE- and IgG-reactive in the 2D immunoblot assay and plasma pools tested. However, our several allergens considered minor based on the prevalence of IgE responses are prominent T cell targets, such as Der p/f 3, Der p/f 6 and Der p/f 20. In particular, while IgE recognition of Der p/f 3, 6, and 9 is observed in a considerable number of allergic subjects, they account for a minor part of overall IgE reactivity; little IgE binding to Der p/f 20 has been reported [72]. Therefore, we conclude that while we find a strong association between T cell and IgE reactivity against the known HDM allergens in general, the hierarchy of immunodominance of B and T cell targets may differ significantly. Interestingly, several novel antigens accounted for significant fractions of the total T cell response, including 10, 75, and 102, despite not being associated with IgE reactivity. The finding that dominant HDM T cell targets may exist that are not IgE-reactive may be relevant for immunotherapy, as they reduce the risk of IgE-mediated adverse reactions.

We report an association between T cell and IgE reactivity against the known allergens, but no correlation was found between IgE and the novel antigens. The most likely explanation for this phenomenon is that all antigens recognized by antibody responses are recognized by T cells, but not all antigens recognized by T cells are also recognized by IgE. This might in turn be due to structural features influencing allergen-binding accessibility. In this light, it is possible to speculate that the situation might be somewhat analogous to antibody and T cell recognition of other antigens such as viruses. In the case of well-studied viruses such as hepatitis B virus (HBV), influenza or dengue virus (DENV), it was shown that the antibody responses are focused on surface and envelope proteins, but T cells recognized also internal parts such as M, NP, NS and core [64]. In addition, we speculate that investigational bias might also contribute in some degree, to these findings since the proteins that have been studied most extensively to date are the ones that are recognized by IgE (the feature classically used to define allergens).

As noted above, Der p/f 1, 2 and 23 were not analyzed in this study, as they were studied previously by our group using the same methodology [10]. The reactivity of a similar and

largely overlapping panel of donors to the Der p/f 1, 2, and 23 allergens was as follows: Der p 1, 3046 SFC; Der f 1, 3450; Der p 2, 1461; Der f 2, 2245 and Der p 23, 344. It should be further noted that these figures result from testing complete sets of overlapping peptides, while the present study utilized a predictive scheme expected to capture about 50% of the response [53]. Overall, even considering the additional reactivity against these allergens, the data presented here represents the first description of several T cell allergens that account for a sizeable fraction of the total reactivity.

Second, we analyzed the potential association of protein conservation and allergic T cell reactivity. Analysis of T cell responses to pollen from multiple allergenic plants suggests that conservation is of high relevance in determining allergenicity [70]. In the case of HDM, no other well-described allergenic species (with the exception of Blo t, which are widespread in subtropical regions and therefore not relevant in this context) are closely related and could contribute to the degree of allergenicity. Nevertheless, we assessed conservation of the sequences corresponding to the various HDM allergens and proteins in three species relatively closely related to HDM and found that the HDM allergens investigated are not preferentially conserved in other Arachnida genomes.

Several studies have indicated a potential correlation between various types of enzymatic activity/protein functions and allergenicity of HDM proteins [73, 74]. Der p/f 1 allergens are cysteine proteases while Der p/f 2 interacts with the innate immune system [75, 76]. Der p/f 1 may enhance overall immunogenicity of HDM by disrupting tight junctions, and Der p/f 2 mimicry of Toll-like receptor 4 could drive airway inflammation [74]. For known allergens, putative hydrolase activity is associated with IgE reactivity, but less so in the case of novel proteins. The reason for this difference is not clear but might reflect less accurate GO annotation for novel proteins. Alternatively, it might reflect different mechanisms influencing immunogenicity for known and novel allergens, consistent with the fact that IgE reactivity correlates with T cell reactivity for the known but not for the novel HDM allergens.

Third, our study has implications for defining the HDM T cell epitope repertoire in HDM-allergic individuals. In this study, 674 epitopes elicited a response in at least one donor. As of February 2016, the Immune Epitope Database (IEDB) contained records associated with 206 different epitopes from 27 references [26-50]. The present study thus increased the number of described epitopes by approximately threefold. Besides being the first account of any epitopes from previously undescribed HDM proteins, this study is the first to identify specific epitopes from Der p/f proteins other than Der p/f 1-5, 9 and 13-14.

In most cases, responses to HDM peptides/epitope regions previously described by others were also observed here with similar or higher frequency, and in addition, new T cell epitopes were found for many HDM allergens. Most previous studies focused on Der p 3, 4 and 14. There was a considerable overlap between the T cell responses observed in the present study and the Der p 3 T cell epitope-containing regions investigated by Zhan et al. [77]. Several Der p 4 epitopes observed by Hinz et al. overlapped with those demonstrated here, though the high frequency of Der p 4 aa 381-401) was not replicated here [10]. Finally, no responses to the Der p 14 epitope-containing regions described by Fujii et al. were

observed [36]. However, these epitopes were not demonstrated by Hinz et al. either, and most Der p 14 epitopes observed in that study were also found here, albeit with low frequency in both [10].

The most dominant 241 peptides recognized by more than one donor accounted for 74% of the total response. The top 106 epitopes, recognized with an average of ≥ 50 SFC/donor, accounted for 78% of the response within the top 241 peptides and 58% of the total response. These data illustrate the remarkable breadth of T cell responses associated with HDM allergies. This remarkable breadth of responses is further enhanced by considering an additional 34 previously identified epitopes derived from Der p/f 1, Der p/f 2 and Der p 23, encompassing a large fraction of the total response directed against those allergens.

Despite this large breadth of responses, it is noteworthy that many of the HLA binding peptides are not recognized by T cell responses in allergic patients. Based on the current data and previous reports, we are exploring what specific characteristics in HDM T cell epitopes might be associated with T cell recognition. We have analyzed the sequences from HDM T cell epitopes and compared them to sequences derived from the same allergen but not recognized by T cells. No significant difference at the level of predicted HLA binding was observed (data not shown). We are interested in addressing the question of what specific characteristics make a defined sequence immunogenic, while others, with similar binding capacity and derived from the same antigens are not recognized. We hypothesize two main factors to be contributing to this issue, namely differential cellular processing [78] and TCR repertoire/prominence [79, 80]. In terms of differential processing, comparison of thousands of data points derived from HLA binding data and elution of natural ligands might enable to discern specific characteristics of HLA class II natural ligands [56, 81]. Relating to the propensity of particular sequences, to be more visible for TCR recognition, we have used a neural network approach to identify residues that are more often recognized by class I restricted T cells [79]. Earlier studies indicate that this is also potentially true for HLA class II T cells [80], and we are currently examining whether an approach similar to what is described by Calis et al.[79] will apply.

In conclusion, hundreds of epitopes have been described in most allergen systems investigated so far, including for example timothy grass, ragweed, and cockroach allergens [11, 12, 51, 52]. In practical terms, the most dominant HDM epitopes identified herein (and the associated novel proteins) could be considered for peptide-based HDM-specific immunotherapy. Our data demonstrates that T cell recognition of HDM sequences is broader than previously thought. In particular we show that an additional 61 new proteins are recognized, and that this recognition is not necessarily associated with Ig E responses. This data demonstrates that helper T cell responses, which are known to be necessary for antibody isotype switching and maturation of antibody responses, including IgE are not necessarily targeting the same allergens recognized by HDM-specific IgE. This result in turn raises the possibility that these T cells might be used for immunotherapeutic interventions not including the major allergens, with potential safety advantages. From the point of view of allergen testing, these data allow to speculate that the proteins and associated epitopes defined herein might be useful to develop diagnostic tests measuring T cell reactivity against HDM antigens.

In summary, the data presented herein suggest that there are two classes of HDM proteins targeted by T-cells from allergic patients. HDM proteins that are strong targets of IgE reactivity have been studied preferentially; for these proteins, IgE reactivity correlates with T cell reactivity (albeit with different hierarchy of immunodominance) and hydrolase activity. By contrast, among novel HDM proteins, which are less frequently targeted by IgE, potent T cell antigens are associated with only marginal IgE and IgG reactivity, and no significant associations with protein function were found.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding was provided in part by ALK-Abello A/S (Horsholm, Denmark) and with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, under grant number U19 AI100275.

References

1. Soyer OU, Akdis M, Ring J, Behrendt H, Crameri R, Lauener R, Akdis CA. Mechanisms of peripheral tolerance to allergens. Allergy. 2013; 68:161–70. [PubMed: 23253293]
2. Stewart GA, Ward LD, Simpson RJ, Thompson PJ. The group III allergen from the house dust mite Dermatophagoides pteronyssinus is a trypsin-like enzyme. Immunology. 1992; 75:29–35. [PubMed: 1537598]
3. Pernas M, Sanchez-Ramos I, Sanchez-Monge R, Lombardero M, Arteaga C, Castanera P, Salcedo G. Der p 1 and Der f 1, the highly related and major allergens from house dust mites, are differentially affected by a plant cystatin. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2000; 30:972–8. [PubMed: 10848919]
4. Deb R, Shakib F, Reid K, Clark H. Major house dust mite allergens Dermatophagoides pteronyssinus 1 and Dermatophagoides farinae 1 degrade and inactivate lung surfactant proteins A and D. The Journal of biological chemistry. 2007; 282:36808–19. [PubMed: 17848554]
5. Machado DC, Horton D, Harrop R, Peachell PT, Helm BA. Potential allergens stimulate the release of mediators of the allergic response from cells of mast cell lineage in the absence of sensitization with antigen-specific IgE. European journal of immunology. 1996; 26:2972–80. [PubMed: 8977293]
6. Nathan AT, Peterson EA, Chakir J, Wills-Karp M. Innate immune responses of airway epithelium to house dust mite are mediated through beta-glucan-dependent pathways. The Journal of allergy and clinical immunology. 2009; 123:612–8. [PubMed: 19178937]
7. Jacquet A. The role of innate immunity activation in house dust mite allergy. Trends in molecular medicine. 2011; 17:604–11. [PubMed: 21741880]
8. Bresciani A, Paul S, Schommer N, Dillon MB, Bancroft T, Greenbaum J, Sette A, Nielsen M, Peters B. T cell recognition is shaped by epitope sequence conservation in the host proteome and microbiome. Immunology. 2016
9. Schulten V, Peters B, Sette A. New strategies for allergen T cell epitope identification: going beyond IgE. International archives of allergy and immunology. 2014; 165:75–82. [PubMed: 25402674]
10. Hinz D, Oseroff C, Pham J, Sidney J, Peters B, Sette A. Definition of a pool of epitopes that recapitulates the T cell reactivity against major house dust mite allergens. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2015; 45:1601–12. [PubMed: 25652035]
11. Oseroff C, Sidney J, Kotturi MF, Kolla R, Alam R, Broide DH, Wasserman SI, Weiskopf D, McKinney DM, Chung JL, Petersen A, Grey H, Peters B, Sette A. Molecular determinants of T cell epitope recognition to the common Timothy grass allergen. Journal of immunology (Baltimore, Md : 1950). 2010; 185:943–55.

12. Oseroff C, Sidney J, Tripple V, Grey H, Wood R, Broide DH, Greenbaum J, Kolla R, Peters B, Pomes A, Sette A. Analysis of T cell responses to the major allergens from German cockroach: epitope specificity and relationship to IgE production. Journal of immunology (Baltimore, Md : 1950). 2012; 189:679–88.
13. Voorhorst R. House dust mite and house dust allergy. Annals of allergy. 1977; 38:71.
14. Platts-Mills TA, Thomas WR, Aalberse RC, Vervloet D, Champman MD. Dust mite allergens and asthma: report of a second international workshop. The Journal of allergy and clinical immunology. 1992; 89:1046–60. [PubMed: 1349902]
15. Platts-Mills TA, Vervloet D, Thomas WR, Aalberse RC, Chapman MD. Indoor allergens and asthma: report of the Third International Workshop. The Journal of allergy and clinical immunology. 1997; 100:S2–24. [PubMed: 9438476]
16. Salo PM, Arbes SJ Jr. Crockett PW, Thorne PS, Cohn RD, Zeldin DC. Exposure to multiple indoor allergens in US homes and its relationship to asthma. The Journal of allergy and clinical immunology. 2008; 121:678–84. e2. [PubMed: 18255132]
17. Chapman MD, Pomes A, Breiteneder H, Ferreira F. Nomenclature and structural biology of allergens. The Journal of allergy and clinical immunology. 2007; 119:414–20. [PubMed: 17166572]
18. Pomes A. Allergen structures and biologic functions: the cutting edge of allergy research. Current allergy and asthma reports. 2008; 8:425–32. [PubMed: 18682111]
19. Tovey ER, Chapman MD, Platts-Mills TA. Mite faeces are a major source of house dust allergens. Nature. 1981; 289:592–3. [PubMed: 7464922]
20. Thomas WR, Smith WA, Hales BJ. The allergenic specificities of the house dust mite. Chang Gung medical journal. 2004; 27:563–9. [PubMed: 15553602]
21. Weghofer M, Grote M, Resch Y, Casset A, Kneidinger M, Kopec J, Thomas WR, Fernandez-Caldas E, Kabesch M, Ferrara R, Mari A, Purohit A, Pauli G, Horak F, Keller W, Valent P, Valenta R, Vrtala S. Identification of Der p 23, a peritrophin-like protein, as a new major Dermatophagoides pteronyssinus allergen associated with the peritrophic matrix of mite fecal pellets. Journal of immunology (Baltimore, Md : 1950). 2013; 190:3059–67.
22. Mueller GA, Randall TA, Glesner J, Pedersen LC, Perera L, Edwards LL, DeRose EF, Chapman MD, London RE, Pomes A. Serological, genomic and structural analyses of the major mite allergen Der p 23. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2016; 46:365–76. [PubMed: 26602749]
23. Kendall M, Mitchell TJ, Costigan G, Armitage M, Lenzo JC, Thomas JA, von Garnier C, Zosky GR, Turner DJ, Stumbles PA, Sly PD, Holt PG, Thomas WR. Downregulation of IgE antibody and allergic responses in the lung by epidermal biolistic microparticle delivery. The Journal of allergy and clinical immunology. 2006; 117:275–82. [PubMed: 16461127]
24. Kidon MI, Chiang WC, Liew WK, Ong TC, Tiong YS, Wong KN, Angus AC, Ong ST, Gao YF, Reginald K, Bi XZ, Shang HS, Chew FT. Mite component-specific IgE repertoire and phenotypes of allergic disease in childhood: the tropical perspective. Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology. 2011; 22:202–10. [PubMed: 21332797]
25. Yasueda H, Mita H, Akiyama K, Shida T, Ando T, Sugiyama S, Yamakawa H. Allergens from Dermatophagoides mites with chymotryptic activity. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 1993; 23:384–90. [PubMed: 8334537]
26. Yssel H, Johnson KE, Schneider PV, Wideman J, Terr A, Kastelein R, De Vries JE. T cell activation-inducing epitopes of the house dust mite allergen Der p I. Proliferation and lymphokine production patterns by Der p I-specific CD4+ T cell clones. Journal of immunology (Baltimore, Md : 1950). 1992; 148:738–45.
27. Okano M, Nagano T, Nakada M, Masuda Y, Kino K, Yasueda H, Nose Y, Nishimura Y, Ohta N. Epitope analysis of HLA-DR-restricted helper T-cell responses to Der p II, a major allergen molecule of Dermatophagoides pteronyssinus. Allergy. 1996; 51:29–35. [PubMed: 8721525]
28. Verhoef A, Higgins JA, Thorpe CJ, Marsh SG, Hayball JD, Lamb JR, O'Hehir RE. Clonal analysis of the atopic immune response to the group 2 allergen of Dermatophagoides spp.: identification of

HLA-DR and -DQ restricted T cell epitopes. International immunology. 1993; 5:1589–97. [PubMed: 7508743]

29. Joost van Neerven R, van t'Hof W, Ringrose JH, Jansen HM, Aalberse RC, Wierenga EA, Kapsenberg ML. T cell epitopes of house dust mite major allergen Der p II. Journal of immunology (Baltimore, Md : 1950). 1993; 151:2326–35.
30. Bonvalet M, Wambre E, Moussu H, Horiot S, Kwok WW, Louise A, Ebo D, Hoarau C, Van Overtvelt L, Baron-Bodo V, Moingeon P. Comparison between major histocompatibility complex class II tetramer staining and surface expression of activation markers for the detection of allergen-specific CD4(+) T cells. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2011; 41:821–9. [PubMed: 21418343]
31. Wambre E, Bonvalet M, Bodo VB, Maillere B, Leclert G, Moussu H, Von Hofe E, Louise A, Balazuc AM, Ebo D, Hoarau C, Garcia G, Van Overtvelt L, Moingeon P. Distinct characteristics of seasonal (Bet v 1) vs. perennial (Der p 1/Der p 2) allergen-specific CD4(+) T cell responses. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2011; 41:192–203. [PubMed: 21105918]
32. Crack LR, Chan HW, McPherson T, Ogg GS. Identification of an immunodominant region of the major house dust mite allergen Der p 2 presented by common human leucocyte antigen alleles. Clinical and experimental dermatology. 2012; 37:266–76. [PubMed: 22188309]
33. Tsitoura DC, Verhoef A, Gelder CM, O'Hehir RE, Lamb JR. Altered T cell ligands derived from a major house dust mite allergen enhance IFN-gamma but not IL-4 production by human CD4+ T cells. Journal of immunology (Baltimore, Md : 1950). 1996; 157:2160–5.
34. Kawamoto S, Ohno K, Tategaki A, Aki T, Shigeta S, Jyo T, Suzuki O, Ono K. T-cell epitope analysis of Mag 3, an important allergen from the house dust mite, Dermatophagoides farinae. Immunology letters. 2000; 72:53–60. [PubMed: 10789682]
35. Takai T, Mori A, Yuuki T, Okudaira H, Okumura Y. Non-anaphylactic combination of partially deleted fragments of the major house dust mite allergen Der f 2 for allergen-specific immunotherapy. Molecular immunology. 1999; 36:1055–65. [PubMed: 10698308]
36. Fujii S, Ono K, Takeuchi A, Aki T, Shigeta S, Suzuki O, Jyo T, Yamashita U. Identification of T-cell epitope sequences on an important mite antigen. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 1997; 27:1086–94. [PubMed: 9678841]
37. Inoue R, Matsuoka T, Kondo N, Nishimura Y, Matsushita S. Identification of Dermatophagoides farinae-2-derived peptides and class II HLA molecules recognized by T cells from atopic individuals. International archives of allergy and immunology. 1997; 114:354–60. [PubMed: 9414139]
38. Hales BJ, Thomas WR. T-cell sensitization to epitopes from the house dust mites Dermatophagoides pteronyssinus and Euroglyphus maynei. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 1997; 27:868–75. [PubMed: 9291282]
39. Matsuoka T, Kohrogi H, Ando M, Nishimura Y, Matsushita S. Dermatophagoides farinae-1-derived peptides and HLA molecules recognized by T cells from atopic individuals. International archives of allergy and immunology. 1997; 112:365–70. [PubMed: 9104792]
40. O'Hehir RE, Verhoef A, Panagiotopoulou E, Keswani S, Hayball JD, Thomas WR, Lamb JR. Analysis of human T cell responses to the group II allergen of Dermatophagoides species: localization of major antigenic sites. The Journal of allergy and clinical immunology. 1993; 92:105–13. [PubMed: 8335846]
41. Smith WA, Hales BJ, Jarnicki AG, Thomas WR. Allergens of wild house dust mites: environmental Der p 1 and Der p 2 sequence polymorphisms. The Journal of allergy and clinical immunology. 2001; 107:985–92. [PubMed: 11398075]
42. van Neerven RJ, van de Pol MM, Wierenga EA, Aalberse RC, Jansen HM, Kapsenberg ML. Peptide specificity and HLA restriction do not dictate lymphokine production by allergen-specific T-lymphocyte clones. Immunology. 1994; 82:351–6. [PubMed: 7525459]
43. Okano M, Nagano T, Kino K, Yasueda H, Baba Y, Saito C, Masuda Y, Ohta N. Population analysis of cellular responses to synthetic peptides of Der p II, a major allergen molecule of Dermatophagoides pteronyssinus, in allergic and nonallergic subjects. Allergy. 1994; 49:436–41. [PubMed: 7521142]

44. Higgins JA, Thorpe CJ, Hayball JD, O'Hehir RE, Lamb JR. Overlapping T-cell epitopes in the group I allergen of Dermatophagoides species restricted by HLA-DP and HLA-DR class II molecules. The Journal of allergy and clinical immunology. 1994; 93:891–9. [PubMed: 7514196]
45. O'Brien RM, Thomas WR, Nicholson I, Lamb JR, Tait BD. An immunogenetic analysis of the T-cell recognition of the major house dust mite allergen Der p 2: identification of high- and low-responder HLA-DQ alleles and localization of T-cell epitopes. Immunology. 1995; 86:176–82. [PubMed: 7490115]
46. O'Brien RM, Thomas WR, Tait BD. An immunogenetic analysis of T-cell reactive regions on the major allergen from the house dust mite, Der p I, with recombinant truncated fragments. The Journal of allergy and clinical immunology. 1994; 93:628–34. [PubMed: 7512103]
47. Higgins JA, Lamb JR, Marsh SG, Tonks S, Hayball JD, Rosen-Bronson S, Bodmer JG, O'Hehir RE. Peptide-induced nonresponsiveness of HLA-DP restricted human T cells reactive with Dermatophagoides spp. (house dust mite). The Journal of allergy and clinical immunology. 1992; 90:749–56. [PubMed: 1385507]
48. Roncarolo MG, Bigler M, Haanen JB, Yssel H, Bacchetta R, de Vries JE, Spits H. Natural killer cell clones can efficiently process and present protein antigens. Journal of immunology (Baltimore, Md : 1950). 1991; 147:781–7.
49. Chen KW, Focke-Tejkl M, Blatt K, Kneidinger M, Gieras A, Dall'Antonia F, Fae I, Fischer G, Keller W, Valent P, Valenta R, Vrtala S. Carrier-bound nonallergenic Der p 2 peptides induce IgG antibodies blocking allergen-induced basophil activation in allergic patients. Allergy. 2012; 67:609–21. [PubMed: 22339348]
50. Banerjee S, Weber M, Blatt K, Swoboda I, Focke-Tejkl M, Valent P, Valenta R, Vrtala S. Conversion of Der p 23, a new major house dust mite allergen, into a hypoallergenic vaccine. Journal of immunology (Baltimore, Md : 1950). 2014; 192:4867–75.
51. Schulten V, Greenbaum JA, Hauser M, McKinney DM, Sidney J, Kolla R, Lindestam Arlehamn CS, Oseroff C, Alam R, Broide DH, Ferreira F, Grey HM, Sette A, Peters B. Previously undescribed grass pollen antigens are the major inducers of T helper 2 cytokine-producing T cells in allergic individuals. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:3459–64. [PubMed: 23401558]
52. Dillon MB, Schulten V, Oseroff C, Paul S, Dullanty LM, Frazier A, Belles X, Piulachs MD, Visness C, Bacharier L, Bloomberg GR, Busse P, Sidney J, Peters B, Sette A. Different Bla-g T cell antigens dominate responses in asthma versus rhinitis subjects. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2015; 45:1856–67. [PubMed: 26414909]
53. Carrasco Pro S, Sidney J, Paul S, Lindestam Arlehamn C, Weiskopf D, Peters B, Sette A. Automatic Generation of Validated Specific Epitope Sets. Journal of immunology research. 2015; 2015:763461. [PubMed: 26568965]
54. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology. 2011; 29:644–52.
55. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols. 2013; 8:1494–512. [PubMed: 23845962]
56. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B. The immune epitope database (IEDB) 3.0. Nucleic acids research. 2015; 43:D405–12. [PubMed: 25300482]
57. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution. 2011; 28:2731–9. [PubMed: 21546353]
58. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson

JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000; 25:25–9. [PubMed: 10802651]

59. Gene Ontology Consortium: going forward. Nucleic acids research. 2015; 43:D1049–56. [PubMed: 25428369]
60. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics (Oxford, England). 2005; 21:3674–6.
61. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S. AmiGO: online access to ontology and annotation data. Bioinformatics (Oxford, England); 2009; 25:288–9.
62. Paul S, Lindestam Arlehamn CS, Scriba TJ, Dillon MB, Oseroff C, Hinz D, McKinney DM, Carrasco Pro S, Sidney J, Peters B, Sette A. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. Journal of immunological methods. 2015; 422:28–34. [PubMed: 25862607]
63. Lindestam Arlehamn CS, McKinney DM, Carpenter C, Paul S, Rozot V, Makgotlho E, Gregg Y, van Rooyen M, Ernst JD, Hatherill M, Hanekom WA, Peters B, Scriba TJ, Sette A. A Quantitative Analysis of Complexity of Human Pathogen-Specific CD4 T Cell Responses in Healthy M. tuberculosis Infected South Africans. PLoS pathogens. 2016; 12:e1005760. [PubMed: 27409590]
64. Weiskopf D, Angelo MA, Grifoni A, O'Rourke PH, Sidney J, Paul S, De Silva AD, Phillips E, Mallal S, Premawansa S, Premawansa G, Wijewickrama A, Peters B, Sette A. HLA DRB1 alleles are associated with different response magnitudes of dengue virus specific CD4+ T cell responses. The Journal of infectious diseases. 2016
65. Moutaftsi M, Peters B, Pasquetto V, Tscharke DC, Sidney J, Bui HH, Grey H, Sette A. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. Nature biotechnology. 2006; 24:817–9.
66. Oseroff C, Peters B, Pasquetto V, Moutaftsi M, Sidney J, Panchanathan V, Tscharke DC, Maillere B, Grey H, Sette A. Dissociation between epitope hierarchy and immunoprevalence in CD8 responses to vaccinia virus western reserve. Journal of immunology (Baltimore, Md : 1950). 2008; 180:7193–202.
67. Arlehamn CS, Sidney J, Henderson R, Greenbaum JA, James EA, Moutaftsi M, Coler R, McKinney DM, Park D, Taplitz R, Kwok WW, Grey H, Peters B, Sette A. Dissecting mechanisms of immunodominance to the common tuberculosis antigens ESAT-6, CFP10, Rv2031c (hspX), Rv2654c (TB7.7), and Rv1038c (EsxJ). Journal of immunology (Baltimore, Md : 1950). 2012; 188:5020–31.
68. Weiskopf D, Yauch LE, Angelo MA, John DV, Greenbaum JA, Sidney J, Kolla RV, De Silva AD, de Silva AM, Grey H, Peters B, Shresta S, Sette A. Insights into HLA-restricted T cell responses in a novel mouse model of dengue virus infection point toward new implications for vaccine design. Journal of immunology (Baltimore, Md : 1950). 2011; 187:4268–79.
69. Goodman R, van Ree R, Vieths S, Ferreira F, Ebisawa M, Sampson HA, Baumert J, Wise J, Taylor SL. Criteria used to categorise proteins as allergens for inclusion in allergenonline.org: a curated database for risk assessment. Clinical and Translational Allergy. 2014; 4:1–2. [PubMed: 24393277]
70. Westernberg L, Schulten V, Greenbaum JA, Natali S, Tripple V, McKinney DM, Frazier A, Hofer H, Wallner M, Sallusto F, Sette A, Peters B. T-cell epitope conservation across allergen species is a major determinant of immunogenicity. The Journal of allergy and clinical immunology. 2016
71. Ladics GS, Fry J, Goodman R, Herouet-Guicheney C, Hoffmann-Sommergruber K, Madsen CB, Penninks A, Pomes A, Roggen EL, Smit J, Wal JM. Allergic sensitization: screening methods. Clin Transl Allergy. 2014; 4:13. [PubMed: 24739743]
72. Calderon MA, Linneberg A, Kleine-Tebbe J, De Blay F. Hernandez Fernandez de Rojas D, Virchow JC, Demoly P, Respiratory allergy caused by house dust mites: What do we really know? The Journal of allergy and clinical immunology. 2015; 136:38–48. [PubMed: 25457152]
73. Gough L, Schulz O, Sewell HF, Shakib F. The cysteine protease activity of the major dust mite allergen Der p 1 selectively enhances the immunoglobulin E antibody response. The Journal of experimental medicine. 1999; 190:1897–902. [PubMed: 10601364]

74. Thomas WR, Hales BJ, Smith WA. House dust mite allergens in asthma and allergy. Trends in molecular medicine. 2010; 16:321–8. [PubMed: 20605742]

75. Ando T, Ino Y, Haida M, Honma R, Maeda H, Yamakawa H, Iwaki M, Okudaira H. Isolation of cysteine protease in the crude mite extract, Dermatophagoides farinae. International archives of allergy and applied immunology. 1991; 96:199–205. [PubMed: 1804791]

76. Stewart GA, Lake FR, Thompson PJ. Faecally derived hydrolytic enzymes from Dermatophagoides pteronyssinus: physicochemical characterisation of potential allergens. International archives of allergy and applied immunology. 1991; 95:248–56. [PubMed: 1718911]

77. Zhan X, Li C, Guo W, Jiang Y. Prokaryotic expression and bioactivity evaluation of the chimeric gene derived from the group 1 allergens of dust mites. Nutricion hospitalaria. 2015; 32:2771–6. [PubMed: 26667733]

78. Mutschlechner S, Egger M, Briza P, Wallner M, Lackner P, Karle A, Vogt AB, Fischer GF, Bohle B, Ferreira F. Naturally processed T cell-activating peptides of the major birch pollen allergen. The Journal of allergy and clinical immunology. 2010; 125:711–8. 18, e1–18, e2. [PubMed: 20132976]

79. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Kesmir C, Peters B. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS computational biology. 2013; 9:e1003266. [PubMed: 24204222]

80. Alexander J, Sidney J, Southwood S, Ruppert J, Oseroff C, Maewal A, Snoke K, Serra HM, Kubo RT, Sette A, et al. Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. Immunity. 1994; 1:751–61. [PubMed: 7895164]

81. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M, Peters B. Immune epitope database analysis resource. Nucleic acids research. 2012; 40:W525–30. [PubMed: 22610854]

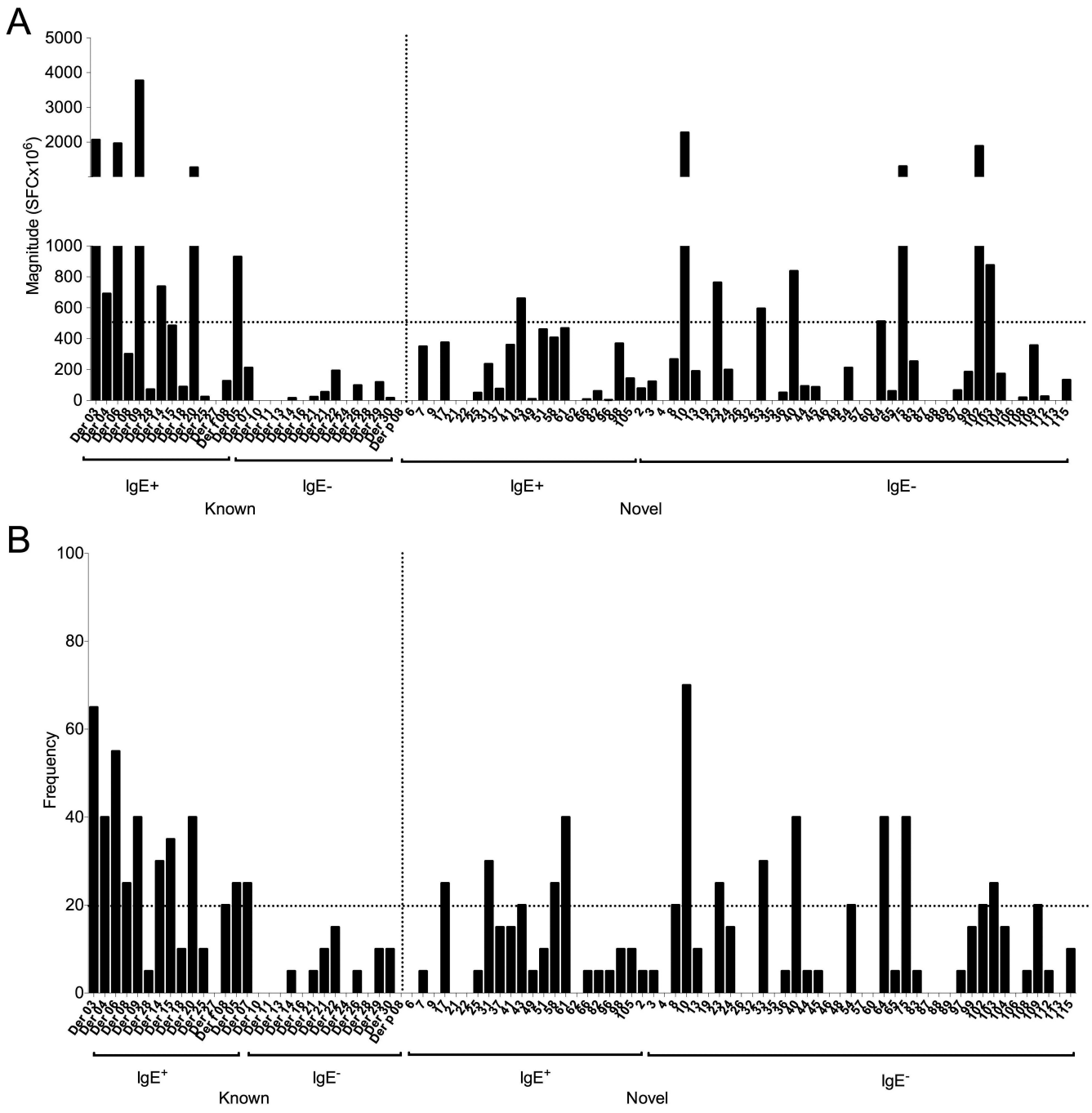

Figure 1. Overall reactivity (A) and frequency (B) of donor recognition of sum of re activity of individual peptides derived from each HDM protein

T cell reactivity was measured by ELISPOT in PBMCs from 20 HDM-allergic subjects in response to pool restimulation after 14-17 day stimulation with Der p/f extract.

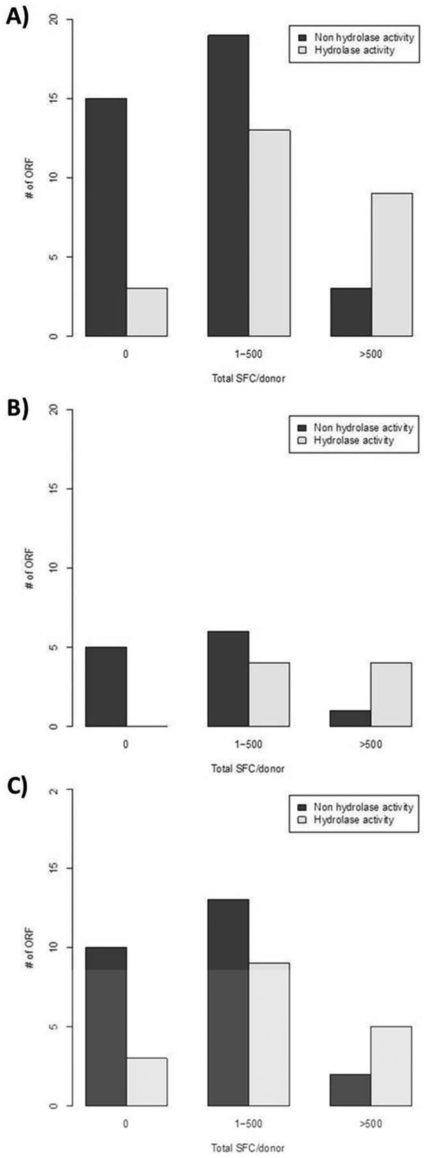

Figure 2. Correlation of T cell reactivity with hydrolase functionality

A, all antigens; **B,** known allergens; **C,** novel antigens. The number of proteins/ORFs in each category of reactivity (in terms of average SFC per donor), divided according to hydrolase or non-hydrolase function. P values: A, 0.001; B, 0.04; C, 0.11.

Table 1

HDM proteins for which T cell responses towards the derived peptides were > 500 total SFC/donor

Cluster	Protein	IgG	IgE	SFC IFNγ/donor	Frequency IFNγ/donor	SFC IL-5/donor	Frequency IL-5/donor	Ratio IFNγ/IL-5	Total freq	Total SFC/donor
72	Der 09	+	+	33	10%	3739	40%	0.0088	40%	3772
42	Der 03	+	+	31	20%	2036	65%	0.0151	65%	2067
74	Der 06	+	+	78	10%	1885	55%	0.0413	55%	1963
29	Der 20	+	+	113	25%	1160	40%	0.0970	40%	1273
77	Der 05	-	-	12	5%	918	25%	0.0131	25%	930
20	Der 14	+	+	180	20%	558	30%	0.3222	30%	738
16	Der 04	+	+	39	10%	652	40%	0.0591	40%	691
10	Novel	+	-	40	20%	2236	70%	0.0179	70%	2276
102	Novel	+	-	16	5%	1871	20%	0.0086	20%	1887
75	Novel	-	-	48	10%	1256	40%	0.0384	40%	1305
103	Novel	+	-	10	5%	867	25%	0.0110	25%	876
40	Novel	-	-	26	15%	811	40%	0.0326	40%	838
23	Novel	-	-	0	0%	763	25%	0.00001	25%	763
43	Novel	+	+	7	5%	653	20%	0.0107	20%	660
33	Novel	-	-	32	5%	564	30%	0.0562	30%	595
64	Novel	-	-	18	10%	495	40%	0.0360	40%	512

Table 2a

Concordance between T cell reactivity and IgE reactivity for known and novel HDM proteins

Magnitude	Known			Novel		
	IgE$^+$	IgE$^-$	%	IgE$^+$	IgE$^-$	%
>500	6	1	86	1	8	11
1-500	6	8	43	15	18	46
0	1	7	13	5	14	26

Table 2b

Correlation between T cell reactivity and IgG reactivity for known and novel HDM proteins

Magnitude	Known			Novel		
	IgG^+	IgG^-	%	IgG^+	IgG^-	%
>500	6	1	86	4	5	44
1-500	9	5	64	17	16	52
0	1	7	13	6	14	30

Table 2c

Distribution of hydrolase activity as a function of IgE and IgG reactivity

	IgE^+		IgG^+	
	Known	**Novel**	**Known**	**Novel**
Hydrolase	7/8 (88%)	6/17 (35%)	8/8 (100%)	8/17 (47%)
Not hydrolase	4/12 (33%)	11/25 (44%)	6/12 (50%)	14/25 (56%)

Table 3

Top 106 epitopes in terms of T cell response magnitude

Cluster #	Protein	Start position	Sequence	% Donors responding	% of average SFC/donor
8	Novel	266	KDSQYLQHFNNAIKQ	15%	0.41%
8	Novel	271	LQHFNNAIKQLNTED	10%	0.05%
10	Novel	771	YDDIDYVFVRGGSII	30%	0.51%
10	Novel	596	ILSDWSSMRWTIPSI	30%	0.12%
10	Novel	471	NATEYWMDMFAEYHK	25%	0.78%
10	Novel	886	PRIIRFNYDEQTNIL	25%	0.55%
10	Novel	591	HWNGDILSDWSSMRW	25%	0.31%
10	Novel	146	DQVLRLKFIDANQKR	20%	0.77%
10	Novel	201	QSIFDINLAYMVYSD	20%	0.24%
10	Novel	676	SSLQYRYRFLAHLYT	20%	0.13%
10	Novel	865	INILGVPKLPTSFKL	20%	0.08%
10	Novel	671	IRAARSSLQYRYRFL	20%	0.07%
10	Novel	643	AFYSFVRNHNTDNAI	20%	0.05%
10	Novel	356	FTNLNTTYTRNRAVG	15%	0.32%
10	Novel	361	TTYTRNRAVGIPMDV	15%	0.27%
10	Novel	859	TQNINFINILGVPKL	15%	0.25%
10	Novel	881	GKPYYQFIYTTNNML	15%	0.15%
10	Novel	636	LCIRWYQLGAFYSFA	15%	0.14%
10	Novel	641	YQLGAFYSFARNHND	15%	0.12%
10	Novel	666	LGESVIRAARSSLQY	15%	0.10%
10	Novel	481	AEYHKTIAFDGAWLD	15%	0.08%
10	Novel	681	RYRFLAHLYTLFYHV	10%	0.55%
16	Der 4	131	AGVRIYVDIVLNHMT	25%	1.18%
17	Novel	266	TFLVYADFLSYKSGV	20%	0.46%
17	Novel	236	DRYFGQSAYHVHSNV	20%	0.20%
17	Novel	241	KSAYRVSSIVSQIQH	15%	0.20%
17	Novel	281	YKRHSVQALGGHAIK	15%	0.13%
23	Novel	76	YDIKYSYNVPAVLPN	25%	0.44%
23	Novel	36	DSEHLKLIISADVNG	20%	0.76%
23	Novel	86	AVLPNIKGTLTAKVI	20%	0.64%
29	Der 20	273	IEKKLPFSRDDRLGF	25%	0.85%
29	Der 20	119	VDPKNEYVISTRVRC	20%	0.60%
29	Der 20	124	EYVISTRVRCGRSLK	15%	0.68%
29	Der 20	258	DLKQVFSRLINGVNH	15%	0.25%
29	Der 20	253	MQKGGDLKQVFSRLI	15%	0.21%
31	Novel	46	LKPEFLKVNPFHKIP	20%	0.13%

Cluster #	Protein	Start position	Sequence	% Donors responding	% of average SFC/donor
31	Novel	56	FHKIPTFVDTDGFTI	15%	0.40%
31	Novel	166	DIAMYFSCNTMEIYS	15%	0.07%
33	Novel	286	KHPWIIVMGHRPLYC	20%	0.32%
33	Novel	351	EHFYARLFPIYKYKM	15%	0.50%
38	Der 22	58	NQLRISFVANENTGN	15%	0.60%
39	Der 8	76	VKITQSMAILRYLAR	15%	0.21%
40	Novel	386	IAVGFEVYKDFMTYR	20%	0.56%
40	Novel	36	GLWLFEESTPINDRT	20%	0.50%
42	Der 3	29	ILDEYWILTAAHCVN	50%	0.80%
42	Der 3	94	PMTLDQTNAKPVPLP	40%	0.81%
42	Der 3	44	GQSAKKLSIRYNTLK	25%	2.04%
42	Der 3	49	KLSIRYNTLKHASGG	25%	1.46%
42	Der 3	79	SYQIDNDIALIKLKS	25%	0.53%
42	Der 3	89	IKLKTPMTLDQTNAK	20%	0.47%
42	Der 3	14	YQISLQSSSHFCGGS	20%	0.07%
42	Der 3	64	EKISVAKIFAHEKYD	15%	0.27%
42	Der 3	84	NDVALIKLKTPMTLD	15%	0.25%
43	Novel	41	AHCVAGQTASKLSIR	15%	0.27%
43	Novel	46	GQSAKKLSIRYNTLK	15%	0.20%
53	Novel	261	YTMHYYLNNGATRDK	25%	0.61%
53	Der 15	91	SWEKRGYERFNNLRL	20%	0.10%
53	Der 15	256	YFNVNYTMHYYLNNG	15%	0.15%
53	Der 15	96	GYERFNNLRLKNPEL	15%	0.09%
53	Der 15	41	GTWSVYHKVDPYTIE	15%	0.08%
53	Der 15	271	ATRDKLVMGVPFYGR	15%	0.06%
61	Novel	106	EQYISGVILFDETVY	20%	1.13%
61	Novel	36	IIMAKFNYLPVDVQE	15%	0.13%
64	Novel	196	PPKPELLVIDTELGR	30%	0.79%
64	Novel	201	LLVIDTELGRLGMQI	20%	0.19%
64	Novel	211	LGMQICFDMIFKTPG	20%	0.10%
72	Der 9	103	STTIDYDVATLILSE	30%	0.62%
72	Der 9	135	YDVATLILSQPFTPS	25%	1.73%
72	Der 9	140	LILSQPFTPSANADI	25%	1.22%
72	Der 9	73	SYFKIRYNTLDRTNG	25%	0.75%
72	Der 9	93	SKIYRHSLYSSTTID	25%	0.24%
72	Der 9	173	SNIWGSVNAITNRML	20%	0.68%
72	Der 9	33	DAIYQIALLRKDSFT	20%	0.64%

Cluster #	Protein	Start position	Sequence	% Donors responding	% of average SFC/donor
72	Der 9	178	SVNAITNRMLCAHSK	20%	0.63%
72	Der 9	1	MKFMILFALIAIGTS	20%	0.45%
72	Der 9	48	CGGSLISSRTVLTAA	15%	3.94%
72	Der 9	185	ILQIASVTKMSRTKC	15%	0.69%
72	Der 9	65	IALFRKDSFTCGGSL	15%	0.43%
72	Der 9	255	PTIYSNVANLRNWII	15%	0.36%
72	Der 9	53	ISSRTVLTAAHCVFG	15%	0.18%
72	Der 9	125	HNLYSSSPIDYDVAT	15%	0.11%
74	Der 6	136	TIQNDISLLILSKPV	35%	1.27%
74	Der 6	66	SLLKDYLIMKSHMCG	35%	0.48%
74	Der 6	61	APFQISLLKDYLIMK	30%	1.23%
74	Der 6	235	GPLVSANRKLTGIVS	30%	1.21%
74	Der 6	131	SYDPDTIQNDISLLI	30%	0.66%
74	Der 6	71	YLIMKSHMCGGSLIS	30%	0.32%
74	Der 6	11	IVITVTVDARFPRSL	20%	0.17%
74	Der 6	256	PPGEYMSVFTRPKYY	15%	0.74%
74	Der 6	206	EKWGSINAIHPGMIC	15%	0.15%
74	Der 6	15	NIWLWSINNSHLKTT	15%	0.09%
75	Novel	69	SVYQGTHKVLARVAS	30%	0.79%
75	Novel	19	YCNGAAIVSAARSQI	30%	0.70%
75	Novel	24	AIVSAARSQIGVPYS	25%	1.26%
75	Novel	102	GDLVFFGNPIHHVGI	15%	0.83%
75	Novel	1	TIMKFFFTLALFCTL	15%	0.08%
77	Der 5	26	YQNEFDFLLMQRIHE	15%	0.76%
77	Der 5	101	FERYNVEIALKSNEI	15%	0.55%
99	Novel	91	VSEFRRMNGLIASKG	15%	0.25%
102	Novel	186	YVPWTTVESRTVDVN	20%	0.18%
102	Novel	166	TGQKGFELIKVMARK	15%	2.09%
102	Novel	81	EYVDIDLVPFGNAHI	15%	1.44%
102	Novel	171	FELIKVMARKTPRHN	15%	1.09%
103	Novel	106	GMIRRMVNLRTYNPH	20%	0.79%
103	Novel	101	LQWGHGMIRRMVNLR	20%	0.78%
103	Novel	111	MVNLRTYNPHLTTMI	20%	0.47%