

In silico prediction of Ara h 2 T cell epitopes in peanut-allergic children

M. Pascal^{1,2*}, G. N. Konstantinou^{1,3*}, M. Masilamani^{1,4}, J. Lieberman¹ and H. A. Sampson^{1,4}

¹Division of Allergy & Immunology and The Jaffe Food Allergy Research Institute, Mount Sinai School of Medicine, New York, NY, USA, ²Department of Immunology, Allergy Unit, Centre de Diagnòstic Biomèdic (CDB), Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain, ³Department of Allergy, 424 General Military Training Hospital, Thessaloniki, Greece and ⁴Immunology Institute, Mount Sinai School of Medicine, New York, NY, USA

Clinical & Experimental Allergy

Summary

Background Despite the frequency and severity of peanut allergy, the only approved treatment is strict avoidance. Different types of immunotherapy with crude peanut extract are not universally effective and have been associated with relatively high adverse reaction rates.

Objective We sought to determine whether *in silico* predictive algorithms were useful in identifying candidate peptides for an Ara h 2 peptide-based vaccine using peanut-allergic patients' peripheral blood mononuclear cells (PBMCs) *in vitro*. A human leucocyte antigen (HLA) distribution analysis was also performed.

Methods Major histocompatibility complex (MHC)-class II-binding peptides were predicted using NetMHCIIpan-2.0 and NetMHCII-2.2 algorithms. PBMCs from 80 peanut-allergic patients were stimulated with overlapping 20-mer Ara h 2 peptides. Cell supernatant cytokine profiles were evaluated by multiplex assays. HLA-DRB1* and HLA-DQB1* typing were performed.

Results Four regions of overlapping sequences induced PBMC proliferation and predominant Th2 cytokine production. HLA genotyping showed 30 different DRB1* allele specificities and eight DQ serological specificities. The *in silico* analysis revealed similar relevant regions and predicted identical or similar core 9-mer epitopes to those identified *in vitro*. If relevant peptides, as determined by either *in vitro* or *in silico* analysis (15 peptides and 9 core epitopes respectively), were used in a peptide-based vaccine, they would cover virtually all subjects in the cohort studied.

Conclusions and Clinical Relevance Four dominant regions in Ara h 2 have been identified, containing sequences that could serve as potential candidates for peptide-based immunotherapy. MHC-class II-based T cell epitope prediction algorithms for HLA-DR and -DQ loci accurately predicted Ara h 2 T cell epitopes in peanut-allergic subjects, suggesting their potential utility in a peptide-based vaccine design for food allergy.

Keywords Ara h 2, MHC class II, NetMHCII-2.2, NetMHCIIpan-2.0, PBMCs

Submitted 14 December 2011; **revised** 18 August 2012; **accepted** 24 August 2012

Correspondence:

Hugh A. Sampson, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1198, New York, NY 10029-6574, USA.

E-mail: hugh.sampson@mssm.edu

Cite this as: M. Pascal, G. N.

Konstantinou, M. Masilamani, J.

Lieberman and H. A. Sampson, *Clinical*

& Experimental Allergy, 2013 (43)

116–127.

Introduction

Peanut allergy is a common and potentially life-threatening condition that considerably compromises the quality of life of many patients. Clinical symptoms may range from mild oropharyngeal pruritus to life-threatening anaphylaxis, although symptoms are often more severe, explaining why peanut allergy is the leading cause of emergency-department visits for food-induced

anaphylaxis in the United States [1]. Peanut allergy is generally life long with only approximately 20% of young peanut-allergic children outgrowing their allergy by school age [2, 3]. Most patients are sensitized to at least one of the three major peanut allergens, Ara h 1, Ara h 2 and Ara h 3. IgE antibodies to Ara h 2 have been shown to correlate most closely with clinical reactivity [4] and Ara h 2-specific IgE is found in more than 90% of peanut-allergic patients in North America [5]. Both Ara h 2 and its homologue, Ara h 6, have been described to be more potent inducers of basophil degranulation *in vitro* than Ara h 1 and Ara h 3 [6, 7].

*Authors contributed equally.

Avoidance of peanut remains the mainstay of treatment for peanut allergy, although it does not seem to change its natural history [3]. Therefore, major efforts have focused on the development of an effective immunotherapeutic approach to modify this IgE-mediated disorder. Unfortunately, conventional subcutaneous immunotherapy for peanut allergy using crude peanut extract has been shown to have an unacceptably high risk of anaphylaxis [8, 9]. One alternative approach to traditional immunotherapy is the use of peptide fragments corresponding to T cell epitopes of the allergen of interest [10]. These peptides retain immunogenicity, but are of insufficient length to cross-link allergen-specific IgE on mast cells and basophils [11]. Modulation of immune responses and reduction of early- and late-phase reactions following short courses of immunotherapy with Fel d 1 for cat-allergic subjects [12] and phospholipase A2 for bee venom-allergic patients [13] have been shown to desensitize affected patients. Reports on CD4⁺ T cell responses to individual peptides spanning whole peanut allergens are limited in the literature [14–17]. Recently, Prickett et al. [18] identified five dominant CD4⁺ T cell epitopes on Ara h 2 that collectively show diverse human leucocyte antigen (HLA) class II restriction, using Ara h 2-specific CD4⁺ T cell lines derived from 16 Caucasian peanut-allergic adults.

During the last decade, several groups have attempted to develop predictive algorithms for identifying T cell epitopes based on computer simulation, known as *in silico* analysis. As binding to major histocompatibility complex (MHC) class II is a prerequisite for an extracellular peptide to stimulate T lymphocyte responses, T cell epitopes may be predicted indirectly by the identification of MHC-binding molecules. In this context, many different methods have been described and used for predicting MHC–peptide binding [19]. In particular, MHC-class II-binding predictions have been widely used for epitope identification on infectious agents, with the intent of developing peptide-based vaccines using these bioinformatics approaches [20]. This pioneering methodology has not yet been widely applied to allergy.

This study focused on the identification of Ara h 2 peptides that can serve as potential candidates for a peptide-based vaccine. Peripheral blood mononuclear cells (PBMCs) from peanut-allergic children were stimulated with consecutive 20-mer peptides with a 15-amino acid overlap (offset of five), spanning the entire Ara h 2 protein and the HLA distribution of the peptide-specific immune responses was evaluated. *In silico* computational algorithms to predict T cell epitopes were utilized to evaluate not only the 20-mers tested *in vitro*, but also all possible theoretical Ara h 2 20-mer peptides (offset of one) to identify the relevant epitopes. These predicted peptides were evaluated and compared with current and previous experimental findings. Population

coverage analysis was performed to identify the most relevant Ara h 2 peptides to be included in a peptide-based vaccine for peanut allergy.

Methods

Subjects

All subjects were recruited from the outpatient clinic at the Jaffe Food Allergy Institute at the Mount Sinai School of Medicine. Subjects were eligible for enrolment if they had a convincing clinical history of an immediate-type reaction to peanut and a positive skin-prick test (SPT) and/or positive serum IgE to peanut; or if they had a peanut-specific serum IgE level > 15 kU_A/L [21]. Exclusion criteria included use of omalizumab within the past 12 months, enrolment in any food allergy interventional trial in the past 12 months and current treatment with immunotherapy (IT) for any allergen in the up-dosing phase. A total of 14 controls were included. The Mount Sinai Program for the Protection of Human Subjects approved the study (MSSM Study ID 08-1314) and informed consent was obtained from all patients.

Antigens

Ara h 2 peptides. Peptides spanning the entire length of Ara h 2.01 allergen (ACN62248.1) (peptides #1 to #30: 20-mers, 15-amino acids overlap; peptide #31: 7-mer; JPT Peptide Technologies, GmbH, Berlin, Germany) were reconstituted in the appropriate solvent [dimethyl sulfoxide (DMSO) for peptide #1, DMSO 50% in AIM-V[®] Medium (Invitrogen by Life Technologies, Grand Island, NY, USA) for peptide #2 or AIM-V[®] Medium alone for peptides #3 to #31] and plated in 96-U-well plates to reach a final concentration of 20 µg/mL in 200 µL of PBMCs for the proliferation assay. Plates were stored at –80°C until used.

Crude peanut extract (CPE). A peanut (*Arachis hypogaea*) extract was prepared as previously described [22]. Two different concentrations were utilized in the proliferation assay (50 µg/mL and 100 µg/mL).

Anti-CD3/CD28 beads (Dynabeads[®] human T-activator CD3/CD28; Invitrogen Inc). One microlitre of reagent was used per 200 µL of PBMCs as a positive control of proliferation.

T cell proliferation assay

T cell proliferation assays were performed with PBMCs from peanut-allergic individuals and controls. PBMCs were separated from 25 mL venous blood by density ficoll centrifugation, washed and resuspended in

AIM-V[®] Medium. A total of 1×10^6 PBMCs were cultured for 4 days in 96-U-well plates with CPE, anti-CD3/CD28 beads or Ara h 2 peptides. On day 4, 100- μ L aliquots of the culture supernatant were removed and stored at -80°C and 100 μ L of fresh AIM-V medium added to the cells. Cultures were then pulsed with tritiated thymidine (1 μ Ci) and cultured for an additional 48 h, harvested on glass fibre filters and incorporated radioactivity measured in counts per minute (cpm) by scintillation spectroscopy. Individual cpm for Ara h 2 peptides, CPE and anti-CD3/CD28 beads culture wells were compared with those containing medium alone (unstimulated cells) plated in triplicates. The within-subject variability of the cpm measured among these triplicates was calculated and used to estimate the per-individual coefficient of variation (CV). T cell responses were considered significant if the Stimulation Index (SI; cpm of antigen-stimulated T cells divided by cpm of unstimulated) exceeded the estimated maximum CV ($> 1 + \text{CV}_{\text{max}}$), otherwise they were regarded as negligible (zero). Peptides inducing positive proliferative responses ($\text{SI} \geq 1.5$) in at least 25% of the recruited subjects were considered significant. A heat map showing the magnitude of the proliferative responses represented by SI values was generated using TIGR Multiexperiment Viewer (www.tm4.org/mev).

Cytokine profile analysis

Concentrations of IL-5, IL-13, TNF- α and IFN- γ secreted into cell culture supernatants collected after 4 days of stimulation were evaluated using the cytometric bead array (CBA) kit (BD[™] CBA Human Soluble Protein Flex Set Assay; BD Biosciences, San Jose, CA, USA) following the manufacturer's protocol. The fold change of cytokine levels measured from supernatants of stimulated cells compared to unstimulated cells was considered significant if > 1 and negligible if ≤ 1 .

HLA-class II genotyping

Genomic DNA was isolated from venous blood (QIAamp DNA Mini Kit; Qiagen, Valencia, CA, USA) and used to perform high-resolution HLA-DRB1* typing (by Luminex technology – Biosynthesis Inc., Lewisville, TX, USA) and single specific primer-PCR (SSP-PCR)-based low-resolution HLA-DQB1* typing (FASTYPE Class II DQB1 Low Resolution SSP Typing Kit; Biosynthesis Inc). Allele frequency (number of copies of the allele) and phenotype frequency (number of individuals who carry the allele) were calculated in our cohort.

In silico T cell epitope prediction

T cell epitopes are principally predicted indirectly by identifying the binding of peptide fragments to the

MHC complexes. Development of accurate predictive algorithms for MHC-class II binding is complicated not only because the binding grooves of MHC-II molecules are open at both ends, allowing peptides with different lengths to bind, but also because the same MHC molecule can accommodate a great variety of different binding sequences. However, it has been reported that a sequence 9-amino acids in length (9-mer), known as a core peptide, is essential for MHC-class II binding [23] (further details are provided in the online repository).

For HLA-DR-based T cell epitope prediction, the artificial neural network-based alignment (NN-align) method NetMHCIIpan-2.0 was applied. This method is a hybrid of an earlier published pan-specific method for MHC-class II binding, NetMHCIIpan [24], and the NN-align method for allele-specific MHC-class II-binding predictions [25]. NetMHCIIpan-2.0 was selected as it has been shown to significantly outperform NetMHCIIpan [24], NN-align [25] and TEPITOPE [26] methods, which comprise the top performing methods in current use [27]. NetMHCIIpan-2.0 provides predictions only for some of the most commonly found HLA-DR alleles. For HLA-DQ alleles, NetMHCII-2.2 was used [25]. Although limited binding-affinity data are available for HLA-DQ, it was recently reported to provide the best performance in predicting this locus [28]. Ara h 2 was not included in 'training and evaluation' data sets for NetMHCIIpan-2.0 and NetMHCII-2.2 development. Both are available publicly online on web-based servers [25, 27].

For the prediction process, all possible 20-mer peptides spanning the full Ara h 2 sequence were tested separately for each HLA allele found in the current cohort. The peptides predicted to be significant binders by at least one allele examined, comprised the 'significant peptides'. The binding significance of each peptide to a given MHC molecule is based on the estimated strength of binding exhibited by a predicted nested core peptide at a set threshold level. Based on the estimated strength of binding, the peptides (and their predicted core 9-mers) are characterized as significant (weak or strong) or non-significant. Each predicted core 9-mer sequence represents the predicted putative T cell epitope within each peptide examined (for more information see the online repository).

Population coverage analysis

A given T cell epitope elicits a significant response only in subjects expressing MHC-class II molecules able to recognize and bind this particular epitope. MHC-class II molecules (HLA-DR, -DQ and -DP) have codominant expression, are highly polymorphic, especially the DRB1* locus, and coexist on the surface of antigen-presenting cells. Given the polymorphic nature of MHC

molecules, a single T cell epitope can bind more than one MHC molecule and hence the fraction of the population of interest expected to respond to a given epitope set is dependent on their HLA genotypes and the HLA genotype frequencies. This calculation can be done using the Population Coverage Calculation program when the HLA genotype frequencies and the existing MHC/peptide recognition patterns are known. This tool, hosted at the Immuno Epitope Database (IEDB) web site [29], has been used to calculate the current cohort coverage afforded by the peptides found to induce significant proliferation in PBMC cultures and by those found to be significant binders according to predictive algorithms.

Statistical analysis

Descriptive statistics of non-normally distributed, continuous variables are presented as medians and interquartile range and/or range. For comparisons of the fold-change in cytokine concentrations and proliferation levels, the Mann–Whitney test was used. Chi square (χ^2) and chi square goodness of fit tests were used to compare HLA allele distributions observed in each of the Ara h 2 regions inducing significant proliferation and to compare the HLA distribution between subjects and controls. All reported *P*-values are based on two-sided tests and compared with a significance level of 5%. Stata 9.1 for Windows (Stata Corp LP, College Station, TX, USA) was used for all statistical calculations and plots.

Results

Demographics

A total of 80 subjects (52 boys) were included in the study. The majority were Caucasians (90%). The median age at enrolment was 10.9 years. Most patients (92.5%) were atopic, with or without atopic eczema, asthma, allergic rhinitis and other food allergies (Table 1). Their median peanut-specific serum IgE level was 78.2 kU_A/L (< 0.35 – > 100 kU_A/L) with a median SPT weal size of 8 mm (3–18 mm). The majority of subjects (82.5%) had at least one past reaction to peanut. A total of 14 controls (seven males) were tested. Six controls reported having either respiratory allergies or food allergies other than peanut, and none were sensitized to peanut. Eight controls were non-atopic; most were Caucasians (79%) with a median age of 32.5 years (22–45 years).

Dominant CD4⁺ T cell epitopes of Ara h 2

T cell proliferation assays were performed with PBMCs from peanut-allergic individuals (*n* = 80) and controls (*n* = 14).

Table 1. Subject demographics and background history

Variable	Result
Male	<i>n</i> = 52 (65.0%)
Age	10.9 years [3.1–41]
Ethnicity	
Caucasian	<i>n</i> = 72 (90.0%)
Asian	<i>n</i> = 4 (5.0%)
Other	<i>n</i> = 4 (5.0%)
Atopic History	
Allergic Rhinitis	<i>n</i> = 53 (66.25%)
Asthma	<i>n</i> = 51 (63.75%)
Atopic eczema	<i>n</i> = 45 (56.25%)
Other food allergies	<i>n</i> = 63 (78.75%)
Peanut Allergy history	
History of reaction to peanut	<i>n</i> = 66 (82.5%)
Peanut-specific serum IgE level	78.2 kU _A /L [< 0.35 – > 100]
SPT weal size	8 mm [3–18]

Age, IgE level, and SPT weal size are reported in medians with ranges.

The within-subject CV in cpms from the unstimulated wells was 0.121 cpm (range 0.006–0.493 cpm). Consequently, a $SI \geq 1.5$ ($1 + CV_{\max} = 1 + 0.493$) was considered positive. Peptides inducing positive proliferative responses ($SI \geq 1.5$) in at least 25% (20 patients) of the study subjects included peptide # 2, 5, 6, 7, 8, 9, 10, 17, 18, 19, 20, 21, 22, 25 and 26. These peptides were further grouped into four regions within Ara h 2: region A (aa #6–25) included peptide #2; region B (aa #21–65), peptides 5–10; region C (aa #81–125), peptides 17–22; and region D (aa #121–145), peptides 25 and 26. The number of patients with a positive response ($SI \geq 1.5$) to each peptide expressed as percentage of all subjects recruited is shown in Fig. 1. Seven of the 80 (8.75%) subjects did not have a positive response to any peptide, although they all had a positive response to CPE and anti-CD3/CD28 beads. Proliferation was found to be significant ($SI \geq 1.5$) in more than 25% of controls following stimulation with peptides # 3 (5/14), 26 (4/14) and 27 (4/14). The magnitude of proliferative responses for all subjects and controls to all peptides tested is shown in Figure S1.

HLA-DRB1* and -DQB1* alleles: frequencies observed

Genomic DNA extraction and HLA-DRB1* and DQB1* loci genotyping was successfully achieved in 77 of the 80 recruited subjects and in all controls (Table S1). Among the 77 genotyped, 30 different HLA-DRB1* alleles and eight different DQB1* alleles were identified. Tables 2a and 2b show the allele and phenotype fre-

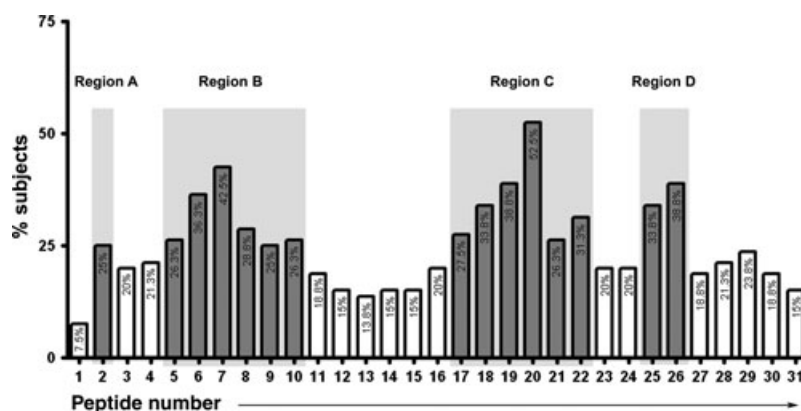


Fig. 1. Percentage of subjects with a significant positive response to each Ara h 2 peptide. Number of peanut-allergic subjects (in%, $n = 80$) with a significant positive response ($SI \geq 1.5$) for each Ara h 2 peptide. Proliferation in $\geq 25\%$ of the subjects is considered to be significant in terms of prevalence.

quencies' for all HLA specificities found in the allergic individuals with a positive proliferative response (i.e., stimulation index above 1.5) in each region (A–D). The

HLA-DR or -DQ allele specificity frequencies were similar among the four regions identified (P -values > 0.700). The HLA distribution frequency of the seven subjects

Table 2a. HLA DRB1* allele frequencies and phenotype frequencies

DRB1*	All Patients		Region A		Region B		Region C		Region D	
	Ph.F	A.F	Ph.F	A.F	Ph.F	A.F	Ph.F	A.F	Ph.F	A.F
Allele	($n = 77$)	($n = 154$)	($n = 18$)	($n = 36$)	($n = 45$)	($n = 90$)	($n = 57$)	($n = 114$)	($n = 35$)	($n = 70$)
*0101	5.2%	2.6%	11.1%	5.6%	2.2%	1.1%	3.5%	1.8%	5.7%	2.9%
*0102	7.8%	3.9%	5.6%	2.8%	8.9%	4.4%	8.8%	4.4%	5.7%	2.9%
*0301	13.0%	6.5%	27.8%	13.9%	6.7%	3.3%	12.3%	6.1%	11.4%	5.7%
*0401	6.5%	3.9%	0.0%	0.0%	6.7%	4.4%	3.5%	2.6%	0.0%	0.0%
*0402	14.3%	8.4%	11.1%	8.3%	11.1%	6.7%	17.5%	10.5%	14.3%	8.6%
*0403	5.2%	2.6%	5.6%	2.8%	6.7%	3.3%	5.3%	2.6%	2.9%	1.4%
*0404	2.6%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.9%	1.4%
*0405	2.6%	1.3%	0.0%	0.0%	2.2%	1.1%	3.5%	1.8%	2.9%	1.4%
*0406	1.3%	0.6%	5.6%	2.8%	0.0%	0.0%	1.8%	0.9%	0.0%	0.0%
*0407	2.6%	1.3%	5.6%	2.8%	2.2%	1.1%	3.5%	1.8%	2.9%	1.4%
*0701	19.5%	11.0%	5.6%	2.8%	22.2%	12.2%	24.6%	14.0%	22.9%	12.9%
*0801	2.6%	1.3%	5.6%	2.8%	4.4%	2.2%	0.0%	0.0%	0.0%	0.0%
*0804	1.3%	0.6%	0.0%	0.0%	0.0%	0.0%	1.8%	0.9%	0.0%	0.0%
*0901	1.3%	0.6%	0.0%	0.0%	2.2%	1.1%	1.8%	0.9%	2.9%	1.4%
*1001	1.3%	0.6%	0.0%	0.0%	2.2%	1.1%	1.8%	0.9%	2.9%	1.4%
*1101	7.8%	3.9%	5.6%	2.8%	11.1%	5.6%	8.8%	4.4%	8.6%	4.3%
*1102	1.3%	0.6%	0.0%	0.0%	2.2%	1.1%	0.0%	0.0%	0.0%	0.0%
*1104	13.0%	6.5%	11.1%	5.6%	8.9%	4.4%	15.8%	7.9%	22.9%	11.4%
*1201	6.5%	3.2%	0.0%	0.0%	8.9%	4.4%	5.3%	2.6%	5.7%	2.9%
*1202	2.6%	1.3%	5.6%	2.8%	4.4%	2.2%	3.5%	1.8%	5.7%	2.9%
*1301	9.1%	4.5%	5.6%	2.8%	13.3%	6.7%	7.0%	3.5%	8.6%	4.3%
*1302	24.7%	12.3%	38.9%	19.4%	24.4%	12.2%	22.8%	11.4%	22.9%	11.4%
*1303	1.3%	0.6%	0.0%	0.0%	2.2%	1.1%	0.0%	0.0%	0.0%	0.0%
*1305	2.6%	1.3%	5.6%	2.8%	4.4%	2.2%	3.5%	1.8%	5.7%	2.9%
*1401	7.8%	3.9%	5.6%	2.8%	8.9%	4.4%	8.8%	4.4%	5.7%	2.9%
*1404	1.3%	0.6%	0.0%	0.0%	2.2%	1.1%	1.8%	0.9%	2.9%	1.4%
*1501	16.9%	8.4%	11.1%	5.6%	13.3%	6.7%	14.0%	7.0%	11.4%	5.7%
*1502	6.5%	3.2%	5.6%	2.8%	4.4%	2.2%	5.3%	2.6%	8.6%	4.3%
*1601	2.6%	1.3%	5.6%	2.8%	2.2%	1.1%	1.8%	0.9%	2.9%	1.4%
*1602	2.6%	1.3%	11.1%	5.6%	4.4%	2.2%	3.5%	1.8%	5.7%	2.9%

A.F, Allele Frequency; Ph.F, Phenotype Frequency.

Table 2b. HLA DQB1* allele frequencies and phenotype frequencies

DQB1*Allele	All Patients		Region A		Region B		Region C		Region D	
	Ph.F (n = 77)	A.F (n = 154)	Ph.F (n = 18)	A.F (n = 36)	Ph.F (n = 45)	A.F (n = 90)	Ph.F (n = 57)	A.F (n = 114)	Ph.F (n = 35)	A.F (n = 70)
DQ2 (0201/0202)	26.0%	14.3%	27.8%	13.9%	24.4%	13.3%	29.8%	16.7%	25.7%	14.3%
DQ2 (0203)	3.9%	1.9%	5.6%	2.8%	2.2%	1.1%	3.5%	1.8%	5.7%	2.9%
DQ4 (0401/0402)	7.8%	3.9%	11.1%	5.6%	6.7%	3.3%	5.3%	2.6%	2.9%	1.4%
DQ5 (050101-0504)	27.3%	14.3%	33.3%	19.4%	28.9%	15.6%	28.1%	14.9%	28.6%	15.7%
DQ6(060101/0620)	49.4%	28.6%	55.6%	30.6%	48.9%	27.8%	42.1%	24.6%	42.9%	25.7%
DQ7 (0301/0309-10/0313)	40.3%	20.8%	33.3%	16.7%	44.4%	23.3%	42.1%	21.1%	48.6%	24.3%
DQ7(0304)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
DQ8 (0302/0307-08/0311)	24.7%	13.6%	16.7%	11.1%	22.2%	12.2%	26.3%	14.9%	20.0%	11.4%
DQ9 (0303/0306)	5.2%	2.6%	0.0%	0.0%	6.7%	3.3%	7.0%	3.5%	8.6%	4.3%

A.F, Allele Frequency; Ph.F, Phenotype Frequency.

that showed no significant proliferation to any peptide did not differ from the rest of the 73 patients who did proliferate to at least one peptide (P -value = 0.787).

Cytokine profile induced by Ara h 2 peptides

The secreted cytokine profiles *in vitro* were measured in 33 randomly selected subjects with significant proliferative responses ($SI \geq 1.5$) and sufficient supernatant to analyse (Figure S2). These 33 subjects did not differ significantly from the 40 subjects not analysed with respect to their HLA frequencies (P -value = 0.254) or in their magnitude of proliferation per peptide (P -values > 0.1). Cytokine profiles were compared with five controls (Table 3). TNF- α responses were negligible in all samples. A dominant Th2 cytokine response (IL-5 and IL-13) was clearly demonstrated for most of the peptides belonging to regions B and C in peanut-allergic subjects, whereas the same regions in controls showed or had a trend towards a Th1 cytokine profile. A similar trend was also observed for the peptides of region D, whereas no particular trend was evident for region A (Table 3). Cytokine levels were negligible in supernatants when no significant proliferation ($SI < 1.5$) was identified.

Ara h 2 epitope prediction

Ara h 2 has 137 theoretical 20-mers (offset of one) plus one 7-mer. The two epitope-prediction algorithms, NetMHCIIpan-2.0 and NetMHCII-2.2, were used to identify significant 20-mers and core 9-mer peptides in the 31 peptides that were utilized in the assays (thirty 20-mers and one 7-mer) and in the 107 additional theoretical 20-mers that were not synthesized (Fig. 2). All estimations were based on the HLA-DRB1* alleles found in the study cohort. For HLA-DQ alleles, only DQB1*0201, *0402, *0501, *0602 were examined as these were the only ones available for the *in silico* analysis.

Peptides #1–4, 5–9, 17–22 and 24–27 (peptide ID based on current experiment) were identified as relevant binders (predicted $IC_{50} < 500$ nM, weak or strong) covering the corresponding four regions. In particular, peptides #1–3, 19–21 and 25–26 were predicted to be strong binders (predicted $IC_{50} < 500$ nM). These results are consistent with the four regions that were experimentally identified as significant in the proliferation studies (Fig. 2).

In the overall analysis using both algorithms, Ara h 2 was predicted to have 21 significant (9 strong and 12 weak binders) putative 9-mer core epitope sequences. These epitopes were clustered in four regions of Ara h 2 (amino acid #): region I (7–23), region II (24–52), region III (91–117) and region IV (126–141) (Fig. 2 and Table 4), which correspond to the sequences identified in the proliferation assays (aa # 7–23 vs. 6–25, 24–52 vs. 21–65, 91–117 vs. 81–125 and 126–141 vs. 121–145 respectively), as well as with previously published epitopes [17, 18]. Likewise, the core nine-mer epitopes identified computationally were similar to previously described epitopes (aa# 32–44 vs. 32–40/34–42, 37–47 vs. 34–42/39–47, 91–102 vs. 91–99/95–103, 95–107 vs. 95–103/101–109 and 128–141 vs. 126–134/130–138/133–141 respectively), although some new epitopes were also predicted in Ara h 2 (aa #7–23) (Table 4). These core peptides are located within the newly identified significant region A (Fig. 2). When analysing the predicted significant epitopes separately for either all examined DR alleles or all available DQ alleles, exactly the same epitopes were identified.

Population coverage analysis

The Population Coverage Calculation program was utilized based on the HLA genotypic frequencies and the MHC/peptide-specific recognition patterns found in our study cohort. All subjects studied (100%) are expected to respond significantly to peptides belonging to at

Table 3. Cytokine fold change from baseline: peptide based comparison between patients and controls

Region	Peptide ID	Patients					Controls			
		IFN- γ	IL-5	<i>P</i> -value	IL-13	<i>P</i> -value	IFN- γ	<i>P</i> -value	IL-5	IL-13
A	2	1.08 (1.49)	1.19 (12.76)		1.04 (7.45)		1 (0)		1.09 (0.04)	1 (0)
B	5	1 (0.25)	10.3 (59.63)*	0.040	8.74 (19.94)		3.35 (4.69)		1.07 (0.01)	1.11 (0.15)
	6	1.07 (0.28)	7.69 (14.85)*	0.039	4.12 (11.5)		6.94 (7.84)		1.1 (0.2)	2.76 (2.72)
	7	1.12 (0.14)	11.34 (34.06)*	0.049	5.84 (23.83)		2.76 (2.43)*	0.035	1.02 (0.01)	1.21 (0.41)
	8	1.06 (0.2)	14.65 (66.55)*	0.047	7.11 (44.21)		3.39 (4.79)		1.04 (0.08)	1.11 (0.22)
	9	1.07 (0.57)	1.63 (14.08)		1.87 (10.21)		6.7 (2.54)*	0.034	1.25 (0.48)	1.64 (0.77)
	10	1.04 (0.09)	15.18 (50.89)*	0.016	11.48 (48.82)*	0.018	1.28 (1.17)		1 (0.05)	1 (0)
C	17	1.03 (0.16)	4.97 (46.46)*	0.034	3.5 (11.75)		1.68 (1.37)		1.07 (0.14)	1.16 (0.2)
	18	1 (0)	8.31 (26.86)		5.31 (9.85)		2.96 (3.92)		1.01 (0.02)	1.96 (1.72)
	19	1.04 (0.1)	11.5 (47.11)*	0.036	8.45 (31.02)		3.49 (0.44)*	0.035	1 (0)	1.54 (0.79)
	20	1.06 (0.14)	1.59 (12.7)*	0.036	1.46 (12.21)		2.36 (2.73)		1 (0)	1 (0)
	21	1.04 (0.46)	6.16 (44.57)*	0.030	3.97 (21.28)		2.67 (3.35)		1.07 (0.12)	1.2 (0.22)
	22	1 (0.19)	1.28 (3.55)		1.44 (2.05)		3.95 (2.84)		1.46 (0.4)	3.68 (3.24)
D	25	1.08 (0.16)	3.1 (22.73)		2.5 (21.75)		3.04 (4.07)		1.17 (0.13)	1.76 (0.91)
	26	1.09 (0.12)	1.65 (2.36)		1.56 (3.66)		4.47 (6.93)		1.03 (0.06)	1.2 (0.4)

*statistically significant (*P*-value < 0.05) higher levels after comparing the same cytokine between patients and control.

least one of the four regions that were identified as significant in the proliferation assays, and almost all of them (99.88%) to those predicted from the *in silico* analysis of significant 20-mers. The population coverage on a per-region per-allele basis is presented in Table S2. Based on this analysis, a vaccine containing at least four peptides (≤ 20 amino acids in length) from the predicted significant Ara h 2 regions (regions I–IV, Table 4 and Fig. 2c) would cover 99.85% of the participating subjects (calculations are based on the significance of the core 9-mer with the existing HLA alleles shown). The Population Coverage Calculation program does not provide standard errors for comparing the estimated percentage coverage between the experimental and *in silico* approaches.

Discussion

Despite significant progress in our understanding of the underlying pathophysiological mechanisms of food allergy in recent years, strict avoidance and treatment of accidental exposures remain the only approved therapeutic option. Several different types of immunotherapy, such as oral (OIT), sublingual (SLIT) and epicutaneous immunotherapy (EPIT) [30] are being studied for the treatment of food allergy. OIT represents a promising therapeutic intervention for peanut allergy [31–34], although adverse reaction rates are high and about 15–20% of patients cannot tolerate this form of therapy. Similar observations have been reported for OIT with milk [35] and egg [36], and recently experts concluded that OIT for food allergy is not yet ready for general clinical use [37, 38]. However, long-term tolerance has not been demonstrated and further studies are needed. An alternative approach that has garnered

increasing interest is the use of peptide-based-immunotherapy for food allergy [39–43].

In the last 3–4 years, *in silico*-derived algorithms have substantially improved their accuracy to predict T cell epitopes. Until now, most of these algorithms have targeted HLA-DR molecules and not HLA-DP and -DQ molecules, even though they are likely important for antigen presentation. NetMHCpan-2.0 has recently been evaluated and shown to have a per-allele mean accuracy of 0.854 (1.0 being 100% accurate, and 0.5 of no significance) [27]. In another recent publication, the per-allele mean prediction accuracy of NN-align, was 0.882 and in particular for DQ, 0.927 [28]. The aforementioned values have been estimated after using a large dataset that covers HLA-DP molecules, as well.

Despite the high accuracy of these predictions, this approach has not yet been applied to peptide-based vaccine development for allergic diseases. To our knowledge, this is the first report to utilize this technique for predicting food-allergen epitopes. In this study, we have used two MHC-class II-based applications to validate their ability to identify significant MHC binders. The results of the peptide predictions were compared with *in vitro* T cell proliferation from a cohort of peanut-allergic patients to select peptide candidates for the development of a peanut allergy vaccine. We found that both NetMHCIIpan-2.0 and NetMHCII-2.2 performed well in predicting Ara h 2 peptides that induced significant T cell proliferation *in vitro* (Fig. 2). While not intending to replace the *in vitro* assays, this bioinformatics approach provides the opportunity to analyse additional theoretical peptides that are not included in the proliferation assays (offset of 1 vs. offset of 5). Moreover, these theoretical peptides can be divided into significant or non-significant peptides by

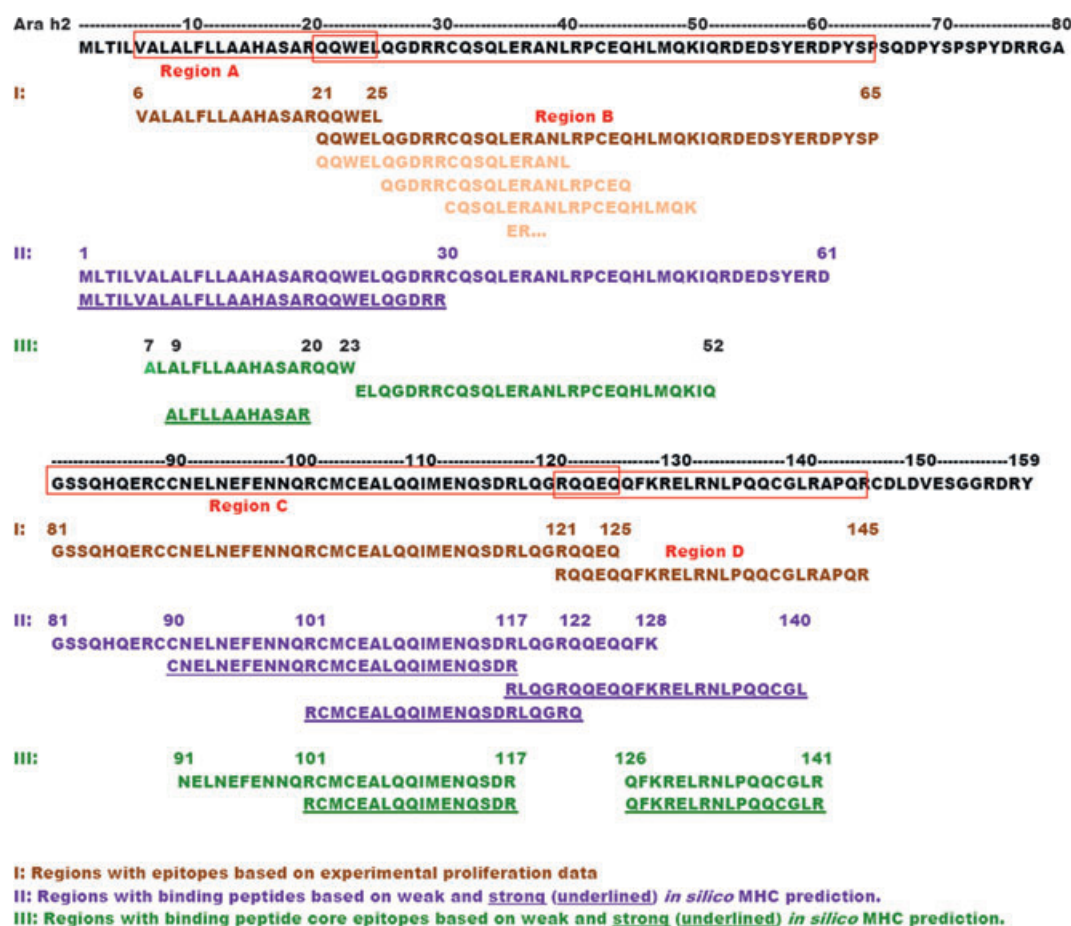


Fig. 2. Schematic representation of dominant regions of proliferation identified experimentally (a) and predicted after *in silico* analysis ones: Major histocompatibility complex (MHC) binding prediction based on weak and strong (bold) binding 20-mer peptides (b) and MHC binding prediction based on weak and strong (bold) binding core epitopes (c). At the top, the entire sequence of Ara h 2 used (ACN62248.1). Sequences provided in the figure represent the whole group of short overlapping peptides that configure each region of proliferation identified either experimentally [region A (Aa #6–25), peptide #2; region B (Aa #21–65), peptides 5–10; region C (Aa #81–125), peptides 17–22; and region D (Aa #121–145), peptides 25 and 26] or through *in silico* predictions.

Table 4. Predicted significant 9-mer core peptides for Ara h 2 selected on the basis of significant HLA binding, clustered in four epitopic sets

	Ara h 2 Aa#			
	Region I (7–23)	Region II (24–52)	Region III (91–117)	Region IV (126–141)
Strong-binding core peptides (predicted IC50 < 50 nM)	9–17	–	101–109	126–134
	10–18		102–110	130–138
	11–19		109–117	
	12–20			
Weak-binding core peptides (predicted IC50 ≤ 500 nM)	7–15	24–32	91–99	133–141
	15–23	31–39	95–103	
		32–40		
		34–42		
		39–47		
		42–50		
		44–52		

IC50, half maximal inhibitory concentration.

the prediction software, which can additionally identify putative 9-mer cores that enable the generation of a shortlist of the most important peptides needed for inclusion in a peptide vaccine [24, 25]. The predicted putative core peptides of Ara h 2 identified using these programs coincided or were very similar to those recently described [18].

We also sought to associate the *in vitro* responses with particular HLA-DRB1* and -DQB1* alleles. Four regions defined by 15 Ara h 2 peptides (20-mers) were found to significantly induce PBMC proliferation. Three of these peptide regions (aa #21–65, #81–125 and #121–145) have been described previously using peanut-specific T cell lines [17, 18]. An additional region (aa #6–25) was identified for the first time (Fig. 2). Interestingly, part of this sequence (aa #6–21) corresponds to the Ara h 2 signal peptide (aa #1–21). The signal peptide is not believed to be secreted, but thought to

remain and be processed inside the endoplasmic reticulum [44]. However, our *in vitro* proliferation results indicate that this peptide is either secreted and recognizable by patient T cells, or is cross-reactive with other non-signal peptides, which seems unlikely.

T cell epitope mapping can be performed through proliferation assays (primary PBMCs, T cell lines or clones), cytokine detection (by flow cytometry or ELISPOT) [45, 46] and more recently with the Tetramer-Guided Epitope Mapping (TGEM) approach [47–50]. Proliferation assays are an easy and relatively rapid mapping technique; however, the low frequency of allergen-specific CD4⁺ T cells requires large numbers of PBMCs to test all potential allergens and obtain sufficient signal for detection. Therefore, most mapping studies use allergen-specific T cell lines or clones [45, 46]. Although T cell lines or clones overcome the frequency limitation, cells are analysed following *in vitro* expansion, which can alter cell phenotypes [47]. In addition, they represent a repertoire potentially biased by prior *in vitro* selection of the rapidly proliferating clones, which may skew the results [51, 52].

When designing peptide-based vaccines, it is very important to ensure that the epitope repertoire is generated from a diverse and representative sample of patients containing MHC-class II types relevant to the population of interest. This was the case in our *in vitro* study, as the HLA genotyping showed 30 different allele specificities for the DRB1* locus and eight serological specificities for DQ in our patient cohort. The HLA-allele profile observed in our cohort is predominantly represented in Caucasian North American populations [53]. Our findings also suggest that the selected T cell epitopes are promiscuous and able to bind to several alleles.

The use of small peptides for immunotherapy is intended to avoid systemic allergic reactions due to cross-linking IgE present on the surface of effector cells [54]. Based on Ara h 2 immunodominant regions for IgE antibody binding reported in the literature [55–58], some B-cell epitopes appear to overlap T cell epitope regions, as determined by proliferation assays and *in silico* predictions. However, preliminary data from basophil activation assays with blood samples from peanut-allergic individuals using these Ara h 2 peptides showed no degranulation above baseline (medium-only control) (unpublished observations). This suggests that the clinical significance of this overlap between B and T cell epitopes is unlikely to be relevant.

The population coverage calculation program is also useful for designing peptide-based vaccines. The HLA genotypic frequencies were analysed to associate them with the MHC/peptide-specific recognition patterns found in our study cohort and to compare them with the *in silico*-predicted data. The expected high coverage of

the *in vitro* cohort data were compared for each region with the coverage indicated by the *in silico* DR and DQ analysis alone, as well as the combined information from both molecules' binding patterns (Table S2). Interestingly, regions A and C had a similar high coverage with that indicated by the *in vitro* data when both estimations from DR and DQ alleles were used. This was not the case for region B and D, as this coverage was underestimated compared to the *in silico* data. This may be explained by the fact that not all existing DR and DQ alleles were available in our cohort, to be utilized for the population coverage calculation based on the *in silico* results. Analysis of additional alleles is expected to increase coverage. It is also possible that significant DP alleles (that were not genotyped) in our cohort contribute to T cell proliferation in response to peptides from region B and D. The *in vitro* results take into account all possible DR, DQ and DP alleles (including those not genotyped in our cohort) that might play an important role in the observed proliferation. The fact that prediction analysis revealed the same significant peptides with those identified *in vitro* suggests that predictions based on either DR alone or DQ alone may be sufficient for identifying significant peptides and core epitopes that 'cover' the majority of peanut-allergic individuals.

As expected, a dominant Th2 cytokine profile was found in the supernatants of allergic subjects' PBMCs following stimulation with most of the peptides identified as significant. On the other hand, the same peptides in control subjects induced or showed a trend towards a Th1 cytokine profile (Table 3). Interestingly, only peptides from regions B and C elicited Th2 cytokines. A lack of statistical power may have precluded the identification of peptides from regions A and D as Th2 inducers, or in fact these peptides may not induce Th2 cytokines, but could still play a role in regulating immune responses.

In control subjects, some peptides (#3, #26 and #27) were also found to generate significant proliferation. Two of the peptides (#3 and #27) did not induce significant responses in peanut-allergic subjects, whereas peptide #26 induced significant proliferation in both controls and peanut-allergic subjects. Interestingly, this peptide induced a Th1 cytokine profile in both.

In conclusion, using primary PBMC cultures from a large cohort of peanut-allergic children, we have identified four dominant regions in Ara h 2 as potential candidates for peptide-based immunotherapy and demonstrated the utility of *in silico* algorithms for predicting Ara h 2 T cell epitopes in peanut-allergic patients. Our findings strongly suggest that the *in silico* approach may be useful in predicting the significant MHC-binding peptides from large major allergens (e.g. Ara h 1 and Ara h 3) with a high degree of certainty and thus provide information to carefully select candi-

date peptides for *in vitro* assay validation and possibly for inclusion in an immunotherapeutic vaccine.

Acknowledgements

We thank Alexander Grishin, PhD, Brittney O'Keefe and Jing Lin, PhD, for their technical assistance. M. Pascal has been funded by Hospital Clínic de Barcelona, Emili Letang Award.

Disclosure

This study has been funded in part by a grant from the Food Allergy Initiative (FAI). HA Sampson serves as a consultant to FAI. The rest of the authors have nothing to disclose.

Conflict of interests: The authors declare no conflicts of interest.

References

- Rudders SA, Banerji A, Vassallo MF, Clark S, Camargo CA Jr. Trends in pediatric emergency department visits for food-induced anaphylaxis. *J Allergy Clin Immunol* 2010; **126**:385–8.
- Sicherer SH, Sampson HA. Food Allergy. *J Allergy Clin Immunol* 2010; **125**:S116–25.
- Skolnick HS, Conover-Walker MK, Koerner CB, Sampson HA, Burks W, Wood RA. The natural history of peanut allergy. *J Allergy Clin Immunol* 2001; **107**:367–74.
- Nicolaou N, Poorafshar M, Murray C *et al.* Allergy or tolerance in children sensitized to peanut: prevalence and differentiation using component-resolved diagnostics. *J Allergy Clin Immunol* 2010; **125**:191–7.
- Burks W, Sampson HA, Bannon GA. Peanut allergens. *Allergy* 1998; **53**: 725–30.
- Blanc F, Adel-Patient K, Drumare MF, Paty E, Wal JM, Bernard H. Capacity of purified peanut allergens to induce degranulation in a functional *in vitro* assay: Ara h 2 and Ara h 6 are the most efficient elicitors. *Clin Exp Allergy* 2009; **39**:1277–85.
- Porterfield HS, Murray KS, Schlichting DG *et al.* Effector activity of peanut allergens: a critical role for Ara h 2, Ara h 6, and their variants. *Clin Exp Allergy* 2009; **39**:1099–108.
- Oppenheimer JJ, Nelson HS, Bock SA, Christensen F, Leung DY. Treatment of peanut allergy with rush immunotherapy. *J Allergy Clin Immunol* 1992; **90**:256–62.
- Nelson HS, Lahr J, Rule R, Bock A, Leung D. Treatment of anaphylactic sensitivity to peanuts by immunotherapy with injections of aqueous peanut extract. *J Allergy Clin Immunol* 1997; **99**:744–51.
- Moldaver D, Larché M. Immunotherapy with peptides. *Allergy* 2011; **66**:784–91.
- Casale TB, Stokes JR. Future forms of immunotherapy. *J Allergy Clin Immunol* 2011; **127**:8–15.
- Grönlund H, Saarne T, Gafvelin G, van Hage M. The major cat allergen, Fel d 1, in diagnosis and therapy. *Int Arch Allergy Immunol* 2010; **151**:265–74.
- Müller U, Akdis CA, Fricker M *et al.* Successful immunotherapy with T cell epitope peptides of bee venom phospholipase A2 induces specific T cell anergy in patients allergic to bee venom. *J Allergy Clin Immunol* 1998; **101**:747–54.
- Hong SJ, Michael JG, Fehring A, Leung DY. Pepsin-digested peanut contains T cell epitopes but no IgE epitopes. *J Allergy Clin Immunol* 1999; **104**:473–7.
- Eiwegger T, Rigby N, Mondoulet L *et al.* Gastro-duodenal digestion products of the major peanut allergen Ara h 1 retain an allergenic potential. *Clin Exp Allergy* 2006; **36**:1281–8.
- King N, Helm R, Stanley JS *et al.* Allergenic characteristics of a modified peanut allergen. *Mol Nutr Food Res* 2005; **49**:963–71.
- Glaspole IN, de Leon MP, Rolland JM, O'Hehir RE. Characterization of the T cell epitopes of a major peanut allergen, Ara h 2. *Allergy* 2005; **60**:35–40.
- Prickett SR, Voskamp AL, Dacumos-Hill A, Symons K, Rolland JM, O'Hehir RE. Ara h 2 peptides containing dominant CD41 T cell epitopes: candidates for a peanut allergy therapeutic. *J Allergy Clin Immunol* 2010; **127**:608–15.
- Nielsen M, Lund O, Buus S, Lundegaard C. MHC class II epitope predictive algorithms. *Immunology* 2010; **130**:319–28.
- Pizza M, Scarlato V, Masignani V *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 2000; **287**:1816–20.
- Sampson HA. Utility of food-specific IgE concentrations in predicting symptomatic food allergy. *J Allergy Clin Immunol* 2001; **107**:891–6.
- Beyer K, Morrow E, Li XM *et al.* Effects of cooking methods on peanut allergenicity. *J Allergy Clin Immunol* 2001; **107**:1077–81.
- Rammensee HG, Friede T, Stevanović S. MHC ligands and peptide motifs: first listing. *Immunogenetics* 1995; **41**:178–228.
- Nielsen M, Lundegaard C, Blicher T *et al.* Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput Biol* 2008; **4**:e1000107.
- Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 2009; **10**:296.
- Sturniolo T, Bono E, Ding J *et al.* Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 1999; **17**:555–61.
- Nielsen M, Justesen S, Lund O, Lundegaard C, Buus S. NetMHCIIpan-2.0 - Improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure. *Immunome Res* 2010; **6**:9.
- Wang P, Sidney J, Kim Y *et al.* Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics* 2010; **11**:568.
- Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 2006; **7**:153.
- Dupont C, Kalach N, Soulaïnes P, Legoué-Morillon S, Piloquet H, Benhamou PH. Cow's milk epicutaneous

- immunotherapy in children: a pilot trial of safety, acceptability, and impact on allergic reactivity. *J Allergy Clin Immunol* 2010; 125:1165–7.
- 31 Clark AT, Islam S, King Y, Deighton J, Anagnostou K, Ewan PW. Successful oral tolerance induction in severe peanut allergy. *Allergy* 2009; 64:1218–20.
 - 32 Jones SM, Pons L, Roberts JL *et al*. Clinical efficacy and immune regulation with peanut oral immunotherapy. *J Allergy Clin Immunol* 2009; 124:292–300.
 - 33 Hofmann AM, Scurlock AM, Jones SM *et al*. Safety of a peanut oral immunotherapy protocol in children with peanut allergy. *J Allergy Clin Immunol* 2009; 124:286–91.
 - 34 Blumchen K, Ulbricht H, Staden U *et al*. Oral peanut immunotherapy in children with peanut anaphylaxis. *J Allergy Clin Immunol* 2010; 126:83–91.
 - 35 Skripak JM, Nash SD, Rowley H *et al*. A randomized, double-blind, placebo-controlled study of milk oral immunotherapy for cow's milk allergy. *J Allergy Clin Immunol* 2008; 122:1154–60.
 - 36 Buchanan AD, Green TD, Jones SM *et al*. Egg oral immunotherapy in non-anaphylactic children with egg allergy. *J Allergy Clin Immunol* 2007; 119:199–205.
 - 37 Thyagarajan A, Varshney P, Jones SM *et al*. Peanut oral immunotherapy is not ready for clinical use. *J Allergy Clin Immunol* 2010; 126:31–2.
 - 38 NIAID-Sponsored Expert Panel, Boyce JA, Assa'ad A *et al*. Guidelines for the diagnosis and management of food allergy in the United States: report of the NIAID-sponsored expert panel. *J Allergy Clin Immunol*. 2010; 126:S1–58.
 - 39 Akdis CA, Akdis M. Mechanisms of allergen-specific immunotherapy. *J Allergy Clin Immunol* 2011; 127:18–27.
 - 40 Focke M, Swoboda I, Marth K, Valenta R. Developments in allergen-specific immunotherapy: from allergen extracts to allergy vaccines bypassing allergen-specific immunoglobulin E and T cell reactivity. *Clin Exp Allergy* 2010; 40:385–97.
 - 41 Rolland JM, Gardner LM, O'Hehir RE. Allergen-related approaches to immunotherapy. *Pharmacol Ther* 2009; 121:273–84.
 - 42 Li XM, Srivastava K, Huleatt JW, Bottomly K, Burks AW, Sampson HA. Engineered recombinant peanut protein and heat-killed listeria monocytogenes co-administration protects against peanut-induced anaphylaxis in a murine model. *J Immunol* 2003; 170:3289–95.
 - 43 Li XM, Srivastava K, Grishin A *et al*. Persistent protective effect of heat-killed *Escherichia coli* producing 'engineered', recombinant peanut proteins in a murine model of peanut allergy. *J Allergy Clin Immunol* 2003; 112:159–67.
 - 44 Von Heijne G. The signal peptide. *J Membr Biol* 1990; 115:195–201.
 - 45 Malherbe L. T cell epitope mapping. *Ann Allergy Asthma Immunol* 2009; 103:76–9.
 - 46 Zeiler T, Virtanen T. Mapping of human T cell epitopes of allergens. *Methods Mol Med* 2008; 138:51–6.
 - 47 Kwok WW, Roti M, DeLong JH *et al*. Direct ex vivo analysis of allergen-specific CD4⁺ T cells. *J Allergy Clin Immunol* 2010; 125:1407–9.
 - 48 Bateman EA, Ardern-Jones MR, Ogg GS. Persistent central memory phenotype of circulating Fel d 1 peptide/DRB1*0101 tetramer-binding CD41 T cells. *J Allergy Clin Immunol* 2006; 118:1350–6.
 - 49 Novak EJ, Liu AW, Gebe JA *et al*. Tetramer-guided epitope mapping: rapid identification and characterization of immunodominant CD4⁺ T cell epitopes from complex antigens. *J Immunol* 2001; 166:6665–70.
 - 50 Day CL, Seth NP, Lucas M *et al*. Ex vivo analysis of human memory CD4 T cells specific for hepatitis C virus using MHC class II tetramers. *J Clin Invest* 2003; 112:831–42.
 - 51 Gammon G, Klotz J, Ando D, Sercarz EE. cell repertoire to a multideterminant antigen. Clonal heterogeneity of the T cell response, variation between syngeneic individuals, and in vitro selection of T cell specificities. *J Immunol* 1990; 144:1571–7.
 - 52 Reece JC, McGregor DL, Geysen HM, Rodda SJ. Scanning for T helper epitopes with human PBMC using pools of short synthetic peptides. *J Immunol Methods* 1994; 172:241–54.
 - 53 Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acid Res* 2011; 39:D913–9.
 - 54 Worm M, Lee HH, Kleine-Tebbe J *et al*. Development and preliminary clinical evaluation of a peptide immunotherapy vaccine for cat allergy. *J Allergy Clin Immunol* 2011; 127:89–97.
 - 55 Stanley JS, King N, Burks AW *et al*. Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. *Arch Biochem Biophys* 1997; 342:244–53.
 - 56 Gruber P, Suhr M, Frey A, Becker WM, Hofmann T. Development of an epitope-specific analytical tool for the major peanut allergen Ara h 2 using a high-density multiple-antigenic peptide strategy. *Mol Nutr Food Res* 2004; 48:449–58.
 - 57 Shreffler WG, Lencer DA, Bardina L, Sampson HA. IgE and IgG4 epitope mapping by microarray immunoassay reveals the diversity of immune response to the peanut allergen, Ara h 2. *J Allergy Clin Immunol* 2005; 116:893–9.
 - 58 Flinterman AE, Knol EF, Lencer DA, Bardina L, den Hartog Jager CF, Lin J *et al*. Peanut epitopes for IgE and IgG4 in peanut-sensitized children in relation to severity of peanut allergy. *J Allergy Clin Immunol* 2008; 121:737–43.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Heat map generated from T cell proliferation responses after stimulation with Ara h 2 peptides.

Ara h 2 peptides are represented in columns (1–31). Subjects (S#, n = 80) and controls [atopic (Ca#, n = 6) and non-atopic (C# = 10)] are shown in rows and the SI is in grading scale, ranging from less than 1.5 (black) to 5 or greater (yellow)

Figure S2. Cytokine profiles from cell culture supernatants of peanut allergic individuals.

Table S1. Demographics, clinical characteristics and HLA genotyping of patients recruited.

Table S2. Population coverage calculations per sig-

nificant region for DR and/or DQ alleles based on *in vitro* and *in silico* data.

Data S1. Stimulation index and raw proliferation data per peptide for each patient.

SUPPLEMENTARY DATA – ONLINE REPOSITORY

SUPPLEMENTARY INFORMATION:

METHODS: T cell Epitope Prediction

The different methodologies used for predicting peptide-MHC class II binding are “trained” with known registered binding sequences in an allele-specific manner. These methods can exploit known peptide binding data to generate predictions for sequences not included in the training set.

The peptide binding strength is estimated by using the predicted half maximal inhibitory concentration (IC₅₀) values. Based on the binding strength of the 9-mer cores, both methods distinguish between significant and non-significant binding 9-mer core peptides, and corresponding 20-mer peptides. Significantly binding core and 20-mer peptides are further subdivided into strong (predicted IC₅₀<50 nM) and weak (IC₅₀≤500 nM) binders. Non-significant peptides have IC₅₀>500 nM [25].

SUPPLEMENTARY TABLES:

Table E1 – Online Repository. Demographics, clinical characteristics and HLA genotyping of patients recruited.

Table E2 – Online Repository. Population coverage calculations per significant region for DR and/or DQ alleles based on *in vitro* and *in silico* data.

SUPPLEMENTARY FIGURES:

Figure S1 – Online Repository. Heat map generated from T cell proliferation responses after stimulation with Ara h 2 peptides.

Ara h 2 peptides are represented in columns (1 to 31). Subjects (S#) and Controls (atopic (Ca#) and non-atopic (C#) are shown in rows. The magnitude of the proliferative response is represented by stimulation index (SI) value, in grading scale ranging from less than 1.5 (black) to 5 or greater (yellow).

Figure S2 – Online Repository. Cytokine profiles from cell culture supernatants of peanut allergic individuals.

Table E1 – Online Repository. Demographics, clinical characteristics and HLA genotyping of patients recruited.

ID	Age (months)	Gender	Ethnicity	Specific IgE	AE	Asthma	AR	Other FAs	HLA Genotyping			
									DRB1 Allele		DQB1 Allele	
									1	2	1	2
S1	77	F	C	>100			x	y	*07:01		DQ2*02:01,02:02	
S2	150	M	C	8.45			x	y	NA			
S3	188	M	C	39.10		x	x	y	*04:04	*15:02	DQ6	DQ8
S4	201	F	C	20.20	x	x	x	y	*13:02	*16:02	DQ5	DQ6
S5	150	F	C	>100		x	x	y	NA			
S6	141	F	I	4.97	x	x	x	y	*07:01	*13:01	DQ6	DQ2*02:01,02:02
S7	84	M	C	56.10	x	x	x	n	*01:01	*13:02	DQ5	DQ6
S8	230	F	C	1.20	x	x		y	*03:01	*04:06	DQ4	DQ2*02:01,02:02
S9	60	M	C	>100	x		x	y	*07:01		DQ2*02:01,02:02	
S10	151	F	A	>100		x	x	y	*09:01	*12:02	DQ7	DQ9
S11	98	M	C	>100				n	*04:01		DQ7	DQ8
S12	141	F	C	2.68	x	x		y	*07:01	*15:02	DQ6	DQ2*02:01,02:02
S13	492	M	C	36.70		x	x	y	*13:02	*14:01	DQ5	DQ6
S14	95	M	C	>100		x	x	n	*03:01	*04:02	DQ8	DQ2*02:01,02:02
S15	130	M	C	>100	x	x	x	y	*03:01	*11:04	DQ7	DQ2*02:01,02:02
S16	59	M	C	>100	x	x	x	y	*07:01	*11:04	DQ7	DQ2*02:01,02:02
S17	57	M	C	29.80	x		x	y	*04:02		DQ8	
S18	131	F	C	>100	x	x	x	n	*08:01	*13:01	DQ4	DQ6
S19	137	M	A	99.90	x	x	x	y	*12:02	*16:02	DQ5	DQ7
S20	130	M	C	>100	x		x	y	*11:01	*14:01	DQ5	DQ7
S21	191	F	P/Am	1.27		x		y	*04:02	*13:02	DQ6	DQ8
S22	237	M	C	>100	x	x	x	y	*01:02	*13:05	DQ5	DQ7
S23	181	F	C	>100	x	x	x	y	*15:01	*15:02	DQ6	
S24	103	M	C	>100	x			n	*04:02	*11:04	DQ8	DQ7
S25	129	M	C	>100	x	x	x	y	*03:01	*15:01	DQ2	DQ6
S26	80	M	C	>100	x	x	x	y	*03:01	*15:01	DQ2	DQ6
S27	259	F	C	9.09		x		y	*08:01	*04:02	DQ4	DQ8

S28	145	M	C	6.50	x			n	*15:01	*13:02	DQ6	
S29	98	M	C	97.50	x		x	y	*13:02	*04:03	DQ4	DQ6
S30	168	M	C	46.60	x	x	x	y	*15:01	*07:01	DQ6	DQ2*02:01,02:02
S31	53	M	C	>100		x		y	*15:01	*11:04	DQ6	DQ7
S32	141	F	C	67.9	x	x	x	y	*08:04	*04:05	DQ4	DQ8
S33	156	M	C	16.5	x			y	*11:01	*01:02	DQ5	DQ7
S34	134	M	C	84.2	x	x	x	y	*07:01	*04:03	DQ8	DQ2*02:01,02:02
S35	107	F	C	9.17			x	y	*03:01	*01:01	DQ5	DQ2*02:01,02:02
S36	160	F	C	>100	x		x	y	*03:01	*01:01	DQ5	DQ2*02:01,02:02
S37	248	M	C	3.24	x	x	x	y	*15:01	*13:02	DQ6	
S38	111	M	C	22.70	x	x	x	y	*11:04	*07:01	DQ7	DQ2
S39	139	M	C	>100		x		n	*15:01	*13:01	DQ6	
S40	217	M	C	43.60	x	x	x	y	*04:02	*03:01	DQ8	DQ2*02:01,02:02
S41	121	M	C	>100		x	x	y	*13:02	*04:07	DQ6	DQ7
S42	195	M	C	1.09	x	x	x	y	*15:01	*13:02	DQ6	
S43	104	F	C	>100		x	x	y	*13:02	*04:07	DQ6	DQ7
S44	114	M	A/C	>100	x	x	x	y	*12:01	*10:01	DQ5	DQ9
S45	218	M	C	27.50			x	y	*04:02	*03:01	DQ8	DQ2*02:01,02:02
S46	83	M	C	>100	x	x		y	*13:02	*11:04	DQ6	DQ7
S47	91	M	C	>100				n	*13:05	*07:01	DQ7	DQ2*02:01,02:02
S48	188	F	C	4.92				n	*13:01	*04:01	DQ6	DQ7
S49	101	F	C	68.30	x	x	x	y	*13:02	*11:04	DQ6	DQ7
S50	147	F	C	>100		x	x	y	*12:01	*07:01	DQ2*02:01,02:02	DQ7
S51	160	M	C	17.50		x	x	y	*16:01	*14:01	DQ5	
S52	38	M	C	6.11	x		x	y	*14:01	*04:01	DQ5	DQ7
S53	259	M	C	69.40	x		x	y	*04:03	*04:02	DQ4	DQ8
S54	80	M	C	8.34	x	x		y	*14:01	*07:01	DQ9	DQ5
S55	216	F	C	>100	x	x		y	*13:02	*11:01	DQ7	DQ6
S56	188	M	C	<0.35				n	*11:04	*07:01	DQ7	DQ2*02:01,02:02
S57	132	M	C	>100	x	x	x	y	*13:02	*11:01	DQ7	DQ6
S58	154	F	I	>100		x	x	y	*14:04	*13:01	DQ6	DQ5
S59	167	M	C	63.50	x	x	x	n	*11:04	*04:02	DQ7	DQ8
S60	65	F	C	55.90				n	*11:04	*04:05	DQ7	DQ8
S61	172	M	C	0.41		x	x	y	*15:01	*07:01	DQ2*02:01,02:02	DQ6
S62	77	F	C	30.10		x		n	*13:02	*04:03	DQ8	DQ6

S63	99	F	A	3.64	x		x	n	NA			
S64	177	F	C	>100		x	x	y	*12:01	*01:02	DQ7	DQ5
S65	156	M	AA	>100		x	x	y	*12:01	*11:02	DQ7	
S66	61	M	C	>100	x			n	*04:02		DQ8	
S67	120	F	C	>100		x		y	*15:01	*01:01	DQ6	DQ5
S68	102	M	C	33.60	x	x		y	*16:01	*15:02	DQ6	DQ5
S69	123	M	C	>100		x	x	y	*13:01	*11:01	DQ7	DQ6
S70	115	M	C	9.36		x	x	y	*03:01	*12:01	DQ7	DQ2*02:01,02:02
S71	134	F	C	>100				y	*04:01	*07:01	DQ8	DQ2*02:01,02:02
S72	128	M	C	>100			x	y	*13:02	*13:03	DQ6	DQ7
S73	150	M	C	>100	x		x	y	*04:04	*13:01	DQ6	DQ8
S74	38	M	C	>100	x		x	y	*01:02	*11:01	DQ5	DQ7
S75	48	M	C	7.50	x			y	*13:02	*14:01	DQ5	DQ6
S76	63	M	C	72.20	x	x	x	y	*04:01	*15:01	DQ6	DQ7
S77	93	F	C	1.95				n	*04:02	*07:01	DQ8	DQ9
S78	37	M	C	<0.35				y	*13:02	*15:01	DQ6	
S79	82	M	C	>100	x	x		n	*01:02	*15:02	DQ6	DQ5
S80	87	F	C	46.20				y	*01:02	*13:02	DQ5	DQ6

Abbreviations used: FA: Food Allergy; AE: Atopic Eczema; AR: Allergic Rhinitis; Ethnicity: C: Caucasian; A: Asian; P/Am: Pakistani/American; I: Indian; AA: African-American; NA: Not available.

Table E2 – Online Repository. Population coverage calculations per significant region for DR and/or DQ alleles based on *in vitro* and *in silico* data.

Alleles	Region A	Region B	Region C	Region D
<i>In vitro data</i>				
All alleles	99.73%	99.94%	99.88%	99.87%
<i>In silico data</i>				
DR	99.46%	34.07%	59.17%	74.70%
DQ	71.7%	49.02%	84.87%	49.02%
DR+DQ	98.72%	66.36%	93.82%	87.1%

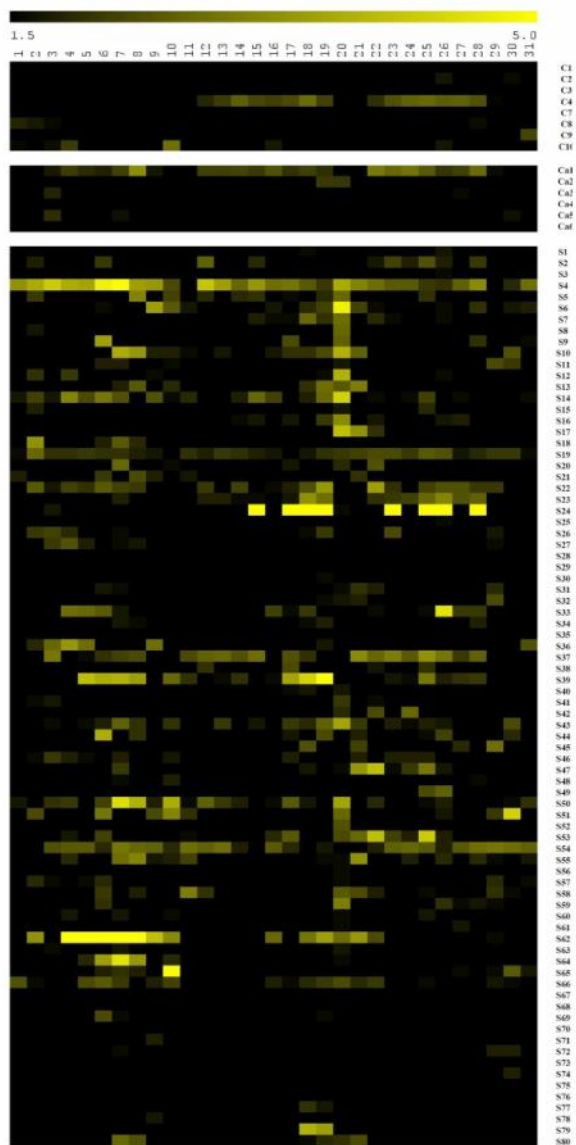


Fig S1

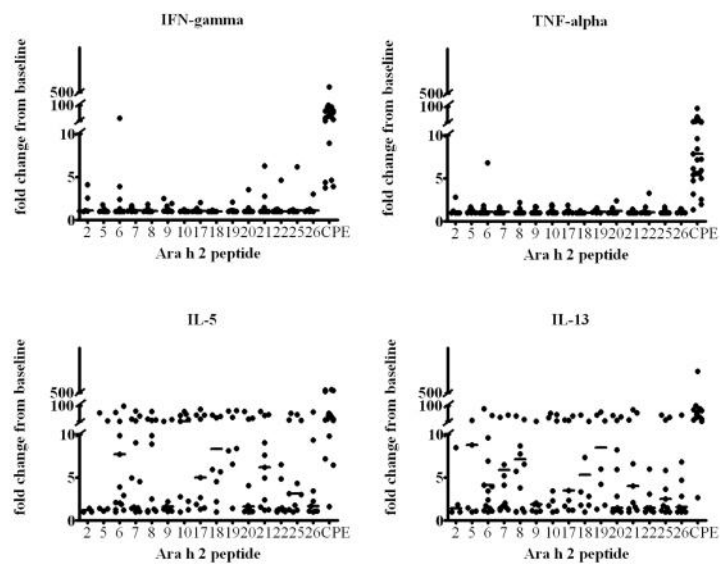


Fig S2