



A novel approach to ameliorate experimental milk allergy based on the oral administration of a short soy cross-reactive peptide

Ángela María Candreva^{a,1}, Paola L. Smaldini^{a,1}, Ana Cauerhff^b, Silvana Petrucci^{c,2}, Guillermo H. Docena^{a,*,2}

^a Instituto de Estudios Inmunológicos y Fisiopatológicos (IIFP), UNLP, CONICET, asociado a CIC PBA, Facultad de Ciencias Exactas, Departamento de Ciencias Biológicas, La Plata, Argentina

^b Departamento Química Biológica, Facultad de Ciencias Exactas y Naturales, CONICET-Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina

^c Centro de Investigación y Desarrollo en Criotecología de Alimentos, CONICET y Universidad Nacional de La Plata, La Plata, Argentina

ARTICLE INFO

Keywords:

Cross-reactivity
Food allergy
IgE
Milk allergy
Soy peptide
Prevention

ABSTRACT

Food allergy is on the rise, and preventive/therapeutic procedures are needed. We explored a preventive protocol for milk allergy with the oral administration of a *Gly-m-Bd-30K* soy-derived peptide that contains cross-reactive epitopes with bovine caseins. B/T-cross-reactive epitopes were mapped using milk-specific human sera and monoclonal antibodies on overlapping and recombinant peptides of *Gly-m-Bd-30K* by SPOT and cell proliferation assays. Bioinformatics tools were used to characterize epitopes on the 3D-modelled molecule, and to predict the binding to HLA alleles. The peptide was orally administrated to mice that were then IgE-sensitized to milk proteins. Immunodominant B-epitopes were mainly located on the surface of the Nt-fragment. The use of a soy-peptide-containing an immunodominant cross-reactive T-epitope, along with a single B epitope, prevents IgE-mediated milk sensitization through the induction of Th1-mediated immunity and induction of blocking IgG. The use of a safe soy-peptide may represent a promising alternative for preventing milk allergy.

1. Introduction

Food allergy is caused by a breakdown in immunologic tolerance, and its prevalence has increased worldwide over time (Tang & Mullins, 2017). Understanding the impaired immune regulatory circuits is crucial for the development of new therapeutic options since the restoration of immune tolerance is a goal in the treatment of allergy. Due to early and frequent exposure to milk in life, cow's milk allergy (CMA) is one of the most common food allergies. On the other hand, soy allergy has not been presented as a high-frequency disorder in atopic patients (Kattan et al., 2011), as it might have been expected due to its high exposure in the general population; additionally, soy proteins are less immunogenic than milk or peanut allergens (Cordle, 2004).

We have previously described the *in vitro* and *in vivo* cross-reactivity between bovine caseins and different soy proteins (Candreva et al., 2016, 2015; Curciarello et al., 2014; Smaldini et al., 2012). The *Gly m Bd 30K* soy allergen showed to contain cross-reactive B and T epitopes

(Candreva, Smaldini, & Curciarello et al., 2015). Based on these findings, we used a short peptide with immunodominant epitopes to control IgE production and allergy induction in an experimental model. When speaking of cross-reactivity, it should be distinguished between sensitization and allergy, since only some cross-reactive IgE give rise to clinical food allergy, depending on the level and affinity of cross-reactive IgE, sensitization of mucosal mast cells, etc. (van Ree, 2004). Clinical cross-reactivity has been usually reported between related allergens; however, cross-reactivity between different protein families of seed storage proteins and mammalian proteins (non-homologous allergens) has also been described (Bublin & Breiteneder, 2020). The efficacy of immunotherapy with peptides derived from the allergen has been mainly investigated in respiratory allergy (Bouchaud et al., 2015). The use of cross-reactive allergens or peptides for immune intervention strategies in food allergy has not been thoroughly investigated. The use of intact food allergens in preventive or therapeutic procedures pose a risk of anaphylaxis (Valenta et al., 2016). Although few reports showed

* Corresponding author at: Instituto de Estudios Inmunológicos y Fisiopatológicos (IIFP), UNLP, CONICET, asociado a CIC PBA, Facultad de Ciencias Exactas, Departamento de Ciencias Biológicas, 24 Argentina Boulevard 120 nro. 1489, CP 1900 La Plata, Argentina.

E-mail addresses: guillermo.docena@gmail.com, guidoc@biol.unlp.edu.ar (G.H. Docena).

¹ These authors contributed equally to this work.

² These authors shared senior authorship.

that peptides from food allergens reduced the allergic responses to the naïve allergens (Meulenbroek et al., 2013; Wai et al., 2016), no studies demonstrated that cross-reactive peptides avoid sensitization and allergic reactions (Lozano-Ojalvo & López-Fandiño, 2018).

Based on findings that show that B cell epitopes are key in therapeutic vaccines (Focke-Tejkl et al., 2015), and T cells play a central role in immunotherapies (Möbs et al., 2010, 2012), we aimed to prove that cross-reactivity can be exploited in a preventive protocol for milk allergy. We evaluated the effect of the mucosal administration of a cross-reactive soy peptide-containing B and T cell epitopes to elicit IgG blocking antibodies and immunomodulatory T cells that prevent potential sensitization and milk allergy.

2. Methods

2.1. Human sera and animal antibodies

Sera were obtained from 12 patients diagnosed with milk allergic according to compatible clinical history, positive skin prick test (SPT) with commercial milk allergen (Allergo Pharma, Buenos Aires, Argentina), elevated total serum IgE and milk-specific IgE performed by EAST (Docena et al., 1996), and positive diagnostic elimination diet. These patients showed symptoms associated with milk exposure. Also, eight sera from control subjects with no serum IgE or a history of food allergy were included. The Ethics Committee of the Children's Hospital of La Plata (Buenos Aires, Argentina) approved the protocols (#389-2014 and #389-2018), and patients or their parents gave written informed consent.

Three monoclonal antibodies (mAb) specific for α -casein (1D5), β -casein (4C3), and κ -casein (3B5) were used (Candreva, Smaldini, & Curciarello et al., 2016).

2.2. Peptide synthesis and SPOTs assay with overlapping peptides of Gly m Bd 30K

Epitopes were mapped using an array of consecutive overlapping peptides (15-mer soluble peptides frameshifted by 5 residues) spanning the whole allergen sequence. Individual peptides were synthesized on a derivatized cellulose membrane using *N*-(9-fluorenyl) methoxycarbonyl (Fmoc) amino-acid-active esters (Frank, 2002). Membranes were blocked with TBS-T 3% of equine serum (ES) (20 mM Tris and 150 mM NaCl, 0.1% Tween) and incubated with the primary antibody (1:5 diluted pool of patient sera or 1:1000 monoclonal antibodies) overnight at 4 °C, followed by biotinylated mouse anti-human IgE monoclonal antibody (VECTOR Lab, Burlingame, CA, USA) and HRP-streptavidin-conjugated (Thermo Scientific, Waltham, MA, USA) for 1 h at 37 °C. Luminol-based chemiluminescent substrate and exposure to X-ray film (Amersham Hyperfilm ECL, GE Healthcare, Chicago, IL, USA) were used for spot visualization.

Sequences of overlapping peptides were included in [Supplementary Table 1](#).

2.3. Production of recombinant Nt and Ct polypeptides and immunoblotting

Two fragments of the Gly m Bd 30K allergen corresponding to the N-terminus (Nt) and the C-terminus (Ct) domains were cloned. The cDNA encoding sequence for Gly m Bd 30K (gi | 195957142 from the GenBank database) was obtained by PCR amplification of the cDNA library (Candreva, Smaldini, & Curciarello et al., 2015). Amplified PCR products were cloned into pET200A vector for expression (Life Technology, S.A. Waltham, MA, USA). E. coli BL21 STAR containing the constructs pET200A His-Nt and pET200A His-Ct were induced, and recombinant proteins were purified as described previously (Curciarello et al., 2014). Sepharose-polymyxin B resin (Pierce, Waltham, MA, USA) was used to deplete in lipopolysaccharide, followed with the Limulus amoebocyte

assay (LONZA, Escobar, Bs As, Argentina). These polypeptides were analysed by 12.5% SDS-PAGE and immunoblotted onto nitrocellulose membranes (electro transference for 60 min). After blocking with TBS-T and ES (2 h at 37 °C), membranes were revealed with different primary human and mouse antibodies by stripping: 1D5, 4C3, and 3B5 mAbs, His-probe (H-3) (Santa Cruz Biotechnology, Inc, Santa Cruz, CA, USA), and pools of human sera. Secondary antibodies were used as described in Peptide synthesis and SPOTs assay.

2.4. Synthetic peptides of Ct

Eleven overlapping 12-mer soluble peptides, frameshifted by three residues covering a small region of the amino acid sequence of Ct, were synthesized (ChinaPeptides Co., China). Sequences of overlapping peptides were included in [Supplementary Table 2](#).

2.5. Bioinformatic analysis

2.5.1. In silico prediction of cross-reactive epitopes

The amino acid similarity in the primary sequences between Gly m Bd 30K (gi | 195957142) and bovine casein peptides representing T and/or B epitopes was analysed by multiple sequence alignment. The amino acid sequence of 68 peptides from α s1- (gi | 162792: 214 aa), α s2- (gi | 27806963:222 aa), β - (gi | 162931: 224 aa) and κ -casein (gi | 1228078:190 aa) designed according to Immune Epitope Database and Analysis Resources IEDB were aligned with the Gly m Bd 30K protein sequence by the ClustalW2 server as described in Curciarello et al. (2014).

2.5.2. Secondary structure prediction

The secondary structure prediction from Gly m Bd 30K primary sequence was performed using JPred 4 server (Drozdetskiy et al., 2015), LOMETS, PSSpred and I-Tasser suite from Zhang lab web page and the Phyre2 web portal (Kelley et al., 2015).

2.5.3. Homology modelling

Models of the 3D structure of the Gly m Bd 30K were generated to localize the residues of the epitopes recognized by the different antibodies. Homology modelling from the primary protein sequence of Gly m Bd 30K was performed. The sequence of the protein Gly m Bd 30K (gi | 195957142) from the GeneBank database was used as a target sequence as a query search in the Protein Data Bank (PDB) (Berman et al., 2002) using the Protein BLAST server. The structure of procaricain from *Carica papaya* (PDB Id = 1PCI; Chain A; gi | 2098464) was used as a template. Then, a tentative model of Gly m Bd 30K was generated with the Modeller 9v7 program (Eswar et al., 2007). The loops of this model were further modelled with the DOPE-based loop modelling protocol. The Nt and Ct domains of Gly m Bd 30K were initially modelled with the Swiss Protein Workspace (Waterhouse et al., 2018). The sequence of the Nt polypeptide sent as input to this server was defined from Gly1 to Ser182; the structure of the precursor of a thermostable variant of papain (*Carica papaya*) (PDB Id = 3tnx) was used as a template. The sequence sent as input to obtain the modelled structure of the Ct polypeptide was a fragment composed by Pro135 to Leu378, and the template was the structure of papain fold-protein without the catalytic cysteines from seeds of *Pachyrhizus erosus* (PDB Id = 2b1m). Additionally, the I-Tasser Suite was employed to model the whole Gly m Bd 30K protein and its Nt fragment (aa 1–182). A similar analysis was carried out using the Phyre2 server, obtaining an output corresponding to the whole Gly m Bd 30K protein and the Nt fragment, respectively.

All models were refined with the Modrefiner algorithm (Xu & Zhang, 2011), the second round of refinement was performed with the 3D refine web server (Bhattacharya et al., 2016). The Yasara Energy Minimization server (Krieger et al., 2009) was also employed.

The quality of the protein structure model was evaluated with the QMEAN server (Qualitative Model Energy Analysis) (Benkert et al.,

2008). For a complete assessment of the protein models, the WHAT IF server was employed (Vriend, 1990). Visualization and representations of all models were made with Phymol 0.99rc6 program. The assignment of secondary structure from three-dimensional coordinates was performed by means of the 2struc Secondary Structure Server (Klose et al., 2010).

2.5.4. Estimation of the isoelectric point of the peptides derived from Ct

We used the Isoelectric Point Calculator (IPC) to estimate the isoelectric point (pI) at pH 7.4 (Kozłowski, 2016).

2.5.5. Estimation of the surface charges of the Gly m Bd 30K allergen

We employed the PBEQ Solver server to estimate the electrostatic potential on the solvent-accessible surface on the molecular model obtained (Jo et al., 2008).

2.5.6. Binding kinetics analysis using the BIACORE assay

The kinetics of binding Ag-Ab was performed using surface plasmon resonance biosensor experiments on an IASys Plus apparatus (Affinity Sensors, Saxon Hill, U.K.) according to Curciarello et al. (2014).

More information about all the procedures and calculations are included in Supplementary Materials and methods.

2.6. Animals, allergic sensitization and analysis of the immune response

Male 6- to 8-week old Balb/c mice (School of Animal Sciences, University of La Plata) kept under pathogen-free conditions were sensitized as described in Oliveira et al. (2019) with intragastric cow's milk protein (CMP) plus cholera toxin (CT) (Sigma Aldrich, St. Louis, MO, USA). Animals were divided on groups of 5 mice each. As controls, a group of naïve mice received P3_{ig} (intragastric) and then CMP (unsensitized), other mice received PBS, and then CT + CMP (sensitized), other received Gly m Bd 30K_{sl} (sublingual) and then CT + CMP, or P3_{ig} and then CT + CMP and finally, mice received CMP_{sl/ig} and were further sensitized. Oral challenges with CMP were done to evidence hypersensitivity reactions. The clinical score was determined 30 min after the oral challenge. The Institutional Committee for the Care and Use of Laboratory Animals of the School of Sciences (University of La Plata) approved the protocol (Protocol Number: 017-00-15).

2.6.1. Assessment of clinical signs

Ten days after the last boost, mice were intragastrically challenged with 20 mg of CMP and observed during the following 15–60 min by two independent observers in a blinded fashion. Clinical scores were: 0 = no symptoms, 1 = scratching and rubbing around the nose and head, 2 = puffiness around the eyes and mouth, diarrhoea, piloerection, reduced activity, and/or decreased activity with increased respiratory rate, 3 = wheezing, laboured respiration, cyanosis around the mouth and the tail, 4 = no activity after prodding, or tremor and convulsion, and 5 = death (Smaldini et al., 2012).

2.6.2. Cutaneous test

Analysis of skin mast cell activation was performed as described in Smaldini et al. (2015). Briefly, mice were injected with 20 µg of CMP in 20 µl of sterile saline in one footpad, and saline in the contra-lateral footpad as a negative control. Then, mice were injected intravenously with 100 µl of 0.1% on saline of Evans blue dye (Anedra, CABA, Argentina). Skin color and the footpad swelling (measured with a digital calibre) were assessed 30 min after the administrations.

2.6.3. Proliferation assay by CFSE dilution

Spleen was removed, and a cell suspension was prepared in RPMI medium with 10% fetal bovine serum and antibiotics. Splenocytes (1×10^5 cells) from CMP-sensitized animals were labelled with 3 µM CFSE (Invitrogen, Waltham, MA, USA) and stimulated with medium, 20 µg/ml of Nt, or 20 µg/ml of Ct for 72 h. Then, cells were labelled with anti-CD4

PerCyP 5.5 (eBioscience, Waltham, MA, USA). Fluorescence acquisition was performed with a FACScalibur cytometer (BD Pharmingen, San Diego, CA, USA) using QuestProCell software. The gating strategy for the cell analysis consisted of a lymphocyte gate based on SSC-H vs. FSC-H parameters, followed by SSC-H vs. CD4 fluorescence. Then CFSE-labelled cells were analysed.

2.6.4. Cytokine quantification

Splenocytes (4×10^6 cells/ml) of sensitized and control mice were stimulated with a-CD3 (1 µg/ml of plate-bound), CMP (350 µg/ml), or eleven overlapping peptides of Ct (see Supplementary Table 2 for the sequence of peptides) for 72 h at 37 °C according to previous optimized assays (Smaldini et al., 2015). Cytokines (IL-5, IL-13 and IFN-γ) were measured in supernatants by ELISA (eBioscience, Waltham, MA, USA) following the manufacturer's instructions.

2.6.5. Indirect enzyme-linked immunosorbent assay for 1D5, 4C3, and 3B5

Microtiter plates were coated with 1 µg/well P3 and blocked with PBS-T with 5% ES for 2 h at 37 °C. Thereafter, 100 µl/well of monoclonal antibodies were added overnight at 4 °C. Finally, the secondary antibody, HRP-conjugated goat anti-mouse IgG (H/L) monoclonal antibody (BIORAD, Philadelphia, PA, USA), was added (1:3000, 1 h at 37 °C) and the colour was developed with o-phenylenediamine and reaction was stopped with 2 N H₂SO₄. Optical density (OD) was measured at 492 nm.

2.6.6. Immunoassays for serum specific antibodies

CMP-specific IgE was measured in serum by EAST, and IgG was quantified by indirect ELISA as previously described (Smaldini et al., 2012). All samples were run simultaneously. Briefly, CNBr-activated cellulose paper discs were coupled with CMP (1.75 mg/ml), discs were blocked with ethanolamine, and incubated overnight at 4 °C with 50 µl of serum samples. They were then incubated with a biotinylated anti-mouse IgE monoclonal antibody (BD Pharmingen, San Diego, CA, USA) at 1:750 dilution for 5 h, at 4 °C. Finally, alkaline phosphatase-streptavidin conjugate (Sigma-Aldrich, St. Louis, MO, USA) was added at 1:1000 dilution and incubated for 30 min at 37 °C. The enzymatic activity was revealed with p-nitrophenyl phosphate (Biochemika, Fluka, St. Louis, MO, USA). The reaction was stopped using 0.1 M EDTA, and the optical density (OD) was measured at 405 nm.

2.6.7. In vivo prevention of experimental food allergy

Prior to the allergic sensitization procedure, mice received intragastric (ig) or sublingual (sl) administration of P3 (100 µg ig), CMP (10 mg ig), Gly m Bd 30K (20 µg sl), CMP (20 µg sl) or PBS. Following sensitization with CMP and CT, mice were orally challenged with CMP, and the immune response was evaluated by *in vitro* (serum antibodies and cytokine secretion) and *in vivo* assays (clinical signs and cutaneous tests) as previously described.

2.7. Statistical analysis

All statistical analysis was carried out using GraphPad Prism 5 software. The significance of the difference was determined using one or two-way ANOVA followed by Bonferroni's test and Student's *t*-test; *p*-value <0.05 was considered as statistically significant.

3. Results

3.1. Identification of B epitopes in Gly m Bd 30K

Dot assays with synthetic overlapping peptides and immunoblotting with recombinant fragments were performed for epitope mapping (Fig. 1A and B). We found positive spots with IgG and IgE epitopes (#3, 7, 8, 11, and 13), while peptides in spots #10, 22, and 36 were only positive with monoclonal antibodies for IgG epitopes. The spot #36 was positive with only pool 3 for IgE; also, the β-casein-specific 4C3 mAb

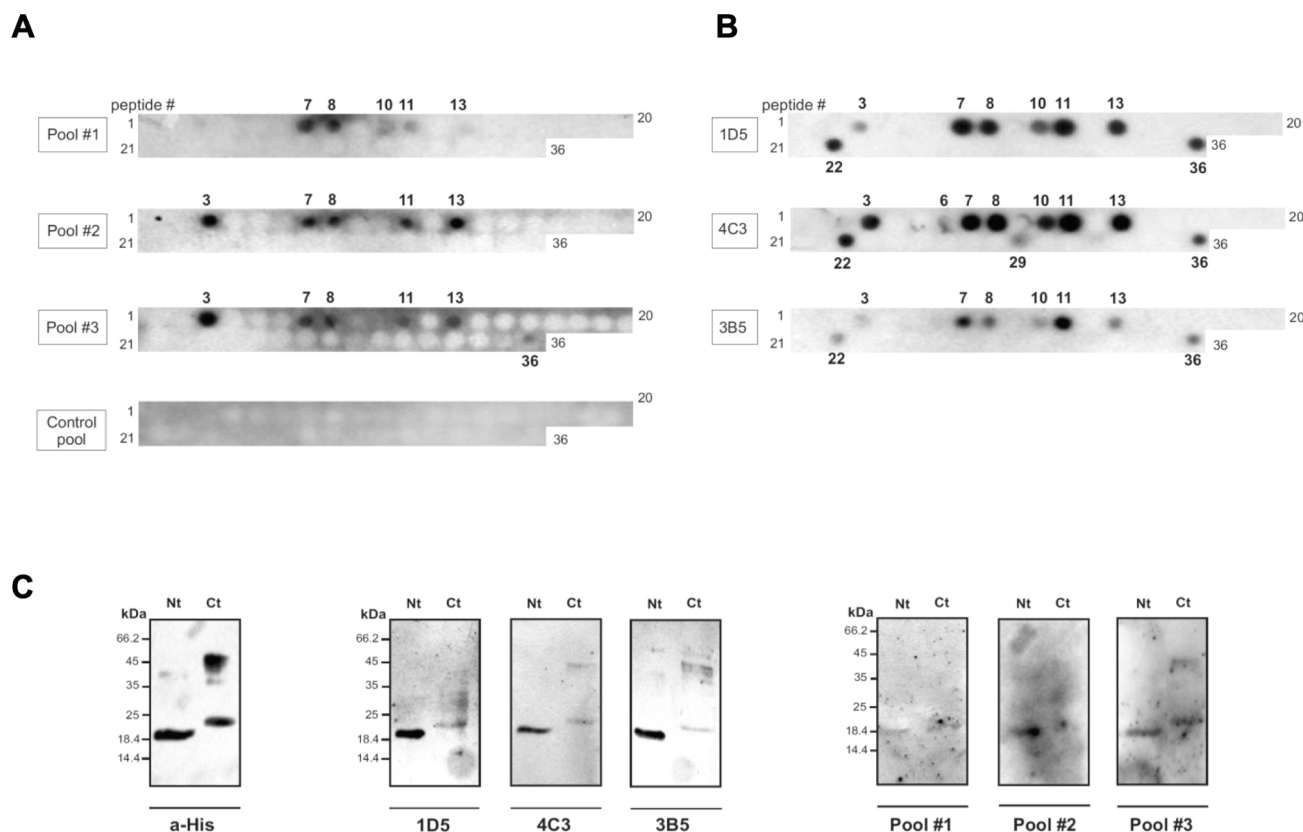


Fig. 1. Epitope mapping of the *Gly m Bd 30K* soybean allergen. A and B, SPOT arrays depicting the overlapping 15-aminoacid peptides covering the complete sequence of *Gly m Bd 30K*. A, Peptide reactivity with four different pools of serum from cow's milk allergic patients (#1, #2, and #3 pools) and a control pool of sera from non-atopic patient. B, Peptide reactivity with three different casein-specific mAbs (1D5 specific for α -casein, 4C3 specific for β -casein and 3B5 specific for γ -casein). C, Immunoblots of recombinant Ct (23 kDa) and Nt (19 kDa) domains of *Gly m Bd 30K* revealed with different specific antibodies: 1D5, 4C3, 3B5 mAbs and pools #1 and #2 of sera from CMA patients. As a control, anti-histidine mAb was used.

identified IgG epitopes on spots #6 and 29. These findings led us to conclude that spots #3, 7, 8, 10, 11, 13, and 36 contained both IgG and IgE epitopes, while spots #6, 10, 22 and 29 contained only IgG epitopes. No spot contained peptides with only IgE epitopes.

The recombinant fragments Nt (aa 1–116, 19 kDa) and Ct (aa 208–360, 23 kDa) were run under reducing conditions (Fig. 1C) and the immunoblotting revealed strong positive bands in Nt for IgG and IgE, which corresponds to peptides of spots #3, 7, 8, 10, and 11; the Ct fragment, which only contains the peptide corresponding to spot #22, showed faint bands for IgG and IgE reactivity. Supplementary Table 1 shows the primary sequence of peptides and the results of the spot array. Overall, these findings suggest that most of B-cell epitopes are located in the Nt domain (spots #1 to 12).

3.2. *In silico* analysis of B epitopes in *Gly m Bd 30K* with similitude with caseins

We analysed the molecular distribution of B epitopes in *Gly m Bd 30K* in *silico* analysis, and Fig. 2A depicts the hit distribution of amino acid similitude of B epitopes in casein and *Gly m Bd 30K* sequences and the spot array results. We found that the *Gly m Bd 30K* sequence positions with high hit values corresponded with the spots that were positive for all antibodies (spots #7, 8, 10, and 11), and are included in Nt. The Ct domain, which only contained IgG epitopes (spots #22, 29, and 36) showed no relevant hit value.

We predicted cross-reactive IgE and IgG epitopes comprised in two domains: a) positions 47–120 with the highest hit values and b) positions 209–236, with low hit values. Also, we carried out a multiple sequence alignment between *Gly m Bd 30K* and epitopes of caseins, and we found multiple amino acid similarities in short sequences (Fig. S1).

The spatial distribution of epitopes and the secondary structure of *Gly m Bd 30K* was 3D-modelled using *Carica papaya* as a template and other structurally similar proteins (Fig. 2B and C). The best model was selected based on results from the prediction and calculation of secondary structure analysis. Three models were assessed: one for the entire *Gly m Bd 30K* molecule, another for the Nt fragment (residues 1–182), both from the Phyre2 server, and a third model of the entire *Gly m Bd 30K* from the I-Tasser server. These models showed high similarity, and their structures were superimposable to each other, giving a root-mean-square deviation value in the range of 1.08–1.13 Å for C-alpha atoms. The main difference between models obtained with the Phyre 2 server, and that from the I-Tasser server, were in the Nt domain (Fig. S2), specifically in the Nt fragment comprised by Gly2 to Ile19.

The structures obtained with the Phyre2 server showed a higher content of α -helix secondary structure than the models from the I-Tasser server. The secondary structure assigned by the Phyre 2 server was in concordance with that predicted with JPred, PSSpred, and LOMETS servers. Fig. 2B shows the ribbon structures with secondary structures with different orientations, and Fig. 2C depicts the putative structure of *Gly m Bd 30K* displaying the two domains and B-epitopes on the surface of the molecule. The Nt domain-containing IgG and IgE epitopes were predominantly composed of α -helix structure, whereas the Ct-domain was formed by a summation of α -helix, β -sheet, and coiled secondary structure. The analysis showed in Fig. S2 also depicts that α -helix is the predominant secondary structure of the cross-reactive IgE/G epitopes. As seen in the figure, most of the cross-reactive B-epitopes were located on the surface of the molecule, highly accessible for antibodies. The grey zone that contained α -helix showed no cross-reactive B-epitopes.

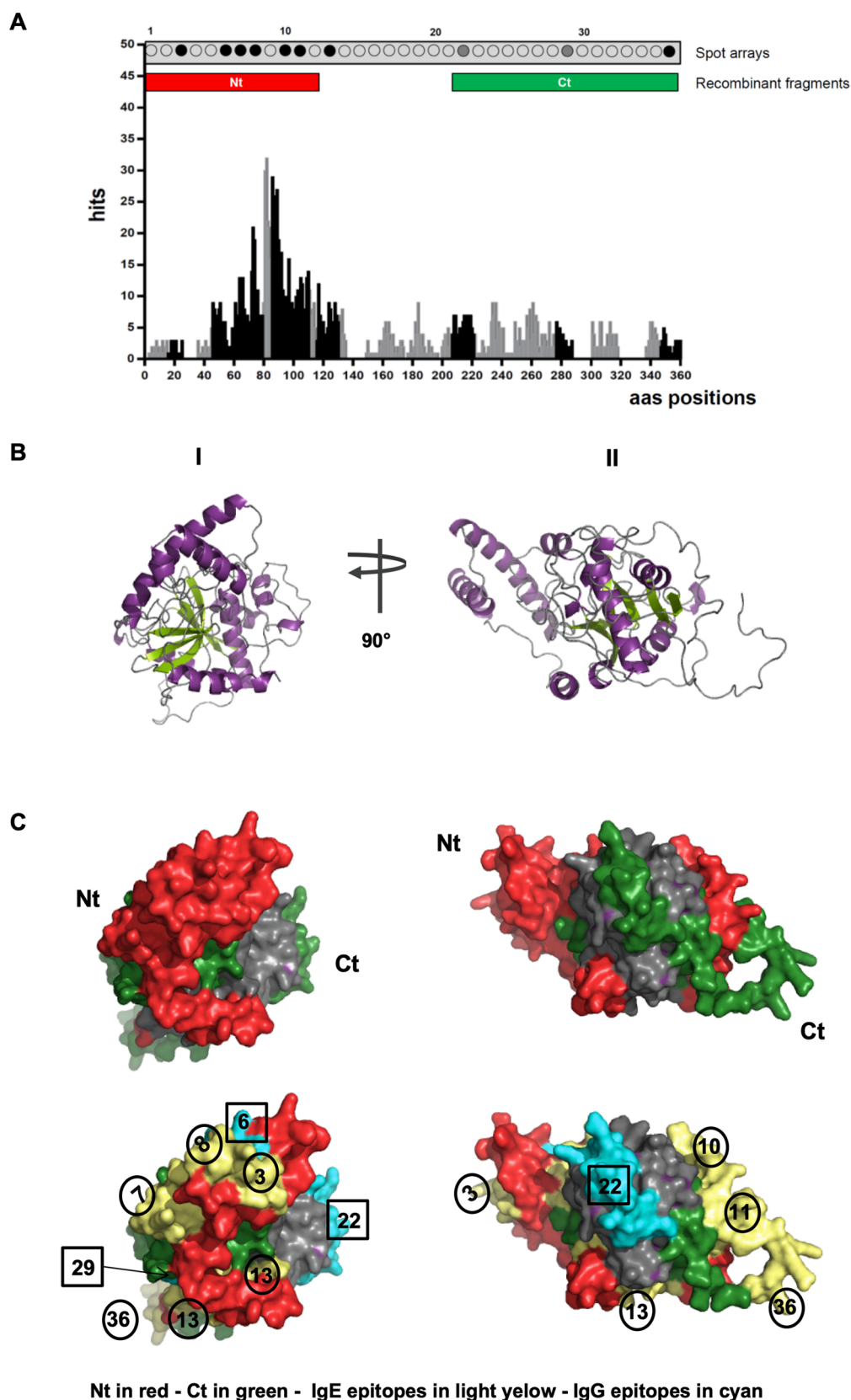


Fig. 2. *In silico* analysis of cross-reactive peptides of *Gly m Bd 30K*. **A**, Hit distribution analysis of the sequence alignment of *Gly m Bd 30K* with B or T epitopes in bovine caseins (IEDB) shown as amino acid similitude per position along the sequence of *Gly m Bd 30K*. Above, positions of *Gly m Bd 30K* with high hit values (black columns) correlate with positive spots of the overlapping assay. Nt (1–116 aa) and Ct (208–360 aa) are depicted. **B**, Secondary structure of *Gly m Bd 30K* predicted from the sequence (consensus structure) and modelled with the Phyre2 web portal. Ribbon diagram of the *Gly-m-bd-30K* is presented in two different orientations; helix is shown in light green, β -sheet in purple, and loops in grey. **C**, 3D-modelled structure of *Gly m Bd 30K*. The orientation of the molecules is the same as shown in **B**. Surfaces of Nt (red), Ct (green), IgE/IgG epitopes (light yellow and circled), and IgG epitopes (cyan and boxed) are indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Binding affinity constants of caseins and *Gly m Bd 30K*

The binding of *Gly m Bd 30K* and α -casein to the α -casein-specific 1D5 mAb was characterized by SPR. Table 1 includes the kinetics and

affinity constants obtained. We found a 74 times greater affinity constant (K_A) for 1D5 with α -casein than with *Gly m Bd 30K*, which was less than two orders of magnitude. The analysis of the k_{ass} and k_{diss} values for both complexes showed that the difference in K_A was mainly due to the

Table 1

Kinetic and equilibrium constants for complexes formed by bovine α -casein and *Gly m Bd 30K* with the 1D5 mAb, specific for α -casein. k_{ass} : Association constant; k_{diss} : Dissociation constant; K_A : Affinity constant; K_D : Equilibrium Dissociation constant.

	k_{ass} ($\text{M}^{-1} \text{s}^{-1}$)	k_{diss} (s^{-1})	K_A (M^{-1})	K_D nM
Bovine α -casein	$1.05 \times 10^4 \pm 0.14 \times 10^4$	$8.88 \times 10^{-4} \pm 0.22 \times 10^{-4}$	$1.18 \times 10^7 \pm 0.13 \times 10^7$	84,4
P34	$3.17 \times 10^3 \pm 0.30 \times 10^3$	$1.97 \times 10^{-2} \pm 0.07 \times 10^{-2}$	$1.60 \times 10^5 \pm 0.16 \times 10^5$	6250

difference in k_{diss} : the 1D5- *Gly m Bd 30K* complex dissociated 22 times faster than the one formed by α -casein.

We also observed that *Gly m Bd 30K* associated 3.3 times slower to 1D5 than α -casein. Recent studies (Schreiber et al., 2010) about the kinetics of protein-protein interactions argued that hotspot residues affecting dissociation rates are both charged and uncharged nature, while hotspots for the association are almost always charged residues. Then, according to our comparative kinetic study, we suggest that the difference in the Ag-Ab binding does not seem to be merely electrostatic.

3.4. Detection of cross-reactive T epitopes in *Gly m Bd 30K* peptides

The sensitization of mice (Fig. 3A) induced intestinal inflammation, milk-specific IgE, and milk-specific T cells (Candreva et al., 2016, 2015; Curciarello et al., 2014; Smaldini et al., 2012). Splenocytes were cultured with Nt and Ct, and we found that milk-specific CD4⁺ cells proliferated mainly with the Ct-domain (CFSE^{low} gate) (Fig. 3B). Hence, cells were then exposed to *Gly m Bd 30K* overlapping peptides (209–241 aa) (Supplementary Table 2) for the detection of peptide-containing T epitopes, and we quantified secreted cytokines (Fig. 3C). As controls, we stimulated cells with anti-CD3/anti-CD28, CMP or medium. We found high levels of IL-5, IL-13, and IFN- γ with the positive control compared with cells exposed to the medium. Cells exposed to CMP induced the secretion of IL-5 and IL-13, while the overlapping peptides generated different patterns of cytokine secretion. In particular, peptides 2 (P2) and 3 (P3) seemed to be the most effective for T-cell recall. We selected P3 (214–225 aa) for immunotherapy because it promoted the highest secretion of IL-5 and IL-13.

To study the presence of B epitopes on P3, we carried out an ELISA with casein-specific mAbs, and skin test in CMP-sensitized mice. As shown in Fig. 3D and E, 4C3 revealed a β -casein cross-reactive IgG epitope on P3, while the negative cutaneous test ruled out the presence of at least 2 IgE epitopes. As controls, sensitized animals showed a positive skin test with CMP, whereas unsensitized animals that received only CMP showed negative skin tests either with CMP or P3. These findings confirmed the presence of B and T epitopes on P3 and discarded the capacity to activate mast cells. We next modelled the structure of *Gly m Bd 30K* (Fig. 3F) with the secondary structure (I and II) and surface charge (IV), and we located P3 (III) on the surface of the molecule. This region was dominated by α -helix structure with positive and negative charged areas. The analysis of the secondary structure of P3 in the whole *Gly m Bd 30K* rendered some α -helix and β -sheet (Fig. S3A); the sequence alignment of P3 and epitopes on caseins showed amino acid similarities with major and minor B epitopes described (Chatchatee et al., 2003) (Fig. S3B).

3.5. Prevention of allergy sensitization

The presence of B and T cross-reactive epitopes in the soy peptide prompted us to investigate the impact of the mucosal administration of P3 in the prevention of IgE sensitization (Fig. 4A). As controls, a group of naïve mice received P3_{ig} and then CMP (unsensitized), other mice received PBS, and then CT + CMP (sensitized), other received *Gly m Bd 30K*_{sl} and then CT + CMP or P3_{ig} and then CT + CMP and finally, mice

received CMP_{sl/ig} and were further sensitized. Oral challenges with CMP were done to evidence hypersensitivity reactions. The clinical score was determined 30 min after the oral challenge. We observed the highest clinical score in sensitized mice that previously received PBS, no signs in unsensitized mice that were administered P3 and then CMP, thus indicating that P3 or CMP do not sensitize the mice, low clinical scores in mice pre-treated with CMP_{sl/ig} and then sensitized, and finally, intermediate but significant low scores were observed in sensitized animals that were pre-treated with P3_{ig} (Fig. 4B). These findings indicate that symptoms can be alleviated in sensitized mice that were pre-treated intragastrically with P3.

The cutaneous test was performed with CMP, and we observed a significant reduction of inflammation in mice pre-treated with P3_{ig}, *Gly m Bd 30K*_{sl} or CMP_{ig/sl} compared with sensitized mice (Fig. 4C). As a control, sensitized mice that were given PBS_{ig/sl} showed a positive skin test.

Based on these findings, we explored the humoral and cellular immune responses, and we found that the administration of P3_{ig} suppressed IgE ($p < 0.01$) with the induction of CMP-specific IgG (IgG1 and IgG2a). Animals that received CMP_{ig/sl} or *Gly m Bd 30K*_{sl} significantly reduced IgE secretion, whereas no significant induction of IgG1 or IgG2a was observed. Finally, spleen cells from mice of the different groups were cultured with CMP, and we observed no secretion of IL-5 or IFN- γ in control mice, while IL-5 was significantly elevated in supernatants of cells from sensitized animals that were pre-treated with PBS. Interestingly, IL-5 production was inhibited in P3-pre-treated animals that were further sensitized ($p < 0.0001$). Furthermore, pre-treatment of mice with CMP_{sl/ig} or *Gly m Bd 30K*_{sl} significantly reduced the production of IL-5 (Fig. 4E). IFN- γ was elevated in supernatants of cells from mice pre-treated with P3_{ig}, although not reaching statistical significance.

3.6. In silico prediction of MHC binding

To gain more insight into the binding capacity of P3 to HLA alleles, we used different bioinformatic prediction tools. We found that the minimum core sequence NKIQDKVTIDGY containing B and T epitopes of P3 fitted into the epitope-binding groove of several frequent HLA class I and class II alleles. Table 2 only includes alleles with the lowest percentile rank or highest estimated affinity binding.

4. Discussion

A promising strategy to reverse or prevent IgE-mediated allergic diseases considers the use of peptides. The efficacy of the early administration of peanut and egg allergens in the diet in at-risk infants has been recently demonstrated to protect them from the onset of food allergy (Du Toit et al., 2016). However, the use of whole allergens could potentially promote sensitization and anaphylaxis overtime. In this respect, the administration of peptides with negligible IgE cross-linking capacity constitutes a safer option.

There is no report showing that cross-reactive allergens may be used for prevention. Our findings support novel evidence for a preventive procedure using a short peptide containing B and T epitopes derived from *Gly m Bd 30K*. We identified a domain (Nt) containing most of the IgG and IgE sequential epitopes, which in silico matched with the sequences of casein epitopes and demonstrated to be on the surface of the 3D modelling of the allergen. Also, our analysis revealed that cross-reactive B epitopes are mainly composed of α -helix secondary structure, and IgE epitopes could be conformational, the latter being critical to define the natural history of food allergy with age (Chatchatee et al., 2003). The primary sequences for IgE epitopes were coincident with some IgG epitopes, which is crucial for immune intervention focused on inducing blocking IgG antibodies using peptidic component-containing these epitopes. The Ct domain of *Gly m Bd 30K* showed low IgE-reactive and contained a 12-amino acid peptide (P3) with an immunodominant cross-reactive T epitope. Interestingly, this peptide had a

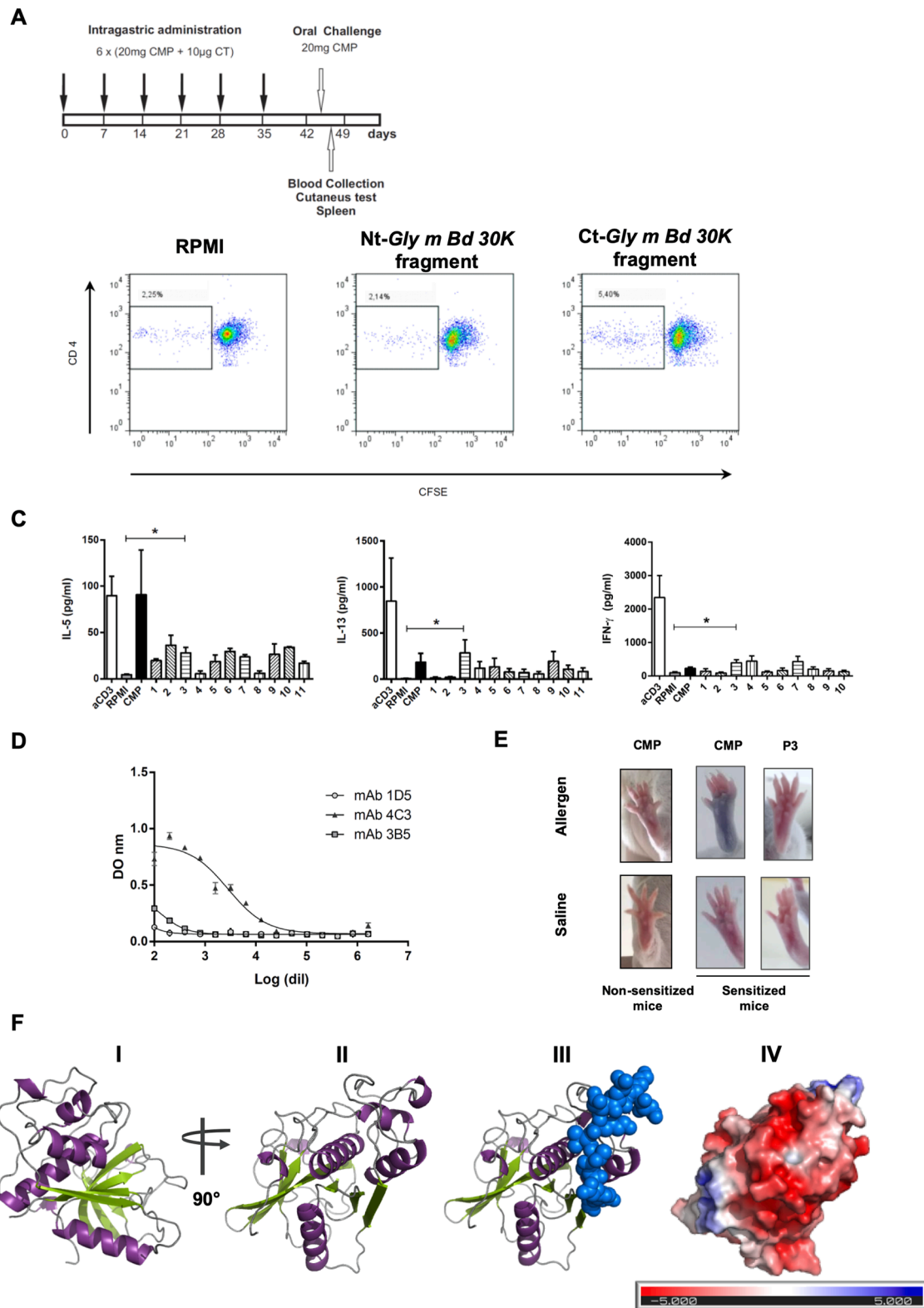
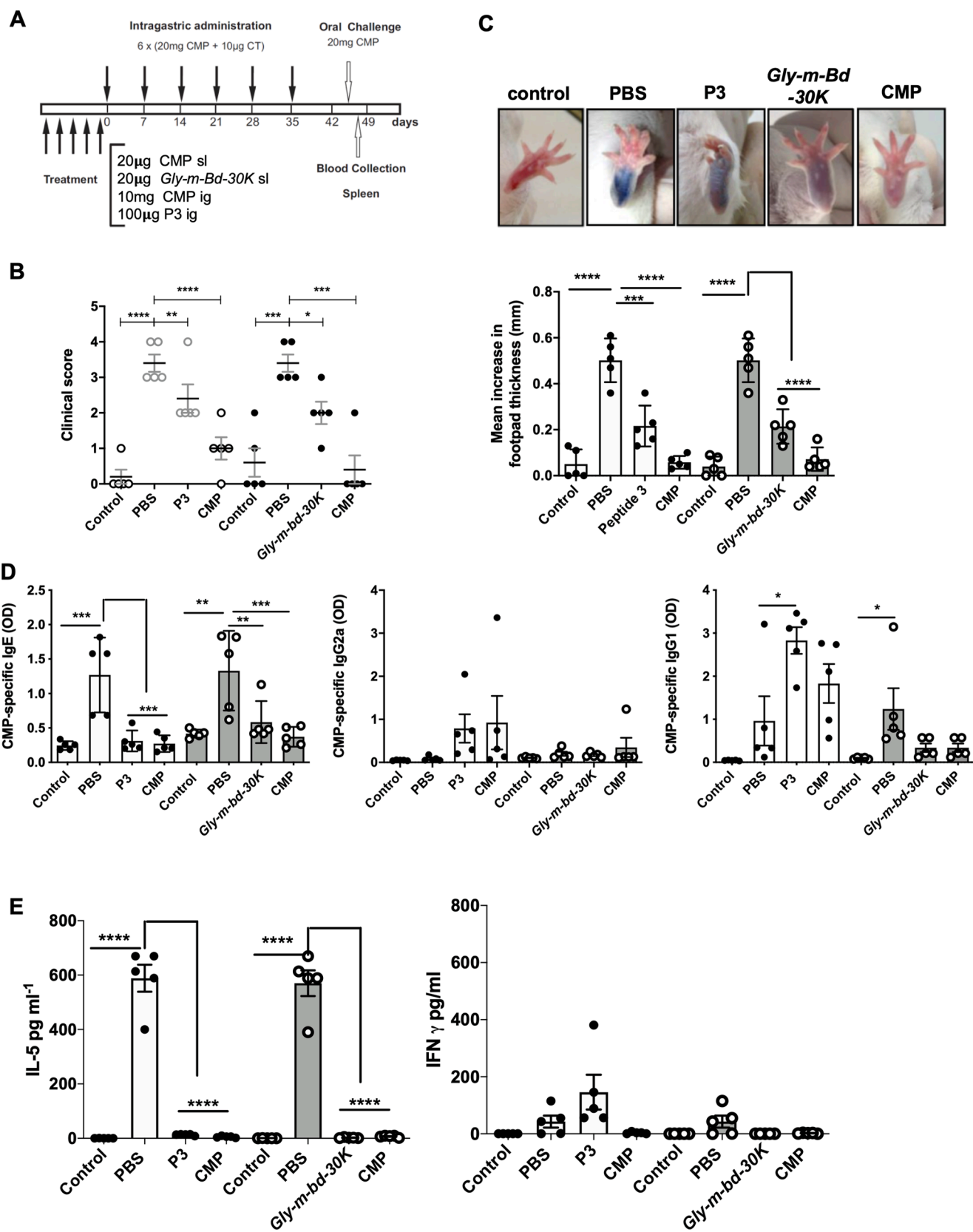


Fig. 3. *In vivo* and *in vitro* epitope mapping of synthetic peptides of *Gly m Bd 30K*. **A**, Schematic drawing of the experimental protocol for CMP sensitization of mice. **B**, CFSE analysis by flow cytometry to evaluate T lymphocyte proliferation. Splenocytes of CMP-allergic mice were incubated with recombinant Nt and Ct fragments from *Gly m Bd 30K* for 72 h and stained with anti-CD4. **C**, T epitope mapping with splenocytes from CMP-sensitized mice incubated with eleven soluble overlapping peptides spanning the Ct domain of *Gly m Bd 30K*, RPMI as a negative control, and CMP and anti-CD3 as positive controls (unpaired *t*-test analysis with $*p < 0.05$). **D**, Titration indirect ELISA of P3 (1 µg) immobilized on microtiter plates and then incubated with the different casein-specific mAbs. **E**, Skin test of sensitized mice with CMP or P3. Non-sensitized mice were included as control. **F**, Ribbon diagram of *Gly m Bd 30K* (I–III) with P3 residues in light blue and surface charge of the 3D-modelled *Gly m Bd 30K*(IV). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).



(caption on next page)

Fig. 4. Use of the P3 soy peptide for prevention in an IgE-mediated food allergy mouse model. A, Schematic drawing of the experimental protocol (n = 5 mice/group). B, Clinical signs scored 30 min following the intragastric challenge with CMP. Control group corresponds to unsensitized mice that received only CMP and were challenged with CMP. C, Cutaneous test in sensitized mice. Animals were injected into the pad of either rear foot with PBS or CMP. The presence of blue colour in the skin within minutes was considered a positive cutaneous test (left panel). Control PBS corresponds to the contralateral pad injected with PBS. The thickness of the feet was measured with a calliper (right panel). D, Humoral immune response specific to CMP. Specific IgE (EAST), specific IgG1 and specific IgG2a (ELISA) were analysed in sera from sensitized or unsensitized mice (control group). E, Systemic cytokines secreted by spleen cells stimulated with CMP. Levels of IL-5 and IFN- γ were quantified (ELISA) in supernatants of spleen cells isolated from sensitized or unsensitized mice (control group). In all graphs, the white points and bars correspond to the intragastric protocol, and the grey points and bars correspond to sublingual protocol. Statistical significant difference with one-way Anova followed by Bonferroni's test: ****p < 0.0001, ***p < 0.001, **p < 0.005, *p < 0.01. CMP, cow's milk proteins; P3, peptide #3 of the Ct domain of *Gly m Bd 30K*; CT, cholera toxin. Mouse groups: Control: untreated and unsensitized mice; PBS: treated with PBS and sensitized; P3: treated with P3 and sensitized; *Gly m Bd 30K*: treated with *Gly m Bd 30K* and sensitized; CMP: treated with CMP and sensitized. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Prediction of HLA binding affinity of P3. The table shows the results obtained by analysing the sequence of the P3 with Epitope Analysis Resource tools to predict the binding of peptide 3 (NKIQDKVTIDGY) to the MHC I and II alleles.

Allele MHC II BINDING				Peptide	Percentile rank
HLA-DRB1*03:01				NKIQDKVTIDGYETL	6.14
HLA-DRB1*08:02				CKANKIQDKVTIDGY	8.04
HLA-DRB3*01:01				NKIQDKVTIDGYETL	9.71
HLA-DRB1*08:02				KANKIQDKVTIDGYE	11
HLA-DRB1*08:02				ANKIQDKVTIDGYET	10.79
HLA-DRB1*08:02				NKIQDKVTIDGYETL	12.99
HLA-DRB1*03:01				CKANKIQDKVTIDGY	17.15
HLA-DRB1*03:01				KANKIQDKVTIDGYE	17.15
HLA-DRB1*03:01				ANKIQDKVTIDGYET	17.15
HLA-DPA1*01:03/DPB1*02:01				NKIQDKVTIDGYETL	19.82
HLA-DQA1*01:01/DQB1*05:01				NKIQDKVTIDGYETL	26.47
HLA-DRB3*01:01				CKANKIQDKVTIDGY	28.81
HLA-DRB3*01:01				KANKIQDKVTIDGYE	28.81
HLA-DRB3*01:01				ANKIQDKVTIDGYET	28.81

Allele MHC I BINDING	Start	End	Length	Peptide	Percentile_rank
HLA-A*01:01	3	12	10	IQDKVTIDGY	2.2
HLA-A*30:02	3	12	10	IQDKVTIDGY	2.35
HLA-B*15:01	3	12	10	IQDKVTIDGY	3.27
HLA-A*30:02	4	12	9	QDKVTIDGY	6.1
HLA-B*53:01	1	9	9	NKIQDKVTI	8.2
HLA-B*44:02	4	12	9	QDKVTIDGY	9.25
HLA-A*26:01	4	12	9	QDKVTIDGY	10.7
HLA-B*51:01	1	9	9	NKIQDKVTI	12.0
HLA-A*32:01	2	10	9	KIQDKVTID	12.0
HLA-B*44:02	1	9	9	NKIQDKVTI	12.5
HLA-A*26:01	3	12	10	IQDKVTIDGY	13.55
HLA-B*15:01	4	12	9	QDKVTIDGY	15.0
HLA-B*44:03	4	12	9	QDKVTIDGY	15.2
HLA-A*03:01	3	12	10	IQDKVTIDGY	16.5
HLA-A*01:01	4	12	9	QDKVTIDGY	17.5
HLA-A*30:01	2	11	10	KIQDKVTIDG	17.75
HLA-B*35:01	3	12	10	IQDKVTIDGY	18.5
HLA-A*23:01	1	9	9	NKIQDKVTI	18.5
HLA-A*11:01	3	12	10	IQDKVTIDGY	21.5

single IgG epitope, and its sequence NKIQDKVT, which is contained in the positive IgG spot #22, is rich in hydrophilic residues shared with casein epitopes (Fig. S3B) (Chatchatee et al., 2003). The negative skin test obtained with P3 test ruled out the possibility of IgE cross-linking on mast cells, which confers clinical safety. Also, and not less relevant, the mucosal administration of this peptide without any adjuvant induced milk-specific IgG1 and IgG2a in pre-treated mice. These antibodies may act as blocking antibodies for milk allergens, although this effect depends on the IgE/IgG ratio and the dose of the allergen. Besides, P3 contains a T epitope that promoted Th1 cell activation with the production of IFN- γ , thus conferring immunoregulatory properties. Finally, we predicted that this core aminoacidic sequence might bind different common HLA Class I and Class II alleles, rendering this peptide as appropriate for clinical use. These points rendered this peptide as a safer and promising candidate for milk allergy treatment, compared with full allergens.

Overall, this targeted *Gly m Bd 30K* peptide that combines the ability to stimulate T cells and to promote IgG production was used in a preventive protocol for experimental IgE-mediated food allergy. The intragastric administration of this peptide with negligible potential to cross-link IgE prevented the oral sensitization with milk allergens and controlled the allergic reaction upon exposure to CMP. This procedure suppressed the IgE secretion, inhibited a positive skin test, reduced the production of Th2 cytokines, and rendered a low clinical score following the oral administration of the allergen. Remarkably, the given peptide redirected the immune response with the induction of blocking milk-specific IgG1, IFN- γ secretion, and lack of inflammation. Our findings represent a proof-of-concept that immunochemical cross-reactivity between unrelated allergens can be exploited to prevent or likely reverse food allergy.

Several mouse models and human trials proved that dominant peptides are valid options for aeroallergen therapies (O'Hehir et al., 2016). Regarding food allergies, the use of peptides is mainly based on the administration of hydrolysates (Giavi et al., 2016). Giavi et al. performed the only clinical trial in a double-blind placebo-controlled randomized immunotherapy with a hydrolysate in egg-allergic patients. Specific IgG4 was induced, IgE was suppressed, tolerance to the hydrolysate was achieved, and no severe side effects were induced. However, no statistical difference with the placebo-treated group was reached upon the final open oral challenge. Several studies in animal models demonstrated the relative efficacy of hydrolysates and synthetic peptides (Lozano-Ojalvo & López-Fandiño, 2018), and the associated cellular immunity that accompanied the therapeutic effect to restore the immunological balance: tolerance induction (Yang et al., 2009) or Th1 immunity (Wai et al., 2016).

Other studies of food allergy used hydrolysed milk proteins to prevent the allergic sensitization with different efficacy depending on the grade of hydrolysis, the composition of the mix of peptides, and the concentration of dominant peptides (Lozano-Ojalvo & López-Fandiño, 2018). The preventive administration of β -lactoglobulin-derived single polypeptide showed a more robust immunomodulatory capacity than a mix of peptides to diminish further sensitization to native β -lactoglobulin (Meulenbroek et al., 2013). Notwithstanding, no study has been performed using cross-reactive peptides to prevent allergen sensitization.

5. Conclusion

We characterized major cross-reactive IgE and IgG epitopes of soy *Gly m Bd 30K*, and we identified a short peptide with an immunodominant T and IgG epitopes, not capable of cross-linking IgE on basophils and mast cells, and with stimulating Th1 cell capacity. This peptide displayed degeneracy for binding to frequent HLA alleles, thus rendering it as a potential clinically relevant safer alternative to full allergens to prevent IgE-mediated milk allergy. The use of this peptide could be considered a novel candidate for the new class of allergy treatment Synthetic Peptide Immuno-Regulatory Epitopes or SPIRE.

CRediT authorship contribution statement

Ángela María Candreva: Methodology, Software, Writing - review & editing. **Paola L. Smaldini:** Conceptualization, Methodology, Writing - review & editing. **Ana Cauerhff:** Methodology, Software, Writing - review & editing. **Silvana Petrucci:** Writing - review & editing. **Guillermo H. Docena:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Agencia Nacional de Promoción Científica y Tecnológica (grants PICT 2015-1648 and PICT 2015-1249), the Consejo Nacional de Investigaciones Científicas y Técnicas (grant PIP 2013-189) and the Universidad Nacional de La Plata (grant 11/X695) to GHD.

AMC is a fellow of CONICET; PS, AC, SP and GHD are researchers of CONICET.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2020.128926>.

References

- Benkert, P., Tosatto, S. C. E., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins*, 71(1), 261–277. <https://doi.org/10.1002/prot.21715>.
- Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., ... Zardecki, C. (2002). The protein data bank. *Acta Crystallographica. Section D, Biological Crystallography*, 58(Pt 6 No. 1), 899–907. <https://doi.org/10.1107/S0907444902003451>.
- Bhattacharya, D., Nowotny, J., Cao, R., & Cheng, J. (2016). 3Drefine: An interactive web server for efficient protein structure refinement. *Nucleic Acids Research*, 44(W1), W406–W409. <https://doi.org/10.1093/nar/gkw336>.
- Bouchaud, G., Braza, F., Chesné, J., Lair, D., Chen, K.-W., Rolland-Debord, C., ... Magnan, A. (2015). Prevention of allergic asthma through Der p 2 peptide vaccination. *Journal of Allergy and Clinical Immunology*. <https://doi.org/10.1016/j.jaci.2014.12.1938>.
- Bublin, M., & Breiteneder, H. (2020). Cross-reactivities of non-homologous allergens. *Allergy*, 75(5), 1019–1022. <https://doi.org/10.1111/all.14120>.
- Candreva, A. M., Smaldini, P. L., Curciarello, R., Fossati, C. A., Docena, G. H., & Petrucci, S. (2016). The major soybean allergen Gly m Bd 28K induces hypersensitivity reactions in mice sensitized to cow's milk proteins. *Journal of Agriculture and Food Chemistry*, 64(7). <https://doi.org/10.1021/acs.jafc.5b05623.s001>.
- Candreva, A. M., Smaldini, P. L., Curciarello, R., Cauerhff, A., Fossati, C. A., Docena, G. H., & Petrucci, S. (2015). Cross-reactivity between the soybean protein P34 and bovine caseins. *Allergy, Asthma & Immunology Research*, 7(1), 60–68.
- Chatchatee, P., Jarvinen, K.-M., Bardina, L., Vila, L., Beyer, K., & Sampson, H. A. (2003). Identification of IgE and IgG binding epitopes on beta- and kappa-casein in cow's milk allergic patients. *Clinical Experimental Allergy*, 31(8), 1256–1262. <https://doi.org/10.1046/j.1365-2222.2001.01167.x>.
- Cordle, C. (2004). Soy protein allergy: incidence and relative severity. *Journal of Nutrition*, 134(5), 1248S–1293S.
- Curciarello, R., Smaldini, P. L., Candreva, A. M., González, V., Parisi, G., Cauerhff, A., ... Docena, G. H. (2014). Targeting a cross-reactive Gly m 5 soy peptide as responsible for hypersensitivity reactions in a milk allergy mouse model. *PLoS One*, 9(1). <https://doi.org/10.1371/journal.pone.0082341>.
- Docena, G. H., Fernandez, R., Chirido, F. G., & Fossati, C. A. (1996). Identification of casein as the major allergenic and antigenic protein of cow's milk. *Allergy*, 51(6), 412–416. <http://www.ncbi.nlm.nih.gov/pubmed/8837665>.
- Drozdzki, A., Cole, C., Procter, J., & Barton, G. J. (2015). JPred4: A protein secondary structure prediction server. *Nucleic Acids Research*, 43(W1), W389–W394. <https://doi.org/10.1093/nar/gkv332>.
- Du Toit, G., Roberts, G., Sayre, P. H., Bahnson, H. T., Radulovic, S., Santos, A. F., ... Lack, G. (2016). Randomized trial of peanut consumption in infants at risk for peanut allergy. *New England Journal of Medicine*, 372(9), 803–813. <https://doi.org/10.1056/NEJMoa1414850>.

- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., ... Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Current Protocols in Protein Science*. <https://doi.org/10.1002/0471140864.ps0209s50>.
- Focke-Tejkl, M., Weber, M., Niespodziana, K., Neubauer, A., Huber, H., Henning, R., ... Valenta, R. (2015). Development and characterization of a recombinant, hypoallergenic, peptide-based vaccine for grass pollen allergy. *Journal of Allergy and Clinical Immunology*, 135(5), 1207–1217.e11.
- Frank, R. (2002). The SPOT-synthesis technique. *Journal of Immunological Methods*, 267(1), 13–26.
- Giavi, S., Visser, Y. M., Muraro, A., Lauener, R., Konstantinopoulos, A. P., Mercenier, A., ... Papadopoulos, N. G. (2016). Oral immunotherapy with low allergenic hydrolysed egg in egg allergic children. *Allergy*, 71(11), 1575–1584. <https://doi.org/10.1111/all.12905>.
- Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11), 1859–1865. <https://doi.org/10.1002/jcc.20945>.
- Kattan, J. D., Cocco, R. R., & Jarvinen, K. M. (2011). Milk and soy allergy. *Pediatric Clinics of North America*, 58(2), 407–426. <https://doi.org/10.1016/j.pcl.2011.02.005>.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The PyR2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858. <https://doi.org/10.1038/nprot.2015.053>.
- Klose, D. P., Wallace, B. A., & Janes, R. W. (2010). 2Struc: The secondary structure server. *Bioinformatics*, 26(20), 2624–2625. <https://doi.org/10.1093/bioinformatics/btq480>.
- Kozlowski, L. P. (2016). IPC – isoelectric point calculator. *Biology Direct*, 11. <https://doi.org/10.1186/s13062-016-0159-9>.
- Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., ... Karplus, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8: High-resolution homology modeling. *Proteins*, 77(Suppl. 9), 114–122. <https://doi.org/10.1002/prot.22570>.
- Lozano-Ojalvo, D., & López-Fandiño, R. (2018). Immunomodulating peptides for food allergy prevention and treatment. *Critical Reviews in Food Science and Nutrition*, 58(10), 1629–1649.
- Meulenbroek, L. A. P. M., van Esch, B. C. A. M., Hofman, G. A., den Hartog Jager, C. F., Nauta, A. J., Willemsen, L. E. M., ... Knipps, L. M. J. (2013). Oral treatment with β -lactoglobulin peptides prevents clinical symptoms in a mouse model for cow's milk allergy. *Pediatric Allergy and Immunology*, 24(7), 656–664. <https://doi.org/10.1111/pai.12120>.
- Möbs, C., Ipsen, H., Mayer, L., Slotsch, C., Petersen, A., Würtzen, P. A., ... Pfützner, W. (2012). Birch pollen immunotherapy results in long-term loss of Bet v 1-specific TH2 responses, transient TR1 activation, and synthesis of IgE-blocking antibodies. *Journal of Allergy and Clinical Immunology*, 130(5), 1108–1116.e6.
- Möbs, C., Slotsch, C., Löffler, H., Jakob, T., Hertl, M., & Pfützner, W. (2010). Birch pollen immunotherapy leads to differential induction of regulatory T cells and delayed helper T cell immune deviation. *Journal of Immunology*, 184(4), 2194–2203.
- O'Hehir, R. E., Prickett, S. R., & Rolland, J. M. (2016). T Cell epitope peptide therapy for allergic diseases. *Current Allergy and Asthma Reports*, 16(2), 1–9. <https://doi.org/10.1007/s11882-015-0587-0>.
- Oliveira, J. P. B., Candreva, A. M., Rizzo, G., Ramos, M. V., Oliveira, J. S., Oliveira, H. D., ... Freitas, C. D. T. (2019). Allergenicity reduction of cow's milk proteins using latex peptidases. *Food Chemistry*, 284(1), 245–253. <https://doi.org/10.1016/j.foodchem.2019.01.123>.
- Schreiber, G., Haran, G., & Zhou, H.-X. (2010). Fundamental aspects of protein-protein association kinetics schreiber. *Chemical Review*, 5(3), 379–390. <https://doi.org/10.2217/FON.09.6.Dendritic>.
- Smaldini, P., Curciarello, R., Candreva, A., Rey, M. A., Fossati, C. A., Petrucci, S., & Docena, G. H. (2012). In vivo evidence of cross-reactivity between cow's milk and soybean proteins in a mouse model of food allergy. *International Archives of Allergy and Immunology*, 158(4), 335–346. <https://doi.org/10.1159/000333562>.
- Smaldini, P. L., Delgado, M. L. O., Fossati, C. A., & Docena, G. H. (2015). Orally-induced intestinal CD4⁺ CD25⁺ FoxP3⁺ treg controlled undesired responses towards oral antigens and effectively dampened food allergic reactions. *PLoS ONE*, 10(10). <https://doi.org/10.1371/journal.pone.0141116>.
- Tang, M., & Mullins, R. (2017). Immunotherapeutic strategies in antiphospholipid syndrome. *Internal Medicine Journal*, 47(3), 250–256. <https://doi.org/10.1111/imj.13245>.
- Valenta, R., Campana, R., Focke-Tejkl, M., & Niederberger, V. (2016). Vaccine development for allergen-specific immunotherapy based on recombinant allergens and synthetic allergen peptides: Lessons from the past and novel mechanisms of action for the future. *Journal of Allergy and Clinical Immunology*, 137(2), 351–357. <https://doi.org/10.1016/j.jaci.2015.12.1299>.
- van Ree, R. (2004). Clinical importance of cross-reactivity in food allergy. *Current Opinion in Allergy and Clinical Immunology*, 4(3), 235–240. <http://www.ncbi.nlm.nih.gov/pubmed/15126948>.
- Vriend, G. (1990). WHAT IF: A molecular modeling and drug design program. *Journal of Molecular Graphics*, 8(1), 52–56.
- Wai, C. Y. Y., Leung, N. Y. H., Leung, P. S. C., & Chu, K. H. (2016). T cell epitope immunotherapy ameliorates allergic responses in a murine model of shrimp allergy. *Clinical and Experimental Allergy*, 46(3), 491–503. <https://doi.org/10.1111/cea.12684>.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and

- complexes. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/nar/gky427>.
- Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, 101(10), 2525–2534. <https://doi.org/10.1016/j.bpj.2011.10.024>.
- Yang, M., Yang, C., Nau, F., Pasco, M., Juneja, L. R., Okubo, T., & Mine, Y. (2009). Immunomodulatory effects of egg white enzymatic hydrolysates containing immunodominant epitopes in a BALB/c mouse model of egg allergy. *Journal of Agriculture and Food Chemistry*, 57(6), 2241–2248. <https://doi.org/10.1021/jf803372b>.

SUPPLEMENTARY MATERIAL

Materials and Methods

Binding kinetics analysis using the BIACORE assay

The kinetics of binding Ag-Ab was performed using surface plasmon resonance biosensor experiments on an IAsys Plus apparatus (Affinity Sensors, Saxon Hill, U.K.). Streptavidin (Sigma Aldrich, USA) (100µg/ml in Acetate buffer, pH 5.0) was covalently coupled to carboxymethyl dextran sensor chips in a final derivatized mass of 5.0 ng of protein per cuvette. A solution of 20µg/ml of the biotinylated mAb 1D5 (the ligand) in PBS buffer was allowed to react with the cuvette surface for 10 minutes, and the antibody excess was removed by adding 1M NaCl and 10mM Acetate buffer at pH 4.5. Binding reactions with the ligate (α -casein or *Gly-m-bd-30K*) were carried out in PBS, 0.05% Tween 20 at 25°C, with constant stirring. Data were collected at intervals of 0.3 sec. Ligand binding to the immobilized ligand was monitored at multiple ligate concentrations, ranging 10-fold below to at least 10-fold above preliminary estimates of equilibrium dissociation constants (K_D) for each reaction. Kinetic and Scatchard analysis was performed by using the FAST FIT software (Affinity Sensors, Saxon Hill, U.K.).

Bioinformatic analysis

-In silico prediction of cross-reactive epitopes: The amino acid similarity in the primary sequences between *Gly m bd 30K* (Sequence ID, gi | 195957142 from the GenBank database) and bovine casein peptides representing T and/or B epitopes was analysed by multiple sequence alignment. The amino acid sequence of 68 peptides from α s1- (gi | 162792: 214 aa), α s2- (gi | 27806963:222 aa), β - (gi | 162931: 224 aa) and κ -casein (gi | 1228078:190 aa) designed according to Immune Epitope Database and Analysis Resources IEDB were aligned with the *Gly m bd 30K* protein sequence by the ClustalW2 server (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) as described in Curciarello et al., (2014). The hits represent the accumulation of consensus amino-acid in every position of the sequence

analysed. Conserved amino-acids (identity) were assigned with a score of "3". Not identical amino acids (similar) were committed with a score of "2" and slightly similar amino acids were designated by a score of "1" (Curciarello et al., 2014).

-Secondary structure prediction: Secondary structure prediction from *Gly m bd 30K* primary sequence was performed using JPred 4 server (Drozdetskiy et al., 2015), LOMETS, PSSpred and I-Tasser suite from Zhang lab web page (<https://zhanglab.ccmb.med.umich.edu/publicdata/benchmark1/>) and the Phyre2 web portal (Kelley et al., 2015) (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>). A consensus prediction was made considering the methods and algorithms included in the JPred4 server: Jnet (final secondary structure prediction), a neural network prediction algorithm that works by applying multiple sequence alignments, alongside PSIBLAST (Jpsssm) and HMM (Jhmm) profiles, and other algorithms from the servers described above. Such consensus secondary structure prediction of each residue was calculated considering its occurrence in more than 60% of the generated predictions.

-Homology modelling: In order to visualize the residues contained in the epitopes recognized by the different antibodies, several models of the 3D structure of the *Gly m bd 30K* protein were generated. Homology modelling from the primary protein sequence of *Gly m bd 30K* was performed. The sequence of the protein *Gly m bd 30K* (gi: 195957142) was obtained from the GeneBank database and was used as a target sequence (without the signal peptide) as a query search in the Protein Data Bank (PDB) (Berman et al., 2002) using the Protein BLAST server. The first model was generated using the structure of procaricain from *Carica papaya* (PDB Id=1PCI; Chain A; gi | 2098464) as a template. This protein belongs to the papain family, shares 39% sequence similarity and the same enzymatic function (cystein-type endopeptidase activity). The first tentative model of *Gly m bd 30K* was generated with the Modeller 9v7 program (Eswar et al., 2007) using the structure of procaricain from *Carica papaya* (PDB

51 Id=1PCI; Chain A; gi | 2098464) as a template (PDB ID: 1PCI, Groves, M.R., Taylor, M.A.J.,
52 Scott, M., Cummings, N.J., Pickersgill R.W., Jenkins, J.A. The prosequence of procaricain
53 forms an alpha-helical domain that prevents access to the substrate-binding cleft. DOI:
54 10.2210/pdb1PCI/pdb). This protein belongs to the papain family, shares 39% sequence
55 similarity and the same enzymatic function (cystein-type endopeptidase activity). The output
56 from that server comprised a structure from residue Thr 31 to Glu 357. The loops of this
57 particular model were further modelled with the DOPE-based loop modelling protocol.

58 The Nt and Ct domains of the *Gly m bd 30K* protein were initially modelled with Swiss Protein
59 Workspace (Waterhouse et al., 2018) (<https://swissmodel.expasy.org/interactive>). The
60 sequence of the Nt polypeptide sent as input to Swiss Model server was defined from Gly 1 to
61 Ser 182; the structure of the precursor of a thermostable variant of papain (*Carica papaya*) at
62 2.6Å resolution (PDB ID: 3TNX. Roy, S., Choudhury, D., Dattagupta, J.K., Biswas, S.
63 Structure of the precursor of a thermostable variant of papain at 2.6 Å resolution. DOI:
64 10.2210/pdb3TNX/pdb) with 38% sequence similarity (55% positive similarity) was used as
65 template. However, the output from Swiss Model comprised the three-dimensional structure of
66 a fragment formed by Lys 32 to Ser182. The sequence sent as input to obtain the modelled
67 structure of the Ct polypeptide, was a fragment composed by Pro 135 to Leu 378, and the
68 template was the structure of papain fold-protein without the catalytic cysteines from seeds of
69 *Pachyrhizus erosus* at 2.0 Å resolution PDB ID: 2B1M. Zhang, M., Wei, Z., Chang, S. Crystal
70 structure of a papain-fold protein without the catalytic cysteines from seeds of *Pachyrhizus*
71 *erosus*. DOI: 10.2210/pdb2B1M/pdb. This protein has an 80% of sequence similarity (89% of
72 positives and 0% gaps) with *Gly m bd 30K* (from the residue Pro135 to Leu 378). Swiss Model
73 Server delivered an output of the same length as the input.

74 Additionally, the I-Tasser Suite was employed to model the whole *Gly m bd 30K* protein and
75 its Nt fragment (aa 1-182). A similar analysis was carried out using the Phyre2 server, obtaining
76 an output corresponding to the whole *Gly m bd 30K* protein and the Nt fragment, respectively.
77 All models were refined with the Modrefiner algorithm (Xu & Zhang, 2011), where the
78 conformational search is guided by a composite of physics-and knowledge-based force field.
79 The second round of refinement was performed with the 3D refine web server (Bhattacharya
80 et al., 2016) that employs iterative optimization of hydrogen bonding network combined with
81 atomic-level energy minimization, along with physics- and knowledge-based force fields. The
82 Yasara Energy Minimization server (Krieger et al., 2009)
83 (<http://www.yasara.org/minimizationserver.htm>) was also employed settled on knowledge-
84 based potentials.

85 The quality of the protein structure models was evaluated with the QMEAN server (Qualitative
86 Model Energy Analysis) (Benkert et al., 2008) (<https://swissmodel.expasy.org/qmean/>). The
87 server reported an output with a global and local quality score. For a complete assessment of
88 the protein models, the WHAT IF server (Build/check/repair model and Structure validation)
89 was employed (Vriend, 1990). Visualization and representations of all models were made with
90 Phymol 0.99rc6 program. The assignment of secondary structure from three-dimensional
91 coordinates was performed by means of the 2struc Secondary Structure Server (Klose et al.,
92 2010) (<https://2struc.cryst.bbk.ac.uk/twostruc>)

93 -*Estimation of the surface charges of the Gly m bd 30K allergen*: We employed the PBEQ
94 Solver server to estimate the electrostatic potentials on the solvent-accessible surface,
95 (<http://www.charmm-gui.org/?doc=input/pbeqsolver>) on the molecular model previously
96 obtained. The PBEQ Solver server was used to calculate and visualize the electrostatic potential
97 of the molecule by solving the Poisson-Boltzmann equation (Jo et al., 2008).

SUPPLEMENTARY TABLES

Supplementary Table 1. Peptide spot sequences. The table summarizes the data obtained from the immunoreactivity tests, the spots corresponding to figure 1. Their sequences are also shown.

Spot #	aa positions	Sequence	IgG/IgE-binding capacity
1	1-10	ISTHRSILDL	negative
2	5-20	SILDLDLTKFTTQKQ	negative
3	15-30	TTQKQVSSLFQLWKS	Pool#1 – Pool#3 - 1D5 - 4C3
4	25-40	QLWKSEHGRVYHNHE	negative
5	35-50	YHNHEEEAKRLEIFK	negative
6	45-60	LEIFKNNLNYIRDMN	4C3
7	55-70	IRDMNANRKS PHSHR	Pool#1 – Pool#2 - Pool#3 1D5 - 4C3 - 3B5
8	65-80	PHSHRLGLNKFADIT	Pool#1 – Pool#2 - Pool#3 - 1D5 - 4C3 - 3B5
9	75-90	FADITPQEFSSKYLQ	negative
10	85-100	KKYLQAPKDV SQQIK	Pool#2 - 1D5 - 4C3 - 3B5
11	95-110	SQQIKMANKKMKKEQ	Pool#1 – Pool#2 - Pool#3 - 1D5 - 4C3 - 3B5
12	105-120	MKKEQYSCDHPASW	negative
13	115-130	PPASWDWRKKGVITQ	Pool#1 – Pool#2 - Pool#3 - 1D5 - 4C3 - 3B5
14	125-140	GVITQVKYQGCGSG	negative
15	135-150	GCGSGWAFSATGAIE	negative
16	145-160	TGAIEAAHAIATGDL	negative
17	155-170	ATGDLVSLSEQLVD	negative
18	165-180	QELVDCEESEGCYN	negative
19	175-190	GCYNGWHYQSFEWVL	negative
20	185-200	FEWVLEHGGIATDDD	negative
21	195-210	ATDDDY PYRAKEGRC	negative
22	205-220	KEGRC KANKIQDKVT	1D5 - 4C3 - 3B5
23	215-230	QDKVTIDGYETLIMS	negative
24	225-240	TLIMSDESTESETE Q	negative
25	235-250	SETEQAFLSAILEQP	negative
26	245-260	ILEQPISVSIDAKDF	negative
27	255-270	DAKDFHLYTGGIYDG	negative
28	265-280	GIYDGENCTSPYGIN	negative
29	275-290	PYGINHFVLLVGYS	4C3
30	285-300	VGYSADGVVDYWIAK	negative
31	295-310	YWIAKNSWGEDWGED	negative
32	305-320	DWGEDGYIWIQRNTG	negative
33	315-330	QRNTGNLLGVCGMNY	negative
34	325-340	CGMNYFASYPTKEES	negative
34	335-350	TKEESETLV SARVKS	negative

36	345-360	ARVKGHRRVDHSPL	Pool#3 - 1D5 - 4C3 - 3B5
----	---------	----------------	-----------------------------

Gray shading: Positive peptides only with monoclonal antibodies (IgG epitopes).

Italic: Sequence represented by 12-aa soluble peptides.

Underlined sequence: P3

Supplementary Table 2. Aminoacidic sequences of overlapping peptides of Ct and charge analysis. pI and charge at pH 7.4 for peptides of the overlapping assay estimated with the Isoelectric Point Calculator (IPC). Bold: P3

Peptide #	Sequence	Iso-electric point pH	Net charge at pH 7,4
1	EGRCKANKIQDK	8,63	1,5
2	CKANKIQDKVTI	8,63	1,5
3	NKIQDKVTIDGY	6,39	-0,1
4	QDKVTIDGYETL	3,8	-2
5	VTIDGYETLIMS	3,29	-2
6	DGYETLIMSDES	3,09	-4
7	ETLIMSDESTES	3,18	-4
8	IMSDESTESETE	3,13	-5
9	DESTESETEQAF	3,13	-5
10	TESETEQAFLSA	3,36	-3
11	ETEQAFLSAILE	3,36	-3

Legends to Supplementary figures

Figure S1. Multiple sequence alignment between the complete sequence of *Gly m Bd 30K* and peptides containing B epitopes of bovine caseins. Bioinformatics analysis of the sequence alignment of *Gly m Bd 30K* with peptides of α -S1, α -S2, β and κ bovine caseins described as B or T epitopes (IEDB-Immune Epitope Data Base) showed as hit distribution of amino acid similitude per position along the sequence of *Gly m Bd 30K* (figure 2). Casein

peptides containing IgG or IgE epitopes are numbered in blue (left), whereas the length of the casein peptide is indicated in black (right); the same casein peptides aligned with *Gly m Bd 30K* are displayed showing as (*) conserved aminoacids (identity), (:) similar, but not identical aminoacids, and (.) slightly similar amino acids. The positions of the *Gly m Bd 30K* sequence with high hit values correlated with positive spots of the overlapping assay displayed in figure 1 A and B and is shadowed in red for Nt. The *Gly m Bd 30K* sequence with letters in red corresponds to the Nt domain (1–116 aa), and with letters in green corresponds to the Ct domain (208-360 aa). The box in the Ct domain contains the sequence analyzed as overlapping peptides for T epitopes; P3 is highlighted in black and aminoacids are numbered. The *Gly m Bd 30K* sequence underlined with dashed line denotes IgE recognition of epitopes in overlapping peptides, and aminoacids underlined with continuous line denotes IgG recognition.

Figure S2. Secondary structure assignment from different homology modeling structures.

The secondary structure was assigned using the Secondary Structure Server and the DSSP definition to the complete molecule of *Gly m Bd 30K*, Nt and Ct domains modeled with different servers. H corresponds to α -helix; S, to β -sheet and “-“, to disordered structure. Abbreviations used are: P34_I_Tasser: whole *Gly m Bd 30K* protein modeled with I-TASSER Server; N-term_I_Tasser: N-terminal fragment (residues 1 to 180) modeled with I-TASSER Server; P34_Phyre_2: whole *Gly m Bd 30K* modeled with Phyre2 Server; N-term_Phyre_2: N-terminal fragment (residues 1 to 180) modeled with Phyre2 Server; N-term_P34_SM: N-terminal fragment (residues 32 to 182) modeled with Swiss Model Workspace using the precursor of a thermostable variant of papain at 2.6 Å resolution (PDB Id: 3TNX); C-term_P34_SM: C-terminal fragment (residues 135-358) modeled with Swiss Model

147 Workspace using a papain-fold protein without the catalytic cysteines from seeds of
148 *Pachyrhizus erosus* at 2.0 Å resolution (PDB Id: 2B1M).

149

150 **Figure S3. Structural bioinformatics analysis of peptide 3. A.** Abbreviations and methods
151 applied in the prediction of the secondary structure of P3: H corresponds to α -helix, S, to β -
152 sheet; “=”, to disordered structure; P34_I_Tasser: whole *Gly m Bd 30K* using I-TASSER
153 Server. P34_Phyre_2: whole P34 protein using Phyre2 server. C-term_P34_SM: C-terminal
154 *Gly m Bd 30K* protein using Swiss Model server with 2B1M as template structure. **B.** Multiple
155 sequence alignments between peptide 3 of *Gly m Bd 30K* and peptides of α S1-bovine casein
156 described by Chatchatee et al. Analysis of the sequence alignment of the peptide 3 with B or T
157 epitopes in α S1-bovine casein. Symbols: “*” for conserved aminoacids (identity), “:” for
158 similar, but not identical amino acids, and “.” for slightly similar amino acids. The P3 sequence
159 underlined with dashed line denotes IgE recognition of epitopes in overlapping peptides, and
160 aminoacids underlined with continuous line denotes IgG recognition.

1 47 116

SecP34 1 ISTHRSILDLDLT~~TKFTTQKQVSSLFOLWKSEHGRVYHNHEEEAKRL~~~~EIFKNNLN~~~~YIRDMNANRKS~~~~SPHSRLGLNKFADITPQ~~~~EFSKKYLOAPKDV~~~~SQOIKMANKKMKKEQ~~~~YSCDHP~~~~PASWDWRKKGVITOVKY~~ 133

Bovpep# 1 -----AMEDIKQMEA----- 10
2 -----AQQKEP-----MIGV----- 10
3 -----AYF-----YP-----ELFRQ----- 10
4 -----DAPS-----FSDI-PN----- 10
5 -----DAY-----PSGAWYY----- 10
6 -----EKNVNELS-----KDI----- 10
7 -----ELAYF-----Y-----PELF----- 10
8 -----GYLEQL-----LRLKKYK-VPO----- 16
9 -----HSMK-----EGIHAAQQKEP-----MIGVNQ----- 20
10 -----IGVNOELAYFYPELF-RQFYQ----- 20
11 -----IHAQQKEPMI----- 10
12 -----IKQMEAEISIS----- 10
13 -----IQKE-----DVPSE-----R----- 10
14 -----ISSSEEIVPN----- 10
15 -----IVPNSAEERL----- 10
16 -----IVPNSVEQKH----- 10
17 -----K-EDV-P-----SERYL----- 10
18 -----KEDVPSEERYLG-----YLEQ-----LLRLK----- 20
19 -----KHPIKHQ-GL-----P-----KHIOKE-----DVPSE----- 10
20 -----KHQ-GL-----PQEV-----L----- 10
21 -----LPQEVLNENL----- 10
22 -----LRF-F-VAP-F-----PE----- 10
23 -----LRLKKYK-VPO----- 10
24 -----PNP-IGSEN-SE----- 10
25 -----PNSAEERLHS----- 10
26 -----PNSVEQKHIO----- 10
27 -----SEKTT-----MPLW----- 9
28 -----SESTEDQAMED-----IKQMEAE----- 18
29 -----TOYT-DAP-SFS----- 10
30 -----VEQKHIOKE-----DVPSE-----RY----- 16
31 -----VFG-----KEKVNEL----- 10
32 -----PNSVEQKHIOK-EDVPS-----ERYL----- 20
33 -----ENL-CS-----TF----- 7
34 -----E-KNRLNPLKKISQR-----YQKFA-LP-----QYLKT----- 26
35 -----KN-TMEHVSS-----SEESIISQETY----- 20
36 -----KTVYQHOKAMKPWIQPKTKVIPYVRY----- 26
37 -----LNPWD-----QVKRNAV-----P-----ITP-TLNR----- 20
38 -----SIISQET-YKQEK-----MAIN-----PSK----- 20
39 -----FPG-----P-----IPNSL-----PQNI----- 23
40 -----TVMFP-----PQS-VLSLSQ-SKVLPV----- 20

Bovpep# 1 :: *:::*
2 *:::* *:
3 : * * *:::
4 .:* *::* *:
5 : * *::* *:
6 : * *::* *:
7 : * *::* *:
8 : * *::* *:
9 : * *::* *:
10 : * *::* *:
11 : * *::* *:
12 : * *::* *:

```

# 13                                     ::**   * *.   *
# 14                                     ..  :*.*:
# 15                                     :   *.*:..  .::
# 16                                     *  *: *  *:**:
# 17                                     :.. *  .: **      **:   ::
# 18                                     *:::**   * *.
# 19                                     **:   ::
# 20                                     :.. *  .: **      *
# 21                                     *  *.*: **      ***.  *
# 22                                     *  *.*: **      ***.  : *
# 23                                     .*: **:*:  .**
# 24                                     *  :::  ::
# 25                                     *  *.*:..:::
# 26                                     *:::..:::
# 27                                     *:..:~::~*
# 28                                     :*:  .   : **
# 29                                     .*:..  :  ::   *::~::~*
# 30                                     **~::~*  *  *  *:.
# 31                                     :  ~::~**   * *.   :*
# 32                                     *:  ~::~~::~*
# 33                                     *:~::~~::~*  ~::~.
# 34                                     *:  *  .   :
# 35                                     *  **~::~~::~.
# 36                                     *  **~::~~::~.
# 37                                     *:  :  *  .   :~::~*  *
# 38                                     *  *  .   :  .  *  **~::~*
# 39                                     :  .   :  **
# 40                                     :  .   :  **
# 41                                     :  .   :  **

```

```

SecP34      134      QGGCGSGWAFSATGAIEAAHAIAATGDLVSLSEQELVDCVEESEGCGYNGWHYQSFQVLEHGGIATDDDPYRAKEGRCRKANKIQDKVTIDGYETLIMSDESTESETEQAFLSAILEQPIS 253
Bovpep# 37      L-----
# 42      -----AWYYVPL-----G---TQ-----
# 43      -----AY--FYPELFRQFYQL-----DAYP-----
# 44      -----D-----IGS-ESTEDQ-----
# 45      -----G--SEN--SEKTTM-----MPL-
# 46      -----G-----S-ESTE---DQAM-----
# 47      -----GVNQELAYFY-----HSMKEG-----IHAQ-----
# 48      -----KEGIHAQQKEPM-----
# 49      -----SESTEDQAMEDIKQMEAESI-----
# 50      -----SG--AWYYVPL-----G-----
# 51      -----SKD-----IGS-ESTE-----
# 52      LG-----YL-----
# 53      -----VF---GKEKVNELSKDI---G-----SESTE-----
# 54      -----FP--EVFGKEKVNELSKDI---G-----SESTE-----
# 55      ---CKE-----
# 56      -----TLN-REQLSTSEENSKKTVDM-----
# 57      -----ESQSLT 6
# 58      -----RINKK--IEKFQ---SEEQQQTEDE-----LQDKIH 26
# 59
# 60
# 61

```

```

# 43          . *  : :  * .  :      * **
# 44
# 45          *      * * ***** :
# 46          * . *  . : * . :      * :
# 47          *      * ***** : * :
# 48          * :  :  * *
# 49          :      * * *      * :
# 50          : *
# 51          * : .  *  * : . : . :
# 52          . *      . * : * :
# 53          : * :      * * *****
# 54          *      :
# 55          :      * : * . * : . : . :      *      . * * *
# 56          : *      * : * . * : . : . :      *      . * * *
# 57          * .
# 58
# 59          * : :  * *  * : * . : . :
# 60          . * . : :
# 61          : * : . *      * : :      * : * . : : * *      * : :

```

SecP34 254 VSIDAKDFHLYTGGIYDGENCTSPYGINHFVLLVGYGSDAGVDYWIAKNSWGEDWGEGDYIWIQRNTGNLLGVCGMNYFASYPTKEESETLV SARVKGHRRVDHSPL 360

```

Bovpep# 58  ITVDDKH---YQKALNE-----IN----- 20
# 60  LT-DVENLHL-----P-----LPLL----- 20
# 62  -----DIPNPIGSEN-----SEKTTMPLW----- 19
# 63  -----ELSKDIGSES----- 10
# 64  -----GYL-----EQLRLK----- 10
# 65  -----MIGVN--QELAY----- 10
# 66  -----NENLL--RFFVA----- 10
# 67  -----NENLLRFFVAP--FPEVFGKEK----- 20
# 68  -----NLLR--FF-----VAPF----- 10

```

```

# 58  : : * * .  *  . : :      * *
# 60  : : * . : : * *      *      : * *
# 62  ..      :  *
# 63  . . : * * . : .
# 64  * * :      : * * :
# 65  : : * .  : :  *
# 66  . * * *      . : * . :
# 67  . : : : : *      : * : * :
# 68  : *      : :      . : * :

```

Figure S2

```

          1          10          20          30          40          50          60          70
          ISTHRSILDL DLTKFTTQKQ VSSLFQLWKS EHGRVYHNHE EEAKRLEIFK>NNLNYIRD MN ANRKSPHSR LG
P34_Modeller: .....--HHHHH HHHHHHHHHH H-----HH HHHHHHHHHH HHHHHHHHHH H-----SS S-
P34_I_Tasser: .....-HHH--HHH HHHHHHHHHH HH-----HH HHHHHHHHHH HHHHHHHHHH HHH----- --
N-term_I_Tasser: .....-HHH-----HHH HHHHHHHHHH H-----HH HHHHHHHHHH HHHHHHHHHH HH----- --
P34_Phyre_2 .....H----- --HHHH-HHH HHHHHHHHHH H-----HH HHHHHHHHHH HHHHHHHHHH HHH----- --
N-term_Phyre_2: .....H----- -HHHH--HHH HHHHHHHHHH HH-----HH HHHHHHHHHH HHHHHHHHHH ----- --
N-term_P34_SM: ..... ---HHH HHHHHHHHHH H-----HH HHHHHHHHHH HHHHHHHHHH- -HHH----- --
C-term_P34_SM:
Cons.S.S.Pred: ..... -HHH HHHHHHHHHH HH-----HH HHHHHHHHHH HHHHHHHHHH H-----S S-

          80          90          100          110          120          130          140          150          160
          LNKFADIT PQEFSKKYLQ APKDVSQQIK MANKKMKKEQ YSCDHPPASW DWRKKG VITQ VKYQGGCGSG WAFSATGAIE AAHAIATGDL VS
P34_Modeller: --HHHH-- HHHHHHH-- -----S S-----H HHHHHHHHH HHHHHHH-- --
P34_I_Tasser: --HHHH-- HHHHHHH-- ----- -HHHH-----H HHHHHHHHH HHHHHHH-- --
N-term_I_Tasser: ----- HHHHHHH-- -----H HHHHHHHHH HHHHHHH-- --
P34_Phyre_2: --HHHH-- HHHHHHH-- -----S SHHHH-----H HHHHHHHHH HHHHHHH-- --
N-term_Phyre_2: --HHHH-- HHHHHHH-- -----H HHHHHHHHH HHHHH-- --
N-term_P34_SM: ----- HHHHHHH-- -HHHH----- -HHHH-----H HHHHHHHHH HHHHHHH-- --
C-term_P34_SM: -----S SHHH-----H HHHHHHHHH HHHHHHH-- --
Cons.S.S.Pred: ----- HHHHHHH-- ----- HHHHHHHHH HHHHH-- --

          170          180          190          200          210          220          230          240          250
          LSEQELVD CVEESEG CYN GWHYQSFEWV LEHGGIATDD DYPYRAKEGR CKANKIQDKV TIDGYETLIM SDESTESETE QAFLSAILEQ PI
P34_Modeller: ---HHHHH ----- --HHHHHHHH HH-----S
P34_I_Tasser: --HHHHHH H----- --HHHHHHHH HHH----- --SSSS-- ----- HHHHHHHHH--S
N-term_I_Tasser:
P34_Phyre_2: --HHHHHH H-----HHH- --HHHHHHHH HHH----- --SSSS-- -----H HHHHHHHHH--S
N-term_Phyre_2:
N-term_P34_SM:
C-term_P34_SM: --HHHHHH H----- --HHHHHHHH HH----- --HHH----- --SSSSS-- -----HHHHH HHHHHHHHH--S
Cons S.S.Pred: -----S- -----HHHHH HHH----- --SSSS-- -----HHH HHHHHHH--S
```

	260	270	280	290	300	310	320	330	340
	SVSIDAKD	FHLYTGGIYD	GENCTSPYGI	NHFVLLVGYG	SADGVDYWIA	KNSWGEDWGE	DGYIWIQRNT	GNLLGVCGMN	YFASYPTKEE SE
P34_Modeller:	SSS-----	HHHH-SSS--	-----	--SSSSSSSS	SS--SSSSSS	S-----	--SSSSS---	-----	--SS-----
P34_I_Tasser:	SSSS-----	HHH-----	-----	-SSSSSS---	SS--SSSSSS	S-----	--SSSSS---	----HHH--	---SSSS---
N-term_I_Tasser:									
P34_Phyre_2:	SSSS-----	---SSS-SS-	-----	-SSSSSSSSS	S---SSSSS	S-----	--SSSSS---	----HHH--	---SSSS---
N-term_Phyre_2:									
N-term_P34_SM:									
C-term_P34_SM:	SSSS-----	HHH---HH-	-----	-SSSSSSSSS	SS--SSSSSS	S-----	--SSSSS---	----HHH--	-SSSSSS---
Cons.S.S.Pred:	SSSSS---	----S--SSS	-----	-SSSSSSSSS	-----SSSS	S-----	--SSSSS---	-----	---SS-SS--

	350	360
	TLVSARVK	GHRRVDHSPL
P34_I_Tasser:	-----	-----
N-term_I_Tasser:		
P34_Phyre_2:	-----	-----
N-term_Phyre_2:		
N-term_P34_SM:		
C-term_P34_SM:		
Cons.S.S.Pred:	-----	-----

Heading 1

Figure S3

A

	210	220	230
P3 sequences:	EGR CKAN	<u>NKIQDKV</u> TIDGY	ETLIM
P34_I_Tasser:	---	-----	---SSSS--
P34_Phyre_2:	---	-----	---SSSS---
C-term_P34_SM:	---	--HH-----	---SSSS--

B

P3	<u>NKIQDKVTIDGY</u>	12
Major IgE/IgG	PEVFGKEKVN--	10
	:: .* :::	
P3	----- <u>NKIQDKVTIDGY</u>	12
Minor IgG	EEIVPNSVEQ-----	10
	*.:::	