


## RESEARCH ARTICLE

# Identification of cross-reactive B-cell epitopes between Bos d 9.0101 (*Bos Taurus*) and Gly m 5.0101 (*Glycine max*) by epitope mapping MALDI-TOF MS

Ángela María Candreva<sup>1,2</sup>, Mario Ferrer-Navarro<sup>3\*\*</sup>, Silvia Bronsoms<sup>4</sup>, Alejandra Quiroga<sup>1</sup>, Renata Curciarello<sup>2</sup>, Ana Cauerhff<sup>6</sup>, Silvana Petrucelli<sup>1</sup>, Guillermo Horacio Docena<sup>2\*</sup> and Sebastián Alejandro Trejo<sup>4,6</sup> 

<sup>1</sup> Centro de Investigación y Desarrollo en Criotecnología de Alimentos (CIDCA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata (UNLP), La Plata, Argentina

<sup>2</sup> Instituto de Estudios Inmunológicos y Fisiopatológicos (IIFP), CONICET, Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata (UNLP), La Plata, Argentina

<sup>3</sup> Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

<sup>4</sup> Servei de Proteomica i Biologia Estructural (SePBioEs), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

<sup>5</sup> Laboratorio de Biología Molecular y Celular, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (UBA), CONICET, Buenos Aires, Argentina

<sup>6</sup> Instituto Multidisciplinario de Biología Celular (IMBICE), CONICET, Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires (CIC), Universidad Nacional de La Plata (UNLP), La Plata, Argentina

Exposure to cow's milk constitutes one of the most common causes of food allergy. In addition, exposure to soy proteins has become relevant in a restricted proportion of milk allergic pediatric patients treated with soy formulae as a dairy substitute, because of the cross-allergenicity described between soy and milk proteins. We have previously identified several cross-reactive allergens between milk and soy that may explain this intolerance. The purpose of the present work was to identify epitopes in the purified  $\alpha$ S1-casein and the recombinant soy allergen Gly m 5.0101 (Gly m 5) using an  $\alpha$ -casein-specific monoclonal antibody (1D5 mAb) through two different approaches for epitope mapping, to understand cross-reactivity between milk and soy. The 1D5 mAb was immobilized onto magnetic beads, incubated with the peptide mixture previously obtained by enzymatic digestion of the allergens, and the captured peptides were identified by MALDI-TOF MS analysis. On a second approach, the peptide mixture was resolved by RP-HPLC and immunodominant peptides were identified by dot blot with the mAb. Finally, recognized peptides were sequenced by MALDI-TOF MS. This novel MS based approach led us to identify and characterize four peptides on  $\alpha$ -casein and three peptides on Gly m 5 with a common core motif. Information obtained from these cross-reactive epitopes allows us to gain valuable insight into the molecular mechanisms of cross-reactivity, to further develop new and more effective vaccines for food allergy.

Received: February 16, 2017

Revised: May 23, 2017

Accepted: June 2, 2017

**Keywords:**

Bos d 9.0101 / Cow's milk / Cross-reactivity / Epitope mapping / Gly m 5.0101 / MALDI-TOF MS / Soybean



Additional supporting information may be found in the online version of this article at the publisher's web-site

**Correspondence:** Dr. Sebastián Alejandro Trejo, IMBICE, Calle 526 entre 10 y 11, (1900) La Plata, Argentina

**E-mail:** sebatrejo@gmail.com

\*Additional corresponding author: Dr. Guillermo Horacio Docena  
E-mail: guidoc@biol.unlp.edu.ar

\*\*Current address: ISGlobal, Institute for Global Health, Hospital Clínic - Universitat de Barcelona, Barcelona, Spain

## Significance of the study

The strategy for epitope mapping described here combines immunoaffinity chromatography and MALDI-TOF MS using two different approaches that enable high resolution and sensitivity, fast analysis and easy automation. This strategy led us to demonstrate that a specific bovine  $\alpha$ -casein monoclonal antibody constitutes an important experimental tool for detecting epitopes in cross-reactive proteins of milk and soy, which resembles the *in vivo* cross-reactivity observed in patients with milk allergy. The characterization of these cross-reactive epitopes provides relevant information for a

better understanding of the mechanisms involved in the IgE-mediated food allergy observed in cow's milk allergic patients treated with soy formulas but never exposed to soy proteins. Moreover, the identification and characterization of cross-reactive epitopes in phylogenetically unrelated allergens may help to understand the immune recognition of antigenic determinants in allergens, which provides relevant information for the genetic manipulation of recombinant allergens that could be used in the development of novel vaccines for food allergy.

## 1 Introduction

The prevalence of allergic diseases is continuously rising and has become a health concern worldwide [1]. Although hundreds of molecules have been described as inducers of inappropriate immune responses in susceptible individuals, the molecular and structural bases of this phenomenon are not fully understood. This information is only available for certain allergens, and fragments of these molecules have been associated with the capability to activate innate and/or adaptive immunity mechanisms (adjuvant capacity and B/T cell epitopes, respectively) [2].

Milk, soy and their derivatives are included in the "the Big-8" group of foods that account for the 90% of all food-allergic reactions. In particular cow's milk is the major food allergen in many regions [3], including Argentina [4], and caseins have been characterized as the main allergens of milk [5]. Moreover, the high nutritional value, acceptability and low-cost formulations have made soy an attractive protein source as a dairy substitute for milk allergic patients older than 6 months of age. For this reason, soy-based formula is commonly used during avoidance diet. However, several publications reported clinical intolerance to this formula in a restricted proportion of IgE-mediated milk allergic patients, when starting the treatment [6–9].

Our group has previously identified four soy allergens that cross-react with bovine caseins [10–15] and the present work aims to test two high-throughput mass spectrometry techniques to identify cross-recognized B-cell epitopes. Up to now, epitope mapping has been performed using synthetic

peptides obtained by SPOT synthesis [16], recombinant peptides obtained using bacteriophage libraries (phage display) [17], or longer polypeptide fragments obtained in bacteria [18]. The main drawbacks of these techniques are their high cost and time-consuming implementation. Therefore, we here propose an alternative strategy to identify cross-reactive epitopes in a mix of peptides obtained by enzymatic digestion of the original proteins using high-performance technologies such as functional proteomics or *immunoproteomics*. The latter term describes the use of mass spectrometry (MS) to study interactions of peptides with the major histocompatibility complex or antibodies. Due to the high sensitivity, accuracy, and performance of MS techniques, MS has become an important tool to identify biological molecules, to study basic immunological mechanisms, and to develop and monitor prophylactic and therapeutic treatments [19–21]. Matrix-assisted laser desorption ionization (MALDI) time-of-flight (TOF) MS is a high-throughput technology for ligand screening due to its extremely high resolution and sensitivity (and thus low sample requirement), fast analysis and easy automation. To take advantage of these properties, we used a MALDI-TOF MS affinity-based screening called Epitope Mapping by MALDI-TOF (EMap-MALDI). We combined MS with immunoaffinity chromatography for epitope mapping of target antigens, with a novel methodology that facilitates data analysis [21–24].

In the present study we identified sequential cross-reactive B cell epitopes between purified bovine  $\alpha$ -casein and the recombinant soy protein Gly m 5.0101 using a monoclonal antibody specific for bovine  $\alpha$ -casein, the major allergen of bovine milk [25] and two mass spectrometry-based epitope mapping techniques. Our findings proved the utility of these approaches to study cross recognized allergens and shed light on understanding the cross-reactivity between phylogenetically unrelated milk and soy allergens and further clinical intolerances observed in milk allergic patients exposed to soy proteins. Furthermore, this information may support the development of a mucosal vaccine using the less immunogenic soy proteins or their fragments to mitigate cow's milk allergy.

**Colour Online:** See the article online to view Figs. 1–5 in colour.

**Abbreviations:** **1D5 mAb**, Specific  $\alpha$ -casein monoclonal antibody; **AEBSF**, 4-(2-Aminoethyl) benzene sulfonyl fluoride hydrochloride (serine protease inhibitor); **CMP**, Cow's milk proteins; **EF**, Eluted fractions; **EMap-MALDI**, Epitope Mapping by MALDI-TOF MS; **FT**, Flow-through or not-retained fractions; **Gly m 5**, Gly m 5.0101 recombinant soybean allergen; **PMF MALDI-TOF MS**, Peptide mass fingerprinting by MALDI-TOF MS; **SE**, Soy crude extract; **SP**, Soy proteins

## 2 Materials and methods

### 2.1 Materials

#### 2.1.1 Proteins

Bovine  $\alpha$ -,  $\beta$ -, and  $\kappa$ -casein, with a purity of approximately 85, 90, and 80% respectively, were purchased from Sigma Aldrich (St. Louis, USA). Soy proteins crude extract (SE) was prepared from *Glycine max* (L.) Merr. (variety: Asgrow) seeds as previously described [12]. The recombinant Gly m 5.0101 (Gly m 5) (GenBank: accession X17698) was cloned from cDNA library from immature soybean seeds [26]. In the text, bovine  $\alpha$ -casein corresponds to Bos d 9.0101 and  $\beta$ -conglycinin  $\alpha$ -subunit to Gly m 5.0101 according to the IUIS nomenclature.

#### 2.1.2 Patient selection and sera

Sera were obtained from six patients (ranged 9 months to 7 years old), diagnosed as milk allergic according to compatible history positive skin prick test (SPT) with commercial milk allergen (Allergo Pharma, Buenos Aires, Argentina), and elevated total serum IgE, and soy- and milk-specific IgE antibodies, performed by EAST according to Docena et al. [27]. Histamine phosphate (10 mg/ml) and physiologic saline solution were used as positive and negative controls in SPT, respectively. In addition, patients were reactive to milk in food challenge after milk deprivation for two weeks. Double-blind placebo-controlled food challenge is not currently performed in Argentina for diagnosis. Soy allergy was ruled out according to previous medical records. Protocols were approved by the Ethics Committee of the Children's Hospital of La Plata (Buenos Aires, Argentina) (# 389–2014) and patients or parents gave informed consent.

#### 2.1.3 Monoclonal antibody

Specific bovine  $\alpha$ -casein monoclonal antibody (1D5mAb) was obtained as described by Docena et al. [25]. This antibody was purified from ascites using affinity chromatography with Protein G (HiTrap Protein G HP, GE Healthcare Life Science, Uppsala, Sweden), and further biotinylated for the surface plasmon resonance (SPR) biosensor assay as described by Curciarello et al. [13].

#### 2.1.4 Reagents

Streptavidin was purchased from Sigma Aldrich (St. Louis, USA). Sequence-grade Trypsin and Glu-C were purchased from Sigma Aldrich (St. Louis, USA). SIGMAFAST Protease Inhibitor cocktail (Components: 2 mM AEBSF, 0.3  $\mu$ M Aprotinin, 130  $\mu$ M Bestatin, 1 mM EDTA, 14  $\mu$ M E-64, 1  $\mu$ M Leu-

peptin) was purchased from Sigma Aldrich (St. Louis, USA). Super-paramagnetic polystyrene beads of 4.5  $\mu$ m in diameter coated with monoclonal human anti-mouse IgG (Dynabeads Pan Mouse IgG) were purchased from Thermo Fisher Scientific (CA, USA). Protein sequences were obtained from GenBank (Bos d 9.0101/*Bos Taurus*, sequence ID: gi: 115646, and Gly m 5.0101/*Glycine max*, sequence ID: gi: 9967357).

### 2.2 Two-dimensional immunoblotting and protein identification

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of SE was performed as previously described [28]. Briefly, for the first dimension, 80  $\mu$ g SE were applied using immobilized pH gradient 7 cm strips (pH 3–10); and for the second dimension, the strips were run on 12.5% (w/v) SDS-PAGE gels. Gels were stained with Coomassie blue G250. 2D-PAGE was transferred to nitrocellulose membrane and blocked with 3% horse serum in PBS. Blocked membranes were incubated with a pool of patient sera (1:5, overnight at 4°C), followed by biotinylated anti-human IgE monoclonal antibody (1:3000, 4 h at 37°C, Vector Laboratories Inc., CA, USA), and horseradish peroxidase (HRP)-streptavidin conjugate (1:3000, 30 min at 37°C, Sigma-Aldrich, MO, USA). Luminol chemiluminescent substrate and exposure to X-ray film (Amersham Hyperfilm ECL, GE Healthcare Bio-Sciences Corp., USA) were used for spot visualization. The protein identification of immunoreactive spots was determined by PMF MALDI-TOF MS [29]. Search parameters were set as described elsewhere (MASCOT 2.6.0, database: SwissProt 2016\_10, enzyme: trypsin; variable modifications: carbamidomethyl (Cys), oxidation (Met), peptide mass tolerance: 35 ppm, taxonomy: other green plants and max missed cleavages: 2).

### 2.3 Binding kinetics analysis

Affinity analysis of 1D5 mAb was done using an IAsysPlus equipment (Affinity Sensors, Saxon Hill, Cambridge, U.K.) as previously described by Curciarello et al. [13]. Briefly, streptavidin (100  $\mu$ g/ml in acetate buffer, pH 5.0) was covalently coupled to carboxymethyl dextran sensor chips (Affinity Sensors, Saxon Hill, Cambridge, U.K.). The resulting amount of streptavidin coupled to the cuvette was 5 ng. Then, a solution of 20 mg/mL in saline buffer of the biotinylated 1D5 mAb (ligand) was added to the cuvette and the antibody excess was removed by adding 1 M NaCl and 10 mM acetate buffer at pH 4.5. Binding reactions with the ligate (caseins and Gly m 5) were carried out in PBS, 0.05% Tween-20 at 25°C, with constant stirring set up at 90%. Data were collected at intervals of 0.3 s. 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 every reaction. Kinetic

and scatchard analyses were performed with the FAST FIT software (Affinity Sensors, Saxon Hill, Cambridge, U.K.).

## 2.4 Enzymatic digestions

The casein peptides were obtained by digestion of 125  $\mu\text{g}$  bovine  $\alpha$ -casein with trypsin or Glu-C at a 20:1 ratio in 50 mM ammonia bicarbonate (previously, Cys residues were reduced and alkylated by DTT and iodoacetamide treatment, respectively). Reactions were performed overnight at 37°C and stopped with 2% trifluoroacetic acid (TFA). The digested peptides were evaluated by MALDI-TOF MS.

The Gly m 5 was run on a 12% (w/v) SDS-PAGE gels. Gels were stained with Coomassie blue G250 and the 70 kDa bands were excised manually from the gel and digested. In gel Glu-C digestion of Gly m 5 was carried out with 50 ng enzyme per band in 50 mM ammonia bicarbonate overnight at 37°C (previously, Cys residues were reduced and alkylated by DTT and iodoacetamide treatment, respectively), then digested peptides were recovered from the gel pieces by sonication. Digested peptides were analyzed by MALDI-TOF MS.

## 2.5 Dot Blot of RP-HPLC-separated peptides

Digested peptides (60  $\mu\text{g}$ ) of  $\alpha$ -casein were separated following Soriani protocol [21]. Briefly, the mixture of peptides was loaded onto a C18 Ascentis<sup>®</sup> Express column (2.7  $\mu\text{m}$  particle size, 160Å pore diameter, 15 cm long and 4.6 mm internal diameter) and separated by RP-HPLC Alliance<sup>®</sup> HPLC System from Waters (Milford, USA).

The flow rate was 0.5 mL/min and we used an ACN gradient with 0.1% TFA (0–10 min, (2% ACN), 40 min (40% ACN), 55 min (98% ACN), and 60 min (98% ACN)). 2-min fractions (1 mL) were collected, and digested with Glu-C. Fractions were vacuum-dried and re-dissolved in 10  $\mu\text{L}$  0.1% TFA. Then, a PVDF membrane was pretreated with 1% glutaraldehyde during 10 min and 1  $\mu\text{L}$  of each fraction was spotted onto the membrane, which was blocked with T-TBS containing 3% fetal bovine serum (16 h at 4°C). Thereafter, the membrane was incubated with 1D5 mAb (2 h at 37°C), followed by alkaline phosphatase-conjugated rabbit anti-mouse IgG (Sigma-Aldrich, MO, USA) for 1 h at 37°C, and finally developed with BCIP/NBT. The fractions with a positive result were analyzed with an ultrafleXtreme MALDI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany).

## 2.6 Immunocapture

The immunocapture of digested peptides from  $\alpha$ -casein or Gly m 5 was carried out following the protocol described by Soriani [21] with modifications. Pan mouse IgG dyn-

abeads (25  $\mu\text{L}$  of the suspension) were washed with PBS, re-suspended in 175  $\mu\text{L}$  PBS and incubated with 3  $\mu\text{g}$  1D5 mAb for 30 min at RT with gentle rotation. Then the beads were washed twice with PBS and incubated with the peptide mixture containing 1  $\mu\text{L}$  protease inhibitor cocktail solution for 2 h at RT with gentle rotation, to inhibit residual enzyme activity. Afterward, the flow-through (FT) of each sample was collected, and the beads were washed nine times with 1 mL PBS and a last wash with 1 mL PBS+0.5 M NaCl. Finally, the peptides were eluted with 50  $\mu\text{L}$  0.2% TFA.

## 2.7 MALDI TOF MS and MALDI TOF MS<sup>2</sup> analysis

All samples, the eluted fractions (EF), washes and FT were concentrated and desalted with C18 ZipTip (Millipore<sup>®</sup>, Merck, Darmstadt, Germany) following the manufacturer's instructions, and finally peptides were eluted from the C18-ZipTip with 3  $\mu\text{L}$  H<sub>2</sub>O:ACN (1:1 containing 0.1% TFA) [21, 30]. For MALDI-TOF target preparation, each sample was mixed (1:1) with alpha-cyano-4-hydroxycinnamic acid (HCCA) matrix (6 mg/mL in H<sub>2</sub>O:ACN at 2:1 proportion containing 0.1% TFA). In all cases, 1  $\mu\text{L}$  of each mix was spotted onto GroundSteel-384 plate using dried droplet method. Profile analyses were performed on an ultrafleXtrem MALDI-TOF mass spectrometer with FlexControl 3.4 build 85 (Bruker Daltonics, Bremen, Germany) data acquisition software. Mass spectra were acquired using the following settings: 600–4500 Da range, reflectron positive mode, ion source 1: 25 kV, ion source 2: 22.35 kV, lens: 7.5 kV, reflector voltage 1: 26.5 kV, reflector voltage 2: 13.4 kV, pulsed ion extraction of 120 ns. Ionization was achieved by irradiation with a solid state 355 nm pulsed laser (Smartbeam-II) operating at 2000 Hz (70% attenuator). All spectra were obtained randomly over the spot surface (4000 shots fired per spot) manually. Mass calibration was performed externally using the Peptide Calibration Standard (Bruker Daltonics: Angiotensin II, Angiotensin I, Substance P, Bombesin, ACTH clip 1–17, ACTH clip 18–39, Somatostatin 28). To obtain MALDI-TOF MS<sup>2</sup> spectra, mass spectrometer was used in Lift mode, ion source 1: 7.5 kV, ion source 2: 6.75 kV, lens: 3.6 kV, reflector voltage 1: 29.5 kV, reflector voltage 2: 14.3 kV, pulsed ion extraction of 100 ns with Smartbeam-II laser operating at 1000 Hz (30% attenuator). Data analysis was performed using FlexAnalysis 3.4 build 76 and BioTools 3.2 build 6.32 software programs (Bruker Daltonics, Bremen, Germany).

## 2.8 Peptide sequence analysis

A first round of sequence analysis of the peptides recognized by 1D5 mAb was performed by sequence alignment using ClustalW 2.0 (set parameters: gap open penalty = 11, gap extension penalty = 0.85, terminal gap penalties = 0.45, bonus score = 0) [31]. In order to simplify the sequence analysis a sequence logo was created from major homology region

using the WebLogo tool [32]. Finally, in order to extend the interpretation of the results, we performed a 3D alignment of these sequences by residues substitution onto the structure of  $\beta$ -conglycinin  $\alpha$ -subunit soybean (PDB ID: 1UIK) using the Maestro software (Schrödinger, LLC, NY, USA).

### 3 Results and discussion

#### 3.1 Soybean extracted Gly m 5.0101 is recognized by CMP-specific IgE

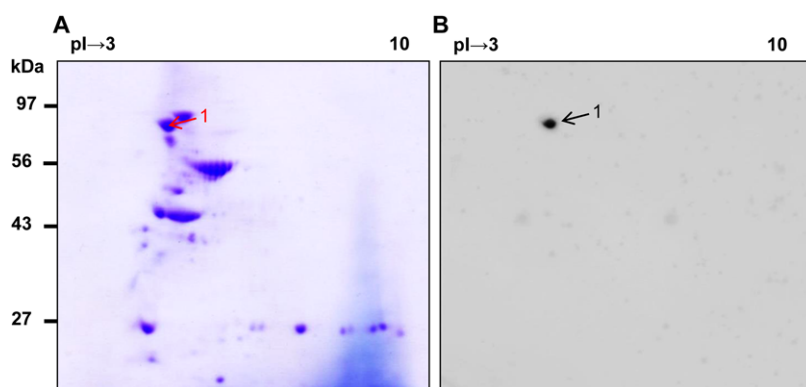
In order to identify the soybean allergenic proteins, soy crude extract was subjected to 2D-PAGE followed by IgE immunoblotting with sera from milk allergic patients. Gel staining revealed the presence of a wide number of proteins (Fig. 1A), whereas the immunoblotting using a pool of sera from patients containing CMP-specific IgE antibodies recognized a single spot (Fig. 1B). Spot 1 corresponds to a 70 kDa protein with an experimental  $pI$  value of 4.9. Spot 1 was excised and analyzed by PMF MALDI-TOF MS (Supporting Information Table 1). The  $\beta$ -conglycinin  $\alpha$ -subunit OS = Glycine max (UniProt id: GLCA\_SOYBN, Gly m 5.0101 allergen) was identified with a score value of 263 (score value is:  $-10 \cdot \log(P)$ ; where  $P$  is the probability that the observed match is a random event and scores greater than 79 are significant ( $p < 0.05$ )), an E-value of  $9.4e-23$ , matches: 24 peaks, an intensity coverage of 93.9%, a sequence coverage of 46.1%, and with the same  $pI$  and Mw values as those previously determined by 2D-PAGE.

#### 3.2 The 1D5 mAb that differentially reacts with bovine $\alpha$ -casein also recognizes soy proteins

The 1D5 mAb specific for  $\alpha$ -casein was used to identify the cross-reactive epitopes with  $\beta$ -conglycinin  $\alpha$ -subunit. The values of kinetic ( $k_{\text{ass}}$ , association rate constant and  $k_{\text{diss}}$ , dissociation rate constant) and affinity parameters ( $K_A$ , equilibrium association constant and  $K_D$ , equilibrium dissociation constant) for bovine caseins and Gly m 5 are shown in Table 1. The analysis of the kinetic parameters revealed that the bovine  $\alpha$ - and  $\beta$ -caseins and Gly m 5 have similar  $k_{\text{ass}}$  values. Also, the reactions between 1D5 mAb and either  $\alpha$ -casein or Gly m 5 showed higher  $K_A$  values than those observed for other caseins. Remarkably, 1D5 mAb revealed a slightly higher  $K_A$  for Gly m 5 than  $\alpha$ -casein, although the 1D5 mAb was obtained using bovine  $\alpha$ -casein as immunogen. Besides, 1D5 mAb also recognized  $\beta$ - and  $\kappa$ -caseins with lower  $k_{\text{ass}}$ . These findings reflect that 1D5 mAb, which is specific for  $\alpha$ -casein, also detects cross-reactive B cell epitopes in  $\alpha$ -casein and Gly m 5 that may be involved in the IgE recognition.

#### 3.3 The use of MS techniques allows to locate epitopes of cross-reactive allergens

Mapping of linear B cell epitopes in  $\alpha$ -casein and Gly m 5 was performed using the 1D5 mAb by MALDI-TOF MS (Supporting Information Fig. 1). We used two complementary approaches for epitope mapping starting out with the digested proteins: RP-HPLC followed by Dot blot, and immunocapture followed by MALDI-TOF MS.



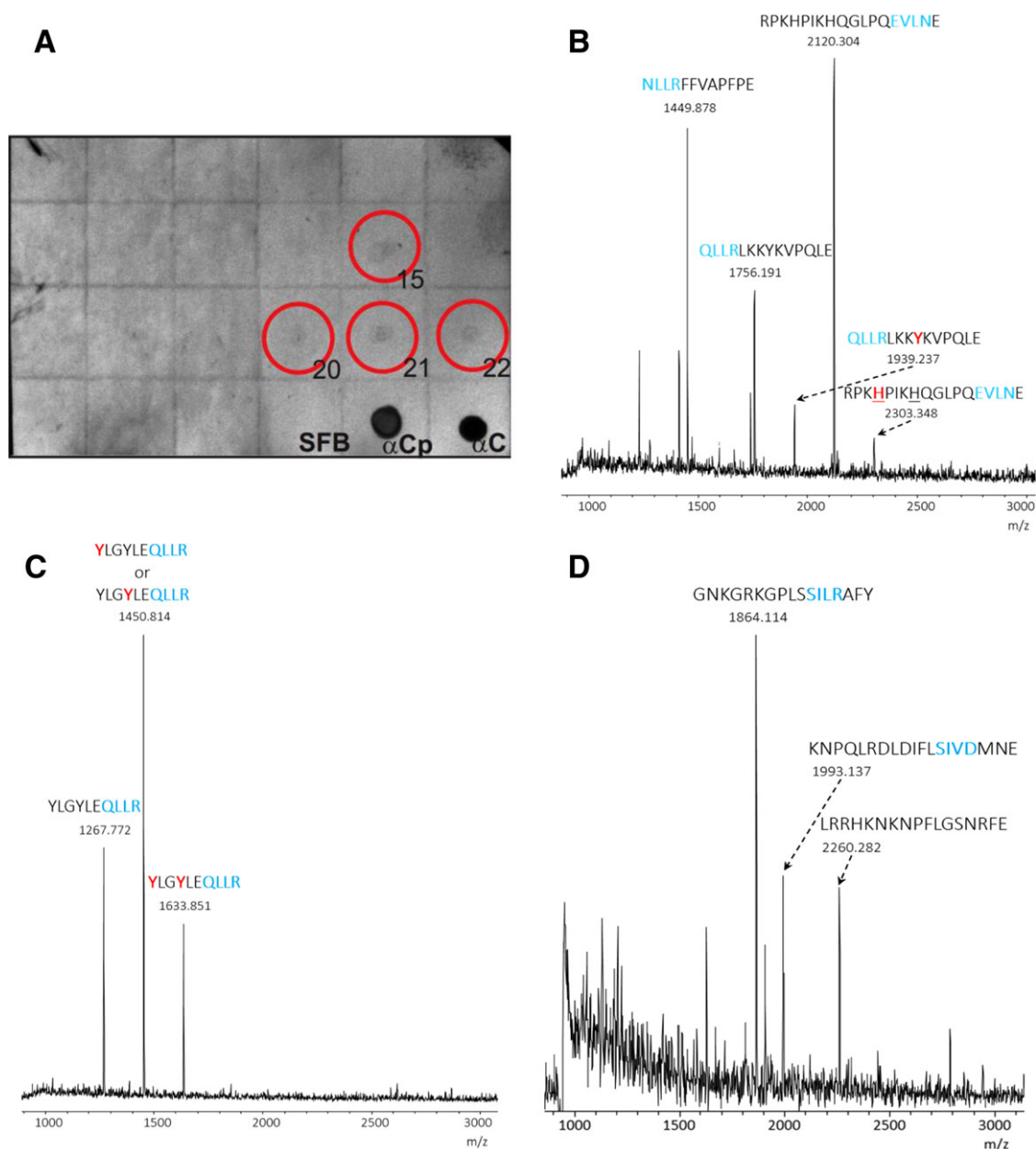
**Figure 1.** IgE immunoblotting of soy crude extract (SE) separated by two-dimensional electrophoresis (2D-PAGE). A. Proteins of the extract were separated using a pH gradient from 3.0 to 10.0 in the first dimension and 12.5% (w/v) gels in the second. 2D-PAGE gel stained with Coomassie Blue G250. B. 2D immunoblotting performed on a replicate gel using pool of human sera from milk allergic patients.

**Table 1.** Kinetic and equilibrium constants for reactions between 1D5 mAb and bovine  $\alpha$ ,  $\beta$  and  $\kappa$ -caseins and soybean Gly m 5.0101

	$k_{\text{ass}}$ ( $M^{-1} s^{-1}$ )	$k_{\text{diss}}$ ( $s^{-1}$ )	$K_A$ ( $M^{-1}$ )	$K_D$ (nM)
Bovine $\alpha$ -casein	$1.05 \times 10^4 \pm 0.14 \times 10^4$	$8.88 \times 10^{-4} \pm 0.22 \times 10^{-4}$	$1.18 \times 10^7 \pm 0.13 \times 10^7$	$84.4 \pm 14.8$
Bovine $\beta$ -casein	$5.36 \times 10^4 \pm 0.52 \times 10^4$	$2.22 \times 10^{-2} \pm 0.11 \times 10^{-2}$	$2.41 \times 10^6 \pm 0.26 \times 10^6$	$415 \pm 26.0$
Bovine $\kappa$ -casein	$3.75 \times 10^3 \pm 0.50 \times 10^3$	$1.97 \times 10^{-2} \pm 0.04 \times 10^{-2}$	$1.90 \times 10^5 \pm 0.25 \times 10^5$	$5250 \pm 250.0$
rGlym5	$1.89 \times 10^4 \pm 0.18 \times 10^4$	$9.79 \times 10^{-4} \pm 1.47 \times 10^{-4}$	$1.93 \times 10^7 \pm 0.22 \times 10^7$	$51.8 \pm 6.0$

$k_{\text{ass}}$ , Association rate constant;  $k_{\text{diss}}$ , Dissociation rate constant;  $K_A$ , Equilibrium association constant;  $K_D$ , Equilibrium dissociation constant.



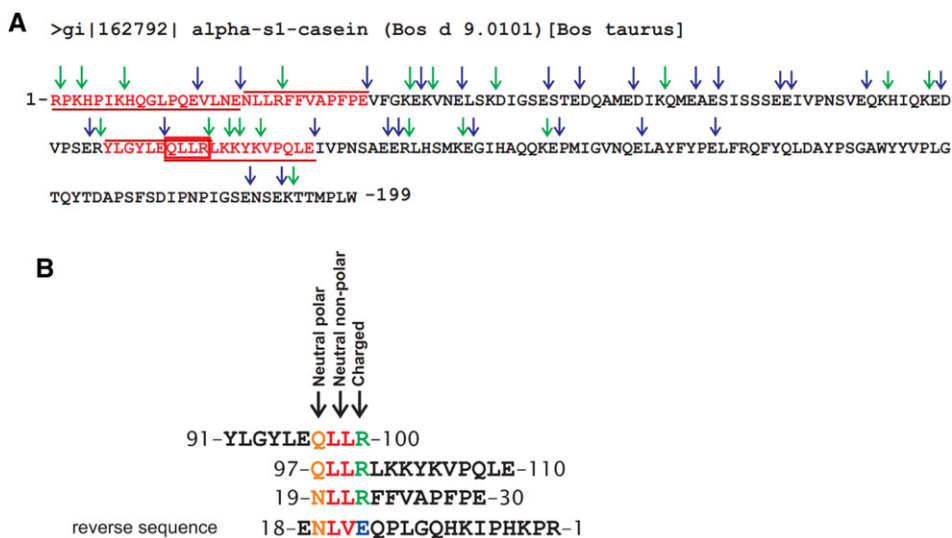


**Figure 2.** Peptide analysis. (A) Dot blot of  $\alpha$ -casein-derived peptides obtained with Glu-C digestion and separated by RP-HPLC, revealed with 1D5 mAb (the number of the chromatographic fractions are indicated). Circles show the fractions containing immunoreactive peptides. Control spots consisted of fetal bovine serum (SFB), digested ( $\alpha$ Cp) and not digested ( $\alpha$ C)  $\alpha$ -casein. (B and C) Peptides obtained from  $\alpha$ -casein digested with Glu-C (B) or trypsin (C), immunocaptured with 1D5 mAb, eluted and analyzed by MALDI-TOF MS (amino acid sequence and mass are indicated). (D) Peptides obtained from Gly m 5.0101 digested with Glu-C, immunocaptured with 1D5 mAb, eluted and analyzed by MALDI TOF MS (amino acid sequence and mass are indicated). The most conserved sequence motifs are shown in cyan; residues modified by AEBF are indicated in red. The experiments were performed in duplicate and the same results were obtained, demonstrating that the method is reproducible.

### 3.3.1 Epitope mapping of soluble $\alpha$ -casein digested peptides identifies four peptides

In a first approach, peptides obtained by digestion of  $\alpha$ -casein with Glu-C were separated by RP-HPLC, isolated and analyzed by dot-blot with 1D5 mAb (Fig. 2A). The positive spots

were analyzed by MALDI-TOF MS and contained peptides of 1449.8 Da (19-NLLRFFVAPFPE-30) in fraction #22, 1756.1 Da (97-QLLRLKKYKVPQLE-110) in fractions #20 and #21, and 2120.2 Da (1-RPKHPIKHQGLPQEVLE-18) in fractions #15 and #21.



**Figure 3.** Identification and analysis of peptides derived from  $\alpha$ -casein that were recognized with 1D5 mAb. (A) The amino acid sequence in one letter code of mature  $\alpha$ S1-casein is depicted with peptide-containing epitopes in red. The red square indicates the overlapping residues of peptides 1267.6 Da (91–100) and 1756.1 Da (97–110). Arrows show the cleavage sites for Glu-C (blue) and trypsin (green). (B) Sequence alignment of the immunoreactive  $\alpha$ -casein-derived peptides. Residues belonging to the epitope are indicated in colors.

In a second approach, we employed EMap-MALDI. Soluble peptides were captured with 1D5 mAb, further eluted and analyzed by MALDI TOF MS. When  $\alpha$ -casein was digested with Glu-C, the following peptides were identified: 1449.8 Da (19-NLLRFFVAPFPE-30), 1756.1 Da (97-QLLRLLKKYKVPQLE-110), and 2120.2 Da (1-RPKHPIKHQGLPQEVLE-18) (2B). Only one peptide was obtained when the sample was digested with trypsin: 1267.6 Da (91-YLGYLEQLLR-100) (Fig. 2C). MALDI TOF MS2 spectra of captured peptides were obtained (Supporting Information Figs. 2–10). All peptides captured with 1D5 belonged to  $\alpha$ S1-casein.

The 1267.6 Da  $\alpha$ -casein tryptic peptide was captured together with 1450.8 Da and 1633.9 Da peptides (Fig. 2C). Several studies have shown that either tyrosine (T) or histidine (H) residues can be modified by sulfonyl fluoride compounds [33, 34]. In particular, AEBF chemical modification (a sulfonyl fluoride derivative) produces a mass shift of 183 Da on T or H residues. Thus, 1450.8 Da and 1633.9 Da peptides observed in Fig. 2C correspond to the 1267.6 Da  $\alpha$ -casein tryptic peptide with one or two AEBF modifications in their T residues. A similar behavior was observed in peptides immunocaptured from  $\alpha$ -casein digested with Glu-C (Fig. 2B), which contained T, H or both amino acids. The 1939.2 Da peptide corresponded to the 1756.2 Da molecule with one AEBF modification, and the 2303.3 Da peptide corresponded to the 2120.3 Da molecule with one AEBF modification. The AEBF chemical modification of T or H residues did not seem to affect the interaction between the 1D5 mAb and the immunoreactive peptides, suggesting that these residues were not involved in the Ag-Ab interaction (Fig. 2B and C).

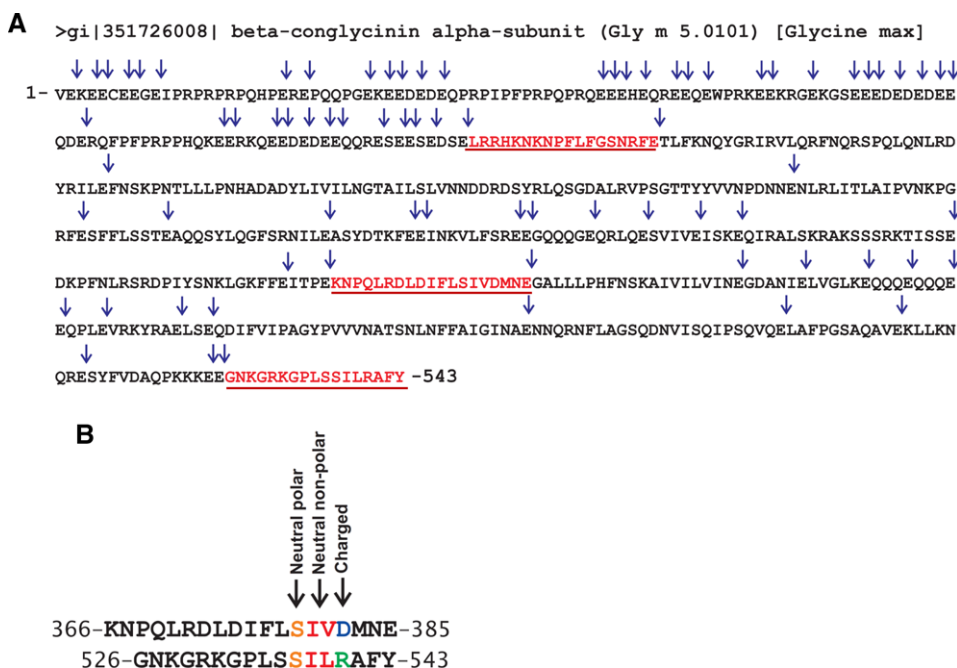
In conclusion, both experimental approaches identified the same peptides using 1D5 mAb, and all of them belonged to  $\alpha$ S1-casein. Figure 3A shows the distribution of the peptides recognized by 1D5 mAb, indicating that epitopes are located in the 1–30 and 91–110 regions of  $\alpha$ S1-casein. The comparison among these four peptides and the anal-

ysis of their sequences reveal that they share a common amino acid sequence and that only four amino acids in  $\alpha$ -casein are structurally relevant for the interaction with 1D5 (Fig. 3B).

Data obtained from this analysis point out that an environment dominated by polar charged amino acids, either positive (R-Arginine) or negative (E-Glutamic acid), and flanked by a neutral core is critical for antibody recognition (Fig. 3B). Remarkably, several groups found the same immunodominant epitopes on  $\alpha$ S1-casein. Chatchatee et al. used overlapping decapeptides representing the entire length of  $\alpha$ S1-casein and sera from 24 children with CMA. The IgE and IgG-binding regions identified included the sequences here mapped by EMap-MALDI (19-NLLRFFVAPFPE-30, 91-YLGYLEQLLR-100, 97-QLLRLLKKYKVPQLE-110) [35]. Cocco et al. identified hydrophobic amino acids F23, F24, F28, and L98 as critical for IgE-binding, using pooled sera from 15 CMA patients and also with eight individual sera [36]. These critical residues were also detected in our study using 1D5 mAb. Furthermore, Enomoto et al. found that the peptide 91-YLGYLEQLLRLLKKYKVPQLE-110 retained a high immunogenic capacity in experimental animal models [37], which coincides with peptides 91–100 and 97–110 detected in this study. In a different study, Spuerger et al. reported that peptides 19-NLLRFFVAPFPE-30 and 93-GYLEQL-98 reacted with 15 sera from CMA patients [38]. Finally, Elsayed et al., using synthetic peptides and 14 sera from CMA patients, determined that 16-LNENLLRFFVAPFPEVFGKE-35 peptide had an IgE epitope, whereas 1-RPKHPIKHQGLPQEVLE-18 peptide had an IgG epitope [39], also containing the motif described in this study for  $\alpha$ -casein.

### 3.3.2 1D5 mAb identifies B cell epitopes in recombinant Gly m 5.0101

Based on the reactivity of Gly m 5.0101 observed with a pool of sera from CMA patients and our previous findings



**Figure 4.** Identification and analysis of peptides derived from Gly m 5.0101 that were recognized with 1D5 mAb. (A) The amino acid sequence of Gly m 5.0101 is depicted with peptide-containing epitopes in red. Blue arrows indicate cleavage sites of Glu-C. (B) Sequence alignment of the immunoreactive Gly m 5-derived peptides. Residues belonging to the epitope are indicated in colors.

showing that this allergen shared common epitopes with bovine caseins [12, 13, 26], we followed the same approach applied to  $\alpha$ -casein to detect B cell epitopes of cross-reactivity in Gly m 5. When Gly m 5 was digested with Glu-C, three peptides of 1864.0 Da (527-GNKGRKGPLSSILRAFY-543), 1993.2 Da (367-KNPQLRDLIDIFLSIVDMNE-385), and 2260.2 Da (125-LRRHKKNKPFLFGSNRFE-142) were captured with 1D5 mAb (Fig. 2D), confirming the existence of cross-reactive epitopes between bovine  $\alpha$ -casein and soy  $\beta$ -conglycinin.

Figure 4A summarizes the position of peptides recognized by 1D5 mAb. Results indicate that one of the epitopes is located in the N-terminal cupin domain: 125–142, and the other two epitopes are located in the C-terminal cupin domain of Gly m 5.0101: 367–385 and 527–543.

Gly m 5.0101 has been described as one of the major allergens from soy. Fu et al. also reported epitope 125-LRRHKKNKPFLFGSNRFE-142 as one of the major IgG-binding region of the Gly m 5.0101 using a pool of plasma from ten piglets [40]. Curciarello et al. used overlapping 15-mer peptides and mapped cross-reactive B cell epitopes in the C-terminal region [14]. Noteworthy, in this study we have been able to map for the first time B cell epitopes in the entire molecule of Gly m 5.0101 using peptides in solution.

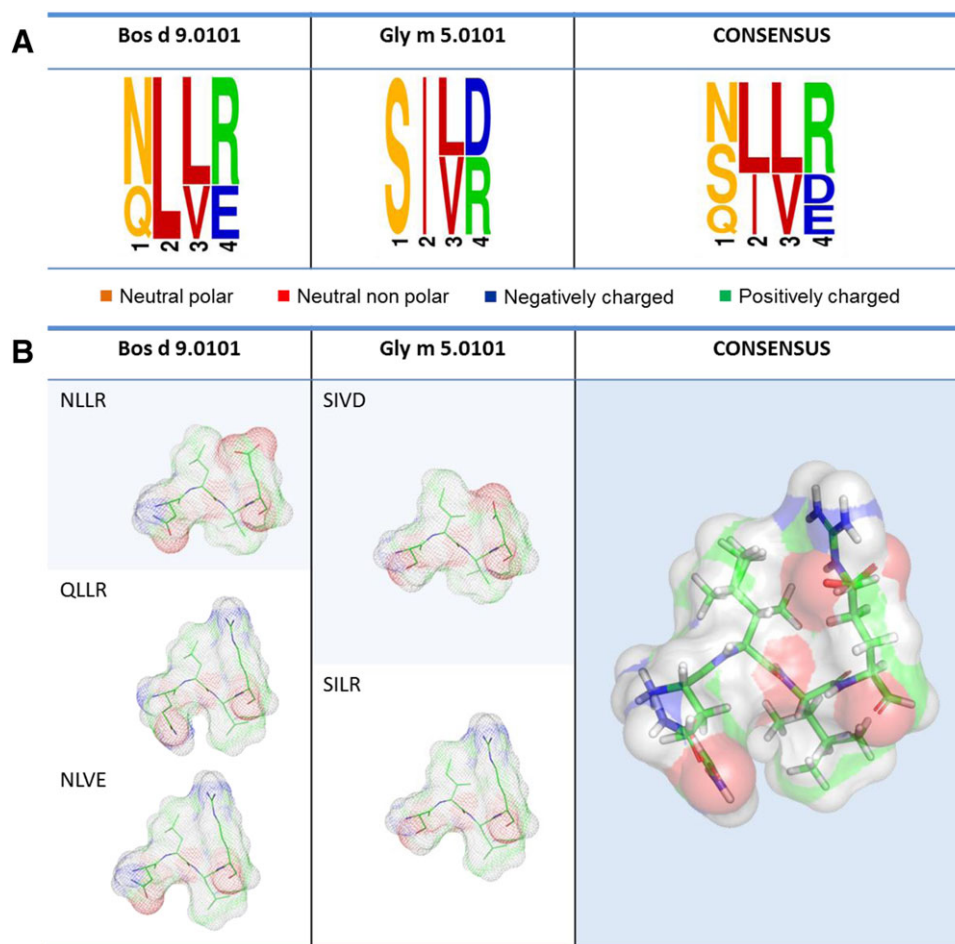
The comparison among the detected sequences led us to define four residues, structurally relevant for the interaction with 1D5 (Fig. 4B). The environment of the antigenic binding site found in Gly m 5 was similar to that found in  $\alpha$ -casein: a polar amino acid, either positively (R-Arginine) or negatively (D Aspartic acid) charged, flanked by neutral residues, polar or non-polar (two contiguous residues) (Fig. 4B).

In addition, mapped epitopes from Gly m 5.0101 and  $\alpha$ S1-casein were analyzed by LOGOS graphic constructions (Fig. 5A) and both graphics were compared. As shown in Fig. 5A, both chart LOGOS are surprisingly similar and overlapping. This observation was confirmed when the same analysis was carried out with the 3D representation obtained with the Maestro software (Fig. 5B). Finally, the analysis revealed that the side chains of the residues and the surface charge distribution around the epitope were very similar in all sequences, regardless of the allergen used in the immunocapture with 1D5 mAb.

## 4 Concluding remarks

The primary structure that defines sequential epitopes on caseins has been previously described as more relevant for allergenicity than the conformational motifs. In this sense, several studies have reported that physical treatments that disrupt the conformational structure of proteins do not affect IgE or IgG reactivity of caseins [41]. For this reason, our investigation was focused on the characterization of linear epitopes using two approaches based on immunocapture and MALDI analysis. We assume that cross-reactivity between these phylogenetically unrelated allergens may only be explained by the presence of short similar sequences. There is no information regarding any biological function associated to these amino acid sequences that could led us to hypothesize that a conserved structure critical for any relevant biological property is maintained in animal and plant allergens. The use of EMap-MALDI methodology to analyze epitopes or allergens by means of a mAb with similar binding parameters





**Figure 5.** (A) Occurrence frequency analysis of the epitope residues by LOGOS. Sequence logo generated from the alignment of peptide residues from: Left panel,  $\alpha$ S1-casein (NLVE:17–14 aa position; NLLR: 19–22 aa position, and QLLR: 97–100 aa position). In the middle, Gly m 5.0101 (SIVD: 379–382 aa position and SILR: 537–540 aa position). Right panel,  $\alpha$ S1-casein and Gly m 5.0101. Residue property color code is as follows: neutral polar, orange; neutral non polar, red; negatively charged, blue; and positively charged, green. (B) Structural representation of the captured epitopes by EMap MALDI-TOF MS. Left panel, sequence motif of  $\alpha$ -casein epitopes. In the middle, sequence motif of Gly m 5.0101. Right panel, the multiple alignment of overall sequence motifs of the epitopes from both allergens. Atom color code is as follows: carbon (C), green; oxygen (O), red; nitrogen (N), blue; and hydrogen (H), white.

for both proteins had not been used before. We consider that this approximation better reproduces the physiological allergen-antibody interaction, since it is performed with the antigen in solution. The analysis of the amino acid sequences of peptides recognized by 1D5 mAb led us to propose a common core amino acid motif for both allergens, which could be critical for paratope–epitope interaction.

Overall, this novel MS approach for epitope mapping on unrelated allergens allowed us to identify cross-reactive B-cell epitopes between soy and milk caseins. The knowledge acquired about the critical residues involved in this cross-allergenicity will contribute to develop novel therapeutic strategies by means of the modification of these key residues to alter the binding capacity of allergens to IgE antibodies, and even to the development of a hypoallergenic protein for a novel vaccine against food allergy.

*This research was supported by Agencia Nacional de Promoción Científica y Tecnológica and the National University of La Plata (grants PICT 2012-1772, PIP 2013-0189 and UNLP 11/X695 to Guillermo H. Docena). The MALDI-TOF MS analyses were carried out at the Proteomics Facility of the Universitat Autònoma de Barcelona (SePBioEs-UAB) by A.M.C. and S.A.T.*

*A.M.C. and S.A.T. conceived and designed the study. A.M.C., S.A.T., and M.F. developed the methodology. A.M.C., S.A.T., A.C., S.B., R.C., and A.Q. acquired the data. A.M.C., S.A.T., and S.B. analysed and interpreted the data. A.M.C., S.A.T., and G.H.D. wrote, reviewed and revised the manuscript. G.H.D., S.P., and S.A.T. supervised the study.*

*The authors have declared no conflict of interest.*

## 5 References

- [1] Kay B., Kaplan A., Bousquet J., Holt P., *Allergy and Allergic Diseases*, Wiley-Blackwell, 2008.
- [2] Wills-Karp, M., Nathan, A., Page, K., Karp, C. L., New insights into innate immune mechanisms underlying allergenicity. *Mucosal Immunol.* 2010, 3, 104–110.
- [3] Järvinen, K. M., Chatchatee, P., Mammalian milk allergy: clinical suspicion, cross-reactivities and diagnosis. *Curr. Opin. Allergy Clin. Immunol.* 2009, 9, 251–258.
- [4] Orsi, O., Fernández, A., Follett, F. R., Marchisone, S., et al., [Cow's milk protein allergy: proposed guidelines for the management of children with cow's milk protein allergy]. *Arch. Argent. Pediatr.* 2009, 107, 459–467.

- [5] Wal, J.-M., Bovine milk allergenicity. *Ann. Allergy Asthma Immunol.* 2004, *93*, S2–S11.
- [6] Szaflarska-Szczepanik, A., Gasiorowska, J., [Prevalence of soya allergy in children with cow's milk allergy]. *Med. Wieku Rozwoj* 2003, *7*, 241–248.
- [7] Ahn, K.-M., Han, Y.-S., Nam, S.-Y., Park, H.-Y., et al., Prevalence of soy protein hypersensitivity in cow's milk protein-sensitive children in Korea. *J. Korean Med. Sci.* 2003, *18*, 473–477.
- [8] Terracciano, L., Isoardi, P., Arrigoni, S., Zoja, A., et al., Use of hydrolysates in the treatment of cow's milk allergy. *Ann. Allergy Asthma Immunol.* 2002, *89*, 86–90.
- [9] American Academy of Pediatrics. Committee on Nutrition. Soy protein-based formulas: recommendations for use in infant feeding. *Pediatrics* 1998, *101*, 148–153.
- [10] Rozenfeld, P., Docena, G. H., Añón, M. C., Fossati, C. A., Detection and identification of a soy protein component that cross-reacts with caseins from cow's milk. *Clin. Exp. Immunol.* 2002, *130*, 49–58.
- [11] Smaldini, P., Curciarello, R., Candreva, A., Rey, M. A., et al., In vivo evidence of cross-reactivity between cow's milk and soybean proteins in a mouse model of food allergy. *Internat. Arch. Allergy Immunol.* 2012, *158*, 335–346.
- [12] Curciarello, R., Lareu, J. F., Fossati, C. A., Docena, G. H., Petruccelli, S., Immunochemical characterization of *Glycine max* L. Merr. var Raiden, as a possible hypoallergenic substitute for cow's milk-allergic patients. *Clin. Exp. Allergy* 2008, *38*, 1559–1565.
- [13] Curciarello, R., Smaldini, P. L., Candreva, A. M., González, V., et al., Targeting a cross-reactive Gly m 5 soy peptide as responsible for hypersensitivity reactions in a milk allergy mouse model. *PLoS ONE* 2014, *9*, e82341.
- [14] Candreva, A. M., Smaldini, P. L., Curciarello, R., Cauherff, A., et al., Cross-reactivity between the soybean protein p34 and bovine caseins. *Allergy Asthma Immunol. Res.* 2015, *7*, 60–68.
- [15] Candreva, Á. M., Smaldini, P. L., Curciarello, R., Fossati, C. A., et al., The major soybean allergen Gly m Bd 28K induces hypersensitivity reactions in mice sensitized to cow's milk proteins. *J. Agric. Food Chem.* 2016, *64*, 1590–1599.
- [16] Carter, J. M., Loomis-Price, L., B cell epitope mapping using synthetic peptides. *Curr. Protoc. Immunol.* 2004, Chapter 9, Unit 9.4.
- [17] Scott, J. K., Discovering peptide ligands using epitope libraries. *Trends Biochem. Sci.* 1992, *17*, 241–245.
- [18] Xia, L., Willison, L. N., Porter, L., Robotham, J. M., et al., Mapping of a conformational epitope on the cashew allergen Ana o 2: a discontinuous large subunit epitope dependent upon homologous or heterologous small subunit association. *Mol. Immunol.* 2010, *47*, 1808–1816.
- [19] Koehler, C., Carlier, L., Veggi, D., Balducci, E., et al., Structural and biochemical characterization of NarE, an iron-containing ADP-ribosyltransferase from *Neisseria meningitidis*. *J. Biol. Chem.* 2011, *286*, 14842–14851.
- [20] Lillehoj, H. S., Kim, C. H., Keeler, C. L., Jr, Zhang, S., Immunogenomic approaches to study host immunity to enteric pathogens. *Poult. Sci.* 2007, *86*, 1491–1500.
- [21] Soriani, M., Petit, P., Grifantini, R., Petracca, R., et al., Exploiting antigenic diversity for vaccine design: the chlamydia ArtJ paradigm. *J. Biol. Chem.* 2010, *285*, 30126–30138.
- [22] Lu, X., DeFelippis, M. R., Huang, L., Linear epitope mapping by native mass spectrometry. *Anal. Biochem.* 2009, *395*, 100–107.
- [23] Zhao, Y., Chalt, B. T., Protein epitope mapping by mass spectrometry. *Anal. Chem.* 1994, *66*, 3723–3726.
- [24] Zhao, C., Trudeau, B., Xie, H., Prostko, J., et al., Epitope mapping and targeted quantitation of the cardiac biomarker troponin by SID-MRM mass spectrometry. *Proteomics* 2014, *14*, 1311–1321.
- [25] Docena, G., Rozenfeld, P., Fernández, R., Fossati, C. A., Evaluation of the residual antigenicity and allergenicity of cow's milk substitutes by in vitro tests. *Allergy* 2002, *57*, 83–91.
- [26] Petruccelli, S., Chirido, F. G., Añón, M. C., Immunochemical reactivity of soybean  $\beta$ -conglycinin subunits. *Food Agricul. Immunol.* 2005, *16*, 17–28.
- [27] Docena, G. H., Fernandez, R., Chirido, F. G., Fossati, C. A., Identification of casein as the major allergenic and antigenic protein of cow's milk. *Allergy* 1996, *51*, 412–416.
- [28] Natarajan, S. S., Xu, C., Bae, H., Caperna, T. J., Garrett, W. M., Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *J. Agric. Food Chem.* 2006, *54*, 3114–3120.
- [29] Mann, M., Højrup, P., Roepstorff, P., Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* 1993, *22*, 338–345.
- [30] Lassaux, P., Peri, C., Ferrer-Navarro, M., Gourlay, L. J., et al., A structure-based strategy for epitope discovery in *Burkholderia pseudomallei* OppA antigen. *Structure* 2013, *21*, 167–175.
- [31] Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., et al., Clustal W and Clustal  $\times$  version 2.0. *Bioinformatics* 2007, *23*, 2947–2948.
- [32] Crooks, G. E., Hon, G., Chandonia, J.-M., Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res.* 2004, *14*, 1188–1190.
- [33] Gu, C., Shannon, D. A., Colby, T., Wang, Z., et al., Chemical proteomics with sulfonyl fluoride probes reveals selective labeling of functional tyrosines in glutathione transferases. *Chem. Biol.* 2013, *20*, 541–548.
- [34] Narayanan, A., Jones, L. H., Sulfonyl fluorides as privileged warheads in chemical biology. *Chem. Sci.* 2015, *6*, 2650–2659.
- [35] Chatchatee, P., Järvinen, K. M., Bardina, L., Beyer, K., Sampson, H.A., Identification of IgE- and IgG-binding epitopes on alpha(s1)-casein: differences in patients with persistent and transient cow's milk allergy. *J. Allergy Clin. Immunol.* 2001, *107*, 379–383.
- [36] Cocco, R. R., Järvinen, K.-M., Sampson, H. A., Beyer, K., Mutational analysis of major, sequential IgE-binding epitopes in

alpha s1-casein, a major cow's milk allergen. *J. Allergy Clin. Immunol.* 2003, *112*, 433–437.

- [37] Enomoto, A., Shon, D. H., Aoki, Y., Yamauchi, K., Kaminogawa, S., Antibodies raised against peptide fragments of bovine alpha s1-casein cross-react with the intact protein only when the peptides contain both B and T cell determinants. *Mol. Immunol.* 1990, *27*, 581–586.
- [38] Spuergin, P., Mueller, H., Walter, M., Schiltz, E., Forster, J., Allergenic epitopes of bovine alpha S1-casein recognized by human IgE and IgG. *Allergy* 1996, *51*, 306–312.
- [39] Elsayed, S., Hill, D. J., Do, T. V., Evaluation of the allergenicity and antigenicity of bovine-milk alphas1-casein using extensively purified synthetic peptides. *Scand. J. Immunol.* 2004, *60*, 486–493.
- [40] Fu, C. J., Jez, J. M., Kerley, M. S., Allee, G. L., Krishnan, H. B., Identification, characterization, epitope mapping, and three-dimensional modeling of the alpha-subunit of beta-conglycinin of soybean, a potential allergen for young pigs. *J. Agric. Food Chem.* 2007, *55*, 4014–4020.
- [41] Kohno, Y., Honma, K., Saito, K., Shimojo, N., et al., Preferential recognition of primary protein structures of alpha-casein by IgG and IgE antibodies of patients with milk allergy. *Ann Allergy* 1994, *73*, 419–422.