Accepted Manuscript

Peptide selection and antibody generation for the prospective immunorecognition of Cry1Ab16 protein of transgenic maize

Joana Costa, Mariela M. Marani, Liliana Grazina, Caterina Villa, Liliana Meira, M. Beatriz P.P. Oliveira, José R.S.A. Leite, Isabel Mafra
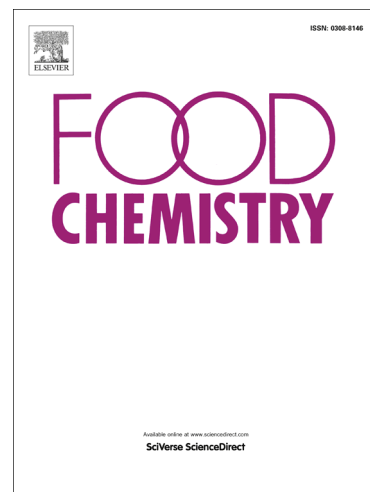
Please cite this article as: Costa, J., Marani, M.M., Grazina, L., Villa, C., Meira, L., Oliveira, M.B.P.P., Leite, J.R.S., Mafra, I., Peptide selection and antibody generation for the prospective immunorecognition of Cry1Ab16 protein of transgenic maize, *Food Chemistry* (2017), doi: http://dx.doi.org/10.1016/j.foodchem.2017.03.154

Peptide selection and antibody generation for the prospective

immunorecognition of Cry1Ab16 protein of transgenic maize

Joana Costa[a,b,1,*], Mariela M. Marani[a,b,1,*], Liliana Grazina[a], Caterina Villa[a], Liliana Meira[a], M. Beatriz P. P. Oliveira[a], José R. S. A. Leite[c,d], Isabel Mafra[a,b]

[a]REQUIMTE-LAQV, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal.

[b]IPEEC-CONICET, Instituto Patagónico para el Estudio de los Ecosistemas Continentales, Consejo Nacional de Investigaciones Científicas y Técnicas, Puerto Madryn, Argentina.

[c]Núcleo de Pesquisa em Biodiversidade e Biotecnologia, Biotec, Campus de Parnaíba, Universidade Federal do Piauí, Parnaíba, Brasil;

[d]Area de Morfologia, Faculdade de Medicina, Universidade de Brasília, Campus Universitário Darcy Ribeiro, Brasília, Brasil.

Corresponding authors

*E-mail: jbcosta@ff.up.pt Phone: +351 220 428 842

*E-mail: mmarani@cenpat-conicet.gob.ar Phone: +54 280 4883184

[1]Joana Costa and Mariela M. Marani contributed equally to this work and should have the first shared authorship.

Running Title: Peptide prediction for prospective biorecognition of Cry1Ab16 protein

1

ABSTRACT

The introduction of genes isolated from different *Bacillus thuringiensis* strains to express Cry-type toxins in transgenic crops is a common strategy to confer insect resistance traits. This work intended to extensively *in silico* analyse Cry1A(b)16 protein for the identification of peptide markers for the biorecognition of transgenic crops. By combining two different strategies based on several bioinformatic tools for linear epitope prediction, a set of seven peptides was successfully selected as potential Cry1A(b)16 immunogens. For the prediction of conformational epitopes, Cry1A(b)16 models were built on the basis of three independent templates of homologue proteins of Cry1A(a) and Cry1A(c) using an integrated approach. PcH_736-746 and PcH_876-886 peptides were selected as the best candidates, being synthesised and used for the production of polyclonal antibodies. To the best of our knowledge, this is the first attempt of selecting and defining linear peptides as immunogenic markers of Cry1A(b)-type toxins in transgenic maize.

Keywords: Cry1A(b)16 toxin, linear and discontinuous epitopes, B- and T-cell epitope prediction, bioinformatics, GMO, processed food, antibody production.

1   Introduction

Maize (*Zea mays* L.) plays a determinant role, both in human and animal nutrition, representing more than 36.5% of the total area of cereal production in 2014 (FAOSTAT, http://faostat3.fao.org/home/E). However, this crop is a common target of insects, nematodes and mites that propagate very rapidly, leading to low production rates of maize and, consequently, to great economic losses. To solve this problem, several biotechnological solutions have been adopted. One frequent strategy to confer insect resistance concerns the introduction of genes isolated from different *Bacillus thuringiensis* strains, which is a group of gram-positive bacteria that produces parasporal crystals (composed of polypeptide protoxins) during sporulation (Guo et al., 2012). In insects, the protoxins are activated by gut proteases to generate Cry toxins that are highly specific to their target species. Cry proteins are considered as innocuous to humans, animals and plants, and are completely biodegradable, being faced as viable alternatives for the control of insect pests in agriculture (Bravo, Gill, & Soberón, 2007). These crystalline (Cry) inclusions are classified as delta-endotoxins, encompassing a high number of toxins that share structural/functional similarities and display high specificity towards particular groups of insects (Ibrahim, Griko, Junker, & Bulla, 2010). So far, genes expressing Cry1, Cry2 and/or Cry3 class-type toxins have been isolated from *B. thuringiensis* strains and used for the generation of different genetically modified (GM) crops (Dehury et al., 2013; ISAAA, http://www.isaaa.org). In the specific case of maize, genetic constructions expressing Cry1 and Cry2 toxins are known to have specific insecticidal activity against insects from Lepidoptera order (e.g. mosquitoes, moths), while Cry3 proteins have toxic activity against insects from Coleoptera order (e.g. beetles) (Ibrahim et al., 2010; ISAAA, http://www.isaaa.org).

3

Maize is the second most cultivated biotech plant, thus contributing for about 30% of the global biotechnological area of 179.7 million hectares in 2015 (James, 2015). Among the GM plants, maize is the crop with the highest number of GM events (142), although only 29 have valid authorisation for food and feed, as well as for importing and processing, inside the European Union (EU) in 2016 (GMO compass, 2016; James, 2015). MON810 and Bt11 transgenic events have been widely cultivated since they present resistance to a specific insect named European corn borer, which is responsible for inducing severe damage in European maize plantations. Both transgenic maize lines were modified to express the Cry1A(b) toxin that is one of the subclasses of Cry1 proteins (ISAAA, http://www.isaaa.org; Kumar et al., 2013). Like other Cry proteins, Cry1A(b) corresponds to the final product of a protoxin of approximately 130 kDa and 1155 amino acids that, after being submitted to different mechanisms occurring at transcriptional, post-transcriptional and post-translational levels, presents a molecular mass of 65 kDa (Ibrahim et al., 2010). The available data regarding this protein suggest that it is composed of 3 domains (I, II and III), also common to other Cry structures (Dehury et al., 2013; kashyap, 2012), in which the domains II and III are believed to be involved in receptor binding and insecticidal activity (Ibrahim et al., 2010).

In spite of all the extensive safety assessments of GM crops about the potential effects of Cry toxins in human health, GM maize crops like Bt11 or MON810 have been widely cultivated for about 20 years. Presently, the consumption of GM plants is still a matter of concern since it continues dividing consumers' opinions. For an informed consumer choice, the mandatory labelling of processed foods containing or produced of GM crops is in force in many countries including EU (Regulation (EC) No. 1829/2003), continuously requiring for simple, fast and highly sensitive solutions to control the presence of genetically modified organisms (GMO) in foods. To answer this challenge,

the interest on biosensor devices has been growing in the field of GMO analysis because of their potential for automation and micro-fabrication of simple and portable detection systems (e.g. visual or electrochemical) (Manzanares-Palenzuela et al., 2016; Plácido et al., 2016). So far, the immunorecognition of proteins expressed by GM plants has only been proposed using enzyme-linked immunosorbent assays (ELISA), but it has not been reported for biosensing devices.

To date, bioinformatics have been pointed out as excellent predicting tools to assess the immunoreactivity of peptides, while reducing the experimental work involved in epitope mapping. The main objective of this study concerned the selection of a set of marker peptides that could function as potential immunogens for the production of novel antibodies for Cry1A(b)16 biorecognition. For this purpose, an extensive *in silico* analysis was performed on Cry1A(b)16 protein, in order to select the best peptide candidates for antibody generation. On the basis of this study, the best peptides were synthesised, purified and used for animal inoculation with subsequent production of functional polyclonal antibodies.

2   Materials and Methods

2.1   Selection of Cry1 sequences and 3D structures

The correspondent protein sequence of Cry1A(b)16 was retrieved from the NCBI database (Table 1) with accession no. AAK55546.1. This protein is a delta-endotoxin produced by *B. thuringiensis* strain AC11 serotype H14, being encoded by a silent *cry* gene with NCBI accession no. AF375608.1. Cry1A(b)16 is the final product of a protoxin with a molecular mass of 130 kDa and an entire primary sequence of protein of 1155 amino acids.

In the absence of an experimental conformational structure of Cry1A(b)16, three

5

homologue proteins, namely one model from Cry1A(a) and two models from Cry1A(c) with RCSB PDB ID: 1CIY, 4ARY and 4W8J, respectively, were selected based on their sequence identity and conformational similarity (Fig. S1, supplementary material). In order to obtain pdb files of Cry1A(b)16 for the analysis of discontinuous B-cell epitopes, the SWISS-MODEL platform was used to generate three models from the experimental predetermined Cry1A(a) and Cry1A(c) structures.

2.2 *Enzymatic cleavage sites*

Cry1A(b)16 was submitted to an *in silico* cleavage site prediction with three enzymes, namely trypsin, chymotrypsin of high specificity (HS) and chymotrypsin of low specificity (LS), using the ExPASy PeptideCutter tool (Table 1). According to the cleavage site map and to the table of cleavage site positions, peptides presenting 10-30 amino acids were comprehensively analysed following two strategies: in the first one, the peptides were primarily submitted to compositional analysis and the resultant candidates were posteriorly evaluated using B-cell and T-cell epitope prediction tools (Table 1); in the second one, the same selected peptides (10-30 residues) were firstly evaluated with T-cell epitope prediction tool. Following the second strategy, only the peptides classified as "high" in the T-cell binding determination for major histocompatibility complex (MHC)-1 or MHC-2 were further assessed for B-cell epitope prediction tools and compositional analysis.

2.3 *Sequence and structure determination of epitopes*

2.3.1 *B- and T-cell Epitope Prediction Tools*

To evaluate the sequence of Cry1A(b)16, several computational tools from the Immune Epitope Data Base (IEDB) were used to predict linear B-cell epitopes (Table 1). For the

assessment of discontinuous B-cell epitopes, ElliPro and DiscoTope 2.0 softwares were used (Table 1) (Kringelum et al., 2012; Ponomarenko et al., 2008).

Additionally, the POPI 2.0 software (Table 1) was used to predict the T-cell epitopes based on the potential immunogenicity of MHC class I and class II (Tung & Ho, 2007).

2.3.2 *Secondary structure prediction of selected peptides*

Selected peptides obtained from each strategy were analysed using the web resources for structural prediction, namely the PEP-FOLD3 software (Table 1) (Shen et al., 2014). This software is a *de novo* approach that allows predicting peptide structures from primary sequences. The referred approach enabled obtaining a predictive secondary structural model for the selected peptides.

2.4 *Peptide synthesis and antibody production*

The selected peptides, as potential candidates for antibody generation were produced by a specialised company (Davids Biotechnologie GmbH, Regensburg, Germany). According to the provided protocol, both peptides were synthesised in solid phase (15 mg/each) with purity ~80%. Quality controls were performed for each peptide using liquid chromatography (LC) and mass spectrometry (MS) and approximately 5 mg of each peptide were chemically conjugated with an immunogenic carrier protein, namely the keyhole limpet hemocyanin (KLH). The immunisation was carried out in rabbit to produce polyclonal antibodies, following a protocol of 28 days with 4 immunisations. ELISA titer was performed to evaluate the amount of antibodies in the blood, being the antibodies submitted to affinity and proteinA purifications.

3 Results and Discussion

7

3.1   Sequence selection

A complete list of the identified *B. thuringiensis* delta-endotoxins can be consulted at the http://www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/toxins2.html. The list has been updated over the last years (Crickmore et al., 2016), currently holding more than 800 delta-endotoxins that are linked to protein and DNA sequences in the NCBI database (Table 1). From this list, the toxin Cry1A(b)10 with NCBI accession no. A29125 was elected because it is the protein expressed in maize MON810 event (Fischhoff et al., 1987), accordingly to its producer (Monsanto Company, St. Louis, MO, USA, http://www.monsanto.com/products/documents/safety-summaries/yieldgard_corn_pss.pdf). However, when accessing the referred protein, its sequence was further updated and replaced by Cry1A(b)16 with NCBI accession no. AAK55546.1. In this sense, the updated Cry1A(b)16 sequence was used in this study for compositional analysis and for B- and T-cell epitope prediction tools.

3.2   Peptide analysis

Prior to peptide analysis, the Cry1A(b)16 was submitted to an *in silico* cleavage with three enzymes, namely trypsin and chymotrypsin of high (HS) and low specificity (LS). ExPASy PeptideCutter tool (Table 1) allows simulating protein cleavage using several proteases. Therefore, some factors such as the size of resultant peptides, as well as the specificity, selectivity and cost of enzymes, were carefully considered before their selection. Considering the cleavage site map and the table of cleavage site positions, peptides presenting 10-30 amino acids were comprehensively analysed according to two strategies.

3.3   First strategy

3.3.1 *Compositional analysis and solubility*

Following a previous study (Marani et al., 2015), 111 resultant peptides of 10-30 amino acids were selected from the *in silico* cleavage of Cry1A(b)16 toxin that is expressed by MON810 maize. Compositional analysis was firstly performed aiming at minimising possible difficulties for future peptide synthesis. Briefly, peptides encompassing residues of cysteine (Cys) and methionine (Met) were side-stepped since Cys is prone to form disulphide bonds and both amino acids (Cys, Met) are commonly susceptible to rapid oxidation (Rao & Kroon, 1993, Sigma-Aldrich, 2016). Duplets of aspartate-proline (Asp-Pro), aspartate-glycine (Asp-Gly) and aspartate-serine (Asp-Ser) were also excluded due to their high potential for hydrolysis under acidic conditions (Balu-Iyer, Miclea, & Purohit, 2007; Peptides International, 2016; Sigma-Aldrich, 2016). To avoid deletions during synthesis and low purity, peptides with multiple Ser residues or containing series of isoleucine (Ile), glutamine (Gln), phenylalanine (Phe), leucine (Leu), valine (Val) or threonine (Thr) were discarded (Peptides International, 2016; Thinkpeptides eBook, 2012). Peptides containing asparagine (Asn) at the N-terminal position, which hamper the removal of the protecting group, were also rejected (Peptide Design Guideline, 2011; Sigma-Aldrich, 2016). Additionally, peptides with at least 1 charged residue out of 5 amino acids were selected to ensure acceptable water solubility during their synthesis (Hruby & Matsunaga, 2002; Peptide Design Guideline, 2011).

From the initial set of 111 *in silico* candidates, 51 peptides were selected after compositional analysis. This number was further reduced to 12 peptides after evaluating the criterion of solubility. At this point, a total of 12 peptides were elected from trypsin, chymotrypsin-LS or chymotrypsin-HS cleavage, being designated as Pt, PcL or PcH, respectively. After enzymatic cleavage with chymotrypsin-LS and chymotrypsin-HS,

two resultant peptides were the same (PcL_876-886 and PcH_876-886), so both were further represented as one peptide under the name of PcH_876-886.

3.3.2   Linear B-cell epitope prediction and T-cell binding analysis

Immunological predictions and simulations have been considered of great interest in the field of applied immunology, being highlighted as efficient tools to reduce the experimental work involved in epitope mapping for vaccine design, immunotherapeutics and development of diagnostic techniques (Lundegaard, Lund, Kesmir, Brunak, & Nielsen, 2007). To date, several computational algorithms mimicking the adaptive immune response have been developed, mostly for medical purposes and vaccine design (Agallou, Athanasiou, Koutsoni, Dotsika, & Karagouni, 2014; Kovjazin et al., 2013; Purcell, MsCluskey, & Rossjohn, 2007), although other areas of research are now beginning to benefit from the available immunoinformatics (Holton, Vijayakumar, & Khaldi, 2013). Peptide epitopes are known to be primarily presented to the immune system by the major histocompatibility complex (MHC) class I and II (T-lymphocytes), leading to the activation of the B-cells and production of antibodies (Flower, 2013). For this reason, different B- and T-cell epitope predicting tools were used to evaluate the immunogenicity of the proposed peptides, which are summarised in Table 2.

B-cell epitope predicting methods are based on different bioinformatic tools that evaluate the primary and the conformational structures of the protein (Kringelum et al., 2012). In primary sequence analysis, the prediction of linear (continuous) epitopes has been highly correlated with several factors, namely flexibility, hydrophilicity, surface accessibility, exposed surface, beta-turns, polarity and antigenic predisposition of polypeptides (Chou & Fasman, 1978, Emini et al., 1985; Karplus & Schulz, 1985;

Kolaskar & Tongaonkar, 1990; Larsen et al., 2006; Parker et al., 1986). Considering the prediction tool based on Emini et al. (1985) calculations (Table 1), 4 out of the 5 peptides, with sizes ranging from 6 to 10 residues, were estimated to present surface accessibility above 1.0, indicating that those residues might belong to epitopes located at protein surface (Table 2). The antigenic propensity of a selected peptide was assessed by Kolaskar and Tongaonkar (1990) software (Table 1), considering a value of at least 1.0 for the allergenicity parameter. From antigenicity analysis, 4 out of 5 peptides exhibited partial content of sequences with allergenicity above 1.0 (Table 2). The antigenicity of amino acids is linked with accessible and flexible surface sites in antibody interaction, thus antigenicity predicting algorithms also include the evaluation of flexibility and hydrophilicity parameters. Chain flexibility or hydrophilicity of the proposed 5 peptides was determined by the Karplus and Schulz (1985) software or Parker et al. (1986) tool (Table 1), respectively. All peptides evidenced high flexibility (≥1.0) and hydrophilicity (≥1.5), with PcH_876-886 presenting the highest value of hydrophilicity (3.8) (data not shown). Combining one of the best propensity scales with the predictions of hidden Markov model, the BepiPred software allows predicting the location of continuous B-cell epitopes (Table 1) (Larsen et al., 2006). Considering an epitope prediction parameter above 0.4, 3 out of the 5 peptides exhibited residues that are predicted to be part of an epitope (Table 2).

For T-cell binding analysis, the software POPI 2.0 (Table 1) was applied to estimate the immunogenicity of peptides using a set of 23 or 21 informative physicochemical properties from MCH class I or class II binding peptides, respectively (Tung & Ho, 2007). The MHC-I molecules bind peptides with 8-11 residues, mainly interacting with cytotoxic T-lymphocytes (CTL), while MHC-II molecules link larger fragments (up to 30 residues) (though maintaining a binding core of 9 amino acids) and interact with

helper T-lymphocytes (HTL) (Lundegaard et al., 2007; Rapin, Lund, Bernaschi, & Castiglione, 2010). According to T-cell predicting tool (POPI 2.0) (Tung & Ho, 2007), 5 candidates presenting "moderate" or "high" response to one of the major histocompatibility complex (CTL or HTL) were further considered (Table 2). Following the steps described in the first strategy, a total of 5 peptides were considered as good candidates for antibody production (Table 2).

3.4   Second strategy

The second strategy followed the inverse order of the steps described for the first one. In this sense, the same initial set of 111 peptides (with 10-30 residues) was primarily submitted to T-cell binding analysis. Using POPI 2.0 software (Table 1), the immunogenicity of peptides was classified as "none", "little", "moderate" or "high" towards responses to MHC-I (CTL) or MHC-II (HTL). From this grid of classification, only the peptides presenting at least "high" response to one of the CTL or HTL were further selected, totalising 18 candidates.

Like in the first strategy, the same linear B-cell epitope predicting methods were also applied to the potential peptide candidates. Considering the same criteria used in the first strategy for surface accessibility, antigenicity, flexibility, hydrophilicity and epitope prediction parameters, 16 out of the 18 peptides were chosen after linear B-cell epitope prediction. Compositional and solubility analysis was performed to the remaining peptides and the number of potential candidates for antibody production was reduced to a final set of 4 peptides. The data from second strategy regarding the T- and B-cell predicting tools, compositional and solubility analysis of the selected 4 peptides is presented in Table 3.

3.5 Structural B-cell epitope prediction

Following the determination of linear epitopes by several B- and T-cell epitope prediction tools, the discontinuous ones were evaluated by different structural B-cell predicting approaches, namely the ElliPro and the DiscoTope 2.0 (Kringelum et al., 2012; Ponomarenko et al., 2008). ElliPro is a computational tool that enables the prediction of linear and discontinuous epitopes, using the protein sequence or structure as template (Ponomarenko et al., 2008). This computational tool is based on two principles: (i) the regions protruding from the globular surface of the protein are more available for interaction with an antibody and (ii) those protrusions can be determined by treating the protein as a simple ellipsoid. For each residue, a protrusion index (PI) is defined as the percentage of the protein atoms enclosed in the ellipsoid at which the residue first becomes lying outside the ellipsoid. A score value of 0.9 in ElliPro shows that all residues are outside the 90% ellipsoid (Ponomarenko et al., 2008).

DiscoTope 2.0 software is based on calculations of surface accessibility (estimated in terms of contact numbers) and a novel epitope propensity amino acid score (Kringelum et al., 2012). By combining the propensity scores of residues in spatial proximity and the contact numbers, the method defines final scores. Different threshold scores can be set for epitope identification, which represent different values for sensitivity/specificity. A defined threshold score of -7.7 corresponds to a sensitivity value of 0.47 and a specificity of 0.75, which means that 47% of the epitope residues (sensitivity) and 25% of the non-epitope residues (specificity) were predicted as part of epitopes.

However, in order to perform analysis using these B-cell epitope prediction programs, a pdb file is required. According to a theoretical structural model proposed by Kashyap (2012) for Cry1A(b)16 generated by homology modelling, this toxin shares high conformational similarity with Cry1A(a) and Cry1A(c) proteins. Accordingly, the

primary sequences of Cry1A(b)16, Cry1A(a) and Cry1A(c) were aligned (Fig. S1, supplementary material) and the SWISS-MODEL platform (Table 1) was used to generate conformational models of Cry1A(b)16 on the basis of experimental predetermined structures of Cry1A(a) (PDB ID: 1CIY) and Cry1A(c) (PDB ID: 4ARY and 4W8J) toxins. The SWISS-MODEL is a fully automated protein structure homology-modelling server that enables modelling a conformational structure of a protein from its primary sequence using homology modelling techniques (Biasini et al., 2014). This server follows four main steps, which include the identification of structural template(s), alignment of target sequence and template structure(s), model building and quality assessment of the proposed model. For model building, the target/template alignment is used as input to generate a conformational structure using the ProMod-II software, or the MODELLER as alternative. The quality of a proposed model is assessed with the local composite scoring function QMEAN, which uses several statistical descriptors expressed as potentials of mean force: geometrical features of the model (pairwise atomic distances, torsion angles, solvent accessibility) are compared to statistical distributions obtained from experimental structures and scored. Global QMEAN scores are indicators of the overall quality of the model, being provided as a Z-score that relates the obtained values to scores calculated from a set of high-resolution X-ray structures. The GMQE (Global Model Quality Estimation) is a quality parameter expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template (Biasini et al., 2014). Using the SWISS-MODEL platform, different molecular structures of Cry1A(b)16 were suggested, according to each template 4W8J, 1CIY and 4ARY, corresponding to model, 1, 2 or 3, respectively (Fig. 1A). Additional data on the alignments of the secondary structures of the proposed models are represented in Figures S2-S4 (supplementary material). Sequence identity

with 1CIY (model 2) and 4ARY (model 3) were 88.78% and 84.20%, respectively, with coverage of approximately 0.50 since both models were based on monomers of Cry1A(a) and Cry1A(c). Sequence identity with 4W8J (model 1) is similar to the other models (87.62%), but with coverage of 0.99 because the structure is a homo-dimer. GMQE scores were 0.79, 0.52 and 0.51 for models 1 to 3, respectively, indicating high reliability of the proposed models (Figs. S5-S7, supplementary material). Table 4 presents the results of potential discontinuous epitopes by ElliPro and DiscoTope 2.0 softwares generated by each model structure. Owing to the fact that the conformational structures proposed by 1CIY (model 2) and 4ARY (model 3) involve only one chain each, peptides located in positions 610 to 1150 were not assessed by ElliPro or Discotope 2.0 in those models (Kringelum et al., 2012; Ponomarenko et al., 2008).

For model 1, ElliPro analysis presents its high protrusion index score for residues that are included in peptides Pt_282-292 and PcH_77-91 (Table 4) (Ponomarenko et al., 2008). Considering that residues with larger scores are associated with greater solvent accessibility, these two peptides are most likely to be on the surface of the conformational structure of the protein and, subsequently, more accessible to antibody interaction. From the 7 selected peptides, only the residues from the PcH_162-178 peptide were not scored according to default values for minimum score (0.5) and maximum distance (6 Å) (Table 4). For models 2 and 3, residues from PcH_77-91 and PcH_162-178 presented the highest scores, followed by some residues from Pt_602-619 and Pt_282-292 peptides (Table 4).

When applying DiscoTope 2.0 software using a threshold value of -7.7, positive scores were attained for residues from Pt_282-292, Pt_602-619 and PcH_77-91 peptides with respect to models 2 and 3 (Table 4) (Kringelum et al., 2012). For model 1, positive scores were obtained for residues from Pt_282-292, PcH_77-91, PcH_690-700,

PcH_736-746 and PcH_876-886 peptides. Considering that all peptides (except PcH_162-178) present residues with positive scores by DiscoTope 2.0, those residues are most likely to be part of epitopes (Table 4) (Kringelum et al., 2012). Figure 1B presents the secondary structures of the seven peptides selected as potential immunogens, where random and α-helix structures are evidenced.

3.6   *Functional peptides and antibodies*

From the seven proposed peptides, PcH_736-746 (QKIDESKLKAY) and PcH_876-886 (KEAKESVDAVF) were suggested by the specialised company as the best candidates for animal immunisation, considering all parameters of antigenicity, solubility and epitope prediction. The selected peptides were in good agreement with all the parameters evaluated during this study since they were both classified with "high" response to one of the major histocompatibility complex (CTL or HTL). Additionally, they were also considered the best candidates concerning the continuous B-cell epitopes (Tables 2 and 3). Thus, the referred peptides were synthesised, purified and used as antigens for the immunisation of rabbit at Davids Biotechnologie GmbH (Regensburg, Germany) facilities. Functional polyclonal antibodies against them were successfully obtained as evidenced by the dot blot presented in Figure 2. The antiserum was tested against each target peptide PcH_736-746 or PcH_876-886, which correspond to QKI or KEA, respectively (Fig. 2). Each polyclonal antibody was considered reactive against the respective peptide in the range of the concentrations tested (1 μg/mL to 111 ng/mL), as evidenced by the presence of dots (Fig. 2). Antiserum collected from rabbit prior to immunisation schedule (day 0) was tested against the target peptides as negative control.

4   Conclusions

In this study, a set of seven peptides with potential immunogenicity was successfully selected by combining two different strategies based on several bioinformatic tools for linear epitope prediction. Additionally, for the prediction of conformational epitopes, three Cry1A(b)16 models were built on the basis of three independent templates of homologue proteins of Cry1A(a) and Cry1A(c) using an integrated approach, enabling a complete analysis of the selected immunoreactive peptides. Two immunogenic peptides PcH_736-746 (QKIDESKLKAY) and PcH_876-886 (KEAKESVDAVF) were selected as the best candidates, being further synthesised, purified and used to inoculate a rabbit for polyclonal antibody production. Both peptides and respective functional polyclonal antibodies are ready to be employed in the development of immunochemical detection systems, such as ELISA and/or immunosensors. To the best of our knowledge, this is the first attempt of selecting and defining linear peptides as immunogenic markers of Cry1A(b)-type toxins in transgenic maize.

respectively) financed by POPH-QREN (subsidised by FSE and MCTES). Mariela M.

Marani is researcher of the CONICET.

Supplementary Material

Detailed data regarding the sequence analysis and conformational models of

Cry1A(b)16 based on three available tri-dimensional structures of Cry1A(a) and

Cry1A(c) were provided as supplementary material.

Conflict of Interests

The authors declare no conflict of interests.

5   References

Agallou, M., Athanasiou, E., Koutsoni, O., Dotsika, E., & Karagouni, E. (2014).
Experimental validation of multi-epitope peptides including promising MHC class I-
and II-restricted epitopes of four known *Leishmania infantum* proteins. *Frontiers in
Immunology, 5*, 268. doi: 10.3389/fimmu.2014.00268

Balu-Iyer, S. V., Miclea, R. D., & Purohit, V. S. (2007). Formulation and delivery
issues of therapeutic proteins. In S. C. Gad (Ed.), *Handbook of pharmaceutical
biotechnology,* (pp. 737-755). New Jersey: John-Wiley & Sons, Inc.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., …
Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary
structure using evolutionary information. *Nucleic Acids Research, 42*, W252-W258.
doi: 10.1093/nar/gku340

Bravo, A., Gill, S. S., & Soberón, M. (2007). Mode of action of Bacillus thuringiensis
Cry and Cyt toxins and their potential for insect control. *Toxicon : official journal of*

18

the International Society on Toxinology, 49, 423-435. doi: 10.1016/j.toxicon.2006.11.022

Chou, P. Y., & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. In A. Meister (Ed.), *Advances in Enzymology and Related Areas of Molecular Biology* (pp. 45-148). New Jersey: John Wiley & Sons.

Crickmore, N., Baum, J., Bravo, A., Lereclus, D., Narva, K., Sampson, K., … Zeigler, D. R. (2016). *Bacillus thuringiensis* toxin nomenclature. Retrieved from http://www.btnomenclature.info/, Accessed 03.03.2016.

Dehury, B., Sahu, M., Sahu, J., Sarma, K., Sen, P., Modi, … Choudhury, M. D. (2013). Structural analysis and molecular dynamics simulations of novel δ-endotoxin Cry1Id from *Bacillus thuringiensis* to pave the way for development of novel fusion proteins against insect pests of crops. *Journal of Molecular Modeling, 19*, 5301-5316. doi: 10.1007/s00894-013-2010-x

Emini, E. A., Hughes, J. V., Perlow, D. S., & Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology, 55*, 836-839. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC255070/

Fischhoff, D. A., Bowdish, K. S., Perlak, F. J., Marrone, P. G., McCormick, S. M., Niedermeyer, … Fraley, R. T. (1987). Insect tolerant transgenic tomato plants. *Nature Biotechnology, 5*, 807-813. doi: 10.1038/nbt0887-807

Flower, D. R. (2013). Designing immunogenic peptides. *Nature Chemical Biology, 9*, 749-753. doi: 10.1038/nchembio.1383

GMO Compass (2016). GMO Database, Genetically modified food and feed: authorization in the EU. Retrieved from http://www.gmo-compass.org/eng/home/, Accessed 12.01.2016.

Guo, C.-H., Zhao, S.-T., Ma, Y., Hu, J.-J., Han, X.-J., Chen, J., & Lu, M.-Z. (2012). *Bacillus thuringiensis* Cry3Aa fused to a cellulase-binding peptide shows increased toxicity against the longhorned beetle. *Applied Microbiology and Biotechnology, 93*, 1249-1256. doi: 10.1007/s00253-011-3523-9

Holton, T. A., Vijayakumar, V., & Khaldi, N. (2013). Bioinformatics: current perspectives and future directions for food and nutritional research facilitated by a Food-Wiki database. *Trends in Food Science & Technology, 34*, 5-17. doi: 10.1016/j.tifs.2013.08.009

Hruby, V. J., & Matsunaga, T. O. (2002). Applications of synthetic peptides. In G. A. Grant (Ed.), *Synthetic peptides: A user's guide*, (pp. 292-376). New York: Oxford University Press.

Ibrahim, M. A., Griko, N., Junker, M., & Bulla, L. A. (2010). *Bacillus thuringiensis*: a genomics and proteomics perspective. *Bioengineered Bugs, 1*, 31-50. doi: 10.4161/bbug.1.1.10519

James, C. (2015). Global status of commercialized biotech/GM crops. Brief No. 51. ISAAA: Ithaca, New York. Retrieved from http://www.isaaa.org

Karplus, P. A., & Schulz, G. E. (1985). Prediction of chain flexibility in proteins. Naturwissenschaften, 72, 212-213. doi: 10.1007/bf01195768

Kashyap, S. (2012). Computational modeling deduced three dimensional structure of Cry1Ab16 toxin from *Bacillus thuringiensis* AC11. *Indian Journal of Microbiology, 52*, 263-269. doi: 10.1007/s12088-011-0191-5

Kolaskar, A. S., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters, 276*, 172-174. doi: 10.1016/0014-5793(90)80535-Q

Kringelum, J. V., Lundegaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol, 8*, e1002829. doi: 10.1371/journal.pcbi.1002829

Kumar, B. S., Ralte, Z., Passari, A. K., Mishra, V. K., Chutia, B. M., Singh, B. P., … Nachimuthu, S. K. (2013). Characterization of *Bacillus thuringiensis* Cry1 class proteins in relation to their insecticidal action. *Interdisciplinary Sciences: Computational Life Sciences, 5*, 127-135. doi: 10.1007/s12539-013-0160-2

Kovjazin, R., Shitrit, D., Preiss, R., Haim, I., Triezer, L., Fuks, L., … Carmon, L. (2013). Characterization of novel multiantigenic vaccine candidates with pan-HLA coverage against *Mycobacterium tuberculosis*. *Clinical and Vaccine Immunology, 20*, 328-340. doi: 10.1128/CVI.00586-12

Larsen, J. E. P., Lund, O., & Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome Research, 2*, 2. doi: 10.1186/1745-7580-2-2

Lundegaard, C., Lund, O., Kesmir, C., Brunak, S., & Nielsen, M. (2007). Modeling the adaptive immune system: predictions and simulations. *Bioinformatics, 23*, 3265-3275. doi: 10.1093/bioinformatics/btm471

Manzanares-Palenzuela, C. L., Mafra, I., Costa, J., Barroso, M. F., de-los-Santos-Álvarez, N., Delerue-Matos, C., … López-Ruiz, B. (2016). Electrochemical magnetoassay coupled to PCR as a quantitative approach to detect the soybean transgenic event GTS40-3-2 in foods. *Sensors and Actuators B: Chemical, 222*, 1050-1057. doi: 10.1016/j.snb.2015.09.013

Marani, M. M., Costa, J., Mafra, I., Oliveira, M. B. P. P., Camperi, S. A., & Leite, J. R. S. A. (2015). In silico peptide prediction for antibody generation to recognize 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) in genetically modified organisms. *Peptide Science, 104*, 91-100. doi: 10.1002/bip.22609
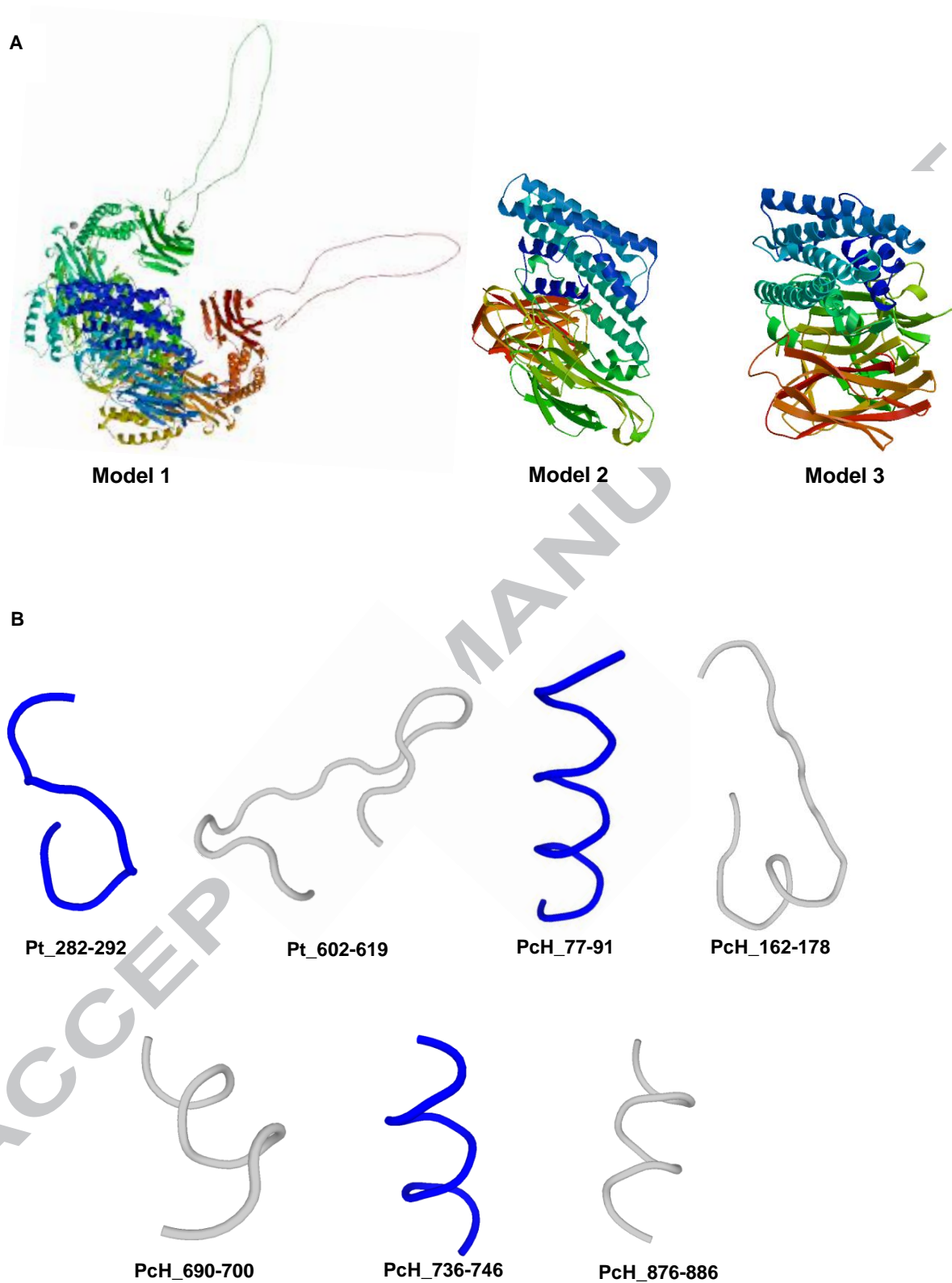
Parker, J. M. R., Guo, D., & Hodges, R. S. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. *Biochemistry, 25*, 5425-5432. doi: 10.1021/bi00367a013

Peptides International (2016). Peptides International FAQs: a general guideline for storage and handling of peptides. Louisville, KY, USA. Retrieved from https://www.pepnet.com/res/uploads/case_studies/peptideguidelines_1.pdf, Accessed 02.02.2016.

Peptide Design Guideline (2011). Peptide Design Guideline version 3.1, revision 2011-05-17. Biomatik Corporation, Ontario, Canada. Retrieved from https://www.biomatik.com/media/service-pdfs/Peptide%20Design%20Guideline.pdf, Accessed 04.02.2016.

Plácido, A., Amaral, J. S., Costa, J., Fernandes, T. J. R., Oliveira, M. B. P. P., Delerue-Matos, C., & Mafra, I. (2016). Novel strategies for genetically modified organism detection. In: R. R. Watson, V. R. Preddy (Eds.), *Genetically modified organisms in food* (pp. 119-131). Waltham: Elsevier.

Ponomarenko, J., Bui, H.-H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics, 9*(1), 1-8. doi: 10.1186/1471-2105-9-514

Purcell, A. W., McCluskey, J., & Rossjohn, J. (2007). More than one reason to rethink the use of peptides in vaccine design. *Nature Reviews Drug Discovery, 6*, 404-414. doi: 10.1038/nrd2224

Rao, P. E., & Kroon, D. J. (1993). Orthoclone OKT3 - chemical mechanisms and functional effects of degradation of a therapeutic monoclonal antibody. In R.
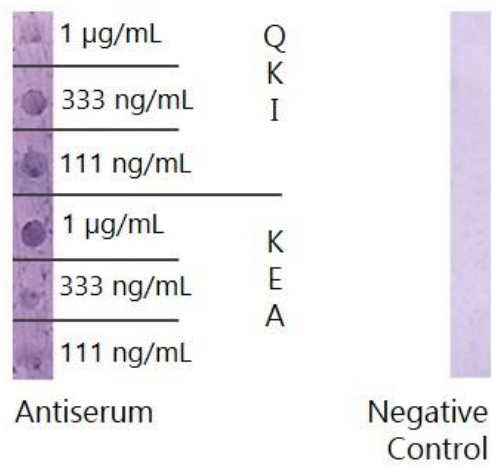
Pearlman, Y. J. Wang (Eds.), *Stability and characterization of protein and peptide drugs* (pp. 135-156). New York: Plenum Press.

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE, 5*, e9862. doi: 10.1371/journal.pone.0009862

Regulation (EC) No. 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. *Official Journal of the European Union*, L 268, 1-37.

Shen, Y., Maupetit, J., Derreumaux, P., & Tufféry, P. (2014). Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *Journal of Chemical Theory and Computation, 10*, 4745-4758. doi: 10.1021/ct500592m

Sigma-Aldrich (2016). Sequence analysis, custom peptide synthesis. St. Louis, MO, USA. Retrieved from http://www.sigmaaldrich.com/life-science/custom-oligos/custom-peptides/learning-center/sequence-analysis.html, Accessed 04.02.2016.

Thinkpeptides eBook (2012). Thinkpeptides eBook: Working with peptides, ST55 V1.0 ProImmune Limited, Sarasota, FL, USA. Retrieved from https://www.proimmune.com/ecommerce/pdf_files/ST55.pdf, Accessed 04.02.2016.

Tung, C.-W., & Ho, S.-Y. (2007). POPI: predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties. *Bioinformatics, 23*, 942-949. doi: 10.1093/bioinformatics/btm061

23

Figure Captions

Fig. 1. Theoretical conformational models of Cry1A(b)16 generated by Swiss-Model software (Table 1) using Cry1A(a) and Cry1A(c) templates with PDB ID: 4W8J (Model 1), 1CIY (Model 2) and 4ARY (Model 3) (A) and secondary structure of the selected peptides using the PEP-FOLD3 software (Shen et al., 2014) (B).

Fig. 2. Dot blot of the antiserum containing the antibodies against each target peptide: QKI and KEA correspond to peptides PcH_736-746 (QKIDESKLKAY) and PcH_876-886 (KEAKESVDAVF), respectively. The negative control was performed with the antiserum collected from rabbit at day 0 of the immunisation schedule (pre-immunisation).

A

Model 1   Model 2   Model 3

B

Pt_282-292   Pt_602-619   PcH_77-91   PcH_162-178

PcH_690-700   PcH_736-746   PcH_876-886

Fig. 1.

Fig. 2

Table 1. List of web server programs used for the prediction assessment.

Name	Function	URL	References
BLAST	Basic Local Alignment Search Tool	http://blast.ncbi.nlm.nih.gov/Blast	-
BioEdit v.7.2.5	Sequence alignment editor	http://www.mbio.ncsu.edu/BioEdit/bioedit.html	-
DiscoTope 2.0	B-cell discontinuous epitope prediction	http://www.cbs.dtu.dk/services/DiscoTope/	Kringelum, Lundegaard, Lund, & Nielsen (2012)
ElliPro	Prediction of linear and discontinuous B-cell epitopes	http://tools.iedb.org/ellipro/	Ponomarenko et al. (2008)
ExPASy Peptide Cutter	Enzymatic cleavage sites	http://web.expasy.org/peptide_cutter/	-
IEDB	B-cell linear epitope prediction	http://tools.iedb.org/bcell/	-
Chou & Fasman Beta-Turn Prediction	Beta-turn prediction	-	Chou and Fasman (1978)
Emini Surface Accessibility Prediction	Accessibility prediction	-	Emini, Hughes, Perlow, and Boger (1985)
Karplus & Schulz Flexibility Prediction	Flexibility prediction	-	Karplus and Schulz (1985)
Kolaskar & Tongaonkar Antigenicity	Antigenicity	-	Kolaskar and Tongaonkar (1990)
Parker Hydrophilicity Prediction	Hydrophobicity prediction	-	Parker, Guo, and Hodges (1986)
Bepipred Linear Epitope Prediction	Linear epitope prediction	-	Larsen, Lund, and Nielsen (2006)
NCBI	DNA and protein sequences	http://www.ncbi.nlm.nih.gov	-
PEP-FOLD3	Peptide structure prediction	http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA#forms::PEP-FOLD3	Shen, Maupetit, Derreumaux, and Tufféry (2014)
POPI 2.0	T-cell epitope prediction	http://e045.life.nctu.edu.tw/POPI/	Tung and Ho (2007)
RCSB PDB	(Research Collaboratory for Structural Bioinformatics-Protein Data Bank. Information of 3D Structure	http://www.rcsb.org/pdb/home/home.do	-
SWISS-MODEL	SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information	http://swissmodel.expasy.org/interactive	Biasini et al. (2014)
UniProt	Protein primary/secondary structures	http://www.uniprot.org/	-

Table 2. Peptides selected by the first strategy based on initial compositional and solubility analyses, followed by B-cell and T-cell epitope predicting tools.

Peptide	Sequence	Predicted B-cell epitopes			T-cell epitope prediction	
		Emini et al. (1985) (surface accessibility >1.0)	Kolaskar and Tongaonkar (1990) (allergenicity > 1.0)	BepiPred (Larsen et al., 2006) (epitope prediction > 0.4)	MHC-I (CTL)	MHC-II (HTL)
Pt_602-619	IEFVPAEVTFEAEYDLER	IEFVPAEVTFEAEYDLER	IEFVPAEVTFEAEYDLER	-	Moderate	Little
PcH_77-91	LVQIEQLINQRIEEF	-	LVQIEQLINQRIEEF	-	None	Moderate
PcH_690-700	RGINRQLDRGW	RGINRQLDRGW	-	RGINRQLDRGW	Moderate	None
PcH_736-746	QKIDESKLKAY	QKIDESKLKAY	QKIDESKLKAY	QKIDESKLKAY	None	High
PcH_876-886	KEAKESVDAVF	KEAKESVDAVF	KEAKESVDAVF	KEAKESVDAVF	High	Little

Shadow letters evidence the residues that are in good agreement with the criteria defined in the respective column, namely surface accessibility >1.0; allergenicity > 1.0 and epitope prediction > 0.4

Table 3. Peptides selected by the second strategy based on initial T-cell and B-cell epitope predicting tools, followed by compositional and solubility analyses.

Peptide	Sequence	T-cell epitope prediction		Predicted B-cell epitopes		
		MHC-I CTL	MHC-II HTL	Emini et al. (1985) (surface accessibility >1.0)	Kolaskar and Tongaonkar (1990) (allergenicity > 1.0)	BepiPred (Larsen et al., 2006) (epitope prediction > 0.4)
Pt_282-292	GSAQGIEGSIR	None	High	-	-	GSAQGIEGSIR
PcH_162-178	VQAVNLHLSVLRDVLVF	High	None	-	VQAVNLHLSVLRDVLVF	-
PcH_736-746	QKIDESKLKAY	None	High	QKIDESKLKAY	QKIDESKLKAY	QKIDESKLKAY
PcH_876-886	KEAKESVDAVF	High	Little	KEAKESVDAVF	KEAKESVDAVF	KEAKESVDAVF

Shadow letters evidence the residues that are in good agreement with the respective criteria defined in the column, namely surface accessibility >1.0; allergenicity > 1.0 and epitope prediction > 0.4

28

Table 4. Determination of discontinuous B-cell epitopes by ElliPro (Ponomarenko et al., 2008) and DiscoTope 2.0 (Kringelum et al., 2012) softwares using the proposed 3 models for Cry1A(b)16 generated with 4W8J, 1CIY and 4ARY as respective templates.

Peptide	Sequence	ElliPro			
		Model 1	Model 2	Model 3	Model
Pt_282-292	GSAQGIEGSIR	282-287; 290; 291[a] (0.715)[b]	282; 283 (0.709)	282; 283 (0.641)	282-289
Pt_602-619	IEFVPAEVTFEAEYDLER	608-610; 613; 614; 617 (0.638)	607-610 (0.718)	606; 607 (0.712)	-
PcH_77-91	LVQIEQLINQRIEEF	77-79; 81-91 (0.715)	81; 82; 86-91 (0.742)	78; 81; 84; 91 (0.743)	82; 85-87; 9
PcH_162-178	VQAVNLHLSVLRDVLVF	-	175; 178 (0.742)	175; 177; 178 (0.743)	-
PcH_690-700	RGINRQLDRGW	691-700 (0.638)	NA[c]	NA	690-69
PcH_736-746	QKIDESKLKAY	739-746 (0.638)	NA	NA	739; 741;
PcH_876-886	KEAKESVDAVF	876; 879; 880; 883-886 (0.637)	NA	NA	876; 87

[a]position of the discontinuous epitopes within the selected peptide; [b]value within brackets refer to the protrusion index in ElliPro software; [c]NA – not applicable due to the absence of a 3D-structure.

Highlights:

- *In silico* analysis of Cry1A(b)16 toxin of transgenic maize

- Identification of peptide markers for the biorecognition of transgenic crops

- T- and B-cell prediction tools to assess linear epitopes of Cry-1-type proteins

- Conformational epitope prediction using three Cry1A(b)16 computational models

- Two peptides were selected, synthesised and used for polyclonal antibody production