In Silico Peptide Prediction for Antibody Generation to Recognize 5-Enolpyruvylshikimate-3-Phosphate Synthase (EPSPS) in Genetically Modified Organisms

Mariela M. Marani,[1] Joana Costa,[1,2] Isabel Mafra,[2] Maria Beatriz P.P. Oliveira,[2] Silvia A. Camperi,[3] José Roberto de Souza Almeida Leite[4]

[1]*CENPAT-CONICET, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas, Puerto Madryn, Argentina*

[2]*REQUIMTE, Departamento de Ciências Químicas, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal*

[3]*Instituto NANOBIOTEC UBA-CONICET, Cátedra de Biotecnología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Buenos Aires, Argentina*

[4]*Núcleo de Pesquisa em Biodiversidade e Biotecnologia, Biotec, Campus Parnaíba, Universidade Federal do Piauí, Parnaíba, PI, Brazil*

Received 30 September 2014; revised 15 December 2014; accepted 5 January 2015
Published online 23 January 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22609

ABSTRACT:

For the prospective immunorecognition of 5-enolpyruvylshikimate-3-phosphate synthase (CP4-EPSPS) as a biomarker protein expressed by transgenic soybean, an extensive in silico evaluation of the referred protein was performed. The main objective of this study was the selection of a set of peptides that could function as potential immunogens for the production of novel antibodies against CP4-EPSPS protein. For this purpose, the protein was in silico cleaved with trypsin/chymotrypsin and the resultant peptides were extensively analyzed for further selection of the best candidates for antibody production. The analysis enabled the successful proposal of four peptides with potential immunogenicity for their future use as screening biomarkers of genetically modified organisms. To our knowledge, this is the first attempt to select and define potential linear epitopes for the immunization of animals and, subsequently, to generate adequate antibodies for CP4-EPSPS recognition. The present work will be followed by the synthesis of the candidate peptides to be incubated in animals for antibody generation and potential applicability for the development of an immunosensor for CP4-EPSPS detection. © 2015 Wiley Periodicals, Inc. Biopolymers (Pept Sci) 104: 91–100, 2015.

Keywords: *GMO; in silico epitope prediction; CP4-EPSPS; bioinformatics*

Correspondence to: Mariela M. Marani, CENPAT-CONICET, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina and Joana Costa, REQUIMTE, Faculty of Pharmacy, University of Porto, Portugal; e-mail: mmarani@cenpat-conicet.gob.ar and joanabcosta@gmail.com
Mariela M. Marani and Joana Costa contributed equally to the work.
Contract grant sponsor: Marie Curie Actions—International Research Staff Exchange Scheme FP7-PEOPLE-2013-IRSES entitled "GMOsensor—Monitoring Genetically Modified Organisms in Food and Feed by Innovative Biosensor Approaches"
Contract grant number: 612545
Contract grant sponsor: Fundação para a Ciência e a Tecnologia (FCT) by POPH-QREN (subsidized by FSE and MCTES)
Contract grant number: PEst-C/EQB/LA0006/2013

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of any preprints from the past two calendar years by emailing the Biopolymers editorial office at biopolymers@wiley.com.

INTRODUCTION

For a long time, the use of genetically modified organisms (GMO) has been a topic of extensive discussion and contradictory opinions, not only among researchers, but also among the general population. The first generation of biotechnological crops concerned the introduction of characteristics such as herbicide tolerance, insect resistance, or a combination of both. So, the benefits of using transgenic crops are mainly focused on agronomical, technological, and utilitarian issues.[1] The second generation of GMO involves enhanced quality traits (e.g. production of higher nutrient contents),[2] with some modifications being recently approved for commercialization, planting, and/or trade in for food and feed use.[3] The third-generation plants was designed to produce special substances for pharmaceutical or industrial purposes.[2] However, the current production of second and third-generation crops is almost negligible compared with the ones with agronomical traits. The potential risks related to immediate and/or delayed adverse effects on human health (e.g. introduction of novel allergenic proteins or synthesis of toxic compounds), environment and farmer's economy are negative issues regarding GMO that cannot be ruled out.[4–7] Despite all the pros and cons associated with the production and use of GMO for food and feed, in 2013 the area of biotech crops continued to increase, totalizing 175.2 million hectares. From the biotech area, more than 85 million hectares concerned to transgenic soybean, corresponding to approximately 79% of the total soybean production in 2013.[8]

Currently, there are 30 events of transgenic soybean presenting one or more traits, which include tolerance to different herbicides (2,4-D, dicamba, glufosinate, glyphosate, isoxaflutole, mesotrione, and sulfonylurea), resistance to antibiotics, resistance to lepidopteran insects, enhanced photosynthesis/yield, modified oil/fatty acid and introduction of visual markers.[3] However, from the existing GM soybean events, only seven have already been approved and other 14 are either submitted or awaiting authorization for food and feed, importation/processing, or cultivation inside the European Union (EU).[9,10]

One of the most representative GM crops is the soybean GTS 40-3-2 event, which is commercially known as Roundup Ready™ (RR) soybean since it was especially developed to be tolerant to the herbicide with the same name. This event contains an introduced construction with a gene of cp4 epsps (aroA:CP4) from the *Agrobacterium tumefaciens* strain CP4 expressing the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) that decreases the binding affinity to glyphosate, conferring increased tolerance to the glyphosate herbicide.[3,11,12] Many other transgenic constructions also include the referred gene expressing the CP4-EPSPS protein, which can be a potential marker for GMO screening. For this purpose, the immunorecognition of proteins relying on the interaction between an antibody and a target antigen (e.g. transgenic proteins or peptides) represents an important tool. Lateral flow devices and plate-based enzyme-linked immunosorbent assays (ELISA) are the most prominent examples for protein recognition.[13] Lately, the application of biosensors as alternative platforms for the detection of GMO in foods has become one of the most emerging and attractive fields. Biosensor technology has been highlighted as one of the most promising ways to solve pertinent issues in relation to simple, fast, reproducible, and low cost multitarget detection, with special focus for potential automation.[14] In general, most attention has been devoted to the detection of transgenic DNA sequences by biosensor technology.[13,15,16] However, its potential application for protein detection has hardly been evaluated. For the specific detection of CP4-EPSPS, some monoclonal antibodies can be commercially acquired, but the CP4-EPSPS protein is currently unavailable, hindering the development of immunodiagnostic-based methods.

In this work, an extensive in silico evaluation of CP4-EPSPS expressed by RR soybean was performed for the future development of an immunosensor as a rapid and cost-effective method for the direct screening of GMO. The main objective of this study was the selection of a set of peptides that could function as potential immunogens for the production of novel antibodies for the CP4-EPSPS biorecognition. Generally, antibodies are capable of recognizing specific regions (epitopes) of intact proteins and/or small peptides. While B-cell epitopes can be linear or discontinuous (spatially in folded proteins), T-cell epitopes are short linear peptides that are presented in the context of major histocompatibility complex (MHC) proteins.[17]

Nowadays, with the advances of bioinformatics, there is a diversified set of tools for the prediction of linear and conformational epitopes in proteins, mostly based on T- and B-cell prediction softwares. So far, the use of bioinformatics to predict the potential immunogenicity of peptides has been mainly focused on medical purposes, with some successful applications already described in the development of vaccines.[18,19] Using bioinformatics, the CP4-EPSPS protein was in silico cleaved with trypsin/chymotrypsin and the resultant peptides were extensively analyzed for further selection of the best candidate peptides for antibody production.

MATERIAL AND METHODS

Sequence Retrieval

The protein sequence corresponding to CP4-EPSPS (3-phosphoshikimate 1-carboxyvinyltransferase or 5-enolpyruvylshikimate-3-phosphate synthase) was retrieved from the NCBI database[20] with the accession number Q9R4E4.2. This enzyme is naturally present in

```
1   MSHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTGKAMQAMGARI  70
71  RKEGDTWIIDGVGNGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFIGDASLTKRPMGRVLNPLREM 140
141 GVQVKSEDGDRLPVTLRGPKTPTPITYRVPMASAQVKSAVLLAGLNTPGITTVIEPIMTRDHTEKMLQGF 210
211 GANLTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGLILT 280
281 LQEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAEGATVMNGLEEL 350
351 RVKESDRLSAVANGLKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFLVMGLVSENP 420
421 VTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA                                     455
```

FIGURE 1 Sequence of 5-enolpyruvylshikimate-3-phosphate synthase of *Agrobacterium* spp. CP4 (CP4-EPSPS) retrieved from NCBI[20] database with accession number Q9R4E4.2.

Agrobacterium spp., being encoded by a gene located in locus AROA_AGRSC. The entire primary sequence of CP4-EPSPS is composed of 455 amino acid residues and is represented in Figure 1.

Trypsin and Chymotrypsin Cleavage Sites

ExPASy PeptideCutter tool[21] was used to submit CP4-EPSPS to an in silico cleavage site prediction by trypsin and chymotrypsin of high (HS) and low (LS) specificity. A cleavage site map and table of cleavage site positions was analyzed and peptides between 10 and 30 amino acids were selected to assess compositional analysis.

Structure Determination of Epitopes

B-Cell Epitope Prediction Tools. Sequence and structured-based in silico methods for epitope prediction were used. Immune Epitope Data Base (IEDB)[22] tools were used to predict linear B-cell epitopes based on sequence characteristics of CP4-EPSPS (Table I), while DiscoTope 2.0[29] server was used to predict B-cell epitopes from protein tridimensional structure. Different tools verified the distinct features of the sequence to predict candidate peptides and/or assign scores (Table I). Selected CP4-EPSPS peptides were compared with predicted peptides, considering the scores and parameters attained in each tool.

T-Cell Epitope Prediction Tools. Physicochemical properties based on web server POPI 2.0[30] were used to predict immunogenicity of MHC class I and class II. Nine peptides selected from CP4-EPSPS enzyme, after its theoretical cleavage with trypsin/chymotrypsin and compositional analysis (Table II), were submitted to POPI 2.0[30] to predict the immunogenicity of MHC class I and class II.

Cross-Reactivity. The cross-reactivity of the selected peptides was determined using the Standard Protein Basic local alignment search tool (Blastp).[31] For a complete evaluation of soybean CP4-EPSPS peptides, all search set databases were used, namely the non-redundant protein sequences nr (GenBank CDS translations + PDB + SwissProt + PIR + PRF, excluding those in env_nr), the NCBI protein reference sequences, the nonredundant UniProtKB/SwissProt sequences, the protein sequences derived from the patent division of GenBank, the proteins from Protein Data Bank (PDB), the proteins from whole genome shotgun (WGS) metagenomic projects and the transcriptome shotgun assembly (TSA) sequences.

Secondary Structure Prediction of Selected Peptides. Each selected in silico sequence was analyzed through the web resources for 3D structure prediction, namely PEP-FOLD (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/). This approach allowed us to obtain a predictive structural model for the selected sequences.

RESULTS AND DISCUSSION

Sequence Retrieval

To select the appropriate protein sequence produced by RR soybean, the keywords 5-enolpyruvylshikimate-3-

Table I List of Web Server Programs used for the Prediction Assessment

Name	Function	URL (Last Accessed on April 2014)
NCBI	Sequence	http://www.ncbi.nlm.nih.gov/protein/
BLAST®	Align sequences	http://blast.ncbi.nlm.nih.gov/Blast
RCSB PDB	3D Structure	http://www.rcsb.org/pdb/home/home.do
ExPASy Peptide Cutter	Potential cleavage sites	http://web.expasy.org/peptide_cutter/
IEDB	B cell linear epitope prediction	http://tools.immuneepitope.org/main/html/bcell_tools.html
Chow and Fastman[23]	B-turn prediction	
Emini et al.[24]	Accessibility prediction	
Karplus and Schulz[25]	Flexibility prediction	
Kolaskar and Tongaonkar[26]	Antigenicity	
Parker et al.[27]	Hydrophobicity prediction	
Larsen et al.[28]	Linear epitope prediction	
DiscoTope 2.0[29]	B cell discontinuous epitope prediction	http://www.cbs.dtu.dk/services/DiscoTope/
POPI 2.0[30]	T cell epitope prediction	http://iclab.life.nctu.edu.tw/POPI/
PEP-FOLD	Peptide structure prediction	http://bioserv.rpbs.univ-paris-diderot.fr/cgi-bin/PEP-FOLD

Table II Enzyme-Cleaved Peptides of CP4-EPSPS

Assigned Peptide Name	Position of Cleavage Site	Name of the Cleaving Enzyme	Resulting Peptide Sequence	Peptide Length (aa)
Pc_19-32	32	Chymotrypsin-LS	SGTVRIPGDKSISH	14
Pc_41-50	50	Chymotrypsin-LS	ASGETRITGL	10
Pt_47-61	61	Trypsin	ITGLLEGEDVINTGK	15
Pc_52-63	63	Chymotrypsin-LS	EGEDVINTGKAM	12
Pc_67-77	77	Chymotrypsin-LS	GARIRKEGDTW	11
Pc_141-156	156	Chymotrypsin-LS	GVQVKSEDGDRLPVTL	16
Pc_157-167	167	Chymotrypsin-LS	RGPKTPTPITY	11
Pc_285-295	295	Chymotrypsin-LS	GADIEVINPRL	11
Pt_295-305	305	Trypsin	LAGGEDVADLR	11
Pc_312-324	324	Chymotrypsin-LS	KGVTVPEDRAPSM	13

phosphate + synthase + CP4 + EPSPS were used to search at NCBI Entrez protein database.[20] From this search, eighteen different accession numbers for protein CP4-EPSPS were found in the database. Comparison among distinct sequences was performed and the selection of the CP4-EPSPS sequence was made based on the available data retrieved from Monsanto's documents.[12]

The 18 sequences were further aligned and the parameters of query cover (QC) and identity (%) were analyzed. Because the genetic modification described in Monsanto's documents corresponded to the expression of the protein sequence with the NCBI accession number Q9R4E4.2, this sequence was selected for further analysis. The alignment of Q9R4E4.2 sequence with the rest of the known sequences showed query covers between 100% and 95% and identity between 99% and 28%.

The CP4-EPSPS enzyme has been the focus of several studies and the information regarding its tridimensional structure is available in the archive of structural data of biological macromolecules.[32,33] With the RCSB-PDB[32] accession number 2GG4, the tridimensional structure of CP4-EPSPS was also used in this study.

Enzyme Cleavage and Peptide Composition Analysis

Synthetic peptides are often involved in the generation of antibodies raised against a specific peptide since they can be used as reagents in various innovative analytical techniques.[34] Inserted in a broad collaborative project (GMOsensor)[35] concerning the development of sensor-based methodologies for the effective detection of soybean CP4-EPSPS in foods, the main objective of this study was to provide the best peptides for antibody production. For the specific case of the development of an immunosensor for CP4-EPSPS identification in foods, the production of an antibody with sufficient affinity and specificity is of utmost importance (Figure 2). To define the best peptide candidates for animal immunization, several parameters had to be carefully considered. Length, solubility, peptide composition, and predicted immunogenicity are only some of the topics of crucial relevance to ensure high-titer custom production of highly-specific antibodies. As one of the aims of the project is the detection of CP4-EPSPS-derived peptides, the in silico cleavage was performed with trypsin, chymotrypsin-HS and chymotrypsin-LS. Among the most common, these enzymes were chosen for their high specificity. For experimental purposes, trypsin and chymotrypsin are more cost-effective than other specific enzymes such as Asp-N protease or endopeptidase Arg-C. With respect to thermolysin or proteinase k, their multisite cleavage action result in many short peptides (<10 amino acids), which are not useful for antibody production.

Singular differences exist in the nature of the peptides that bind the MHC-I or MCH-II. Peptides that are released from class I molecules have a typical length of 9 to 11 amino acids, while those released from class II molecules could oscillate between 10 and 30 amino acids or more.[36,37] According to this criterion, 10 peptides were obtained simulating trypsin cleavage. Eighteen and four peptides were obtained by chymotrypsin-LS and chymotrypsin-HS cleavage, respectively.

To minimize any potential difficulties during future peptide synthesis, as well as to allow reducing the cost and increase the yields of the empiric process, peptide composition was extensively analyzed. Accordingly, certain residues and regions were avoided to select the best peptide candidate for animal immunization. Peptides presenting residues of cysteine (Cys) and methionine (Met) were avoided, as these amino acids are susceptible of rapid oxidation and because Cys is involved in disulphide formation.[38,39] Sequences containing asparagine (Asn) at the N-terminal position were also excluded due to the difficulty of removing the N-terminal protecting group.[39,40] Peptides with aspartate-glycine (Asp-Gly), aspartate-proline (Asp-Pro), and aspartate-serine (Asp-Ser) pairs were also not selected because they are prone to hydrolysis under acidic conditions.[39,41,42] Peptides exhibiting glutamine (Gln), isoleucine

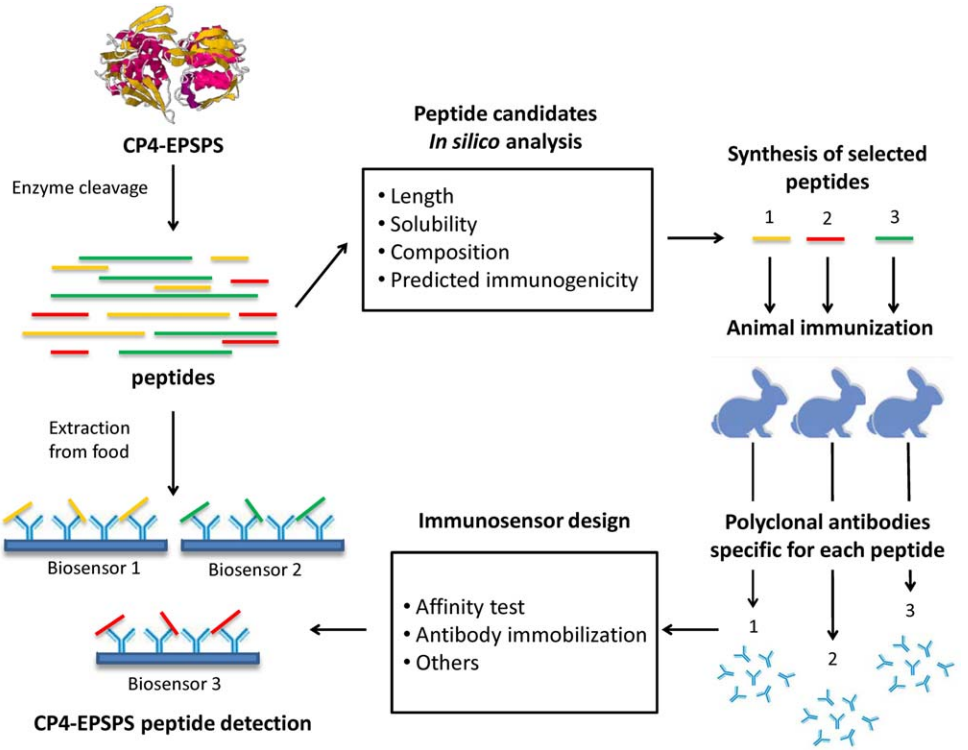

FIGURE 2 Schematic procedure for the development of sensor-based methodologies for soybean CP4-EPSPS detection in foods.

(Ile), leucine (Leu), phenylalanine (Phe), threonine (Thr), or valine (Val) series or multiple Ser residues were discarded to avoid low peptide purity and deletions during synthesis.[42,43]

Solubility was another parameter evaluated and considered of great importance in the selection of best candidate peptides. Peptides with one charged residue out of five amino acids were chosen since over weighted peptides with hydrophobic residues are likely to difficult their water solubility.[40,44,45]

From the initial set of 32 peptides, 10 were selected, being 2 from protein cleavage with trypsin and 8 from chymotrypsin-LS (Table II). Peptides obtained from trypsin cleavage are designated as Pt and from Chymotrypsin-LS as Pc. None of the four peptides obtained after the in silico CP4-EPSPS cleavage with chymotrypsin-HS was selected after the composition analysis.

Structure Determination of B-Cell Epitopes

The methods of predicting B-cell epitopes are commonly divided in two categories, depending on the information used to do the prediction: the protein sequence or the protein tridimensional structure.[29] In general, sequence-based methods include calculations from hydrophilicity, flexibility, beta-turns, and surface accessibility parameters,[27,46,47] and, more recently, amino acid composition and cooperativeness.[48–50] Although the referred methods perform adequate epitope predictions of continuous stretch of amino acid (linear epitopes), in terms of conformational epitopes these methods fail to provide proper peptide prediction.[29] To overcome the drawbacks of the methods based on protein sequence, recent methods such as DiscoTope[29] or ElliPro[22] have contributed with the inclusion of structural data (protein tridimensional structure) to enhance in silico epitope prediction.

Based on Sequence. The results from the in silico study of CP4-EPSPS sequence using different tools to predict linear B-cell epitopes were compared with the previously selected enzyme-cleaved peptides (Table III). Nine out of the 10 in silico predicted peptides have at least one predicted B-cell epitope in their sequence. Ten peptides from CP4-EPSPS with sizes between 6 and 13 amino acids were estimated to have surface accessibility above 1.0 by Emini et al.[24] tool, indicating a strong probability for the epitope to be found on protein surface. From this analysis, a total of five sequences were completely contained in Pc_19–32, Pc_67–77, Pc_141–156, Pc_157–167, and Pc_312–324 peptides (Table III). The antigenicity of the CP4-EPSPS was evaluated with Kolaskar and Tongaonkar software,[26] enabling predicting 21 peptides with variable sizes (6 to 23 amino acids). Considering a value of at least 1.0 for the antigenicity parameter,[26] those estimated fragments were

Table III Predicted B-Cell Epitopes Contained in Enzyme-Cleaved Selected Peptides of CP4-EPSPS

Protein position	Enzyme-cleaved selected peptides of CP4-EPSPS.	Predicted B-cell epitopes completely or partially contained in enzyme-cleaved peptides of CP4-EPSPS		
		Emini et al. tool[24] (surface accessibility > 1.0)	Kolaskar and Tongaonkar[26] (allergenicity > 1.0)	Larsen et al.[28] (epitope prediction > 0.4)
Pc_19-32	SGTVRIPGDKSISH	PGDKSISH (25-32)[a]	SGTVRIP (19-25)	VRIPGDKSI (22-30)
Pc_41-50	ASGETRITGL	-	-	ASGETR (41-46)
Pt_47-61	ITGLLEGEDVINTGK	-	-	DVIN (55-58)
Pc_52-63	EGEDVINTGKAM	-	-	-
Pc_67-77	GARIRKEGDTW	RIRKEG (69-74)	-	IRKE (70-73) TW (76-77)
Pc_141-156	GVQVKSEDGDRLPVTL	VKSEDGD (144-150)	EMGVQVK (139-145), RLPVTLR (151-157)	QVKSEDGDRL (143-152)
Pc_157-167	RGPKTPTPITY	RGPKTPTPITYRV (157-169)	PITYRVPMASAQVKSAVLLAGL (164-185)	LRGPKTPTPITY (156-167)
Pc_285-295	GADIEVINPRL	-	DIEVINPR (287-294)	-
Pt_295-305	LAGGEDVADLR	-	VADLRVRSS (301-309)	LAGGEDVA (295-302)
Pc_312-324	KGVTVPEDRAPSM	VPEDRA (316-321)	-	VTVPEDRAPSMI (314-325)

Emini et al.,[24] Kolaskar and Tongaonkar,[26] and Larsen et al.[28] completely or partially predicted peptide sequences contained in enzyme-cleaved selected peptides sequences are indicated in shadows letters.

[a] Values in brackets represent amino acid positions within the protein.

compared with the selected candidate peptide. From those, two sequences were found to be completely contained in the Pc_19-32, and Pc_285-295 peptides; and another four sequences were partially included in the Pc_141-156, Pc_157-167, and Pt_295-305 peptides (Table III).

Hydrophilicity and flexibility are two of the most extensively used algorithms to predict the antigenicity of the amino acid residues once they correlate accessible and flexible surface sites with antibody interaction. Parker et al.[27] tool is used to estimate hydrophilicity, while Karplus and Schulz[25] tool evaluates chain flexibility by analyzing protein sequence. All selected enzyme-cleaved peptides showed high hydrophilicity ($\geq$1.5) and flexibility (1.0) scores with the exception of peptide Pc_285-295, with the highest hydrophilicity score (7.1) belonging to peptide Pc_141-156.

Using a different tool for linear epitope prediction[28] that combines the hidden Markov model with one of the best propensity scale methods described by Parker et al.,[27] 18 candidate peptides with 4 to 19 amino acids in length were proposed. After comparison with the selected candidate peptides and considering an epitope prediction parameter above 0.4, seven peptides were found to be completely, and two almost completely, contained in enzyme-cleaved selected sequences: Pc_19–32, Pc_41–50, Pt_47–61, Pc_67-77, Pc 141-156, Pc_157-167, Pt_295-305 and Pc_312-324 (Table III).

Based on Three-Dimensional Structure. Seven three-dimensional structure hits were available at RSCB-PDB[32] when 5-enolpyruvylshikimate-3-phosphate+synthase+CP4+EPSPS keywords were entered for search. Structure 2GG4 was selected to predict B-cell epitopes because it was the only conformation available as unliganded form and the only sequence obtained by X-ray diffraction with a resolution of 2.10 Å. Figure 3 presents some of the enzyme-cleaved peptides on the surface of the CP4-EPSPS structure with the PDB accession number 2GG4.

DiscoTope 2.0 server[29] was applied to predict the discontinuous B-cell epitopes from the three dimensional structure of CP4-EPSPS. The method uses the calculation of surface accessibility (estimated in terms of contact numbers) and the novel epitope propensity amino acid score. Final scores were assessed by combining the propensity scores of residues in spatial proximity and the contact numbers. A threshold of −3.7 was used, corresponding to a sensitivity of 0.47 (47% of the epitope residues were predicted as part of epitopes) and a specificity of

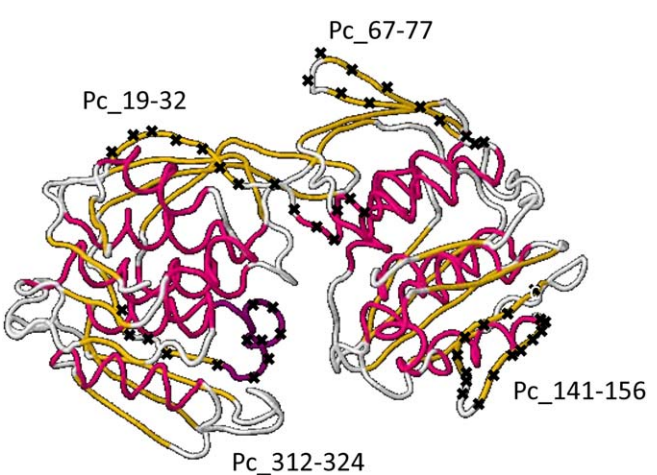

FIGURE 3 Pc_19-32, Pc_67-77, Pc_141-156, and Pc_312-324 enzyme-cleaved peptides on the surface of the tridimensional structure model of CP4-EPSPS (RCSB-PDB[32] accession number 2GG4).

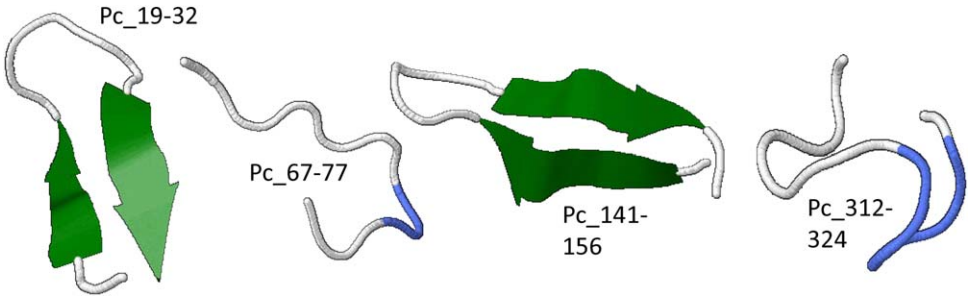

FIGURE 4 Structural models of the peptides Pc_19-32, Pc_67-77, Pc_141–156, and Pc_312-324. The pdb secondary structure models were obtained using the Web resources PEP-FOLD (http://bioserv.rpbs.univparis-diderot.fr/cgi-bin/PEP-FOLD).

0.75 (25% of the nonepitope residues were predicted as part of epitopes). According to these parameters, eight residues out of a total of 443 were identified as likely to be involved in B-cell epitopes, from which three (positions 73–75) were located in Pc_66-77 sequence. When the threshold was decreased to −7.7, 35 residues were identified as likely to be involved in B-cell epitopes, from which 13 were included in 4 of the selected peptides, namely Pc_66-77 (positions 71–75), Pc_141-156 (positions 145, 147-149), Pc_157-167 (positions 160–161), and Pc_312-324 (positions 319, 322).

Besides the DiscoTope 2.0,[29] the ElliPro (antibody epitope prediction) software[22] was also used. It predicts linear and discontinuous epitopes for antibody recognition based on a protein antigen's tridimensional structure. ElliPro accepts either a protein structure or a protein sequence as an input, although if a protein sequence is used the software will predict its tridimensional conformation by homology modeling. ElliPro associates each predicted epitope with a score, defined as a PI (Protrusion Index) value averaged over epitope residues. For each amino acid, a PI value is defined as percentage of the protein atoms enclosed in the ellipsoid that approximates the protein surface, at which the residue first becomes lying outside the ellipsoid. The tridimensional structure of CP4-EPSPS with PDB accession number 2GG4 was analyzed with ElliPro, setting the epitope prediction parameters for a minimum score of 0.6, as recommended by the software. From this analysis six out of the ten selected candidate peptides, namely Pc_19-32 (positions 19–21), Pc_41–50 (positions 42–46), Pc_67-77 (positions 67–77), Pc_141–156 (positions 141–156), Pc_157-167 (positions 157–167), and Pc_312-324 (positions 312–324), exhibited several residues with high probability of being located near the surface of the protein (Figure 3).

Structure Determination of T-Cell Binding Peptides

The induction of an effective T-cell response against immunogens requires antigen processing and peptide presentation by antigen-presenting cells. Additionally, it is well-established that T-cells recognize peptide sequences coupled with appropriate MHC molecules.[51] The recognition of peptide-MHC by the T-cell receptors is part of the regulated effective immune response, whose interaction depends on both the MHC (I and II) and the presented peptide. The discovery of MHC-binding complexes in proteins contributed to the development of several algorithms predicting MHC class I- and II-restricted epitopes for presentation to CD8+ or CD4+ T-cells, respectively, leading to recent advances in the research related to peptide-based approach.[51,52] Immunogenicity (T-cell recognition) is an important factor that determines if a certain peptide-MHC can be targeted in an immune response. So far, several studies demonstrate that some peptides are more immunogenic than others, indicating that they are more likely to be T-cell epitopes.[53] Based on this concept, the immunogenicity of the peptides seems to be increased by the presence of certain amino acids in the sequence (e.g. large and aromatic residues). In addition, positions 4 to 6 of peptides are thought to be highly relevant for their immunogenicity.[53] With respect to the selection of best candidate peptides, the interaction of the B-cells with T-helper cells is considered of great importance to produce the full activation of the B-cells and consequently generating specific antibodies. Until now, most of the studies focusing on the design and consequent use of synthetic peptides for antibody production mainly regard the development of effective vaccines.[18,51,54] Using three algorithms, SYFPEITHI, BIMAS and NetMHCII to predict sequences capable of binding to MHC class I and II molecules in compatible BALB/c mice, the authors were able to determine epitopes in each protein and design multi-epitope peptides capable of inducing peptide-specific T-cell proliferation and cytokine production by CD4+ and/or CD8+ T-cells.[18]

The selected peptides were analyzed with POPI 2.0[30] program that utilizes physicochemical properties to predict immunogenic sequences. After the analysis, 4 of the 10 selected

Table IV Predicted T-Cell Epitopes by POPI 2.0[30] and Cross-Reactivity Analysis

Protein Position	Enzyme	T-Cell Epitope Prediction (Immunogenicity) MHC-1 (CTL)	MHC-2 (HTL)	Cross-Reactivity
Pc_19-32	Chymotrypsin-LS	Moderate	None	100% of sequence identity with soybean CP4-EPSPS and wheat EPSPS (accession no AEK64772.1), ≤92% of sequence identity with other EPSPS resistant to glyphosate
Pc_41-50	Chymotrypsin-LS	None	None	100% identity with soybean CP4-EPSPS and wheat EPSPS (accession no AEK64772.1), as well as other EPSPS from different organisms
Pt_47-61	Trypsin	None	None	100% of sequence identity with soybean CP4-EPSPS and wheat EPSPS (accession no AEK64772.1), ≤92% of sequence identity with other EPSPS resistant to glyphosate
Pc_52-63	Chymotrypsin-LS	None	None	100% of sequence identity with soybean CP4-EPSPS and wheat EPSPS (accession no AEK64772.1), ≤92% of sequence identity with other EPSPS resistant to glyphosate
Pc_67-77	Chymotrypsin-LS	Moderate	None	100% of sequence identity with soybean CP4-EPSPS and wheat EPSPS (accession no AEK64772.1), ≤91% of sequence identity with other EPSPS resistant to glyphosate
Pc_141–156	Chymotrypsin-LS	Moderate	Moderate	100% of sequence identity with soybean CP4-EPSPS and wheat EPSPS (accession no AEK64772.1), ≤94% of sequence identity with other EPSPS resistant to glyphosate
Pc_157-167	Chymotrypsin-LS	None	None	100% of sequence identity with soybean CP4-EPSPS and wheat EPSPS (accession no AEK64772.1), ≤91% of sequence identity with other EPSPS resistant to glyphosate
Pc_285-295	Chymotrypsin-LS	None	None	100% of sequence identity with soybean CP4-EPSPS, ≤82% of sequence identity with other EPSPS resistant to glyphosate
Pt_295-305	Trypsin	None	None	100% of sequence identity with soybean CP4-EPSPS, wheat EPSPS (accession no AEK64772.1) and with other EPSPS resistant to glyphosate
Pc_312-324	Chymotrypsin-LS	Moderate	None	100% of sequence identity with soybean CP4-EPSPS, ≤92% of sequence identity with wheat EPSPS (accession no AEK64772.1) and other EPSPS resistant to glyphosate

peptides presented moderate immunogenicity for the MHC-I cytotoxic T-cell (CTL) (Figure 4) and only one peptide exhibited moderate immunogenicity to MHC-II helper T lymphocytes (HTL) (Table III). Using the prediction tool reported by Calis et al.[53] that evaluates MHC-I immunogenicity, seven out of ten peptides had positive scores, being the highest for the Pt_47–61, Pc_285-295, and Pc_41-50 peptides.

Cross-Reactivity

Data from cross-reactivity is included in Table IV. Information retrieved from Blastp search suggests high sequence identity of all selected peptides with wheat CP4-EPSPS (NCBI[20] accession no AEK64772.1). Besides this protein, data also evidence high sequence identity with other organisms, mostly bacteria that produces the EPSPS enzyme and other genetically modified plants, although sequence homology is below 100%. The high sequence identity with EPSPS enzymes of other transgenic plants makes the proposed peptides as adequate candidates for GMO screening.

CONCLUSION

Antibodies bind to antigens (e.g. proteins) at sites known as B-cell epitopes. Therefore, the knowledge/identification of the areas on the surface of the proteins capable of binding to the B-cell epitopes is considered a major contribute for the development of various immune related applications, such as protein detection techniques based on immunological interactions (e.g. ELISA, lateral flow devices, immunosensors). Since the experimental identification of B-cell epitopes is an expensive and intensive hard task, bioinformatics has become a much appealing complementary approach.[55] In the present work, a set of peptides with potential immunogenicity was successfully selected by the use of bioinformatics, which enabled defining biomarker peptides of CP4-EPSPS protein of RR soybean.

Considering that this protein is also present in many other transgenic crops, whose high sequence identity was also highlighted, the candidate peptides can be very useful as screening markers for the presence of GMO.

In summary, four peptides, namely Pc_19-32, Pc_67-77, Pc_141-156, and Pc_312-324 were proposed as the best candidates for antibody production. To our knowledge, this is the first attempt to select and define potential linear epitopes for the immunization of animals and subsequently, to generate adequate antibodies for CP4-EPSPS recognition. The following steps of this work will include the synthesis of the candidate peptides to be incubated in animals for antibody generation and potential applicability for the development of an immunosensor for CP4-EPSPS detection.

This work was supported by Marie Curie Actions – International Research Staff Exchange Scheme FP7-PEOPLE-2013-IRSES through the project no. 612545 entitled "GMOsensor – Monitoring Genetically Modified Organisms in Food and Feed by Innovative Biosensor Approaches" and by Fundação para a Ciência e a Tecnologia (FCT) through grant no. PEst-C/EQB/LA0006/2013 financed by POPH-QREN (subsidised by FSE and MCTES). M.M.M. and S.A.C. are researchers of the CONICET.

REFERENCES

1. Labrada, R. Weed Management for Developing Countries—Addendum 1; FAO Corporate Document Repository: Rome, 2003; Chapter 3, Benefits and Risks of the Use of Herbicide-Resistant Crops. Available at: http://www.fao.org/docrep/006/y5031e/y5031e00.htm#Contents. Last accessed on 10 July 2014.
2. Quaim, M. Annu Rev Resour Econ 2009, 1, 665–693.
3. GM Approval Database at International Service for the Acquisition of Agri-Biotech Applications (ISAAA). Available at: http://www.isaaa.org/. Last accessed on 24 July 2014.
4. Craig, W.; Tepfer, M.; Degrassi, G.; Ripandelli, D. Euphytica 2008, 164, 853–880.
5. Kramkowska, M.; Grzelak, T.; Czyżewska, K. Ann Agr Env Med 2013, 20, 413–419.
6. Macek, T.; Kotrba, P.; Svatos, A.; Novakova, M.; Demnerova, K.; Mackova, M. Trends Biotechnol 2008, 26, 146–152.
7. Nap, J. P.; Metz, P. L. J.; Escaler, M.; Conner, A. J. Plant J 2003, 33, 1–18.
8. James, C. Global Status of Commercialized Biotech/GM Crops: 2013; ISAAA Brief no 46; ISAAA: Ithaca, NY, 2013.
9. Food and Feed Safety, Genetically Modified Food and Feed at European Commission. Available at: http://ec.europa.eu/dgs/health_consumer/index_en.htm. Last accessed on April 2014.
10. GMO database at GMO Compass. Available at: http://www.gmo-compass.org/eng/home/. Last accessed on 24 July 2014.
11. GM crop database at GM Crop Database. Available at: http://www.cera-gmc.org/GMCropDatabase. Last accessed on 24 July 2014.
12. Product Safety Summaries at Monsanto. Available at: http://www.monsanto.com/products/pages/product-safety-summaries.aspx. Last accessed on 15 April 2014.
13. Oliveira, M. B. P. P.; Mafra, I.; Amaral, J. S. Current Topics on Food Authentication; Transworld Research Network: Kerala, 2011; Chapter 10, pp 211–236.
14. Pilolli, R.; Monaci, L.; Visconti, A. Trends Anal Chem 2013, 47, 12–26.
15. Kamle, S.; Ali, S. Gene 2013, 522, 123–132.
16. Minunni, M.; Mascini, M.; Cozzani, I. Anal Lett 2000, 33, 3093–3125.
17. Korber, B.; LaBute, M.; Yusim K. PLoS Comput Biol 2006, 2, e71.
18. Agallou M, Athanasiou E, Koutsoni O, Dotsika E, Karagouni E. Front Immunol 2014, 5, 268.
19. Purcell AW, McCluskey J, RossjohnJ. Nat Rev Drug Discov 2007, 6, 404–414.
20. National Centre for Biotechnology Information (NCBI) Entrez Protein Database. Available at: http://www.ncbi.nlm.nih.gov/protein. Last accessed on April 2014.
21. Walker, J. M. The Proteomics Protocols Handbook; Humana Press: Totowa, 2005; Chapter 52, pp 571–607.
22. Immune Epitope Data Base (IEDB). Available at: http://tools.immuneepitope.org/main/index.html. Last accessed on April 2014.
23. Chou, P. Y.; Fasman, G. D. Adv Enzymol Relat Areas Mol Biol 1978, 47, 45–148.
24. Emini, E. A.; Hughes, J. V.; Perlow, D. S.; Boger, J. J Virol 1985, 55, 836–839.
25. Karplus, P. A.; Schulz, G. E. Naturwissenschaften 1985, 72, 212–213.
26. Kolaskar, A. S.; Tongaonkar, P. C. FEBS Lett 1990, 276, 172–174.
27. Parker, J. M.; Guo, D.; Hodges, R. S. Biochemistry 1986, 25, 5425–5432.
28. Larsen, J. E. P.; Lund, O.; Nielsen, M. Immunome Res 2006, 2, 2.
29. Kringelum J. V.; Lundegaard, C.; Lund, O. G.; Nielsen, M. PLoS Comput Biol 2012, 8, 1–10.
30. Tung, C. -W.; Ho, S. -Y. Bioinformatics 2007, 23, 942–949.
31. Standard Protein Basic Local Alignment Search Tool (Blastp). Available at: http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?-PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=-blasthome. Last accessed on March 2014.
32. RCSB-PDB, Research Collaboratory for Structural Bioinformatics-Protein Data Bank. Available at: http://www.rcsb.org/pdb/home/home.do. Last accessed on May 2014.
33. Berman, H. M.; Westbrook, J.; Feng, Z.; Guilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Nucleic Acids Res 2000, 28, 235–242.
34. McMurray, J. S. Biopolymers (Peptide Sci) 1998, 47, 405–411.
35. GMOsensor. Monitoring Genetically Modified Organisms in Food and Feed by Innovative Biosensor Approaches. Available at: http://www.isep.ipp.pt/gmosensor/index.php?page=seminars-meetings.
36. Brusic, V.; Rudy, G.; Honeyman, M.; Hammer, J.; Harrison, L. Bioinformatics 1998, 14, 121–130.
37. Chicz, R. M.; Urban, R. G.; Gorga, J. C.; Vignali, D. A. A.; Lane, W. S.; Strominger, J. L. J Exp Med 1993, 178, 27–47.
38. Pearlman, R.; Wang, Y. J. Stability and Characterization of Protein and Peptide Drugs; Plenum Press: New York, 1993; Chapter 4, pp 135–156.

39. Sigma-Aldrich. Sequence Analysis; Sigma-Aldrich: St. Louis, MO. Available at: http://www.sigmaaldrich.com/life-science/custom-oligos/custom-peptides/learning-center/sequence-analysis.html. Last accessed on December 2014.
40. Peptide Design Guideline, Biomatik Corporation, Cambridge, Ontario, Canada, Version 3.1, revision 2011-05-17. Available at: http://www.biomatik.com/linkfiles/Peptide%20Design%20Guideline.pdf. Last accessed on December 2014.
41. Gad, S. C. Handbook of Pharmaceutical Biotechnology; John-Wiley & Sons, Inc.: New Jersey, 2007; Chapter 6.2, pp 737–756.
42. Peptides International FAQs: A General Guideline for Storage and Handling of Peptides. Peptides International Inc., Louisville, KY, USA. Available at: http://pepnet.com/products/peptideguidelines.pdf. Last accessed on December 2014.
43. Thinkpeptides; ProImmune: Sarasota, FL. Available at: https://www.proimmune.com/ecommerce/pdf_files/ST55.pdf. Last accessed on December 2014.
44. Iyengar, R. Methods in Enzymology, Heterotrimeric G-Protein Effectors; Academic Press, Inc.: San Diego, 1994; Chapter 2, pp 13–30.
45. Grant, G. A. Synthetic Peptides: A User's Guide; Oxford University Press: New York, 2002; Chapter 5.
46. Hopp, T. P.; Woods, K. R. Proc Natl Acad Sci USA 1981, 78, 3824–3828.
47. Jameson, B. A.; Wolf, H. Comput Appl Biosci 1988, 4, 181–186.
48. Chen J.; Liu H.; Yang J.; Chou K. -C. Amino Acids 2007, 33, 423–428.
49. Wee, L. J. K.; Simarmata, D.; Kam, Y. -W.; Ng, L. F. P.; Tong, J. C. BMC Genomics 2010, 11, S21.
50. Zhao, L.; Li, J. BMC Struct Biol 2010, 10, S6.
51. Davies, M. N.; Flower; D. R. Drug Discov Today 2007, 12, 389–395.
52. Koch, C. P.; Perna, A. M.; Pillong, M.; Todoroff, N. K.; Wrede, P.; Folkers, G.; Hiss, J. A.; Schneider, G. PLoS Comput Biol 2013, 9, e1003088.
53. Calis, J. J.; Maybeno, M.; Greenbaum, J. A.; Weiskopf, D.; De Silva, A. D.; Sette, A.; Keşmir, C.; Peters, B. PLoS Comput Biol 2013, 9, e1003266.
54. Miconnet, I.; Koenig, S.; Speise, D.; Krieg, A.; Guillaume, P.; Cerottini, J. C.; Romero, P. J Immunol 2002, 168, 1212–1218.
55. Gu, J.; Bourne, P. E. Structural Bioinformatics; Wiley: New York, 2009; Chapter 35, pp 849–879.