Cry1A(b)16 toxin from *Bacillus thuringiensis*: Theoretical refinement of three-dimensional structure and prediction of peptides as molecular markers for detection of genetically modified organisms

Alexandra Plácido,[1] Andreia Coelho,[1] Lucas Abreu Nascimento,[2,5] Andreanne Gomes Vasconcelos,[3] Maria Fátima Barroso,[1] Joilson Ramos-Jesus,[3] Vladimir Costa,[4] Francisco das Chagas Alves Lima,[2] Cristina Delerue-Matos,[1] Ricardo Martins Ramos,[2,5] Mariela M. Marani,[6] and José Roberto de Souza de Almeida Leite [7]*

[1] REQUIMTE-LAQV, Instituto Superior de Engenharia do Porto, Instituto Politécnico do Porto, Porto, Portugal

[2] Grupo de Química Quântica Computacional e Planejamento de Fármaco, GQQCPF, Departamento de Química, Universidade Estadual do Piauí, UESPI, Teresina, Brasil

[3] Núcleo de Pesquisa em Biodiversidade e Biotecnologia, Biotec, Campus Ministro Reis Velloso, Universidade Federal do Piauí, Parnaíba, Brasil

[4] Laboratório de Pesquisas em Leishmanioses, Instituto de Doenças Tropicais Natan Portela, IDTNP, Teresina, Piauí, Brasil

[5] Laboratório de Pesquisa em Sistemas de Informação, LaPeSI, Departamento de Informação, Ambiente, Saúde e Produção Alimentícia, Instituto Federal do Piauí, Teresina, Brasil

[6] IPEEC-CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas, Puerto Madryn, Argentina

[7] Área de Morfologia, Faculdade de Medicina, FM, Campus Universitário Darcy Ribeiro, Universidade de Brasília, UnB, Brasília, Brasil

ABSTRACT

Transgenic maize produced by the insertion of the *Cry* transgene into its genome became the second most cultivated crop worldwide. Cry gene from *Bacillus thuringiensis kurstaki* expresses protein derivatives of crystalline endotoxins which confer insect resistance onto the maize crop. Mandatory labeling of processed food containing or made by genetically modified organisms is in force in many countries, so, it is very urgent to develop fast and practical methods for GMO identification, for example, biosensors. In the absence of an available empirical structure of Cry1A(b)16 protein, a theoretical model was effectively generated, in this work, by homology modeling and molecular dynamics simulations based on two available homologous protein structures. Molecular dynamics simulations were carried out to refine the selected model, and an analysis of its global structure was performed. The refined models of Cry1A(b)16 showed a standard fold and structural characteristics similar to those seen in *Bacillus thuringiensis* Cry1A(a) insecticidal toxin and *Bacillus thuringiensis* serovar *kurstaki* Cry1A(c) toxin. After *in silico* analysis of Cry1A(b)16, two immunoreactive candidate peptides were selected and specific polyclonal antibodies were produced resulting in antibody–peptide interaction. Biosensing devices are expected to be developed for detection of the Cry1A(b) protein as a marker of transgenic maize in food.

Key words: Cry1A(b)16; *Bacillus thuringiensis*; homology modeling; molecular dynamics simulation.

INTRODUCTION

In the past decade, since the middle of the XX century, agriculture underwent a remarkable revolution because of the development, creation, and cultivation of genetically modified organisms (GMO) by "intentional manipulation",[1,2] where genetic material has been altered in such a way that does not occur naturally or by natural recombination.[3] Maize is the

*Correspondence to. José Roberto de Souza de Almeida Leite; Faculdade de Medicina (FM), Universidade de Brasília (UnB), Campus Universitário Darcy Ribeiro, Asa Norte, Brasília, Distrito Federal (DF). Brazil 70910900.
E-mails: jrleite@pq.cnpq.br and jrsaleite@gmail.com

second most cultivated GMO crop, corresponding to approximately 30% of the global cultivated biotechnological area (179.7 million hectares) in 2015.[4] In the European Union (EU) and worldwide, maize has the highest number of authorized GM events (142).[3]

The insertion of a foreign gene (*cry*) into the maize genome from a gram-positive spore-forming bacterium, *Bacillus thuringiensis kurstaki*, to express protein derivatives of crystalline endotoxins (δ-endotoxins) became the first successful creation of a GMO maize crop that is resistant to the European corn borer (*Ostrinia nubilalis*).[5]

δ-Endotoxins act at the intestinal level, and after ingestion, the alkaline environment of the intestine and the presence of proteases promote a release of internal peptides (of δ-endotoxins) that bind to specific receptors in the plasma membrane of the insect's epithelium, resulting in pore formation in the intestinal wall. This process induces severe inflammation, which causes severe illness in the insect and eventually its death.[6,7]

This effect coupled with the fact that Cry proteins are considered innocuous to humans, animals, and plants, and are completely biodegradable makes these proteins an attractive choice for genetic improvement of crops to provide protection from insect pests.[7–12]

GM maize lines with Bt11 and MON810 transgenic events are, by far, the most widely cultivated crops, and both express Cry1A(b).[3] Because the expressed protein is a toxin, several concerns have been raised regarding food safety and environmental issues. Therefore, it is necessary to have proper methods for toxin monitoring and for the detection of GMO.

The conventional methods for identification of GMO relies mostly on molecular biology methods, such as the polymerase chain reaction (PCR) technique for amplification and detection of the transgene sequence or on immunochemical methods such as the enzyme-linked immunosorbent assays (ELISA).[13] Nevertheless, this kind of methods require extensive sample preparation high associated costs, are time consuming, and are not suitable for *in situ* assays. Consequently, it is important to develop techniques that allow fast, accurate, and easy identification and detection of GMO, for example, biosensors.

The Cry toxins have been widely used in insect pest control studies, but the three-dimensional (3D) structure of the Cry1A(b)16 protein is not still available in protein databases. The knowledge of the 3D structure is very important for better understanding of the structure–function relations of the protein and the underlying mechanisms.

Cry proteins are mostly composed of three domains with defined structure and specific functions: a seven-helix-bundle domain (DI) containing an α-helical bundle, which is certainly involved in membrane insertion and pore formation; a three-antiparallel β-sheet domain (DII) which forms a "Greek key" topology organized to form a β-prism fold and contains surface-exposed loops that are considered the most probable candidates for receptor binding;[5] and a β-sandwich domain (DIII) consisting of two twisted antiparallel β-sheets, forming a β-sandwich with a jelly roll topology.

Additionally, DIII is considered a multifunctional domain that plays a fundamental role in the biochemistry, structural integrity, membrane penetration, ion channel function, and receptor binding.[14]

Some reported studies based on homology modeling and molecular dynamics simulations have already been conducted on other Cry proteins, with a previously proposed structural model for protein Cry1A(b)16.[14,15] On the other hand, the structure–function research on this protein has been based on its closest structural homologs, Cry1A(a) and Cry1A(c), which represent a more probable and realistic model than the model studied before.

The aim of the present study was to develop a new approach to monitor GMO in processed foods. The developed approach consists of prediction of a homology modeling and molecular dynamics simulations of the 3D structure of the Cry1A(b)16 protein in accordance with bioinformatic data and extensive *in silico* analysis of immunogenic peptides to obtain marker peptides that can function as immunogens for production of novel antibodies for biorecognition of Cry1A(b)16.

Two peptides of interest were selected on the basis of their immunogenicity. The novel 3D structure of the target toxin obtained by homology modeling and molecular dynamics simulations is expected to not only form the basis for the development of a protein-based biosensor but should also facilitate the design of hybrid Cry proteins and new fusion proteins with novel features against several species of insects, thereby improving management of pests.[5,16]

MATERIAL AND METHODS

Selection of Cry1A sequences and 3D structures

Cry1A(b)16 protein sequence, accession No. AAK55546, was retrieved from the National Center for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov), and homologous empirical 3D structures of Cry1A(b)16 were explored at RCSB Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do).

Two homologue protein structures were chosen as templates: Cry1A(a) from *Bacillus thuringiensis* (PDB ID: 1CIY; 2.25 Å resolution, 88% sequence identity, and 92% of conserved residues) and Cry1A(c) from *Bacillus thuringiensis* (PDB ID: 4ARY; 2.95 Å resolution, 83% sequence identity, and 86% of conserved residues).

Homology modeling

By means of the 3D structures of Cry1A(a) and Cry1A(c) that have already been empirically determined by X-ray diffraction, structural models of Cry1A(b)16

Table I
Comparison of Quality of Templates Structures with the Homology Model

	Procheck						
Template or model	Core[a]	Allow.[b]	Gener.[c]	Desall.[d]	Overall G-factor[e]	Errat	Z-Score
BthCry1A(a) (1CIY)	88.8%	10.6%	0.4%	0.2%	0.17	95.75	−9.47
BthCry1A(c) (4ARY)	91.5%	8.3%	0.0%	0.2%	0.09	96.59	−9.81
HM-Cry1A(b)16	93.2%	6.2%	0.4%	0.2%	0.02	90.73	−9.78

[a]Percentage of residues in most favored regions.
[b]Percentage of residues in permitted areas.
[c]Percentage of residues in regions generously allowed.
[d]Percentage of residues in areas not permitted.
[e]dihedral G-factor [37].

were generated by homology modeling in the MODELLER 9v13 software.[17,18] The model with the most negative value for the discrete optimized protein energy (DOPE) function[19] was chosen for an additional stage of refinement using molecular dynamics simulations.

For verification and validation of protein structures during and after model refinement, different software packages were used, namely PROCHECK[20] and ERRAT[21] from the Structure Analysis and Verification Server (SAVES) and the Z-Score software (ProSA-web Protein Structure Analysis) (Table I).[22,23]

Molecular dynamics simulations

Simulations were carried out using GROMOS96 53a6 force field[24] implemented in the GROMACS package[25], version 4.5.5. All systems were simulated in the NPT ensemble (wherein the number of particles, pressure, and temperature were all constant) and periodic boundary conditions. The dimensions of the central box were chosen in such a way that the minimal distance of any protein atom to the closest box wall was 12 Å. The simulations were conducted using explicit solvent water molecules described by the simple point charge (SPC) model.[26] The total charge of the systems was −3 for homology modeled (HM)-Cry1A(b)16. Sodium ions were added to neutralize each system.

Protein structure was subjected to a maximum of 500 steps of steepest-descent energy minimization. One picosecond of a molecular dynamics simulation was performed restraining the protein structure to relax strong solvent–solvent and solvent–protein nonbonded interactions. Initial velocities were assigned according to the Maxwell distribution. The simulations were performed for 30 ns using an integration time step of 2 fs.

Each system was heated with gradual increments at the following temperatures: 100 K (10 ps), 150 K (5 ps), 200 K (5 ps), and 250 K (5 ps). The temperatures of the systems were adjusted to 310 K. The first 15 ns of each simulation was considered a part of the heating (0.025 ns) and equilibration (14.975 ns) steps; therefore, they were not used for data analysis. The temperatures of the solvent and solutes (protein and sodium ions) were independently coupled to a thermal bath with relaxation time of 0.1 ps by means of the Berendsen thermostat.[26] The pressure in the systems was weakly coupled to a pressure bath of 1 atm with isotropic scaling and 0.5 ps relaxation time using the Parrinello-Rahman barostat.[27,28]

Bond lengths were constrained by the LINCS algorithm[29] with fourth-order expansion. Electrostatic interactions among nonligand atoms were evaluated by the particle mesh Ewald (PME) method[30] with charge grid spacing of approximately 1.0 Å. The charge grid was interpolated on a cubic grid with the direct sum tolerance set. Lennard-Jones interactions were evaluated by means of a 14 Å atom-based cutoff. The pair list was updated every 10 steps.

Peptide synthesis and purification

Peptides were selected by *in silico* evaluation of the Cry1A(b)16 protein (accession No. AAK55546; Fig. 1), on the basis of their sequence and immunogenicity.[31,32] The parameters used *in silico* involved theoretical digestion at trypsin and chymotrypsin cleavage sites, sequence and structure determination of epitopes (B-Cell and T-cell epitope prediction tools), cross-reactivity, and secondary-structure prediction for the selected peptides.[31–33]

Final selection of the peptides for synthesis, purification, and production of polyclonal antibodies was based on the exposure of amino acid residues to the solvent in the native protein.

The peptides were manually synthesized by means of a solid-phase approach using Fmoc/*t*-butyl chemistry.[34] Peptide elongation was carried out in polypropylene syringes fitted with a polyethylene porous disk. Solvents and soluble reagents were removed by suction. The rink amide (4-Methylbenzhydrylamine hydrochloride) MBHA resin was used for the synthesis of the peptides. A Cys was incorporated at the C-terminus in order to couple the peptides to a carrier protein prior to their inoculation into animals. Samples were treated with the mixture trifluoroacetic acid/triisopropylsilane/water (TFA/TIS/H_2O; 95:2.5:2.5, v/v/v) to remove the protective groups.

Peptide purification was carried out by means of an analytical reverse-phase high-performance liquid chromatography (RP-HPLC) system (Shimadzu Prominence

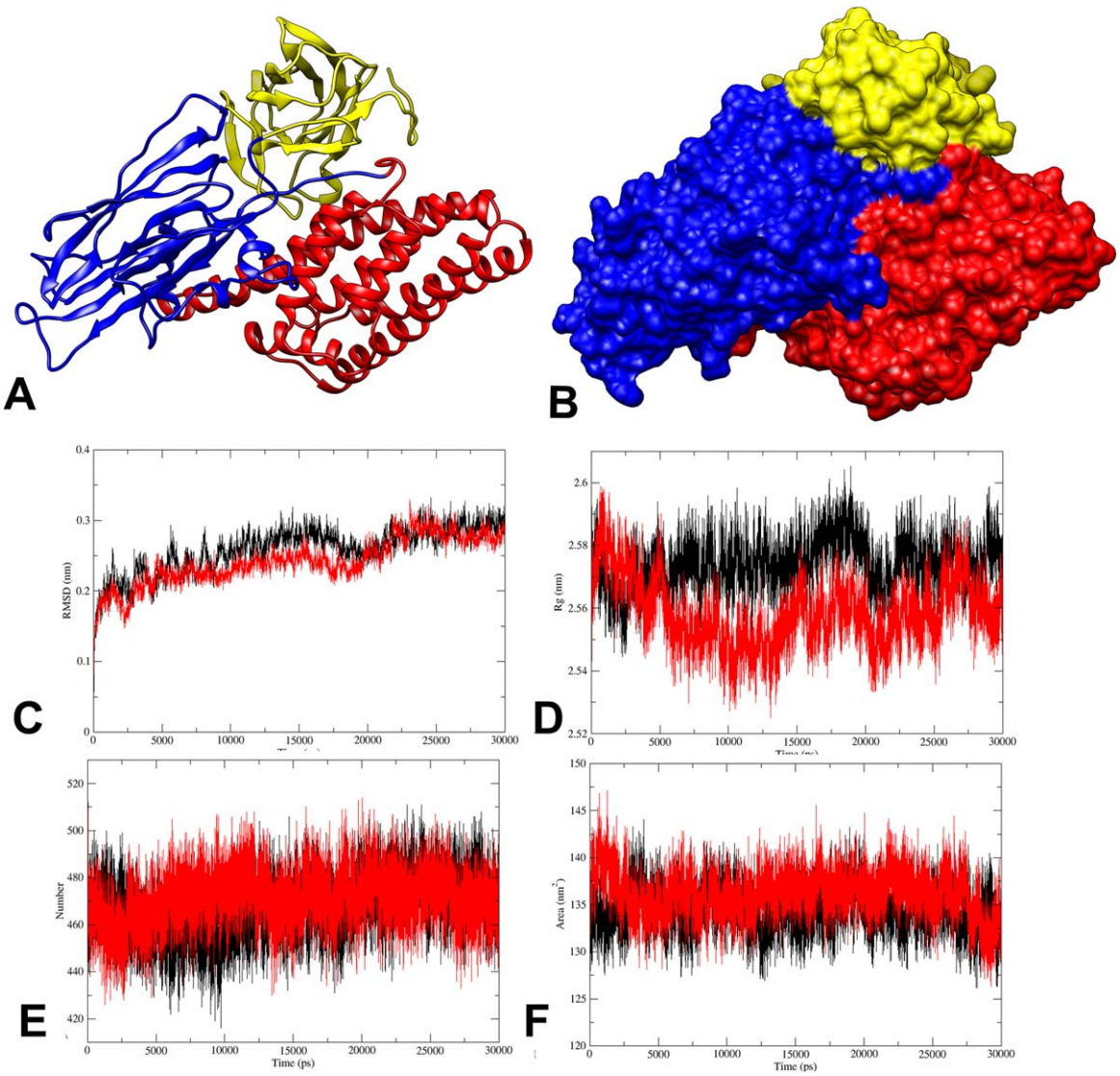

Figure 1

(**A**) Schematic representations of the HM-Cry1A(b)16 model. Domain I is red, domain II is blue, and domain III is yellow. (**A**) Ribbon trace. (**B**) Molecular surface. The image was generated using UCSF Chimera (https://www.cgl.ucsf.edu/chimera/).[37] (B) Time dependence of the structural parameters from molecular dynamics simulations with respect to the initial structure. (**C**) Atom-positional root mean-square deviation (RMSD) of Cα atoms; (**D**) radius of gyration (Rg); (**E**) the total number of hydrogen bonds (Nhb); (**F**) total solvent-accessible surface area (SASA). *Color legend*: Cry1A(b)16_S01 (black) and Cry1A(b)16_S02 (red). The image was generated using Xmgrace (http://plasma-gate.weizmann.ac.il/Grace/).[38]

Instrument, Shimadzu, Kyoto, Japan) on a Phenomenex RP chromatography column (Phenomenex columns Kinetex, 5 µm C18 50 × 21.20 mm). Each peptide was dissolved in the mixture H_2O:CH_3CN (6:4, v/v) and loaded into the RP-HPLC system using a gradient of CH_3CN, starting with 0.1% trifluoroacetic acid in water and then increasing CH_3CN concentration to 100% during 60 min at a flow rate of 1.0 mL/min. Peptides were monitored at 216 and 280 nm. The formula $(A_{215} - A_{225}) \times 144$ (µg/mL) was applied to peptide quantification.[35]

Determination of purity and molecular masses of the synthetic peptides was performed by matrix-assisted laser desorption ionization tandem time-of-flight (MALDI-TOF/TOF) mass spectrometry (MS, ultrafleXtreme™, Bruker Daltonics, Bremen, Germany). The instrument was operated in positive ion mode and controlled by the Compass for Flex software, version 1.3 (FlexControl 3.3, FlexAnalysis 3.3, Bruker Daltonics, Bremen, Germany); 5000 laser shots were accumulated per spectrum in the MS and MS/MS modes.

One-microliter aliquots of the chromatographic fractions dissolved in an α-cyano-4-hydroxycinnamic acid matrix solution (1:3, v/v) were applied to a stainless steel plate and dried at room temperature for 30 min. The peptide monoisotopic mass was determined in reflector mode with external calibration, using the peptide calibration standard for

Table II
Comparison Between Secondary Structures of Cry1A(a) (PDB ID 1CIY) and the Model Generated by Homology Modeling (HM-Cry1A(b)16)

Domain I			Domain II			Domain III		
	Cry1A(a)[a]	HM-Cry1A(b)16		Cry1A(a)[a]	HM-Cry1A(b)16		Cry1A(a)[a]	HM-Cry1A(b)16
α1	Pro35-Ser48	Pro7-Ser20	β1	Thr259-Val260	Arg230-Val232	β16	Ile464-Ile465	Asn436-Ile438
α2	Ala54-Il63	Gly27-Trp37	β2	Glu266-Thr269	Ile239-Thr241	β17	Ile470-Leu474	Ile443-Leu448
α3	Pro70-Ile84	Ser43-Ile56	α13	Pro271-Glu274	Pro243-Glu246	α18[b]	Leu475-Lys477	–
α4	Glu90-Ala119	Glu62-Ala91	α14	Ala284-Gln289	Ala256-Gly261	β18	Asn480-Leu481	Asn453-Gly455
α5	Pro124-Ala144	Pro96-Phe120	β3	Asp298-His310	Ile271-His282	β19	Ser486-Val488	Ser459-Val461
α6[b]	Pro146-Leu148	–	β4	Phe313-Pro325	Glu285-Ser296	β20	Ile498-Arg501	Ile471-Arg474
α7	Gln154-Phe178	Arg126-Trp154	α15[b]	Val326-Phe328	–	β21	Gly505-Ile514	Gly478-Thr488
α8[b]	Gln180-Trp182	–	β5	Phe333-Phe334	Phe305-Phe307	β22	Tyr522-Ser530	Arg494-Thr504
α9	Ala186-Val218	Ala158-Arg189	β6[b]	Phe338-Gly339	–	β23	Leu534-Ile540	Leu507-Asp514
α10	Ser223-Thr239	Ser195-Asn221	β7	Ala345-Ser351	Tyr310-Thr312	β24	Arg543-Phe550	Arg516-Phe523
α11[b]	Leu241-Val244	–	β[c]	–	Gln320-Ala324	α19[b]	Ser562-Ser564	–
α12[b]	Ser248-Tyr250	–	β8	Ile357-Arg367	Gly329-Arg340	β25	Arg566-Gly569	Arg539-Phe543
			β9	Leu380-Phe390	Gln351-Gly363	β26	Ser580-His588	Gly552-His561
			β10	Thr400-Tyr402	Ala371-Tyr373	β27	Val596-Pro605	Glu568-Ala579
			β11	Val408-Asp409	Val379-Ser381			
			α16	Ser410-Asp412	Leu382-Glu384			
			α17	Pro423-Gly426	Pro394-Gly397			
			β12	His429-Val434	Ser399-Ser410			
			β13[b]	Leu437-Ser438	-			
			β14[b]	Thr446-Ala449	-			
			β15	Phe452-His456	Val417-Arg430			

[a]Structures (α-helix and β-strand only) obtained from http://www.rcsb.org/pdb/explore/remediatedSequence.do?structureId = 1CIY.
[b]Asecondary structures that appear only in Cry1A(a),
[c]Bsecondary structures that appear only in HM-Cry1A(b)16.

an MS mixture (mass range up to 4000 Da, Bruker Daltonics). Peptides were loaded onto an automatic sequencer for *De novo* sequencing in "LIFT™" technology mode.

Antibody production and immunoreactivity

The specific antibodies were produced by a specialized biotech company (GenScript, NJ, USA). Briefly, synthetic peptides PcH77-91 and Pt282-292 were coupled to a carrier protein called keyhole limpet hemocyanin (KLH) (for immunization) using a single-step coupling protocol.[36] Peptide conjugation to KLH via the C-terminal Cys residue was carried out using *m*-maleimidobenzoic acid N-hydroxysuccinimide ester (MBS)-activated KLH (Pierce Chemical Company, Rockford, Illinois, USA). For each peptide, rats were immunized according to standard procedures.[36]

Immune responses were screened using a dot blot assay as follows: peptides coupled to KLH as well as carrier-only controls (all at 20 mg/mL in 0.2 *M* phosphate buffer pH 7.4), were incubated with nitrocellulose membrane strips for 2 h at 37°C. The following steps were performed at room temperature. After blockage in 10% fetal calf serum (FCS) in D-MEM (Dulbecco's Modified Eagle's Medium) for 2 h, 2 mL aliquots of serially diluted serum samples were spotted onto antigen-coated membranes, incubated for 2 h, washed 3× in 0.01*M* phosphate buffer pH 7.4, incubated with horseradish peroxidase-conjugated goat anti-rat IgG antibodies (Sigma-Aldrich, cat. # A9037) diluted 1:100 with DMEM plus 10% FCS, washed 3× for 5 min in PBS, and developed using 1,4-chloronaphthol. To remove polyclonal antibodies against the KLH carrier, selected antisera were affinity-purified on a column with a KLH-conjugated resin (Hena s.r.o., Czech Republic).

RESULTS AND DISCUSSION

Homology modeling

Three-dimensional information on Cry1A(b)16 was obtained by homology modeling; for this purpose, two proteins homologous to Cry1A(b)16 (NCBI accession No. AAK55546.1) with available empirical 3D structures in the PDB were used as model templates: Cry1A(a) from *Bacillus thuringiensis* (PDB ID 1CIY) and Cry1A(c) from *Bacillus thuringiensis* (PDB ID 4ARY). From the primary sequence of 1155 aa, a structural model of the Cry1A(b)16 core protein, designated as HM-Cry1A(b)16,

Table III
Average Structural Parameters of HM-Cry1A(b)16 from the Last 5 ns of the Simulations

HM-Cry1A(b)16	RMSD (nm)[a]	Rg (nm^3)[b]	NHb-intra[c]	SASA (nm^2)[d]
Simulation 01	0.29 ± 0.01	2.57 ± 0.006	475.3 ± 10.0	133.8 ± 2.2
Simulation 02	0.27 ± 0.01	2.56 ± 0.008	471.1 ± 10.3	134.7 ± 2.7

[a]Root mean square deviation of HM-Cry1A(b)16 Cα atoms from starting coordinates (homology model).
[b]Radius of gyration.
[c]The number of hydrogen bonds (intramolecular).
[d]Solvent-accessible surface area.

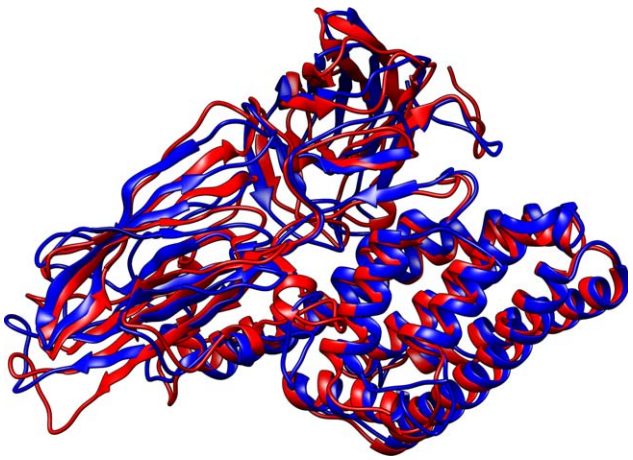

Figure 2

Superposition of the C^{α} backbone of molecules DM-Cry1A(b)16 (blue, obtained by molecular dynamics simulations) and HM-Cry1A(b)16 (red, obtained by homology modeling). [Color figure can be viewed at wileyonlinelibrary.com]

was proposed by homology modeling based on chain A that is 590 aa long. Sequence identity of each template and respective alignments are depicted in Supporting Information Figure S1. Cry1A(b)16 alignment with the empirical structures from PDB shows that domain I (residues 1 to 240) and domain 2 (residues 241 to 446) mostly contain conserved residues, while domain III (residues 447 to 591) shows only 33% of conserved residues (48 amino acid residues; Supporting Information Fig. S1).[35,36]

One hundred models were independently generated for HM-Cry1A(b)16. Quality of the templates and of the homology model (model chosen because of the best DOPE score) was evaluated using software applications PROCHECK[20] and Errat[21] implemented on SAVES (Structural Analysis and Verification Server, http://nih-server.mbi.ucla.edu/SAVES/) and Z-Score (ProSA-web Protein Structure Analysis).[22,23] The first software package examines residue-by-residue geometry and the overall structure geometry, whereas the *Z*-score was used for the refinement and validation of empirical protein structures and for structure prediction/modeling.

Table I shows a comparison of HM-Cry1A(b)16 structural parameters with templates Cry1A(a) and Cry1A(c). The HM-Cry1A(b)16 structure contains 93.2% of residues in the most favored regions, which is higher than the 90% defined as a criterion for a good-quality model.[20] All the remaining parameters were also in good agreement with the ones determined for empirical conformations of Cry1A(a) and Cry1A(c), showing only 0.2% of residues in disallowed regions (Table I). The *Z*-

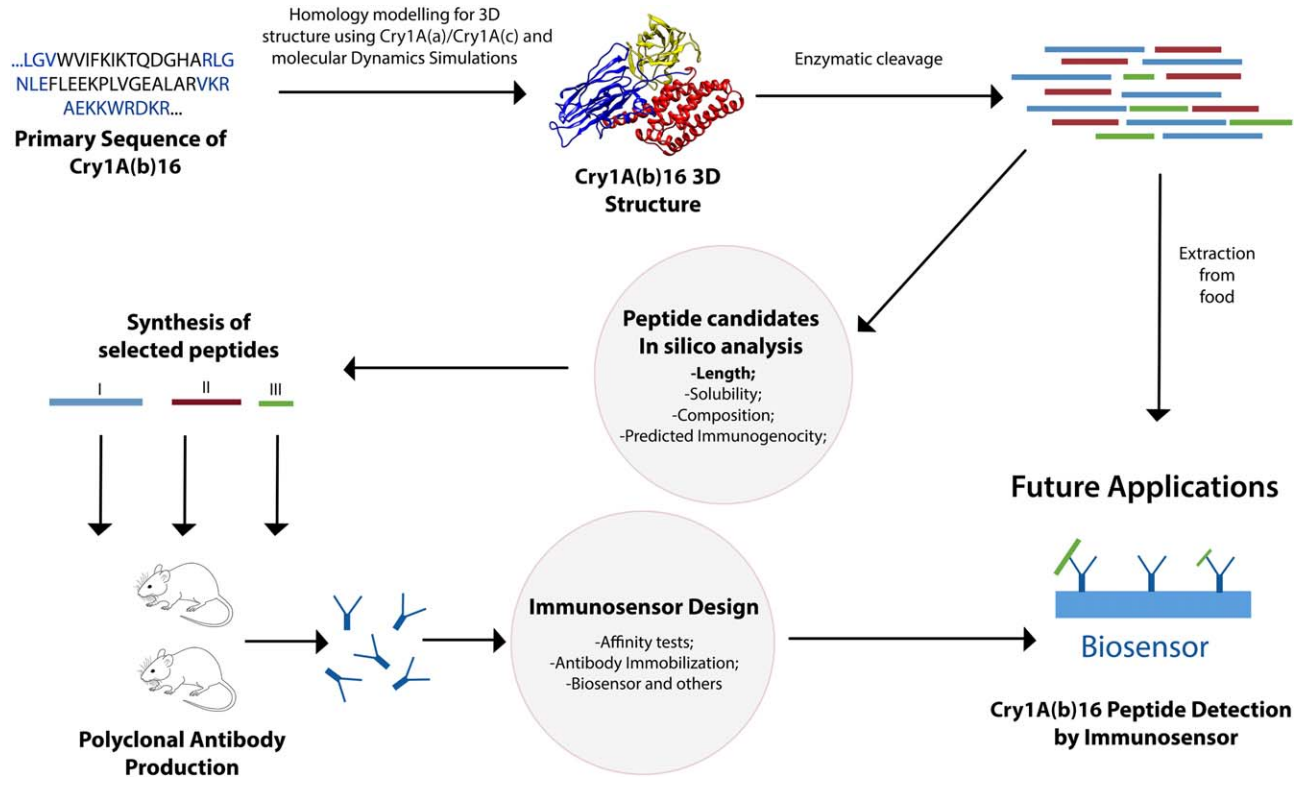

Figure 3

The outline of the procedures in this study and the development of sensor-based methods for detection in foods. [Color figure can be viewed at wileyonlinelibrary.com]

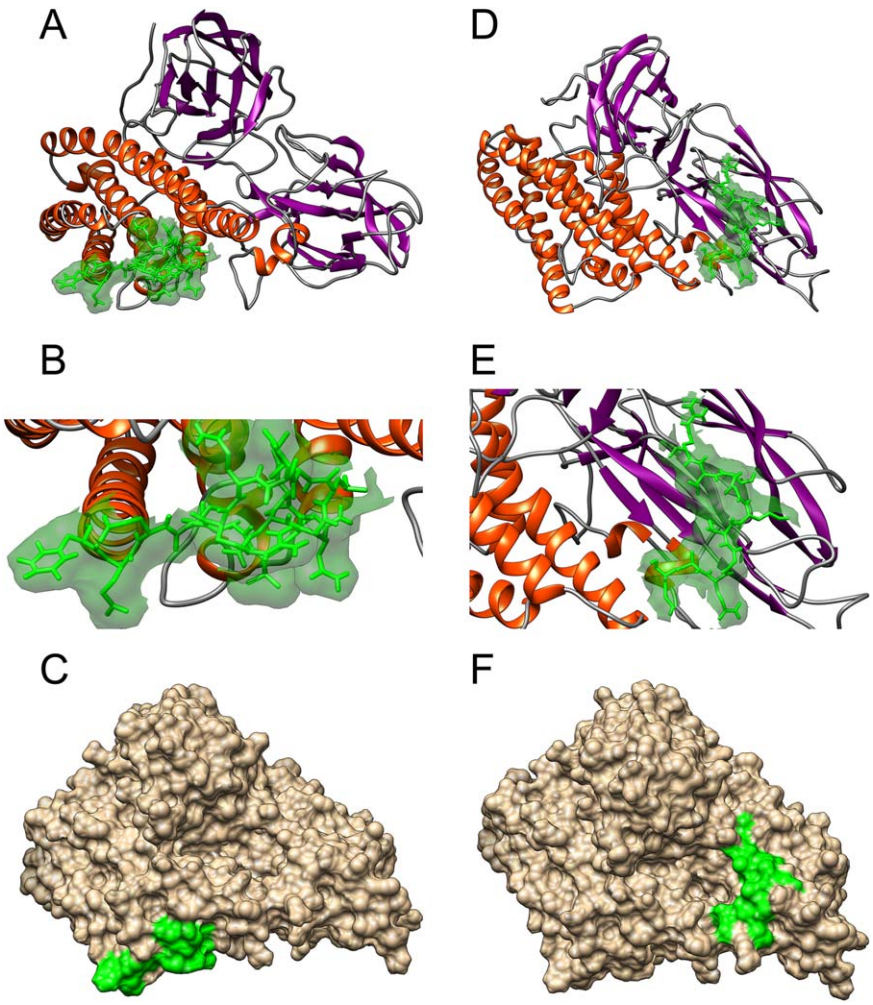

Figure 4

FPcH_77-91 (**A, B**, and **C**). Global structure for the last frame of molecular dynamics simulations for DM-BTCry1A(b)16. (A) DM-BTCry1A(b)16 as ribbon representation and peptide PcH_77-91 as stick representation (green). (B) Expansion of the peptide shown as a stick model (green). (C) DM-BTCry1A(b)16 represented by the surface and peptide region (green). Pt_282-292 (**D**, **E**, and **F**). Global structure for the last frame of molecular dynamics simulations for DM-BTCry1A(b)16. (D) DM-BTCry1A(b)16 represented by ribbons and peptide Pt_282-292 by sticks (green); (E) expansion of the peptide represented by sticks (green); and (F) DM-BTCry1A(b)16 represented by the surface and peptide region (green). [Color figure can be viewed at wileyonlinelibrary.com]

score for TB-Cry1A(b)16 was −9.78, which is very similar to the *Z*-scores of the empirical structures of homolog proteins Cry1A(a) and Cry1A(c) (obtained by X-ray diffraction; Table I). Therefore, the proposed HM-Cry1A(b)16 can be considered a reliable 3D model of the native toxin.[22]

As expected, the fold of the HM-Cry1A(b)16 model resembles the general fold observed in Cry1A(a) and Cry1A(c) [Fig. 1(A) and Table II]. Table II shows the secondary structures in HM-Cry1A(b)16. Domain I (residues 1–229) consists of eight helices (α1–α5, α7, α9–α10) in opposition to the 12 α-helices of the Cry1A(a) secondary structure.

In domain II (residues 230–435)—which consists of four helices (α13–α14, α16–α17) and 13 β-strands—three β-sheets (6, 12, and 13) and α-Helix 15 from the Cry1A(a) sequence were not identified in TB-Cry1A(b)16. In the third domain (residues 436–590)—which contains 12 β-strands—the α-helices (18 and 19) were also not detected in the secondary structure of TB-Cry1A(b)16. The region of domain III in the alignment has a small proportion of conserved residues (Supporting Information Fig. S1); therefore, an additional step of model refinement was carried out using molecular dynamics simulations.

Molecular dynamics simulations

In order to confirm stability of the target HM-Cry1A(b)16 structure, to avoid artifacts and to increase

the sampling, two molecular dynamics simulations with duration of 30 ns were carried out using different starting atomic velocities from a Maxwellian distribution. To ensure that the simulations were stable, some structural parameters such as the atom-positional root-mean-square deviation (RMSD), the radius of gyration (Rg), total number of intramolecular hydrogen bonds (NHb), and the solvent-accessible surface area (SASA) were monitored in the course of the molecular dynamics simulations. Figure 1(B) presents two simulations in each panel; black and red curves correspond to simulations 1 [Cry1A(b)16_S1] and 2 [Cry1A(b)16_S2], respectively. From these data, it is possible to verify that the stability of the systems was attained during the last 5 ns of the simulation. The average values of the parameters calculated for the last 5 ns during the simulations of each system (the period defined as the production stage) are shown in Table III. These data indicate that when each parameter is considered independently, no significant variations were observed among the simulations, thus indicating conformational stability of the proposed structure.

The system achieved an adequate amount of sampling, and no significant structural deviations could occur after 25 ns of the simulation. Because of the lower average value of RMSD (0.27 ± 0.01; Table III) in relation to the structure generated by homology [HM-Cry1A(b)16], the refined model with a lower energy obtained by simulation 02 (Cry1A(b)16_S02, named DM-Cry1A(b)16) was chosen to study the immunogenic peptides and for the development of biosensors. As expected, the fold of the refined DM-Cry1A(b)16 model resembles the general fold observed in HM-Cry1A(b)16 (Fig. 2).

Prediction of peptides as molecular markers for detection of GMO

The protein sequence corresponding to Cry1A(b)16 was retrieved from the NCBI database (accession No. AAK55546). Cry1A(b)16 is a δ-endotoxin produced by *B. thuringiensis* AC11 (H14). The entire primary sequence of this protein is composed of 1155 aa (Supporting Information Fig. S1). The ExPASy PeptideCutter tool was used to subject Cry1A(b)16 to *in silico* cleavage site prediction using trypsin and chymotrypsin for high (HS) and low (LS) specificity.[31] (Fig. 3)

A cleavage site map and the table of cleavage site positions were analyzed, and peptides 10 and 30 aa length were selected for compositional analysis. The two selected peptides after theoretical digestion were chosen after *in silico* analysis of the antibody production capacity, such as structure determination of epitopes and T-cell epitope prediction.[31]

The nomenclature of the peptides takes into account positioning characteristics of the sequence in the native protein and how the peptide was generated: for example,

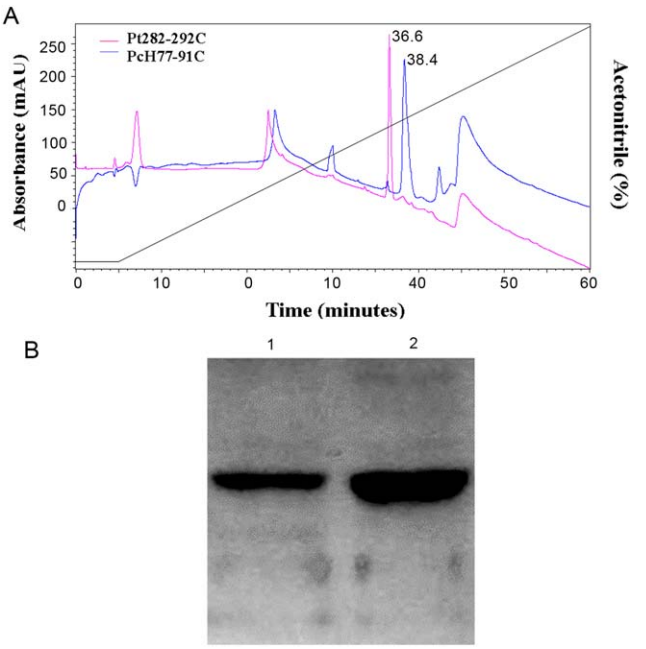

Figure 5

(**A**) Representative chromatograms of purified synthetic peptides PcH_77-91 and Pt_282-292. The experiments were conducted independently and are grouped in the figure for comparison purposes. The numbers above the fractions correspond to the retention time (T_r). The fractions were collected manually, and the curves correspond to monitoring at 216 nm; mAU: arbitrary units. (**B**) An immunoblot of polyclonal anti-PcH_77-91 (**Lane 1**) and anti-Pt_282-292 antibodies (**Lane 2**) as the primary antibody. Antigens were analyzed by SDS-PAGE and blotted onto a nitrocellulose membrane. [Color figure can be viewed at wileyonlinelibrary.com]

PcH_77-91 (P = peptide; cH = chymotrypsin with high specificity; positions 77–91 in the native protein) and Pt_282-292 (P = peptide; t = trypsin; positions 282–292 in the native protein), these peptides are indicated in the native protein structure in Figure 4.

In some situations, we could use the nomenclature PcH_77-91 C (C = cysteine residue at the C terminal), because the peptides were synthesized with a C-terminal cysteine to enable conjugation to KLH for polyclonal-antibody production.

Peptides PcH_77-91 and Pt_282-292 were selected, synthesized, and purified by RP-HPLC, and the resulting peptides showed estimated purity >98% [Figure 5(A)]. After synthesis, they were designated as PcH_77-91 C and Pt_282-292 C.

Molecular masses were determined and confirmed for pure peptides by MALDI-TOF MS, with $[M + H]^+ = 1976.30$ Da for PcH77-91 and $[M + H]^+ = 1178.36$ Da for Pt282-292. The *de novo* sequencing by MS/MS was performed to confirm the primary sequence (Supporting Information Figs. S2 and S3).

After in-house assessment of the potential peptides for antibody production, bioinformatics analyses of the

peptides were also carried out by a specialized company (GenScript, NJ, USA) that produced the antibodies. PcH_77-91 and Pt_282-292 [Fig. 5(A)] were also suggested by GenScript as good candidates for animal immunization, and these two peptides were selected for antibody production. Immunoblots [Fig. 5(B)] illustrate the successful development of anti-PcH_77-91 and anti-Pt_282-292 polyclonal antibodies against the respective target peptides. Specific binding was observed for each peptide, pointing to good quality of the antibodies.

CONCLUSION

In this work, a novel homology modeling and structural analysis of 3D models obtained from Cry1A(b)16 (*Bacillus thuringiensis*) were performed. Molecular dynamics simulations were carried out to refine the chosen model, and using the obtained data, we analyzed its global structure. The new DM-BTCry1A(b)16 model can be used in research on new biosensors.

In general, two peptides, namely, PcH_77-91 and Pt_282-292, were identified as the best candidates for antibody production and were successfully used as immunogens, resulting in successful production of the respective reactive polyclonal antibodies. To the best of our knowledge, this is the first attempt to select and define potential linear epitopes for immunization of animals and subsequently to generate adequate antibodies for Cry1A(b)16 recognition. The next steps will include the development of an protein-based sensor for Cry1A(b)16 detection as a marker of GM maize in food.

ACKNOWLEDGMENTS

This work was partially supported by grants from the National Council for Scientific and Technological Research (CONICET) (PIP No. 11220120100050OC, and the National Agency for Scientific and Technological Promotion, ANPCyT, PICT N° 1199) Mariela Marani is a researcher at CONICET. This work was supported by Marie Curie Actions – International Research Staff Exchange Scheme FP7-PEOPLE-2013-IRSES with grant No. 612545 entitled "GMOsensor – Monitoring Genetically Modified Organisms in Food and Feed by Innovative Biosensor Approaches," FCT/MEC through national funds and cofinanced by FEDER, within the framework of the Partnership Agreement PT2020 with grant No. UID/QUI/50006/2013 – POCI/01/0145/FEDER/007265 and by the CNPq/CBAB 2014 Nanobiotec Project. The authors also acknowledge CENAPAD-UFC for the computational resources used for the development of theoretical calculations. Alexandra Plácido and M. Fátima Barroso are grateful for the FCT grants (SFRH/BD/97995/2013 and SFRH/BPD/78845/2011, respectively) financed by POPHQREN (subsidized by FSE and MCTES).

REFERENCES

1. Premanandh J. Global consensus–Need of the hour for genetically modified organisms (GMO) labeling. J Commer Biotechnol 2011, 17, 37–44;
2. Singh A, Kumar V, Poonam GH. Genetically modified food: A review on mechanism of production and labeling concern. Adv Plants Agric Res 2014;1.
3. Commission E. Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. Off J Eur Union 2001;106:1–38.
4. James C. 20th Anniversary of the global commercialization of biotech crops (1996 to 2015) and biotech crop highlights in 2015. ISAAA Brief No. 51 2015.
5. Dehury B, Sahu M, Sahu J, Sarma K, Sen P, Modi MK, Barooah M, Choudhury MD. 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. J Mol Model 2013;19:5301–5316.
6. Qi Y-B, Ye S-h, Lu Y-T, Jin Q-S, Zhang X-M. Development of marker-free transgenic Cry1Ab rice with Lepidopteran pest resistance by Agrobacterium mixture-mediated co-transformation. Rice Sci 2009;16:181–186
7. Van Rie J. Bacillus thuringiensis and its use in transgenic insect control technologies. Int J Med Microbiol 2000;290:463–469
8. Douville M, Gagne F, Masson L, McKay J, Blaise C. Tracking the source of Bacillus thuringiensis Cry1Ab endotoxin in the environment. Biochem Syst Ecol 2005;33:219–232
9. Douville M, Gagné F, Blaise C, Andre C. Occurrence and persistence of Bacillus thuringiensis (Bt) and transgenic Bt corn cry1Ab gene from an aquatic environment. Ecotoxicol Environ Saf 2007;66:195–203
10. Flores S, Saxena D, Stotzky G. Transgenic Bt plants decompose less in soil than non-Bt plants. Soil Biol Biochem 2005;37:1073–1082
11. Peng D, Chen S, Ruan L, Li L, Yu Z, Sun M. Safety assessment of transgenic Bacillus thuringiensis with VIP insecticidal protein gene by feeding studies. Food Chem Toxicol 2007;45:1179–1185.
12. Singh OV, Ghai S, Paul D, Jain RK. Genetically modified crops: success, safety assessment, and public concern. Appl Microbiol Biotechnol 2006;71:598–607.
13. Marmiroli N, Maestri E, Gullì M, Malcevschi A, Peano C, Bordoni R, De Bellis G. Methods for detection of GMO in food and feed. Anal Bioanal Chem 2008;392:369–384.
14. Kashyap S. Computational modeling deduced three dimensional structure of Cry1Ab16 toxin from Bacillus thuringiensis AC11. Ind J Microbiol 2012;52:263–269.
15. Kashyap S, Singh B, Amla D. Homology modeling deduced 3D structure of the Cry1Ab22 toxin. Ind J Biotechnol 2011;10:202–206.
16. Maithri SK, Ramesh KV, Dieudonné M, Deshmukh S. Molecular modeling and docking studies of PirB fusion protein from Photorhabdus luminescens. Int Res J Biol Sci 2012;1:7–18.
17. Eswar N, Webb B, Marti-Renom M, Madhusudhan M, Eramian D, Shen M, Pieper U, Sali A. Curr Protoc Protein Sci Chapter 5. Unit: 2006; Comparative protein structure modeling using Modeller. doi: 10.1002/0471250953.bi0506s15.
18. Sali A, Blundell T. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
19. Shen M. y, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006;15:2507–2524.
20. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
21. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 1993;2:1511–1519.
22. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35 (Suppl 2): W407–W410

23. Sippl MJ, Recognition of errors in three-dimensional structures of proteins. Proteins Struct Funct Bioinform 1993;17: 3–3.
24. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 2004;25:1656–1676.
25. D, van der Spoel EL, B, Hess AR, van Buuren E, Apol PJ, Meulenhoff DP, Tieleman ALTM, Sijbers KA, Feenstra R, van Drunen HJC. Berendsen Gromacs User Manual version 3.3. 2005.
26. Berendsen HJ, Postma J. v, van Gunsteren WF, DiNola A, Haak J. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
27. Nosé S, Klein M. Constant pressure molecular dynamics for molecular systems. Mol Phys 1983;50:1055–1076.
28. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 1981;52:7182–7190.
29. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997;18: 1463–1472.
30. Darden T, York D, Pedersen L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
31. Plácido A, de Oliveira Farias EA, Marani MM, Vasconcelos AG, Mafud AC, Mascarenhas YP, Eiras C, Leite JR, Delerue-Matos C. Layer-by-layer films containing peptides of the Cry1Ab16 toxin from Bacillus thuringiensis for potential biotechnological applications. Mater Sci Eng C 2016;61:832–841
32. Plácido A, de Oliveira Farias EA, Marani MM, Vasconcelos AG, Leite JR, Delerue-Matos C. Peptide isolated from Cry1Ab16 toxin present in Bacillus thuringiensis: Synthesis and morphology data for layer-by-layer films studied by atomic force microscopy. Data Brief 2016;8:114–119.
33. Marani MM, Costa J, Mafra I, Oliveira MBP, Camperi SA, de Souza Almeida Leite JR. In silico peptide prediction for antibody generation to recognize 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) in genetically modified organisms. Pept Sci 2015;104: 91–100.
34. Merrifield RB. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. J Am Chem Soc 1963;85:2149–2154.
35. Wolf P. A critical reappraisal of Waddell's technique for ultraviolet spectrophotometric protein estimation. Anal Biochem 1983;129: 145–155.
36. Harlow E, Lane D. A laboratory manual. New York: Cold Spring Harbor Laboratory; 1988, 579
37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612.
38. Turner P, Team G, Stambulchik E. xmGrace Modeling Software. 2004.