In silico Structural Characterization of Class II Plant Defensins from *Arabidopsis thaliana*

Laura S.M. Costa[1,2], Állan S. Pires[1], Neila B. Damaceno[1], Pietra O. Rigueiras[1], Mariana R. Maximiano[1], Octavio L. Franco[1,2,3], William F. Porto[3,4]*

[1] Centro de Análises Proteômicas e Bioquímicas. Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília-DF, Brazil.

[2] Departamento de Biologia, Programa de Pós-Graduação em Genética e Biotecnologia, Universidade Federal de Juiz de Fora, Campus Universitário, Juiz de Fora-MG, Brazil.

[3] S-Inova Biotech, Pós-Graduação em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande-MS, Brazil.

[4] Porto Reports, Brasília-DF, Brazil – www.portoreports.com

*Corresponding author: williamfp7@yahoo.com.br

Abstract

Defensins compose a polyphyletic group of multifunctional defense peptides. The cis-defensins, also known as cysteine stabilized αβ (CSαβ) defensins, are one of the most ancient defense peptide families. In plants, these peptides have been divided in two classes, according to their precursor organization. Class I defensins are composed by the signal peptide and the mature sequence, while the class II defensins have an additional C-terminal prodomain, which is posteriorly cleaved. The class II defensins have been described only in Solanaceae species, which indicated that this class is restricted to this family. In this work, a search by regular expression (RegEx) was applied to *Arabidopsis thaliana* proteome, a model plant with more than 300 predicted defensin genes. Two sequences were identified, A7REG2 and A7REG4, which have a typical plant defensin structure and an additional C-terminal prodomain. The evolutionary distance between Brassicaceae and Solanaceae and the presence class II defensin sequences in both families suggest that class II may be derived from a common eudicots ancestor. The discovery of class II defensins in other plants could shed some light in the plant physiology, as this class plays multiple roles in such context.

Keywords: Defensin evolution; Gene Duplication; Multifunctional Peptides; Structural Prediction; Regular Expression.

1 Introduction

Defensins compose a polyphyletic group of multifunctional defense peptides, with a clear division between cis- and trans-defensins. Currently, it is not clear if these classes share a common ancestor, mainly due to their distributions, while the trans-defensins are mainly present in vertebrates, the cis-defensins are present in invertebrates, plant and fungi (Shafee et al., 2017).

The cis-defensins, also known as cysteine stabilized αβ (CSαβ) defensins are one of the most ancient defense peptide families (Zhu, 2008). Usually, the CSαβ defensins are composed by 50 to 60 amino acids residues with three to five disulfide bridges. Their secondary structure is composed of an α-helix and a β-sheet, formed by two or three β-strands (Lacerda et al., 2014). They also present two conserved domains including (i) the α-core, which consists in a loop that connects the first β-strain to the α-helix, and (ii) the γ-core, a hook harboring the GXC sequence, that connects the second and third β-strands (Yount et al., 2007; Yount and Yeaman, 2004). The γ-core is important to be highlighted because it is shared with trans-defensins and also other classes of defense peptides stabilized by disulfide bonds, such as heveins (Porto et al., 2012b), cyclotides (Porto et al., 2016) and knottins (Cammue et al., 1992). These conserved features allow their identification in sequence databases (Porto et al., 2017), as demonstrated by Porto et al. (Porto et al., 2014), who has found a new defensin from *Mayetiola destructor* (MdesDEF-2) between 12 sequences classified as hypothetical (Porto et al., 2014); and by Zhu, in the identification of 25 new defensins from 18 genes of 25 species of fungus (Zhu, 2008).

According to Zhu (Zhu et al., 2005), the CSαβ defensins can be divided in three subtypes: (I) ancient invertebrate-type defensins (AITDs); (II) classical insect-type defensins (CITDs); (III) plant/insect-type defensins (PITDs). The PITDs are known to have three β-strands in their structures, and also at least four disulfide bridges, regardless of the discovery of an *Arabidopis thaliana* defensin with only three disulfide bridges (Omidvar et al., 2016). The plant defensins could present diverse functions (van der Weerden and Anderson, 2013), resulting from gene differentiation after gene duplication events, process also known as peptide promiscuity (Franco, 2011). In plants, the PITDs could be divided in two major classes, depending on their precursor organization. class I defensins are composed of a signal peptide and a mature defensin; while class II defensins present an additional C-terminal prodomain (Lay and Anderson, 2005).

This classification of plant defensins has a little bias due to the dependence on the precursor sequence, which is cleaved to release the mature defensin. Because of that number of class II defensins would be classified as class I, mainly in cases where the precursor sequence is not available. Therefore, the class I defensins end up being the largest class, while class II have only been characterized in solanaceous species (Lay and Anderson, 2005).

Thus, assuming this scenario, we hypothesized that other plants also produce class II defensins and once the classification is dependent on the precursor sequence, which can be obtained by cDNA sequences, we can identify those class II defensins using the large amount of biological data available in public-access databases (Porto et al., 2017). In the post-genomic era, several sequences resulting from automatic annotations and without functional annotations can be found in biological sequences

databases. Therefore, those can be a source of uncharacterized defensins, and a number of studies have been showing the possibility to identify cysteine stabilized peptides using the sequences available in databases, mainly those annotated as hypothetical, unnamed or unknown proteins (Porto et al., 2017).

Consequently, here we used the predicted proteome from *Arabidopsis thaliana* (Brassicaceae), a model plant which has at least 300 predicted defensin-like peptides (Silverstein et al., 2005), to identify the class II defensins and then characterize their structures by means of comparative modeling followed by molecular dynamics.

2 Results

2.1 Defensin identification

In order to identify unusual defensin sequences we designed a semiautomatic pipeline (Figure 1). For that, we initially download all proteins from *Arabidopsis thaliana* Uniprot database. The set consists of 86.486 sequences (March 2017). From this dataset we performed a search by using regular expression (RegEx) which resulted in 387 sequences (step 2, Figure 1). From these, 285 had up to 130 amino acids residues (step 3, Figure 1). This criterion was used since, generally, AMPs have up to 100 amino acid residues in its mature chain (Yount and Yeaman, 2013) and, in this way, it allows the identification of C-terminal prodomains. Then, we used a PERL script to select the sequences with the flags: hypothetical, unknown, unnamed and uncharacterized (step 4, Figure 1), resulting in 15 sequences. From 15 final sequences, seven were incomplete and therefore were discarded (step 5, Figure 1). From the remaining sequences, were discarded those without signal peptide or with transmembrane domains (step 6, Figure

1). Finally, the sequences with C-Terminal prodomain were selected, resulting in two final sequences with accession codes A7REG2 and A7REG4.

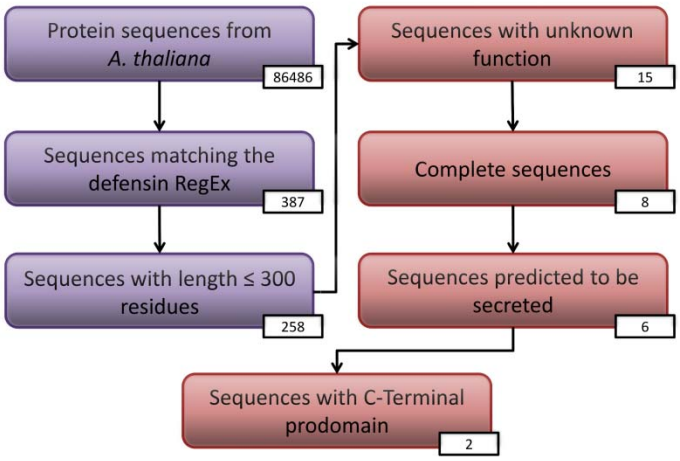

Figure 1. Flowchart of identification of class II defensins. The indigo boxes indicated steps performed by Perl scripts and the red boxes, steps curated by hand. The black boxes indicated the number of sequences for each step. The sequences from A. thaliana were retrieved from UniProt database; the defensin RegEx "$CX_{2-18}CX_3CX_{2-10}$[GAPSIDERYW]X_1CX_{4-17}CXC" was determined by Zhu (Zhu, 2008); The complete sequences were retrieved according the UniProt annotations; and the sequences predicted to be secreted were selected according to the phobius prediction, presenting signal peptide and no transmembrane domains.

2.2 Characterization of A7REG2 and A7REG4 as class II defensins

At the end of the search process (Figure 1), two sequences had fit in the parameters, corresponding to A7REG2 and A7REG4 that contained 71 and 64 amino acid residues in its mature chain, respectively (Figure 2). In order to assess the signal peptide from the obtained sequences, Phobius (Käll et al., 2007) and SignalP 4.0 (Petersen et al., 2011) were used. Phobius was used in the pipeline for sequence discovery due to its dual function of identify signal peptides and also transmembrane regions, however for a more accurate prediction SignalP was also included, as described by Porto et al. (Porto et al., 2012b). Phobius indicated signal peptides from A7REG2 and A7REG4 with 19 and 23 amino acid residues, respectively, while SignalP indicated both sequences with signal peptides with 23 residues. Because these

sequences are paralogous, we considered the prediction of SignalP for defining the signal sequences (Figure 2). Next, the sequences from *A. thaliana* were aligned with three class II defensins that had their cDNA and protein sequences characterized (Lay and Anderson, 2005), which allows the identification of the probable cleavage site of *A. thaliana* class II defensins. By using the alignment, depicted in Figure 2, was noted the conservation of two charged residues in the C-terminal prodomain could be observed. Those residues are located two amino acids far from the last cysteine residue, similarly to the cleavage point of solanaceous sequences (Figure 2). The first charged residue is characterized by an acidic residue, while the second could be an acidic or a basic one (Figure 2).

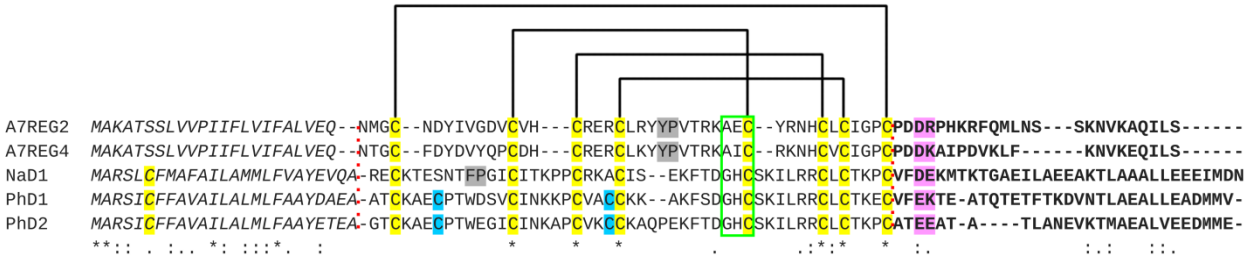

Figure 2. Alignment between sequences obtained at the end of automatic search and known plant defensins. The red dots indicate the cleavage points of signal peptide and pro-region. The □-core location is highlighted by green box; none of the identified sequences have the characteristic □-core GXCX3-9C between the IV and VI cysteines. Cysteines are highlighted in yellow and the lines above the sequences indicate the bond pattern between them. The cis-proline motif is marked in grey. The predicted cleavage site is marked as pink. This site presents a pair of charged residues, being the first always negative.

2.3 Three-Dimensional Structure Modelling

The mature sequences of both defensins presented a βαββ formation in their structures and also the four disulfide bonds that stabilize the structure, following the bond pattern between Cys^{I}-Cys^{VIII}, Cys^{II}-Cys^{V}, Cys^{III}-Cys^{VI} and Cys^{IV}-Cys^{VII} residues, which is common in plant defensins (PITD's) (Figure 3).

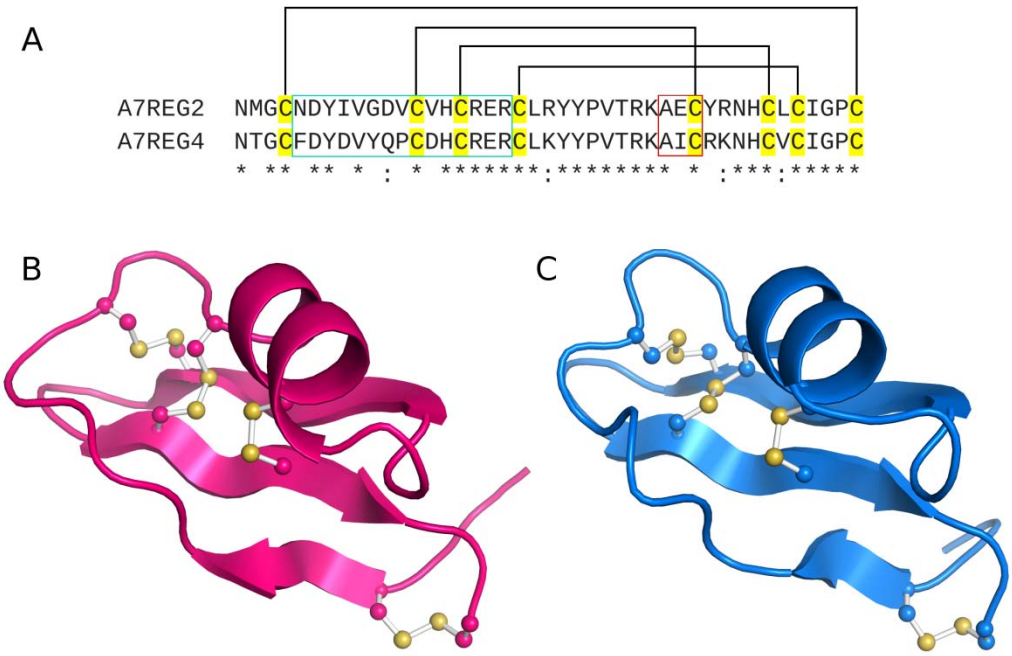

Figure 3. Tridimensional molecular models of the sequences. (A) After the removal of signal peptide and C-terminal prodomain, both sequences resulted in a 43 amino acid residues mature chain with 72% of identity. (B) A7REG2 and (C) A7REG4 structures. Disulfide bonds are represented in ball and sticks. Both models were generated using the sugar cane (*Saccharum officinarum*) defensin 5 (SD5, PDB ID: 2KSK) (de Paula et al., 2011). On the Ramachandran plot, A7REG2 model presented 83.3% of residues in favored regions, 13.9% in allowed regions and 2.8% in generously allowed regions; while A7REG4 model presented model presented 77.8% of residues in favored regions, 13.9% in allowed regions and 8.3% in generously allowed regions. In addition, the models also presented a Z-Score on ProSa II of -4.47 and -5.74, respectively.

In order to evaluate the peptides structural maintenance their models were submitted to 300 ns of molecular dynamics simulations. From each trajectory, we analyze the backbone root mean square deviation (RMSD) and the residue root mean square fluctuation (Figure S1A). These analyses showed that A7REG2 and A7REG4 presented an average deviation of 3Å, indicating that the initial topology is maintained.

Due to the fact that the sequences do not present the classical γ-core sequence (where a glycine residue was expected in the 30th position, both sequences presented an alanine residue, depicted in Figure 3), we performed the DSSP analysis to evaluate the secondary structure evolution during the simulation (Figure 4). Despite the maintenance of the secondary structure at the end of the simulation, the first β-strand in

A7REG2 structure could fold and unfold during the simulation period; while in A7REG4, this portion alternates between a β-strand and a β-bridge (Figure 4). However, this behavior is not related to the γ-core sequence. Because the γ-core is close to the α-helix, the minimum distance between the Ala^{30} and the Arg^{17} was measured. Ala^{30} is in the same axis that Arg^{17}, thus any stereochemical clash with would involve these residues, however, no clashes were observed as the minimum distance between these residues was >2 Å (Figure S1B).

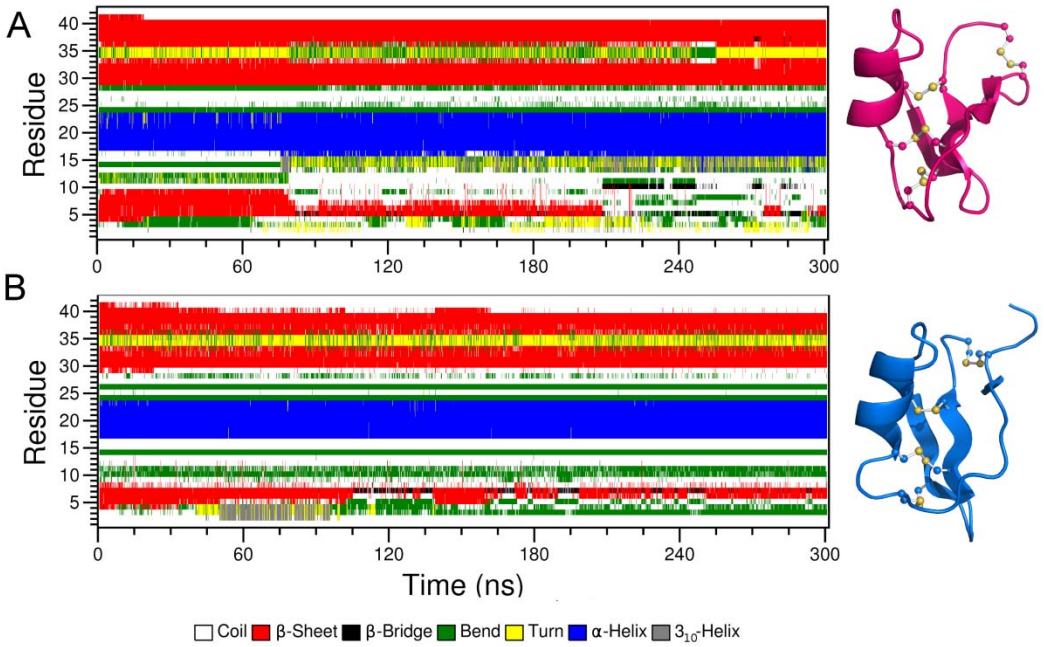

Figure 4. Secondary structure evolution during the simulations. The overall secondary structures of (A) A7REG2 and (B) A7REG4 are maintained during the simulations, with exception of the first β-strand, which could transit to β-bridge and/or coils. The final three dimensional structures at 300 ns of simulation are displayed in the right side of DSSP. Disulfide bridges are represented in ball and sticks.

The peptides identified here present a high sequence similarity, being a possible evidence of recent gene duplication (Figure 3A). Thus, we took advantage of molecular dynamics simulations to analyze the structural changes generated by the amino acid residues exchanges between the sequences. The region between the first and second

cysteines presented the highest number of mutations (Figure 3A). This portion comprises the first β-strand, which, as demonstrated by DSSP, could fold and unfold (Figure 4), however, the 10th position draws attention because there is a change involving a glycine and a tyrosine residue, this change alters directly the movement of Arg^{17}, because the Tyr residue in A7REG4 makes a sandwich with Tyr^{23}; while in A7REG2, the Gly residue allows the free moment of Arg^{17} (Figure S2).

3 Discussion

During the evolution process, plants were (and are still being) constantly exposed to biotic stresses. The CSαβ defensins play a pivotal role in plant defense against such stresses, showing activity against fungal, insects and bacteria (van der Weerden and Anderson, 2013). And these multiple activities are related to events of gene duplication, which are common during plant evolution. Multiple copy cysteine-rich peptides genes have been reported in many plant organisms such as *Triticum aestivum*, *Medicago truncatula* and *Arabidopsis thaliana* (Silverstein et al., 2007, 2005). Such events allow a phenomenon also known as “peptide promiscuity” (Franco, 2011), where one copy is maintained intact while the other could accumulate some mutations, which could confer multiple functions to the same peptide family (Franco, 2011).

Considering these gene duplication events together the fact that, in other CSαβ defensin producing organisms, the precursor organization is similar to class I, the class II defensins may be derived of a class I gene. In fact, the majority of plant defensins belong to the class I, even with the bias that some of them do not have the precursor sequence elucidated. However, this brings up the question: when did this gene duplication event occur? Considering only the previous knowledge, this event would be restricted to Solanaceae family, as the class II defensins were restricted to such.

However, the identification in *A. thaliana* (Brassicaceae) we could infer that this event that generated the class II defensins occurred in an common ancestor of eudicots, due to the evolutionary distance between Solanaceae and Brassicaceae families.

Despite this distance between these plant families, the mechanism of precursor processing seems to be conserved. In the N-terminal they presented the signal peptide, which contains ~25 amino acid residues (Figure 2), while in the C-terminal, there is a two charged residues signature near the cleavage point (Figure 2). Moreover, the mature peptide presents the classical structure of PITDs, with a β-sheet formed by three β-strands, interconnected to an α-helix, being stabilized by three disulfide bonds (Figure 3).

Still from the point of view of gene duplication and accumulation of mutations, there are two positions in the mature peptide that should be highlighted, being the 30th residue in a global perspective; and the 10th residue in a local one. In A7REG2 and A7REG4, the 30th position is filled up by an alanine residue, which means that both sequences do not have the classical γ-core, harboring the sequence “GXC”, instead they present “AXC”. This mutation is allowed by the RegEx “$CX_{2-18}CX_3CX_{2-10}$[GAPSIDERYW]$X_1CX_{4-17}CXC$”, determined by Zhu (Zhu, 2008), and thus, the sequences passed by our search system (Figure 1), however this position present an enormous structural restriction, because any bulker side chain would result in a stereochemical clash with the α-helix (Shafee et al., 2017). Nevertheless, despite rarely, alanine or serine residues could take the glycine place in the γ-core (Shafee et al., 2017). Besides, the DSSP analysis (Figure 4) indicated that the α-helix is kept during the simulation time; and the distance between the Ala^{30} and Arg^{17} indicated that there are no stereochemical clashes.

By the local perspective, the 10th residue in the mature chains present a remarkable difference between A7REG2 and A7REG4, while A7REG4 presents a tyrosine residue, A7REG2 presents a glycine at this position (Figure 3A). This mutation generates structural differences that affect the moves of Arg^{17}. In A7REG4, the Tyr^{10} makes an arginine sandwich together with Tyr^{23}, which implies a spatial restraint, forcing the Arg^{17} to move only in same axis (Figure S2), while in A7REG2, the Gly^{10} removes this restriction, allowing Arg^{17} to more in more directions (Figure S2). Even with the predicted antimicrobial activity by CS-AMPPred (Porto et al., 2012a) and CAMP (Waghu et al., 2014) (Table S1), we don't know, in fact, the function of these defensins; and the position of positive charged residues could be important to the activity, as observed in VuD1 (Pelegrini et al., 2008) and Cp-Thionin (Melo et al., 2002).

4 Conclusion

In terms of duplicated defensins genes *A. thaliana* deserves a special highlight. Silverstein and co-workers described more than 300 sequences of defensin and defensin-like peptides for this organism, more than any plant described in the literature (Silverstein et al., 2005). Despite being a model organism with a well annotated genome, novel information on defensins has been increasingly discovered in this organism, including a typical plant defensin with only three disulfide bridges, and now, in the present manuscript, novel class II defensins. Despite the method applied here has been extensively used in the last decade (Porto et al., 2017), the application of emerging methods, such as the identification directly by the structure (Pires et al., 2019) and/or structure prediction by contact maps (Zhang et al., 2018) could bring novel information about the distribution and evolution of this intriguing peptide family. Indeed, the

discovery of class II defensins in other plants could shed some light in the plant physiology, as this class plays multiple roles in such context. Also, considering that these defensins were unnoticed in a well-annotated genome such as *A. thaliana*, there must be sequences like that in much more plants.

5 Material and methods

5.1 RegEx and sequence analysis

Figure 1 describes the search system that was used. Primarily the *A. thaliana* proteome was obtained from the UniProt (Universal Protein Resource -//http://www.uniprot.org/proteomes) protein data bank. Then, the RegEx: $CX_{2\text{-}18}CX_3CX_{2\text{-}10}$[GAPSIDERYW]$X^1CX_{4\text{-}17}$CXC was used in the group of obtained sequences, where each amino acid is presented by its one letter code; the “X” means that any proteinogenic amino acid can fit the position; and the brackets indicate that only one of those amino acids between the brackets fit in that position, elaborated by Zhu (Zhu, 2008). From the resulting sequences, were selected the peptides with 130 amino acids residues or less and those that didn’t had functional validation. The group without validation includes sequences with the tags: hypothetical, uncharacterized, unnamed and unknow. From the remaining group, were removed all partial sequences, and with no signal peptide or with transmembrane region. The predictions of signal peptide and transmembrane region were made by Phobius (Käll et al., 2007). The selected sequences were evaluated for the presence of C-terminal prodomains after the last cysteine residue.

5.2 Molecular Modelling

The selection of structural templates was performed by using the LOMETS server. Then, the peptides were modelled using MODELLER 9.19 (Webb and Sali, 2014). The models were constructed using the default methods of environ class and a modified automodel class to include the cis-peptide restraints. One hundred models were generated for each sequence and the best model was selected by the DOPE (Discrete Optimized Protein Structure) score that indicates the most probable structure. The selected models were also evaluated by Prosa II (Wiederstein and Sippl, 2007), that analyzes the model's quality by comparing it to proteins from PDB; and PROCHECK (Laskowski et al., 1993), which evaluates the model's stereochemistry quality by using the Ramachandran plot. PyMOL (www.pymol.org) was used to visualize the models.

5.3 Molecular Dynamics Simulations

GROMACS 4.6 (Hess et al., 2008) was used for the molecular dynamics simulations, under the all atom CHARMM36 force field (Vanommeslaeghe et al., 2009). Each structure was immersed in a water cubic box that had a distance of 8 Å to the edges of the box. The cubic box was filled with the single point charge (Berendsen et al., 1981) water model and a NaCl concentration of 0.2M. Additional counter ions were added to the system to neutralize the charges. The geometry of water molecules was forced through SETTLE (Miyamoto and Kollman, 1992) algorithm, the atomic bonds were made by LINCS (Hess et al., 1997) algorithm, and the electrostatic correlations were calculated by Particle Mesh Ewald (Darden et al., 1993) algorithm, with a threshold of 1.4 nm to minimize computational time. The same threshold was used to van der Waals interactions. The neighbor searching was done with Verlet cutoff scheme. The

steepest descent algorithm was used to minimize the system's energy for 50.000 steps. After the energy minimization, the temperature (NVT) and the pressure (NPT) of the system were normalized to 300 K and 1 bar, respectively, for 100 ps each. The complete simulation of the system lasted for 300 ns, using the leap-frog algorithm as an integrator.

5.4 Molecular Dynamics Simulation Analysis

The simulations were evaluated for their root mean square deviation (RMSD) and root mean square fluctuation (RMSF) of each amino acid residue - of the peptides throughout the simulation using, respectively, the g_rms and g_rmsf tools from GROMACS pack. RMDS calculations were done using the initial structure at 0 ns of the simulation. The peptides were also evaluated for their structure conservation by DSSP 2.0, using do_dssp tool from GROMACS.

6 Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Apoio a Pesquisa do Distrito Federal (FAPDF) and Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (FUNDECT).

7 References

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J., 1981. Interaction Models for Water in Relation To Protein Hydration, in: Pullman, B. (Ed.), Intermolecular Forces. Springer, pp. 331–338.

Cammue, B.P., De Bolle, M.F., Terras, F.R., Proost, P., Van Damme, J., Rees, S.B., Vanderleyden, J., Broekaert, W.F., 1992. Isolation and characterization of a novel

class of plant antimicrobial peptides form Mirabilis jalapa L. seeds. J. Biol. Chem. 267, 2228–33.

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089. doi:10.1063/1.464397

de Paula, V.S., Razzera, G., Barreto-Bergter, E., Almeida, F.C.L., Valente, A.P., 2011. Portrayal of Complex Dynamic Properties of Sugarcane Defensin 5 by NMR: Multiple Motions Associated with Membrane Interaction. Structure 19, 26–36. doi:10.1016/J.STR.2010.11.011

Franco, O.L., 2011. Peptide promiscuity: an evolutionary concept for plant defense. FEBS Lett. 585, 995–1000. doi:10.1016/j.febslet.2011.03.008

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 18, 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput. 4, 435–447. doi:10.1021/ct700301q

Käll, L., Krogh, A., Sonnhammer, E.L.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res. 35, W429-32. doi:10.1093/nar/gkm256

Lacerda, A.F., Vasconcelos, Ã.A.R., Pelegrini, P.B., Grossi de Sa, M.F., 2014. Antifungal defensins and their role in plant defense. Front. Microbiol. 5, 116. doi:10.3389/fmicb.2014.00116

Laskowski, R., Macarthur, M., Moss, D., Thornton, J., 1993. {PROCHECK}: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291.

Lay, F.T., Anderson, M.A., 2005. Defensins--components of the innate immune system in plants. Curr. Protein Pept. Sci. 6, 85–101.

Melo, F.R., Rigden, D.J., Franco, O.L., Mello, L. V., Ary, M.B., Grossi de Sá, M.F., Bloch, C., 2002. Inhibition of trypsin by cowpea thionin: Characterization, molecular modeling, and docking. Proteins Struct. Funct. Bioinforma. 48, 311–319. doi:10.1002/prot.10142

Miyamoto, S., Kollman, P.A., 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. 13, 952–962. doi:10.1002/jcc.540130805

Omidvar, R., Xia, Y., Porcelli, F., Bohlmann, H., Veglia, G., 2016. NMR structure and conformational dynamics of AtPDFL2.1, a defensin-like peptide from Arabidopsis thaliana. Biochim. Biophys. Acta - Proteins Proteomics 1864, 1739–1747. doi:10.1016/j.bbapap.2016.08.017

Pelegrini, P.B., Lay, F.T., Murad, A.M., Anderson, M.A., Franco, O.L., 2008. Novel insights on the mechanism of action of alpha-amylase inhibitors from the plant defensin family. Proteins 73, 719–29. doi:10.1002/prot.22086

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–6. doi:10.1038/nmeth.1701

Pires, Á.S., Rigueiras, P.O., Dohms, S.M., Porto, W.F., Franco, O.L., 2019. Structure-

guided identification of antimicrobial peptides in the spathe transcriptome of the non-model plant, arum lily (*Zantedeschia aethiopica*). Chem. Biol. Drug Des. doi:10.1111/cbdd.13498

Porto, W.F., Fensterseifer, G.M., Franco, O.L., 2014. In silico identification, structural characterization, and phylogenetic analysis of MdesDEF-2: a novel defensin from the Hessian fly, Mayetiola destructor. J. Mol. Model. 20, 2339. doi:10.1007/s00894-014-2339-9

Porto, W.F., Miranda, V.J., Pinto, M.F.S., Dohms, S.M., Franco, O.L., 2016. High-performance computational analysis and peptide screening from databases of cyclotides from poaceae. Biopolymers 106, 109–118. doi:10.1002/bip.22771

Porto, W.F., Pires, A.S., Franco, O.L., 2017. Computational tools for exploring sequence databases as a resource for antimicrobial peptides. Biotechnol. Adv. 35. doi:10.1016/j.biotechadv.2017.02.001

Porto, W.F., Pires, Á.S., Franco, O.L., 2012a. CS-AMPPred: an updated SVM model for antimicrobial activity prediction in cysteine-stabilized peptides. PLoS One 7, e51444. doi:10.1371/journal.pone.0051444

Porto, W.F., Souza, V.A., Nolasco, D.O., Franco, O.L., 2012b. In silico identification of novel hevein-like peptide precursors. Peptides 38, 127–136. doi:10.1016/j.peptides.2012.07.025

Shafee, T.M.A., Lay, F.T., Phan, T.K., Anderson, M.A., Hulett, M.D., 2017. Convergent evolution of defensin sequence, structure and function. Cell. Mol. Life Sci. 74, 663–682. doi:10.1007/s00018-016-2344-5

Silverstein, K.A.T., Graham, M.A., Paape, T.D., VandenBosch, K.A., 2005. Genome organization of more than 300 defensin-like genes in Arabidopsis. Plant Physiol. 138, 600–10. doi:10.1104/pp.105.060079

Silverstein, K.A.T., Moskal, W.A., Wu, H.C., Underwood, B.A., Graham, M.A., Town, C.D., VandenBosch, K.A., 2007. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J. 51, 262–280. doi:10.1111/j.1365-313X.2007.03136.x

van der Weerden, N.L., Anderson, M.A., 2013. Plant defensins: Common fold, multiple functions. Fungal Biol. Rev. 26, 121–131. doi:10.1016/J.FBR.2012.08.004

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., Mackerell, A.D., 2009. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. 31, NA-NA. doi:10.1002/jcc.21367

Waghu, F.H., Gopi, L., Barai, R.S., Ramteke, P., Nizami, B., Idicula-Thomas, S., 2014. CAMP: Collection of sequences and structures of antimicrobial peptides. Nucleic Acids Res. 42, D1154-8. doi:10.1093/nar/gkt1157

Webb, B., Sali, A., 2014. Comparative Protein Structure Modeling Using MODELLER. Curr. Protoc. Bioinformatics 47, 5.6.1-5.6.32. doi:10.1002/0471250953.bi0506s47

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407-10. doi:10.1093/nar/gkm290

Yount, N.Y., Andrés, M.T., Fierro, J.F., Yeaman, M.R., 2007. The gamma-core motif correlates with antimicrobial activity in cysteine-containing kaliocin-1 originating from transferrins. Biochim. Biophys. Acta 1768, 2862–72. doi:10.1016/j.bbamem.2007.07.024

Yount, N.Y., Yeaman, M.R., 2013. Peptide antimicrobials: cell wall as a bacterial target. Ann. N. Y. Acad. Sci. 1277, 127–138. doi:10.1111/nyas.12005

Yount, N.Y., Yeaman, M.R., 2004. Multidimensional signatures in antimicrobial peptides. Proc. Natl. Acad. Sci. U. S. A. 101, 7363–8. doi:10.1073/pnas.0401567101

Zhang, C., Mortuza, S.M., He, B., Wang, Y., Zhang, Y., 2018. Template-based and free modeling of I-TASSER and QUARK pipelines using predicted contact maps in CASP12. Proteins 86 Suppl 1, 136–151. doi:10.1002/prot.25414

Zhu, S., 2008. Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CSalphabeta defensins. Mol. Immunol. 45, 828–38. doi:10.1016/j.molimm.2007.06.354

Zhu, S., Gao, B., Tytgat, J., 2005. Phylogenetic distribution, functional epitopes and evolution of the CSalphabeta superfamily. Cell. Mol. Life Sci. 62, 2257–69. doi:10.1007/s00018-005-5200-6