Received: 6 August 2018 | Revised: 29 December 2018 | Accepted: 31 January 2019

DOI: 10.1111/cbdd.13498

WILEY

SPECIAL ISSUE ARTICLE

Structure-guided identification of antimicrobial peptides in the spathe transcriptome of the non-model plant, arum lily (*Zantedeschia aethiopica*)

Állan S. Pires[1] | **Pietra O. Rigueiras**[1] | **Stephan M. Dohms**[1] | **William F. Porto**[2,3] | **Octavio L. Franco**[1,3]

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

[2]Porto Reports, Brasília, Brazil

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

Correspondence
Octavio L. Franco, Centro de Análises Proteômicas e Bioquímicas, Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília, Brazil.
Email: ocfranco@gmail.com

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico; Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul; Fundação de Apoio à Pesquisa do Distrito Federal; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

Abstract

Antimicrobial peptides (AMPs) are small molecules present in all living beings. Despite their huge sequence variability, AMPs present great structural conservation, mainly in cysteine-stabilized families. Moreover, in non-model plants, it is possible to detect cysteine-stabilized AMPs (cs-AMPs) with different sequences not covered by conventional searches. Here, we described a threading application for cs-AMP identification in the non-model arum lily (*Zantedeschia aethiopica*) plant, exploring the spathe transcriptome. By using the predicted proteins from the *Z. aethiopica* transcriptome as our primary source of sequences, we have filtered by using structural alignments of 12 putative cs-AMP sequences. The two unreported sequences were submitted to PCR validation, and ZaLTP7 gene was confirmed. By using the structure alignments, we classified ZaLTP7 as an LTP type 2-like. The successful threading application for cs-AMP identification is an important advance in transcriptomic and proteomic data mining. Besides, the same approach could be applied to the use of NGS public data to discover molecules to combat multidrug-resistant bacteria.

KEYWORDS

cysteine-stabilized antimicrobial peptides, molecular modeling, next-generation sequencing, threading techniques

1 | INTRODUCTION

A wide number of bacteria can be harmful to human health, either by the production of metabolites and toxins or by direct contagion (Corona & Martinez, 2013). Due to the continuous use of antibiotics and the development of microorganism resistance, there are currently no effective treatments for many infections, generating the need for identification of new antimicrobial molecules (Corona & Martinez, 2013). In this context, antimicrobial peptides (AMPs) have been proposed as an alternative for controlling resistant microorganism infections.

The AMPs consist of small molecules active against bacteria, fungi, and viruses, being virtually present in all living beings (Ganz, 2003). Currently, there are two major groups of AMPs, the disulfide-free and the cysteine-stabilized peptides (cs-AMPs) (Brogden, 2005; Porto, Pires, & Franco, 2017a). While the disulfide-free peptides are mainly composed of helical or disordered structures, cs-AMPs present a wide structural variation, such as β-sheets, α-helical, and mixed structures (Porto et al., 2017a).

Until now, cs-AMPs have comprised a great portion of the natural AMPs described. These peptides are identified in a many organisms such as Big defensins in mollusks (Kouno

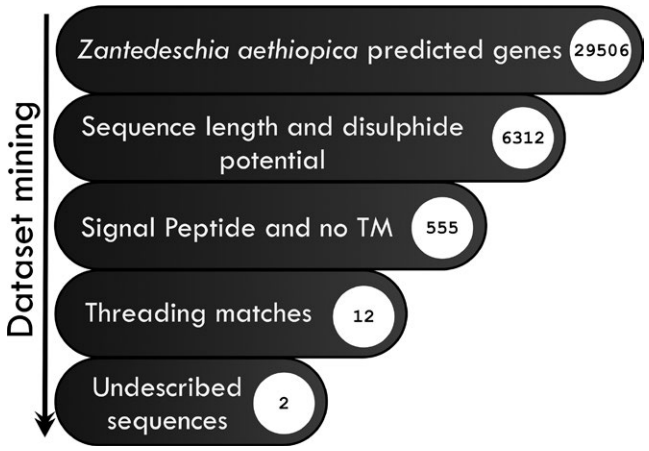

FIGURE 1 Summary of the data mining process. The number of sequences of each step of filtering is shown inside the white circles. The *Zantedeschia aethiopica* transcriptome contains 29,506 predicted proteins (step 1). Using this initial dataset, we obtained 6,312 sequences with 350 amino acids or less (step 2), and from these, 555 sequences with signal peptide and no transmembrane signature according to Phobius (step 3). These sequences were submitted to structural alignments by HHPred in order to select sequences with structural resemblance with plant cysteine-stabilized-antimicrobial peptides structure dataset (Supporting Information Table S1). The signal peptide and transmembrane regions (TM) recognition were performed by Phobius. The threading matches evaluation was performed by using HHPred alignments (step 4). The undescribed sequences were retrieved from the threading matches by comparisons with Cândido, Fernandes et al. (2014) (step 5)

et al., 2008), inhibitory cysteine knots in arthropods (Reily, Thanabal, & Adams, 1995), α- (Schibli et al., 2002), β- (Wei et al., 2009), and θ-defensins in vertebrates (Conibear, Rosengren, Harvey, & Craik, 2012). In plants, cs-AMPs comprise a major portion of known AMP families, including α-helical hairpins, cyclotides, cysteine-stabilized αβ-defensins (CSαβ defensins), Greek key β-barrel peptides, α- and β-thionin, lipid transfer proteins, and snakin families (Cândido, E. S., Cardoso, M. H. S., Cândido, Cardoso et al., 2014; Silverstein et al., 2007). Moreover, despite the structural and disulphide motif conservation, these families present a huge sequence variation which confers on them a wide range of activities, even for members of the same family (Tam, Wang, Wong, & Tan, 2015).

Commonly, natural extracts are the main sources for cs-AMPs identification (Kumar, Kizhakkedathu, & Straus, 2018). Despite of this, the creation of publicly available web servers and databases following the Chou's five-step rule (Chou, 2011) has been determinant in medical science. As pointed out by Chou and Shen (2009), web servers represent the future for many important computational analyses (Cheng, Zhao, Xiao, & Chou, 2017; Hamra et al., 1993; Meher, Sahu, Saini, & Rao, 2017; Wang et al., 2011; Xiao, Wang, Lin, Jia, & Chou, 2013). Indeed, they have been driving medical science into an unprecedented revolution (Chou, 2017). Moreover, for AMPs search and their activity prediction, these approaches represent a great advance and increased impact of computational science (Porto et al., 2017a). Further, given the great increase of computational resources and the next-generation sequencing technologies at reduced costs, the search for cs-AMPs in transcriptomic and genomic data has become more accessible (Ke et al., 2015; Porto et al., 2017a; Zhang et al., 2015). In this way, our group used previously described regular expressions (RegEx) by Silverstein et al. (2007) to identify novel cs-AMPs in the non-model plant aurum lily (*Zantedeschia aethiopica*) transcriptome (Cândido, Fernandes et al., 2014). Ten potential AMPs were identified from a total of 29,506 proteins. They matched with cs-AMP patterns and present cs-AMP domains, including three LTPs, three snakin/GASA peptides, and one chimerolectin containing a hevein domain. These peptides were associated with the antimicrobial activity of the protein-rich fraction of the arum lily's spathe (Cândido, Fernandes et al., 2014).

However, *Z. aethiopica*, as a non-model plant, might present peptide sequences which could not be covered by RegEx searches. In that case, a direct use of molecular modeling in data mining seems to be a reasonable alternative (Porto et al., 2017a). Threading techniques predict the secondary structure of a sequence and compare it with experimentally determined structures. Due to the fact that threading is based on secondary structure, this approach could find structural similarities even in cases of extremely divergent sequences (Gille, Goede, Preißner, Rother, & Frömmel, 2000; Porto et al., 2017a). Thus, we hypothesized that such approach could be applied to identify new cs-AMP sequences in transcriptome data, based on structural conservation. Moreover, we applied HHPred for searching AMPs in the previously described arum lily (*Z. aethiopica*) spathe transcriptome (Cândido, Fernandes et al., 2014).

2 | METHODS AND MATERIALS

2.1 | Databases

The *Z. aethiopica* sequence data (reads, contigs, and protein sequences) were retrieved from the previous report by our group (Cândido, Fernandes et al., 2014). This dataset comprises 91,218,320 total paired-end reads, 24,469,876 clean reads, 83,578 contigs, and 29,506 proteins (Cândido, Fernandes et al., 2014). For structural alignments with solved structures, we downloaded structures from Protein Data Bank (PDB November, 2017; https://www.rcsb.org/).

2.2 | Representative cysteine-stabilized AMP sets

An initial set of structure queries was submitted to DALI server (Holm & Rosenström, 2010), in order to collect a

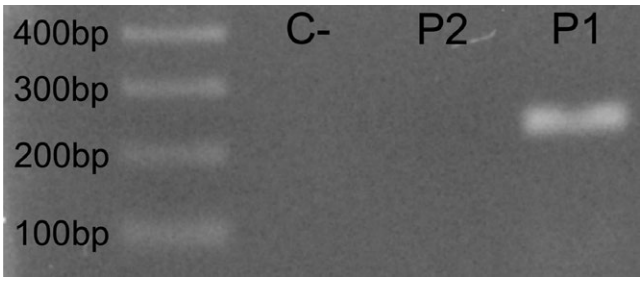

FIGURE 2 PCR amplified samples electrophoresis gel of. C− is the negative control, P2 and P1 are respectively, the pair two and one of primers (Supporting Information Table S2). The leader Kb Plus was used. The 212 pb band identify the lipid transfer peptide (LTP) gene in the plant *Zantedeschia aethiopica* DNA

representative set of plant cs-AMP families. The cysteine-rich AMP database was formed by all resulting structures with a *Z*-score above 2 and <3 Å root mean square deviation (RMSD) (Shafee, Lay, Hulett, & Anderson, 2016). According to this, we assembled a data set of 439 representative cs-AMP PDB IDs divided into eight families. The initial queries selected for each family as well the resulting database are listed in Supporting Information Table S1.

2.3 | Data mining

The overview of data mining is shown in Figure 1. From the protein sequences from *Z. aethiopica,* we used Perl scripts to filter sequences with length of 350 amino acids or less, and containing four or more cysteine residues. After that, we selected peptides with the presence of signal peptide and without signatures of transmembrane regions by using Phobius (Käll, Krogh, & Sonnhammer, 2007). These steps were performed according to Cândido, Fernandes et al. (2014). To find the AMPs from the *Z. aethiopica* transcriptome, the remaining sequences were submitted to HHPred (Söding, Biegert, & Lupas, 2005) for secondary structure alignments. Sequences that matched with the defined representative cs-AMP structures with an e-value <0.05 were collected. The matching peptides were compared with the sequences described by Cândido, Fernandes et al. (2014), and the uncharacterized cs-AMPs were selected.

2.4 | DNA extraction

Plant material of *Z. aethiopica* in adult growth phase was acquired from ornamental collections located in Brasília—DF (Brazil). Genomic DNA was extracted using 0.1 g of *Z. aethiopica* spathe samples and 500 µl preheated CTAB extraction buffer (2% CTAB [w/v], 100 mM Tris–HCl; pH 8.0, 20 mM EDTA; pH 8.0, 1.4 M NaCl, 0.2% 2-mercaptoethanol [v/v], added before use) was added and placed in a 60°C water bath for 30 min. Samples were separated by centrifugation, and supernatant was transferred to fresh tubes containing 500 µl chloroform:isoamyl alcohol (24:1, v/v), followed by centrifugation at 5,000 *g* for 10 min. Upper phase was transferred to new tubes containing 0.6 volume of isopropanol and incubated at −20°C for 20 min before centrifugation at 10,000 *g* for 20 min. Supernatant was discarded and pellet washed with 500 µl alcohol 70% (v/v). After centrifugation, supernatant was removed and pellet dried and resuspended in 50 µl sterile ultrapure water.

2.5 | PCR analysis

The primers were designed using PRIMER3PLUS (Untergasser et al., 2007) by using default parameters. Selected sequences were 19–21 bp, 57° of annealing temperature, and GC content close ranging from 50% to 60%. The reactions were conducted with 1.5 ng/µl genomic DNA, 2 µl of buffer at 10× dilution, 0.3 nM of each primer, 250 mM of dNTPs, and 2.5 U of recombinant Taq DNA polymerase (Invitrogen). The PCR program used was an initial step of 95°C for 2 min followed by 30 cycles of 95°C for 30 s, 57°C for 30 s, and 72°C for 30 s, with an extension time of 7 min at 72°C. The results were analyzed in 1% agarose–gel electrophoresis (Figure 2). The PCR product was purified in silica columns and submitted to Sanger sequencing. The sequences of the primers used are listed in Supporting Information Table S2.

2.6 | Domain assignment

New identified peptides were submitted to domain assignment by INTERPRO SCAN (Jones et al., 2014). This step was used to evaluate sequences for the presence of conserved domains and their extension.

2.7 | Contig reassembly

Raw Illumina data were processed according to Cândido, Fernandes et al. (2014) with minor modifications. The Trimmomatic tool was used to remove the sequence adaptors, using the provided fasta file with paired-end adaptors. The low-quality reads were also removed using Trimmomatic, by scanning the read with five-base sliding window, cutting when the average quality per base dropped below 15. The resulting sequence reads below 36 bp long were excluded. Then, the processed reads were mapped against the individual contig sequences of each peptide by using bowtie2, under the defaults (Langmead, Trapnell, Pop, & Salzberg, 2009). The mapped reads were then de novo assembled using the trinity tool (Grabherr et al., 2011), in order to obtain the complete sequences.

2.8 | Antimicrobial activity prediction

In order to evaluate the potential antimicrobial activity of the reassembled ZaLTP7, we used an array of antimicrobial peptide predictors, including CS-AMPPred (Porto, Pires, & Franco, 2012), CAMP (Waghu, Barai, Gurung, & Idicula-Thomas, 2016), iAMP-2L (Xiao et al., 2013), ADAM (Lee et al., 2015), and AntiBP2 (Lata, Mishra, & Raghava, 2010).

2.9 | Comparative molecular modeling

The molecular modeling was done by MODELLER v.9.16 (Fiser & Šali, 2003) using the best hit from HHPred analyses and the final sequence after transcriptome reassembly. The best model was selected according to DOPE (Discrete Optimized Protein Structure) score from 100 generated models. DOPE score indicates the most probable structure by measures of the energy corresponding to each model. The ligand coordinates were imported from template pdb file by setting as true the property io.hetatm from the class environ from MODELLER 9.16. The best model was evaluated by PROSA II (Wiederstein & Sippl, 2007) and PROCHECK (Laskowski, MacArthur, Moss, & Thornton, 1993). PROSA II evaluates the model quality, while PROCHECK evaluates the stereochemical quality of the model by means of a Ramachandran plot. Good quality models were selected by more than 90% of residues in most favored and additional allowed regions. The visualization of the structures was done using PYMOL (http://www.pymol.org).

TABLE 1 List of cysteine-stabilized-antimicrobial peptides sequences identified by structural alignments

Peptide Name	Sequence	Reference
ZaSNK1	NEEELYAQVSHPPTPSPAPAAAPVHHIVPVDEKECPGL CEVRCSKHSRPNHCHRVCQTCCRRCRCVPPGTAGNREMCGAC YTDMTTHRNATKCPKNVKCKVELGIINDKAAWLTVLYQCDG LHKKSKDKYVTGNCSGSCDFYVFAFHSLTHNNHAYGLDALIPLLPVLDNQR	Cândido, Fernandes et al. (2014)[a]
ZaSNK2	AMAGSGIYTITLRLPSQRYCCTPTTTVVSPPRGGRGVPRLRVSAAAL TEQMTSFGAEFCDTKCKARCSKASVHDRCLFNCGVCCKECNCV PSGTYGNKDECPCYRDKVTKDEKKKPKCP	Cândido, Fernandes et al. (2014)[a]
ZaSNK3	AGQSPAPAAQPTVPGPAAQPKLPMYGFTEGSLQPQECSGRCTGR CSATQYKKPCLFFCQKCCAKCLCVPPGTYGNKQFCPCYNDWKTKRGGPKCP	Cândido, Fernandes et al. (2014)[a]
ZaLTP1	ADDPPCSYVKSNLRPCGPYVFGRVAGPSPQCCAAVRQLNRIARTTAQ RRNVCRCLRGLEVEKAAAGRPFSLNAVRSLPARCGVNLGFPITRYVDCNRVP	Cândido, Fernandes et al. (2014)
ZaLTP2	AIMCGQVYRTLTPCMGYLRSGGKPSPPCCSGVRELVRLAGTTADRQATC RCLKAAAQGLSGNFGPAASLPSSCGVTISYKISTDTDCNKYLAPVPIPWTVIRSRDWTTYVT	Cândido, Fernandes et al. (2014)
ZaLTP3	TYCPPPTKKPPPKKPPPVKPPPAYPSCPRDTVKLGGCANVLNLLKVKLGA PPKETCCPLLKGLTDVEAAMCLCTVLKASIMGINLNLPVDLSLLLSYCGKSVPSGFKCA	Cândido, Fernandes et al. (2014)
ZaLTP4	QTGCTAMMVDLNPCSNYITGNISAPPVTCCTQLDNLVQTQMGCLCAFLA NGAPFGIPLNLAQVLSLPGACNIKAGPLSQCYGVATTPAGPSPVAPGPASP AMPSTPGSALAPSGTPATTPMTPSMPNRPSGAGSKTVPGGPVVGSSTRLG THPLLFLLLVAITSYASSGVGF	Cândido, Fernandes et al. (2014)[a]
ZaLTP5	SPIITKVTCLPLSRGNNPKSKGGIPALPPVVKPLPVPVVPPIITIPP VVGTVTCPVDAVKIGSCVDQLGGLVDVQLGEPAANVCCPVLEG LLEAEVAVCLCTTLQLKLCNLGIYVPVCLKLLSGLKYGGNHDIRQASDALY	Cândido, Fernandes et al. (2014)[a]
ZaLTP6	YLDCEDDDIYGLQGHCGKAVWVGAARAAPSTDCCHFVKERTSLACVCEN VVTPRHEKYISMKKLAYVAGYCGAPLKPGTKLPGPAKALRLILLPREL KQPQGGAWKCSWGFGKCQGKIL	Cândido, Fernandes et al. (2014)[a]
ZaLTP7	ASWAPTAESVTCNPYELLPCAGAISGGSPSRECCARLRAQQPCLCGYARNPN LVRLVNIPKALQVASACKFLQRVFSLDTVVKPLPPVMAYNVSRNLSFFTRIFT QFFDPEGIANAQKSLGLGQEQKDRRFAVILKLWLKRIKHWHVLSLAA	This work
ZaβB1	EGSYMTAWDGPGCNNSAERYSACGCSSINLHGGYEFVYQGQTAATYNQ PDCQGNNFSGSKLVNIPATRQIVRIISRGNL	This work
ZaChL1	EQCGSQAGGALCPGGLCCSRFGWCGDTAAYCDPAQGCQSQCRGGPTP TPSPPTGGGGSGVGSIVSQSLFDRMLMHRNDAACPARNFYTYNAFIS AANSFGGFGTTDSEKARSFCGSGVSDVCSSDGSRGEWERE	Cândido, Fernandes et al. (2014)[a]

[a]Sequences described but without structural analysis.

2.10 | Molecular dynamics simulation

The molecular dynamics simulations of the model and template structures were performed by GROMACS 4 computational package using the GROMOS96 43A1 for ce field (Hess et al., 2008). Ligand parameterization was performed by PRODRG2 web ser ver at chirality, full charges, and no energy minimization conditions (Schüttelkopf & Van Aalten, 2004). Structures were immersed in cubic water boxes with 8 Å distance between the edge of the box and the structure. The simulations were done under ionic strength conditions (0.2 M NaCl). The box was filled using the Single Point Charge water model (Berendsen, Postma, van Gunsteren, & Hermans, 1981). Additional chlorine and sodium ions were also inserted into the complexes with positive and negative charges to neutralize the system charge, respectively. Geometry of water molecules was constrained by using the SETTLE algorithm (Miyamoto & Kollman, 1992). Atomic connections were made through LINCS algorithm (Hess, Bekker, Berendsen, & Fraaije, 1997). Electrostatic corrections were made by Particle Mesh Ewald algorithm (Darden, York, & Pedersen, 1993), with 1.4-nm cutoff to minimize the computational time. The same cutoff radius was applied for van der Waals interactions. The steepest descent algorithm was applied to minimize system energy for 50,000 steps. After the energy minimization, the temperature (NVT ensemble) and pressure (NPT ensemble) systems were normalized to 300 K and 1 bar, respectively, each per 100 steps. The velocity-rescaling thermostat and the Parrinello–Rahman barostat were used for normalization of temperature and pressure, respectively. Full simulation of the system was made by 200 ns using the leap–frog algorithm as the integrator. Molecular dynamics simulations were analyzed by means of the backbone RMSD using the g_rms in functions of the GROMACS package. Visualizations were done using PYMOL Molecular Graphics System, version 1.6 Schrödinger, LLC.

3 | RESULTS

3.1 | Threading as an efficient method for cs-AMP identification

Zantedeschia aethiopica is an evergreen ornamental plant used in traditional medicine. Previously, our group found antimicrobial activity in floral tissues and ten potential cs-AMP sequences were described as agents of this function (Cândido, Fernandes et al., 2014) (Table 1). Applying HHPred as a search engine, we retrieved 12 putative cs-AMP sequences, seven LTPs (ZaLTP1-7), three snakins (ZaSnakin1-3), one hevein (ZaHev1), and one β-barrelin (ZaβB1; Figure 1). From the data mining, we found two new sequences corresponding to a class II LTP and a β-barrelin (Table 1). Both peptides were submitted to PCR analysis, and only the ZaLTP7gene was confirmed (Figure 2).

TABLE 2 Antimicrobial prediction of ZaLTP7 in publicly available AMP predictors

Predictor	Algorithm	Prediction	Reference
iAMP2L	–	AMP[a]	Xiao et al. (2013)
CAMP	SVM	AMP	Waghu et al. (2016)
	RF	AMP	
	ANN	AMP	
	DA	AMP	
CS-AMPPred	Polynomial	AMP	Porto et al. (2012)
AntiBP2	–	NAMP	Lata et al. (2010)
ADAM	SVM	AMP	Lee et al. (2015)
	HMM	AMP	

Notes. AMP: antimicrobial peptides; ANN: Artificial Neural Network; DA: discriminant analysis; HMM: Hidden Markov Model; RF: random forest; SVM: support vector machine.

[a]The sequence was predicted as antibacterial, antifungal, and antiviral (scores: 0.65, 0.78, and 0.79, respectively).

3.2 | Reassembled ZaLTP7 sequence does not present the unusual C-Terminal portion

Za-LTP7 was submitted to domain recognition by Interpro Scan. The peptide presented the non-specific lipid transfer protein type 2 domain (PF14368) at N-terminal portion. Given the results of domain assignment, we mapped the reads against the individual contigs for the ZaLTP7 sequence and then performed a new de novo assembly. The ZaLTP7 contig had 331 mapped reads. The de novo assembly of mapped reads indicated some divergences in relation to the initial sequences. The ZaLTP7 revised version did not have the unusual C-terminal portion, but was instead a complete LTP sequence. Therefore, the initial sequence of ZaLTP7 was a result from an erroneous chimeric transcript (Grabherr et al., 2011).

3.3 | ZaLTP7 is predicted to have antimicrobial activity

In order to evaluate the antimicrobial peptide activity, we submit ZaLTP7 sequence to a series of free web available predictors. The peptide presents positive prediction in almost all predictors gives a reliable indicative of its activity (Table 2).

3.4 | ZaLTP7 is an LTP type 2-like and presents the same family topology

Usually, type I and II LTPs present different disulfide pattern (Finkina, Melnikova, Bogdanov, & Ovchinnikova, 2016). Despite of that, all cysteines could be aligned. In this

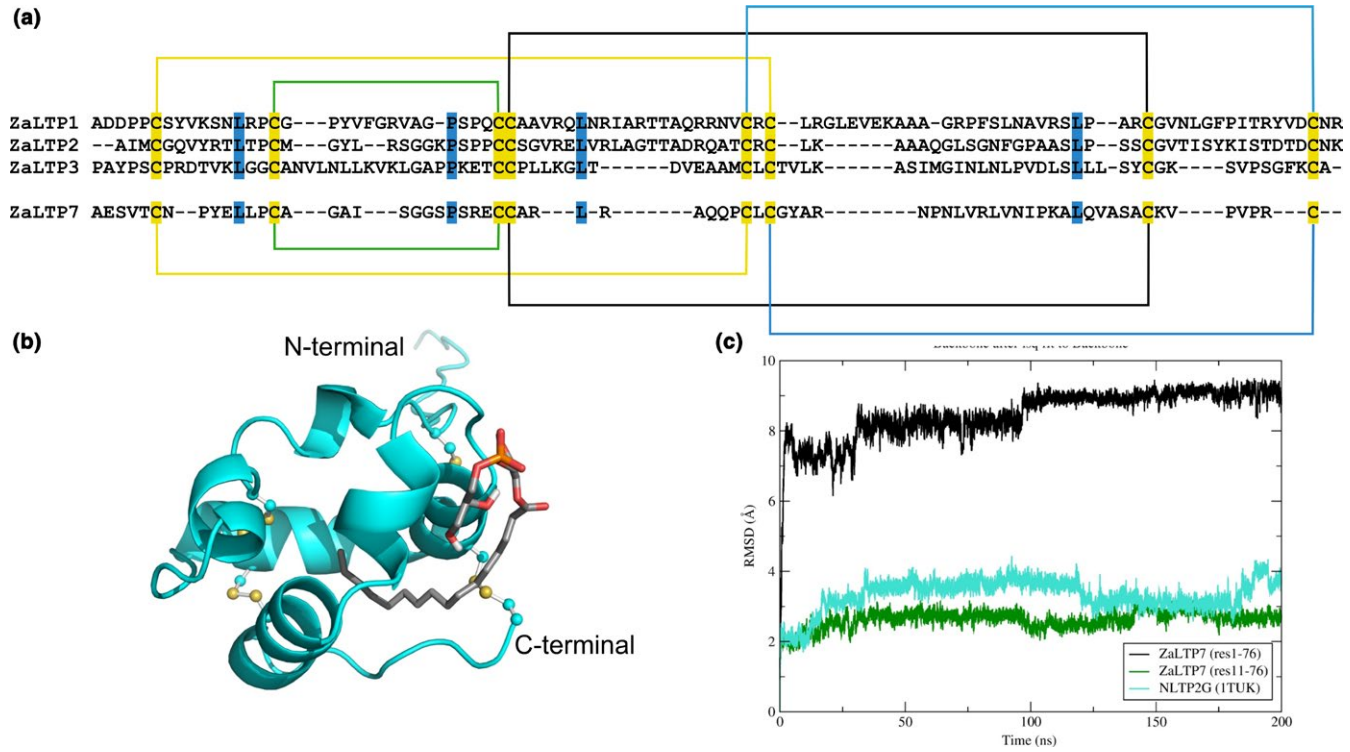

FIGURE 3 (a) Sequence alignment of previously analyzed ZaLTP1-3 (Cândido, Fernandes et al., 2014) and ZaLTP7 (this work). Some N- and C-terminal amino acids were omitted in order to facilitate the visualization. Each disulphide bridge is present with different colors. Cysteine residues are highlighted in yellow and conserved residues in cyan. (b) Three dimensional structure of ZaLTP7. The disulphide bridges are shown in ball and sticks, and the ligand (lysophosphatidylglycerol) is present in gray sticks. (c) The root mean square deviation plot of ZaLTP7 and NLTP2G simulations. In black, the fluctuation of full structure backbone, in green, without the first ten loop residues, and in cyan, the template trajectory. The N-terminal loop is the major responsible for variation on the analyses

case, the use of molecular modeling could be essential in classification of the peptides. The sequence of ZaLTP7 was retrieved from reassembled contig. The type 2 non-specific LTP from *Triticum aestivum* (PDB ID: 1TUK) (Hoh, Pons, Gautier, De Lamotte, & Dumas, 2005) was chosen as the best template which indicate that ZaLTP7 is a LTP type 2-like. This protein presents the typical LTP fold, with a four-helix bundle, stabilized by four disulfide bonds (Hoh et al., 2005). However, in contrast to ZaLTP7, ZaLTP1-3 (Cândido, Fernandes et al., 2014) were classified as a type 1 LTP-like (Figure 3a). In addition, the template structure contains a lysophosphatidylglycerol as ligand. For ZaLTP7 model, the ligands were imported from template coordinates by MODELLER (Figure 3b).

The both structures (template and predicted model) were submitted to molecular dynamics simulations of 200 ns. After the simulations, we analyzed the backbone RMSDs of the peptides. The analyses showed the behavior of the template and the ZaLTP7 without the N-terminal loop were very similar, ranging from 2 to 3 Å for both (Figure 3c). The RMSD ranged from 2 to 3 Å without the N-terminal loop region, which had no coverage to modeling (^{1}ASWAPTAESV^{10}), while the complete structure ranged from 6 to 9 Å (Figure 3c). Thus, despite the initial residues, the ZaLTP7 core is very stable, reinforcing the classification as type 2 LTP-like.

4 | DISCUSSION

Commonly, AMPs have been identified and studied by traditional purification experiments, screening different targets (Mandal et al., 2013, 2014; Pinto et al., 2016). Molecular biology could also be an alternative for this identification (Cunha et al., 2016; Porto, Miranda, Pinto, Dohms, & Franco, 2016). However, nowadays, with the reduced costs of next-generation sequencing technologies, the use of computational data mining approaches for identification of AMPs has become more common (Ganz, 2003; Porto et al., 2017a). The use of RegEx should be highlighted among the common computational data mining techniques, and this has been used in a number of works, especially to find cs-AMP (Cândido, Fernandes et al., 2014; Cunha et al., 2016; Porto et al., 2016; Silverstein et al., 2007). Based on this, in a previous work, our group used pattern recognition to find 10 new potential cs-AMPs in the *Z. aethiopica* transcriptome (Cândido, Fernandes et al., 2014). However, in these cases, molecular modeling has been used only as an additional step

of validation (Cândido, Fernandes et al., 2014; Cunha et al., 2016; Porto et al., 2016; Silverstein et al., 2007). Here, we propose the use of threading as the search engine for AMP search instead of an additional step.

Despite the high performance of RegEx, this kind of search identifies sequences by the presence of expressions with fixed, ambiguous, and wildcard characters (Mulvenna et al., 2006; Porto et al., 2017a; Thompson, 1968), and for that reason this technique presents a limitation related to the absence of fixed characters (e.g., the initial ZaLTP7 sequence did not present the last cysteine residue, which probably hindered its identification by Cândido, Fernandes et al. 2014). Another limitation for the application of RegEx is its requirement of the sequence pattern (Porto et al., 2017a), which may become an obstacle in view of poorly studied cs-AMP families, such as β-barrelins, which do not have RegExs described so far. For that reason, this family was not sought in the previous report (Cândido, Fernandes et al., 2014).

However, the use of threading is limited by the existence of at least one solved structure deposited in databases such as PDB. The case of the snakin/GASA family exemplifies this situation. The first X-ray structure was recently developed (Yeung et al., 2016), and thus, until then, the threading application to find novel snakins had not been possible without a representative structure of the family (Porto et al., 2017a). Sequence alignments, in turn, do not have this obstacle as a limiting factor.

The existence of this unnoticed sequence in the transcriptome gives the possibility of new findings in databases with uncharacterized sequences, such as the NCBI non-redundant database, which has been used as a resource for the discovery of new AMP sequences (Cândido, Fernandes et al., 2014; Porto, Fensterseifer, & Franco, 2014; Porto, Fensterseifer, Ribeiro, & Franco, 2018; Porto, Souza, Nolasco, & Franco, 2012; Zhu, 2008). This information could be exploited to search for different sequences from those already reported in the literature. Besides, it could be helpful in the discovery of sequences with new disulfide patterns and unusual amino acid compositions (Ganz, 2003; Porto et al., 2017a). In this way, with the application of this method in the *Arabidopsis thaliana* proteome, for example, it would be possible to find new defensin-like peptides even with the 300 sequences already identified using RegEx by Silverstein, Graham, Paape, and VandenBosch (2005). Thus, the identification of new features that do not influence the maintenance of the structural topology could be possible. However, this kind of approach requires more computational processing than local alignments or ReGex, due to the application of structural predictions in addition to sequence alignment itself.

For searches in non-model plants, for example, this approach could reveal distinct cs-AMPs, with possible new targets and functions. ZaLTP7 in the initial data presents seven cysteines, a different composition of LTPs already described (Table 1). Although the unusual composition has been demonstrated as an assembly artifact, this fact can be used to exemplify the capacity of the method. In addition, peptides from this family have been reported to be active against bacteria (Lin et al., 2007; Tam et al., 2015), fungi (Finkina et al., 2016; Lin et al., 2007), and HIV (Finkina et al., 2016). Besides, the protein-rich fraction of aurum lily floral tissues present antimicrobial activity against *Escherichia coli*, which could be a good sign of the possible application of peptides from this family as antibacterial substances (Cândido, Fernandes et al., 2014). Moreover, despite the antimicrobial activity prediction algorithms are not quite accurate for synthetic peptides (Porto, Pires, & Franco, 2017b), they have a reasonable performance to natural peptides (Mandal et al., 2014), and ZaLTP7 presents antimicrobial prediction in almost all predictors used in the study (Table 2).

Three dimensional structures have been essential for classification of plant AMPs. With that kind of approach, even in groups such as LTP that presents the same cysteine disposal, it is possible to evaluate their impact on structure (Figure 3). For that reason, both X-ray crystallography and NMR have been used for that purpose. However, despite the great power of these techniques, they are time-consuming and expensive. Moreover, not all structures can be successfully solved by these methods (Porto, Nolasco, & Franco, 2014). To acquire the structural information in a timely manner, a series of 3D protein structures were developed by means of comparative molecular modeling and were found very useful for drug development (Chou, 2005; Chou, Jones, & Heinrikson, 1997; Chou, Tomasselli, & Heinrikson, 2000; Li, Wang, Xu, Wang, & Chou, 2011; Ma, Wang, Xu, Wang, & Chou, 2012; Wang, Du, & Chou, 2007; Wang, Du, Huang, Zhang, & Chou, 2009). In view of this, the threading technique was also adopted for molecular modeling in the current study.

Moreover, since 1997, many works have been done to investigate the dynamics of biomolecules. Such studies have analyzed the occurrence of low-frequency motions in proteins and their influence in protein function (Chou, 1988; Martel, 1992). Furthermore, many biological functions such as cooperative effects (Chou, 1988) and allosteric transition (Chou, 1988) have been demonstrated by such analyses. Moreover, the investigation of the internal motion of proteins and other macromolecules as well as their function in biological activity is one of the new frontiers of biological physics. In that way, to understand the activity and mechanisms of action of biomacromolecules such as proteins, we need to analyze not only their static structure but their dynamics in biological solvents as well. In view of this, molecular dynamics has become a powerful tool in the study this kind of motions of proteins. In our work, by using molecular dynamics approach, we evaluate the template and the peptide model motions

(Figure 3c). Based on the ZaLTP7 structure and behavior in contrast with its template, we could classify ZaLTP7 as an LTP type 2-like.

5 | CONCLUSION

Here, we identified two unreported and ten known cs-AMPs in *Z. aethiopica* transcriptomic data previously analyzed by RegEx. This demonstrated that threading is an efficient tool for the identification of AMPs in NGS data. Although the data presented here characterize an advance in the mining of transcriptomic data, in vitro tests with the identified peptide are still necessary. However, ZaLTP7 could be used as a structural scaffold for novel synthetic peptides (Porto et al., 2018). Moreover, threading can be applied in larger datasets, exploring the large amount of data in public databases as well as NGS data, turning the data into new knowledge to combat multidrug-resistant bacteria.

ACKNOWLEDGMENTS

The authors are grateful to Mr. Onoyama for providing the arum lily plants. This work was supported by CAPES, CNPq, FAPDF, and FUNDECT.

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

ORCID

Állan S. Pires https://orcid.org/0000-0002-5633-3838
William F. Porto https://orcid.org/0000-0002-4739-8250
Octavio L. Franco https://orcid.org/0000-0001-9546-0525

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. *Intermolecular Forces*, *31*, 331–338. https://doi.org/10.1007/978-94-015-7658-1_21

Brogden, K. A. (2005). Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology*, *3*(3), 238–250. https://doi.org/10.1038/nrmicro1098

Cândido, E. S., Cardoso, M. H. S., Sousa, D. A., Viana, J. C., de Oliveira- Júnior, N. G., Miranda, V., & Franco, O. L. (2014). The use of versatile plant antimicrobial peptides in agribusiness and human health. *Peptides*, *55*, 65–78. https://doi.org/10.1016/j.peptides.2014.02.003

Cândido, S., Fernandes, G., de Alencar, S. A., Cardoso, M. H., Lima, S. M., Miranda Vde, J., … Franco, O. L. (2014). Shedding some light over the floral metabolism by arum lily (*Zantedeschia aethiopica*) spathe de novo transcriptome assembly. *PLoS ONE*, *9*(3), e90487. https://doi.org/10.1371/journal.pone.0090487

Cheng, X., Zhao, S. G., Xiao, X., & Chou, K. C. (2017). iATC-mISF: A multi-label classifier for predicting the classes of anatomical therapeutic chemicals. *Bioinformatics*, *436*(2), 168–177. https://doi.org/10.1093/bioinformatics/btw644

Chou, K. C. (1988). Low-frequency collective motion in biomacromolecules and its biological functions. *Biophysical Chemistry*, *30*(1), 3–48. https://doi.org/10.1016/0301-4622(88)85002-6

Chou, K. C. (2005). Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein. *Journal of Proteome Research*, *4*(5), 1681–1686. https://doi.org/10.1021/pr050145a

Chou, K. C. (2011). Some remarks on protein attribute prediction and pseudo amino acid composition. *Journal of Theoretical Biology*, *273*(1), 236–247. https://doi.org/10.1016/j.jtbi.2010.12.024

Chou, K.-C. (2017). An unprecedented revolution in medicinal chemistry driven by the progress of biological science. *Current Topics in Medicinal Chemistry*, *17*(21), 2337–2358. https://doi.org/10.2174/1568026617666170414145508

Chou, K. C., Jones, D., & Heinrikson, R. L. (1997). Prediction of the tertiary structure and substrate binding site of caspase-8. *FEBS Letters*, *419*(1), 49–54. https://doi.org/10.1016/S0014-5793(97)01246-5

Chou, K.-C., & Shen, H.-B. (2009). REVIEW : Recent advances in developing web-servers for predicting protein attributes. *Natural Science*, *1*, 63–92. https://doi.org/10.4236/ns.2009.12011

Chou, K. C., Tomasselli, A. G., & Heinrikson, R. L. (2000). Prediction of the tertiary structure of a caspase-9/inhibitor complex. *FEBS Letters*, *470*(3), 249–256. https://doi.org/10.1016/S0014-5793(00)01333-8

Conibear, A. C., Rosengren, K. J., Harvey, P. J., & Craik, D. J. (2012). Structural characterization of the cyclic cystine ladder motif of θ-defensins. *Biochemistry*, *51*(48), 9718–9726. https://doi.org/10.1021/bi301363a

Corona, F., & Martinez, J. (2013). Phenotypic resistance to antibiotics. *Antibiotics*, *2*, 237–255. https://doi.org/10.3390/antibiotics2020237

Cunha, N. B., Barbosa, A. E., de Almeida, R. G., Porto, W. F., Maximiano, M. R., Álvares, L. C., … Dias, S. C. (2016). Cloning and characterization of novel cyclotides genes from South American plants. *Biopolymers*, *106*(6), 784–795. https://doi.org/10.1002/bip.22938

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, *98*(12), 10089. https://doi.org/10.1063/1.464397

Finkina, E. I., Melnikova, D. N., Bogdanov, I. V., & Ovchinnikova, T. V. (2016). Lipid transfer proteins as components of the plant innate immune system: Structure, functions, and applications. *Acta Naturae*, *8*(2), 47–61

Fiser, A., & Šali, A. (2003). MODELLER: Generation and refinement of homology-based protein structure models. *Methods in Enzymology*, *374*, 461–491. https://doi.org/10.1016/S0076-6879(03)74020-8

Ganz, T. (2003). Defensins: Antimicrobial peptides of innate immunity. *Nature Reviews Immunology*, *3*(9), 710–720. https://doi.org/10.1038/nri1180

Gille, C., Goede, A., Preißner, R., Rother, K., & Frömmel, C. (2000). Conservation of substructures in proteins: Interfaces of secondary

structural elements in proteasomal subunits. *Journal of Molecular Biology*, *299*(4), 1147–1154. https://doi.org/10.1006/jmbi.2000.3763

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Hamra, F. K., Forte, L. R., Eber, S. L., Pidhorodeckyj, N. V., Krause, W. J., Freeman, R. H., … Smith, C. E. (1993). Uroguanylin: Structure and activity of a second endogenous peptide that stimulates intestinal guanylate cyclase. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(22), 10464–10468. https://doi.org/10.1073/pnas.90.22.10464

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463–1472. https://doi.org/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 simulations. *J Chem Theory Comput*, *4*, 435–447. https://doi.org/10.1021/ct700301q

Hoh, F., Pons, J. L., Gautier, M. F., De Lamotte, F., & Dumas, C. (2005). Structure of a liganded type 2 non-specific lipid-transfer protein from wheat and the molecular basis of lipid binding. *Acta Crystallographica. Section D, Biological Crystallography*, *61*(Pt 4), 397–406. https://doi.org/10.1107/S0907444905000417

Holm, L., & Rosenström, P. (2010). Dali server: Conservation mapping in 3D. *Nucleic Acids Research*, *38*(SUPPL. 2), W545–W549. https://doi.org/10.1093/nar/gkq366

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Research*, *35*(SUPPL.2), W429–W432. https://doi.org/10.1093/nar/gkm256

Ke, T., Cao, H., Huang, J., Hu, F., Huang, J., Dong, C., … Liu, S. (2015). EST-based in silico identification and in vitro test of antimicrobial peptides in Brassica napus. *BMC Genomics*, *16*, 653. https://doi.org/10.1186/s12864-015-1849-x

Kouno, T., Fujitani, N., Mizuguchi, M., Osaki, T., Nishimura, S., Kawabata, S., … Kawano, K. (2008). A novel β-defensin structure: A potential strategy of big defensin for overcoming resistance by gram-positive bacteria. *Biochemistry*, *47*(40), 10611–10619. https://doi.org/10.1021/bi800957n

Kumar, P., Kizhakkedathu, J. N., & Straus, S. K. (2018). Antimicrobial peptides: Diversity, mechanism of action and strategies to improve the activity and biocompatibility in vivo. *Biomolecules*, *8*, 1–24. https://doi.org/10.3390/biom8010004

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. https://doi.org/10.1107/S0021889892009944

Lata, S., Mishra, N. K., & Raghava, G. P. S. (2010). AntiBP2: Improved version of antibacterial peptide prediction. *BMC Bioinformatics*, *11*(Suppl 1), S19. https://doi.org/10.1186/1471-2105-11-S1-S19

Lee, H. T., Lee, C. C., Yang, J. R., Lai, J. Z. C., Chang, K. Y., & Ray, O. (2015). A large-scale structural classification of antimicrobial peptides. *BioMed Research International*, *2015*, 475062. https://doi.org/10.1155/2015/475062

Li, X. B., Wang, S. Q., Xu, W. R., Wang, R. L., & Chou, K. C. (2011). Novel inhibitor design for hemagglutinin against H1N1 influenza virus by core hopping method. *PLoS ONE*, *6*, e28111. https://doi.org/10.1371/journal.pone.0028111

Lin, P., Xia, L., Wong, J. H., Ng, T. B., Ye, X., Wang, S., & Shi, X. (2007). Lipid transfer proteins from Brassica campestris and mung bean surpass mung bean chitinase in exploitability. *Journal of Peptide Science*, *13*(10), 642–648. https://doi.org/10.1002/psc.893

Ma, Y., Wang, S. Q., Xu, W. R., Wang, R. L., & Chou, K. C. (2012). Design novel dual agonists for treating type-2 diabetes by targeting peroxisome proliferator-activated receptors with core hopping approach. *PLoS ONE*, *7*(6), e38546. https://doi.org/10.1371/journal.pone.0038546

Mandal, S. M., Bharti, R., Porto, W. F., Gauri, S. S., Mandal, M., Franco, O. L., & Ghosh, A. K. (2014). Identification of multifunctional peptides from human milk. *Peptides*, *56*, 84–93. https://doi.org/10.1016/j.peptides.2014.03.017

Mandal, S. M., Porto, W. F., Dey, P., Maiti, M. K., Ghosh, A. K., & Franco, O. L. (2013). The attack of the phytopathogens and the trumpet solo: Identification of a novel plant antifungal peptide with distinct fold and disulfide bond pattern. *Biochimie*, *95*(10), 1939–1948. https://doi.org/10.1016/j.biochi.2013.06.027

Martel, P. (1992). Biophysical aspects of neutron scattering from vibrational modes of proteins. *Progress in Biophysics and Molecular Biology*, *57*(3), 129–179. https://doi.org/10.1016/0079-6107(92)90023-Y

Meher, P. K., Sahu, T. K., Saini, V., & Rao, A. R. (2017). Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. *Scientific Reports*, *7*, 42362. https://doi.org/10.1038/srep42362

Miyamoto, S., & Kollman, P. A. (1992). SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, *13*(8), 952–962. https://doi.org/10.1002/jcc.540130805

Mulvenna, J. P., Mylne, J. S., Bharathi, R., Burton, R. A., Shirley, N. J., Fincher, G. B., … Craik, D. J. (2006). Discovery of cyclotide-like protein sequences in graminaceous crop plants: Ancestral precursors of circular proteins? *Plant Cell*, *18*(9), 2134–2144. https://doi.org/10.1105/tpc.106.042812

Pinto, M. F., Silva, O. N., Viana, J. C., Porto, W. F., Migliolo, L., B da Cunha, N., … Franco, O. L. (2016). Characterization of a bioactive acyclotide from *Palicourea rigida*. *Journal of Natural Products*, *79*(11), 2767–2773. https://doi.org/10.1021/acs.jnatprod.6b00270

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. *Journal of Molecular Modeling*, *20*(7), 1–8. https://doi.org/10.1007/s00894-014-2339-9

Porto, W. F., Fensterseifer, I. C. M., Ribeiro, S. M., & Franco, O. L. (2018). Joker: An algorithm to insert patterns into sequences for

designing antimicrobial peptides. *Biochimica et Biophysica Acta - General Subjects*, *1862*(9), 2043–2052. https://doi.org/10.1016/j.bbagen.2018.06.011

Porto, W. F., Irazazabal, L., Alves, E. S. F., Ribeiro, S. M., Matos, C. O., Pires, Á. S., … Franco, O. L. (2018). In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design. *Nature Communications*, *9*(1), 1490. https://doi.org/10.1038/s41467-018-03746-3

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*(1), 109–118. https://doi.org/10.1002/bip.22771

Porto, W. F., Nolasco, D. O., & Franco, O. L. (2014). Native and recombinant Pg-AMP1 show different antibacterial activity spectrum but similar folding behavior. *Peptides*, *55*, 92–97. https://doi.org/10.1016/j.peptides.2014.02.010

Porto, W. F., Pires, Á. S., & Franco, O. L. (2012). CS-AMPPred: An updated SVM model for antimicrobial activity prediction in cysteine-stabilized peptides. *PLoS ONE*, *7*(12), e51444. https://doi.org/10.1371/journal.pone.0051444

Porto, W. F., Pires, A. S., & Franco, O. L. (2017a). Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnology Advances*, *35*(3), 337–349. https://doi.org/10.1016/j.biotechadv.2017.02.001

Porto, W. F., Pires, Á. S., & Franco, O. L. (2017b). Antimicrobial activity predictors benchmarking analysis using shuffled and designed synthetic peptides. *Journal of Theoretical Biology*, *426*, 96–103. https://doi.org/10.1016/j.jtbi.2017.05.011

Porto, W. F., Souza, V. A., Nolasco, D. O., & Franco, O. L. (2012). In silico identification of novel hevein-like peptide precursors. *Peptides*, *38*(1), 127–136. https://doi.org/10.1016/j.peptides.2012.07.025

Reily, M. D., Thanabal, V., & Adams, M. E. (1995). The solution structure of ω-Aga-IVB, a P-type calcium channel antagonist from venom of the funnel web spider, *Agelenopsis aperta*. *Journal of Biomolecular NMR*, *5(2)*, 122–132. https://doi.org/10.1007/BF00208803

Schibli, D. J., Hunter, H. N., Aseyev, V., Starner, T. D., Wiencek, J. M., McCray, P. B. Jr, … Vogel, H. J. (2002). The solution structures of the human β-defensins lead to a better understanding of the potent bactericidal activity of HBD3 against Staphylococcus aureus. *Journal of Biological Chemistry*, *277*(10), 8279–8289. https://doi.org/10.1074/jbc.M108830200

Schüttelkopf, A. W., & Van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(Pt 8), 1355–1363. https://doi.org/10.1107/S0907444904011679

Shafee, T. M. A., Lay, F. T., Hulett, M. D., & Anderson, M. A. (2016). The defensins consist of two independent, convergent protein superfamilies. *Molecular Biology and Evolution*, *33*(9), 2345–2356. https://doi.org/10.1093/molbev/msw106

Silverstein, K., Graham, M., Paape, T., & VandenBosch, K. (2005). Genome organization of more than 300 defensin-like genes in Arabidopsis. *Plant Physiology*, *138*(2), 600–610. https://doi.org/10.1104/pp.105.060079

Silverstein, K. A., Moskal, W. A. Jr, 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 Journal*, *51*(2), 262–280. https://doi.org/10.1111/j.1365-313X.2007.03136.x

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(SUPPL. 2), W244–W248. https://doi.org/10.1093/nar/gki408

Tam, J. P., Wang, S., Wong, K. H., & Tan, W. L. (2015). Antimicrobial peptides from plants. *Pharmaceuticals*, *8*(4), 711–757. https://doi.org/10.3390/ph8040711

Thompson, K. (1968). Programming techniques: Regular expression search algorithm. *Communications of the ACM*, *11*, 419–422. https://doi.org/10.1145/363347.363387

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, *35*(SUPPL.2), W71–W74. https://doi.org/10.1093/nar/gkm306

Waghu, F. H., Barai, R. S., Gurung, P., & Idicula-Thomas, S. (2016). CAMPR3: A database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Research*, *44*(D1), D1094–D1097. https://doi.org/10.1093/nar/gkv1051

Wang, S. Q., Du, Q. S., & Chou, K. C. (2007). Study of drug resistance of chicken influenza A virus (H5N1) from homology-modeled 3D structures of neuraminidases. *Biochemical and Biophysical Research Communications*, *354*(3), 634–640. https://doi.org/10.1016/j.bbrc.2006.12.235

Wang, S. Q., Du, Q. S., Huang, R. B., Zhang, D. W., & Chou, K. C. (2009). Insights from investigating the interaction of oseltamivir (Tamiflu) with neuraminidase of the 2009 H1N1 swine flu virus. *Biochemical and Biophysical Research Communications*, *386*(3), 432–436. https://doi.org/10.1016/j.bbrc.2009.06.016

Wang, P., Hu, L., Liu, G., Jiang, N., Chen, X., Xu, J., … Chou, K. C. (2011). Prediction of antimicrobial peptides based on sequence alignment and feature selection methods. *PLoS ONE*, *6*(4), e18476. https://doi.org/10.1371/journal.pone.0018476

Wei, G., de Leeuw, E., Pazgier, M., Yuan, W., Zou, G., Wang, J., … Lu, W. (2009). Through the looking glass, mechanistic insights from enantiomeric human defensins. *Journal of Biological Chemistry*, *284*(42), 29180–29192. https://doi.org/10.1074/jbc.M109.018085

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(SUPPL.2), W407–W410. https://doi.org/10.1093/nar/gkm290

Xiao, X., Wang, P., Lin, W. Z., Jia, J. H., & Chou, K. C. (2013). IAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Analytical Biochemistry*, *436*(2), 168–177. https://doi.org/10.1016/j.ab.2013.01.019

Yeung, H., Squire, C. J., Yosaatmadja, Y., Panjikar, S., López, G., Molina, A., … Brimble, M. A. (2016). Radiation damage and racemic protein crystallography reveal the unique structure of the GASA/Snakin protein superfamily. *Angewandte Chemie International Edition*, *55*(28), 7930–7933. https://doi.org/10.1002/anie.201602719

Zhang, J., Li, J., Huang, Z., Yang, B., Zhang, X., Li, D., … Liao, B. (2015). Transcriptomic screening for cyclotides and other cysteine-rich proteins in the metallophyte Viola baoshanensis. *Journal of Plant Physiology*, *178*, 17–26. https://doi.org/10.1016/j.jplph.2015.01.017

Zhu, S. (2008). Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CSalphabeta defensins. *Molecular Immunology*, *45*(3), 828–838. https://doi.org/10.1016/j.molimm.2007.06.354

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Pires ÁS, Rigueiras PO, Dohms SM, Porto WF, Franco OL. Structure-guided identification of antimicrobial peptides in the spathe transcriptome of the non-model plant, arum lily (*Zantedeschia aethiopica*). *Chem Biol Drug Des*. 2019;00:1–11. https://doi.org/10.1111/cbdd.13498