UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
TESE DE DOUTORADO

MARX OLIVEIRA DE LIMA

CARACTERIZAÇÃO E ANÁLISE ESTRUTURAL DE PEPTÍDEOS ANTIMICROBIANOS DE PLANTAS

RECIFE
2018

MARX OLIVEIRA DE LIMA

CARACTERIZAÇÃO E ANÁLISE ESTRUTURAL DE PEPTÍDEOS ANTIMICROBIANOS DE PLANTAS

Tese apresentada como parte dos requisitos para a obtenção do título de Doutor em Ciências Biológicas.
Área de concentração: Biotecnologia/Biologia Celular e Molecular.

Orientadora: **Profª Drª Ana Maria Benko-Iseppon**
Coorientadora: **Profª Drª Valesca Pandolfi**

RECIFE
2018

Catalogação na fonte:
Bibliotecária Elaine C Barroso (CRB4/1728)

Lima, Marx Oliveira de
Caracterização e análise estrutural de peptídeos antimicrobianos de plantas / Marx Oliveira de Lima - 2018.
203 folhas: il., fig., tab.

Orientadora: Ana Maria Benko-Iseppon
Coorientadora: Valeska Pandolfi
Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-Graduação em Biotecnologia. Recife, 2018.

Inclui referências e anexos.

1. Peptídeos antimicrobianos 2. Bioinformática 3. Mineração de dados I. Benko-Iseppon, Ana Maria (orient.) II. Pandolfi, Valeska (coorient.) III. Título

572.65 CDD (22.ed.) UFPE/CB-2019-340

MARX OLIVEIRA DE LIMA

CARACTERIZAÇÃO E ANÁLISE ESTRUTURAL DE ESNAQUINAS EM PLANTAS DE INTERESSE ECONÔMICO

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como parte dos requisitos obrigatórios para a obtenção do título de Doutor em Ciências Biológicas.

APROVADO EM: 19/10/2018

COMISSÃO EXAMINADORA

MEMBROS TITULARES

_____

Drª Ana Maria Benko Iseppon
(Orientadora – Departamento de Genética/UFPE)

_____

Dr. Valdir Queiroz Balbino
(Membro Interno - Departamento de Genética – UFPE)

_____

Dr. João Pacífico Bezerra Neto
(Membro Externo - Departamento de Genética – UFPE)

_____

Dr. Antônio Mauro Rezende
(Membro Externo - Departamento de Genética – UFPE)

_____

Dr. Roberto Dias Lins Neto
(Membro Externo - Departamento de Química Fundamental – UFPE)

AGRADECIMENTOS

Agradeço à minha orientadora Profª Ana Benko, por toda dedicação e paciência, por sempre acreditar em mim e neste projeto e acima de tudo por ser esse grande exemplo que tenho a seguir.

À minha coorientadora e amiga Drª. Valesca Pandolfi, a quem sempre pude recorrer especialmente nas horas mais difíceis, onde sempre encontrei disposição e acima de tudo competência.

À Universidade Federal de Pernambuco (UFPE), por meio do Programa de Pós-Graduação em Ciências Biológicas (PPGCB), pela oportunidade e suporte durante todo o período do Mestrado.

À Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) pelo apoio financeiro durante todo o processo de Doutorado. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte através do Programa Nacional de Pós-Doutorado – PNPD / e Auxílio Financeiro– no âmbito do Edital Biologia Computacional 51/2013 no âmbito da Rede InterSys: Biologia Sistêmica no Estudo de Função Gênica em Interações Bióticas.

Aos demais colaboradores, em especial os membros do Centro de Análises Proteômicas e Bioquímicas da Universidade Católica de Brasília - DF, especialmente ao Professor Dr. Octávio Franco (bolsa FACEPE AMD), bem como do *Luxembourg Centre for Systems Biomedicine* da Universidade de Luxemburgo a meus supervisores Dr. Reinhard Schneider e Dr. Adriano Barbosa (Bolsa doutorado-sanduíche CAPES Biologia Computacional).

MuitoObrigado.

RESUMO

As esnaquinas são AMPs membros da família Snakin/GASA, devido ao seu domínio conservado que é caracterizado por um péptido sinal, uma região variável e um domínio C-terminal, o domínio GASA, que é composto por 12 cisteínas formando seis pontes dissulfeto. Este estudo objetivou a caracterização desta família de peptídeos no transcriptoma da soja, descrevendo sua estrutura, abundância e distribuição genômica. Para caracterizar os genes, foram utilizadas sequências *seed* de 10 famílias de plantas, assim, 33 sequências (não redundantes) de snakins foram identificadas no transcriptoma da soja, 20 das quais possuíam o domínio GASA completo, ponto isoelétrico variando de 5,53 a 9,32 para o peptídeo maduro e peso molecular de 6,88 a 18,04 KDa. Todas as sequências foram endereçadas ao meio extracelular e estão relacionadas tanto ao desenvolvimento natural da planta quanto aos estresses bióticos e abióticos, estes genes podem ser distribuídos em três subfamílias de acordo com os motivos internos. Uma nova esnaquina (GmSN2) tem aqui sua primeira descrição, isolamento gênico, sequenciamento genômico, validação da expressão RT-qPCR e análise de sua estrutura. O péptido mostra uma estrutura semelhante a outros AMPs como a tionina e é estabilizado por seis ligações dissulfeto com o padrão: Cis1-7, Cis2-5, Cis3-4, Cis6-12, Cis8-11 e Cis9-10. O isolado revelou expressão constitutiva, mesmo após inoculação com o fungo *Phakopsora pachyrhizi*. O gene possui três éxons e 40 homólogos distribuídos em 16 de 20 cromossomos de soja e em seis cromossomos de *P. vulgaris* e *M. truncatula*, com um padrão de distribuição diferente de outros genes de defesa, apresentando uma distribuição heterogênea. Aqui fornecemos uma fonte importante para mapear essa família de AMPs e também inferir sobre sua função.

Palavras-chave: Peptídeos antimicrobianos. Bioinformática. GENOSOJA. NordEST. Mineração de dados.

ABSTRACT

Snakins are members of the Snakin/GASA family, because of their conserved domain which is characterized by a signal peptide, a variable region and a C-terminal domain, the GASA domain, which is composed by 12 cysteines connecting six disulfide bonds. This study aimed the characterization of this peptide family in the soybean transcriptome describing structure, abundance and genomic distribution. To characterize the genes, we used seed sequences from 10 plant families, 33 (non-redundant) sequences of snakins were identified in the soybean transcriptome, 20 of which had the GASA domain complete, isoelectric point ranging from 5.53 to 9.32 for the mature peptide and molecular weight 6.88 to 18.04. All sequences were addressed to the extracellular environment and they are related to the natural development of the plant and to biotic and abiotic stresses, these genes can be distributed in three subfamilies according to internal motifs. A new snakin (GmSN2) has its first description, gene isolation, genomic sequencing, RT-qPCR expression validation and analysis of its structure. The peptide shows a similar structure to other AMPs like thionin and it is stabilized by six disulfide bonds with the pattern: Cys1-7, Cys2-5, Cys3-4, Cys6-12, Cys8-11, and Cys9-10. The isolated GmSN2 revealed constitutive expression, even after inoculation with the fungus *Phakopsora pachyrhizi*. The gene has three exons and 40 homologues distributed along 16 of 20 soybean chromosomes and in six chromosomes of *P. vulgaris* and *M. truncatula*, with a distribution pattern different from other defense genes, presenting a heterogenous distribution. Here we provide an important source to map this family of AMPs and also to infer about their function.

Keywords: Antimicrobial peptides. Bioinformatic. GENOSOJA. NordEST. Data mining.

LISTA DE FIGURAS

Figura 1- Hipótese do Receptor-Ligante *vs. guard hypothesis*. **(a)** Hipótese do Receptor-Ligante: proteínas de resistência (R) detectam o patógeno via interação direta com as proteínas de avirulência (avr), desencadeando a sinalização de defesa; **(b)** *Guard hypothesis:* os *guardees* são os alvos das proteínas avr, assim, as proteínas *guard* e *guardee* interagem dinamicamente podendo haver outras proteínas no complexo; **(c)** Situação onde ocorre a modificação das proteínas *guardee* (pelas proteínas avr), alterando sua interação e desencadeando os sinais de defesa (Soosaar et al. 2005)..... 19

Figura 2 - Representação de um alinhamento progressivo. Os alinhamentos construídos par a par passam por etapas qualitativas e de refinamento até o alinhamento final (Pirovani and Heringa 2008)..... 27

Figura 3 - Quantidade de estruturas depositadas no PDB (azul claro) em comparação com o número de sequências manualmente anotadas (Swiss-Prot – Azul escuro) e automaticamente anotadas (TREMBL - Amarelo) até Junho de 2018..... 30

Figura 4 - Os princípios de classificação fenética e filogenética podem convergir (a) ou divergir (b e c) entre si (Ridley 2004)..... 31

Figure 1 - Multiple sequence alignment of snakin/GASA peptides. The triangles show de cleavage site for the signal peptide in *Solanum tuberosum* Snakin-1 (St-Snakin-1), *Arabidopsis thaliana* GASA1 (At_GASA1), *Solanum tuberosum* Snakin-2 (St_Snakin-2) and *Capsicum annuum* Snakin (Ca_CaSN), which is in the same position for At_Snakin-1, Ca_Casn and St_Snakin-2. The vertical bar indicates the limit between the variable region (VR) and the start of the GASA domain that exhibits twelve conserved cysteines and other nine conserved residues (dark gray in the alignment). Bars below the alignment indicate the levels of conservation, quality and, at the bottom, the consensus sequence……. 36

Figure 1 - Hierarchical clustering of transcriptional profile in soybean performed using the CLUSTER software. The red color indicates higher prevalence; the dark red indicates intermediate prevalence and black means absence of snakins in the respective tissue. Libraries: B01 (vegetable buds; field-plants); C0Y (young cotyledons); C06 (wounded

cotyledons); C07 (degenerating cotyledons; 10 days old, etiolated seedlings); EN1 (endosperm - developing seeds); EP2 (seedling epicotyl); F01 (floral meristem); F0M (mature flowers); F04 (floral meristem mRNA); F05 (immature flowers; field-plants); H03 (hypocotyl/plumule; germinating seeds); H04 (*P. sojae*; infected hypocotyl); H0E (etiolated hypocotyls); L01 (senescing leaf; greenhouse-plants); L03 (fully expanded leaves; greenhouse-plants); L04 (immature leaves; greenhouse-plants); L0M (mature leaves); L06 (leave; drought stressed); R0R (roots); R03 (seedling roots); S02 (mature seed pods; greenhouse-plants); S04 (young seeds); S0C (seed coats); S0S (seedlings); S12 (seeds; globular stage embryos); SHS (shoots); SO1 (somatic embryos; cultured on MSD 20); S0T (stem) and UK1 (unknown tissue)…….. 51

Figure 2 - Venn diagram showing the number of DeepSuperSAGE UR tags matching soybean snakin exclusive in each contrast. (**A**) TD *vs*. TC, roots of drought-tolerant accession (Embrapa 48); stressed *vs*. non-stressed control; (**B**) SD *vs*. SC, roots of drought-susceptible accession (BR16); stressed *vs*. non-stressed control and (**C**) RR *vs*. RC, leaves of rust-resistant accession (PI561356); inoculated *vs*. non-inoculated - Mock…….. 52

Figure 3 - *Gm*SN2 gene sequence and the putative amino acid sequence. VR: Variable region; GASA: GASA domain.……..……. 53

Figure 4 - Multiple sequence alignment performed with peptide sequences of Snakin/GASA family members, the *Gm*SN2 and the other soybean peptides (GmSN-like) identified showing the main conserved residues represented by the 12 cysteine motif (blue) and other eight residues, Arginine (R) – yellow; Lysine (K): pink; Valine (V): light green; Proline (P): red; Glycine (G): green and Tyrosine (Y): grey, that compose the general motif "CX3CX2RCX8CX3CX2CCX2CXCVPXGX2GX4CXCY X10/13KCP"..... 54

Figure 5 - Phenetic analysis of Snakin/GASA proteins including *Gm*SN2 and the GmSN-like peptides. The branches are supported by the bootstrap values showed in percent (%). The three subfamilies are evidenced by the bracket, at the right side. The scale length (0.1) indicates distance..... 56

Figure 6 - Genomic distribution of *Gm*SN2 homologs in soybean (GM), *M. truncatula,* (Mt) and *P. vulgaris* (Pv). The orange circles, blue triangles and green squares indicate the soybean transcripts mapped and their gene structures with two, three

and four exons respectively.…… 57

Figure 7 - GmSN2 structure showing the disulfide bonds (in yellow) (A). (B) Ramachandran plot where the blue dots are the aminoacids, the red regions are the most favored regions, in brown the additional allowed regions and in yellow the generously allowed regions and (C) Quality of folding showed by the Z-Score where the black spot is the GmSN2 model.…… 59

Supplementary Fig.1- Taxonomic distribution of the representative genes from Snakin/GASA family found by SeedServer according to their plant taxonomy families. The y-axis represents the number of genes and the x-axis represents the plant family……..... 73

Supplementary Fig. 2 - geNorm analysis, indicating the average expression stability (M value) (A) and pairwise variation (B), calculated by pair-wise variation analysis between normalization factors (NF_n and NF_{n+1}) ……. 74

Supplementary Fig. 3 - Expression profile of GmSN2 gene from soybean plants (PI561356 accession) in response to *Phakopsora pachyrhizi*, 48 hai (hours after inoculation) in comparison with those in the mock-inoculated plants (control). Soybean genes encoding CYP2 (cytochromes P4502 gene family, 26P (26S proteasome) and ZMT (Zinc metallopeptidase) were used as reference genes. The error bars represent the set of three independent biological replicates and two technical repeats……..... 75

LISTA DE TABELAS

Tabela 1 - Classificação das proteínas relacionadas à patogênese (proteínas PR)..... 20

Tabela 2 - Bancos de dados e ferramentas importantes no estudo das interações proteína-proteína..... 25

Table 1 - Snakin/GASA peptides previously described in plants…….. 37

Table 1 - Soybean clusters identified through tBLASTn tool in the GENOSOJA database. A search conducted with seed sequences (*e-value* ≤ e^{-4}) reported by the SeedServer (only those with complete GASA-domain) and annotated via BLASTx tool (GenBank).. ……. 50

Table S1. Primer sequences of the seven candidate reference genes and of the gene of interest (*Gm*SN2) used in the RT- qPCR analysis……..……. 72

Table S2. Soybean transcripts anchored in the virtual chromosomes……. 76

Table S3. Qualitative aspects of the soybean homolog structures, the Z-Score and the Ramachandran plot in % comparing with snakin-1 model (PDB ID 5e5q)..... 77

LISTA DE ABREVIAÇÕES

ABA - Ácido Abscísico (*Abscisic Acid*)

AMP - Peptídeos antimicrobianos (*Antimicrobial Peptide*)

Avr - Gene de Avirulência

BLAST - Ferramenta Básica de Busca e Alinhamento Local (*Basic Local Alignment Search Tool*)

CaMV - Vírus do Mosaico da Couve-Flor (*Cauliflower mosaic virus*)

CaSN - *Capsicum annuum snakin*

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária

EST - Etiquetas de Sequencias Expressas (*Expressed Sequence Tags*)

FaGASA - *Fragaria ananassa* GASA

FBCBP - *French bean chitin-binding protein*

FsGASA - *Fagus sylvatica* GASA

GA - Ácido Giberélico (*Gibberellic acid*)

GASA - *Gibberellin acid-stimulated arabidopsis*

GAST - *Gibberellic acid-stimulated transcript*

GEG - *Gerbera homolog of GAST1* gene

Gip - *GA-induced proteins*

GmSN - *Glycine max Snakin*

GsGASA - *Glycine soja* GASA

GSL - Proteína estimulada por giberelina (*Gibberellin Stimulated Like*)

HR - Resposta Hipersensível (*Hypersensitive Response*)

LTP - Proteínas transferidoras de lipídeos (*Lipid Transfer Proteins*)

LRR - Repetições ricas em leucina (*Leucine rich repeat*)

NBS - Sítio de ligação de nucleotídeos (*Nucleotide binding site*)

NCBI - Centro Nacional de Informações sobre Biotecnologia (*National Center for Biotechnology Information*)

OsGASR - *Oryza sativa GA-stimulated transcript-related gene*

PDB - Banco de Dados de proteínas (*Protein Data Bank*)

PDC - Morte celular programada (Programmed cell death)

PR - Proteínas Relacionadas à Patogênese (*Pathogenesis Related Protein*)

R - Genes de Resistência

ROS - Espécies Reativas de Oxigênio (*Reactive Oxygen Species*)

RSI-1 - *Root system inducible*-1

AS - Ácido Salicílico (*Salicilic acid*)

SAGE - Análise Serial de Expressão Gênica (*Serial Analysis of Gene Expression*)

SAR - Resposta Sistêmica Adquirida (*Systemic Acquired Resistance*)

SNK - Snakin

StSN - *Solanum tuberosum* snakin

ZmGSL - *Zea mays gibberellin-stimulated like*

SUMÁRIO

1	**INTRODUÇÃO ……..**	**13**
2	**OBJETIVOS ……...**	**15**
2.1	**Objetivo Geral ……**	**15**
2.2	**Objetivos Específicos …….**	**15**
3	**REVISÃO BIBLIOGRÁFICA ……..**	**16**
3.1	**Estresses ambientais e mecanismos de defesa das plantas .....**	**16**
3.1.1	Resposta Hipersensível (HR) e Resistência Sistêmica Adquirida (SAR)	**16**
3.1.2	Genes R e PR	**17**
3.1.3	Peptídeos antimicrobianos de plantas ricos em cisteínas	**21**
3.1.4	Esnaquinas (*Snakin*) ……	**21**
3.2	**Bioinformática: do armazenamento ao processamento de dados biológicos .....**	**23**
3.2.1	Bancos de dados biológicos ……	**24**
3.2.2	Recuperação de sequências e Anotação	**25**
3.2.3	Alinhamento Múltiplo ……	**26**
3.2.4	Domínios e motivos conservados …….....	**27**
3.2.5	Predição *in silico* de estruturas proteicas	**28**
3.2.6	Análises Filogenéticas e Fenéticas ……..	**30**
4	**Snakin: structure, roles and applications of a plant antimicrobial peptide ……**	**35**
5	**Structural and functional analysis of snakin/GASA in soybean (*Glycine max*) and description of a new member (*Gm*SN2) ……..... **	**42**
6	**CONCLUSÕES GERAIS .....**	**78**
	REFERÊNCIAS	**79**
	ANEXOS	**89**

1. INTRODUÇÃO

Ao longo dos processos evolutivos os vegetais desenvolveram diversos mecanismos de defesa que variam desde barreiras físicas, como a rigidez do caule e a cobertura serosa das folhas que ajudam a impedir a predação por insetos, a invasão por agentes microbianos e a perda de água por transpiração. Contudo, em se tratando de agentes antimicrobianos, muitas vezes tais moléculas são capazes de ir além destas barreiras, incitando a liberação de agentes químicos de defesa como as proteínas PR (*Pathogenesis-Related*) que incluem os peptídeos antimicrobianos - AMPs (*Antimicrobial peptides*) ricos em cisteínas (AMP) (Silverstein et al. 2007; Benko-Iseppon et al., 2010).

Encontrados em todas as famílias de mono e eudicotiledôneas, AMPs são codificados por famílias multigênicas, incluindo defensinas, thioninas, ciclotídeos, heveínas, esnaquinas, entre outros. Possuem em comum a carga positiva e a natureza anfipática, características que estão relacionadas à sua atividade de interação com membranas biológicas, além da presença de 4-12 resíduos cisteína, cujo motivo resulta na formação de 2-6 pontes dissulfeto – contribuindo para a compactação da estrutura, além de conferir estabilidade e resistência à degradação química e proteolítica. A expressão destes peptídeos pode ser constitutiva ou induzida por patógenos e predadores de diferentes naturezas, como insetos e nematoides, agindo como inibidores de proteases ou desestabilizadores de membranas (Tam et al. 2015).

Neste sentido, as esnaquinas – classificadas como membros da família snakin/GASA (Segura et al. 1999; Berrocal-lobo et al. 2002) – vêm recebendo crescente destaque por sua diversidade funcional, com ação contra fungos, bactérias e nematoides, além de atuarem como efetores do crescimento e de divisão celular. Tais peptídeos são expressos em vários tecidos (desde as raízes até as flores). Sua estrutura primária é composta por três motivos principais: (1) um peptídeo sinal na região N-Terminal, que pode conter de 18-29 resíduos, (2) uma região variável que possui entre 7-31 resíduos e o domínio C-Terminal GASA, composto por 60 aminoácidos, incluindo as 12 cisteínas. A estrutura terciária é formada por um domínio hélice-volta-hélice, uma hélice 3_{10} e uma terceira α-hélice. Usualmente, estes peptídeos são subcelularmente localizados na parede celular, mas podem também ser direcionados para o apoplasto ou para o núcleo (quando traduzidos sem o peptídeo sinal). Assim, devido à diversidade funcional e às suas características estruturais, as esnaquinas destacam-se como alvos biotecnológicos interessantes para uso tanto na transformação de plantas, como no desenvolvimento de novos fármacos.

Outra pequena família de AMPs ricos em cisteínas que apresenta interesse biotecnológico é a família das heveínas (*hevein-like*), tendo em vista sua capacidade de se ligar à quitina (componente presente em nematoides, na parede celular de fungos e no exoesqueleto de insetos). Além disso, estes peptídeos também demonstraram atividade contra bactérias gram-positivas e gram-negativas. Estruturalmente, heveínas são peptídeos básicos, pequenos (variando de 29-45 aminoácidos), ricos em cisteínas, glicinas, bem como alguns aminoácidos aromáticos, cujo domínio conservado (*hevein*, ligante de quitina) pode variar quanto à quantidade de cisteínas (de 6-10). A sua estrutura terciária é composta pelo motivo estrutural α- hélice–β1–β2–α-hélice–β3, onde as folhas-beta são antiparalelas. Estes peptídeos têm sido encontrados preferencialmente nos fluidos intracelulares de vegetais superiores, sendo explorado com sucesso na transformação de plantas, resultando em variedades resistentes a diversos fitopatógenos (Slavokhotova et al. 2017).

Tendo em vista o grande potencial destes peptídeos em diversas áreas biotecnológicas, um número cada vez maior de sequências tem sido depositado em bancos de dados biológicos. Entretanto, a identificação e caracterização destes peptídeos ainda é um grande desafio, em vista da carência de protocolos que recuperem, de forma rápida e apurada, todos os peptídeos antimicrobianos de interesse. Nesse contexto, a bioinformática tem sido indispensável na prospecção de alvos, tendo em vista o desenvolvimento de softwares e ferramentas direcionados a cada tipo de questão biológica, proporcionando as "pistas" necessárias para a montagem desse "quebra-cabeças". Contudo, frequentemente é preciso complementar os dados compilados e tratados com as metodologias de *wet lab* para que tenham uma maior robustez e confiabilidade.

O presente trabalho teve como objetivo identificar e caracterizar esnaquinas e heveínas em plantas de interesse econômico e selecionar os candidatos de maior potencial biotecnológico para uso no melhoramento genético vegetal e/ou uso farmacológico.

2. OBJETIVOS

2.1. Objetivo Geral:

- Identificar e caracterizar genes codificadores de peptídeos antimicrobianos de plantas, em especial esnaquinas, com dados gerados a partir de transcriptomas e genomas disponíveis em bancos de dados públicos e de acesso restrito, com ênfase em elementos desconhecidos do ponto de vista funcional, visando ao entendimento dos processos de defesa vegetal e à futura exploração do seu potencial biotecnológico.

2.2. Objetivos Específicos:

- Identificar, quantificar e caracterizar domínios conservados de genes codificantes para esnaquinas nos bancos de dados públicos, comparativamente àqueles gerados por nossos projetos para soja e de feijão-caupi (GENOSOJA e NordEST), inferindo sua diversidade;
- Elucidar aspectos estruturais e funcionais de natureza gênica e proteica desses peptídeos com base em vias metabólicas, modelagem molecular e identificação de possíveis interações;
- Identificar regiões filogeneticamente informativas, inferindo sobre a evolução destes grupos;
- Estabelecer o perfil de expressão dos genes candidatos (*in silico* e experimentalmente), validando seus níveis de expressão em situações contrastantes (controle e sob estresse biótico e/ou abiótico, quando presentes), inferindo sobre sua função espacial e temporal;

3. REVISÃO BIBLIOGRÁFICA

3.1. Estresses ambientais e mecanismos de defesa das plantas

As plantas são constantemente submetidas a mudanças ambientais, as quais muitas vezes são estressantes. Estas condições incluem estresses bióticos, tais como infecção por patógenos e ataque de herbívoros, além dos estresses abióticos, como seca, temperaturas extremas, solos pobres em nutrientes, excesso de sais ou metais pesados no solo, entre outros. Como forma de defesa, os vegetais respondem através de mecanismos físicos e químicos, a fim de mitigar os danos e conservar recursos necessários ao seu crescimento e à reprodução (Rizhsky et al. 2004; Atkinson and Urwin 2012; Dorantes-Acosta et al. 2012).

Entre as barreiras desenvolvidas que servem tanto para a detecção quanto para a proteção contra estresses ambientais, destaca-se a cutícula, uma camada composta de cutina (um polímero de ácidos graxos) que ajuda tanto na defesa contra a penetração de microrganismos, quanto na minimização da perda de água (Serrano et al. 2014). Outro exemplo é a deposição de lignina na parede celular, uma barreira química que possui efeitos antimicrobianos (Vanetten et al. 1994). Apesar de eficientes, muitas vezes estas barreiras primárias não são suficientes para a defesa da planta e podem ser superadas, resultando em danos aos vegetais (Sels et al. 2008).

No caso de estresses bióticos, quando os agentes microbianos ultrapassam a primeira barreira (física), a planta aciona uma cascata de sinalização ativando moléculas de defesa (ex. metabólitos secundários e/ou proteínas), ou seja, uma forma de "contra-ataque" mais eficiente. Nesse mecanismo, hormônios – como o ácido salicílico, o ácido jasmônico e o etileno – agem como sinalizadores (Robert-Seilaniantz et al. 2011; Dorantes-Acosta et al. 2012), ao passo que as proteínas relacionadas à patogênese (PR, *Pathogenesis-Related*) atuam diretamente sobre o agente invasor (Saboki et al. 2011). Em conjunto, esses dois mecanismos constituem uma resposta de defesa mais complexa, eficiente e segura (Fu and Dong 2013; Pandey et al. 2016).

3.1.1 Resposta Hipersensível (HR) e Resistência Sistêmica Adquirida (SAR)

Quando um patógeno ou um herbívoro ataca a planta, superando suas barreiras primárias, esta responde ativando sinais nos locais de infecção, resultando em uma resposta

de hipersensibilidade (*Hypersensitive response*, HR), o que leva à morte celular programada (*Programmed cell death* - PCD) no entorno da região em contato com o patógeno (Dixon and Harrison, 1994; Morel and Dangl, 1997). O reconhecimento de padrões moleculares e/ou de efetores dos fitopatógenos (fungos, vírus e bactérias) pode levar à sinalização celular através da produção de espécies reativas de oxigênio (*Reactive Oxygen Species* - ROS) de uma série de hormônios vegetais (ex. giberelina e ácido salicílico), além de proteínas PR relacionadas à defesa contra patógenos, as quais fazem parte da segunda via das respostas químicas, levando a um tipo de resistência em plantas conhecida como Resistência Sistêmica Adquirida (*System Acquired Resistance* – SAR) (Liu and Ekramoddoullah 2006; van Verk et al. 2009; Zurbriggen et al. 2010; Kachroo and Robin 2013).

A HR pode ser desencadeada em resposta a vários patógenos (ocorrendo pouco tempo após a infecção), sendo condicionada pela interação dos produtos dos genes de resistência (*R*) do hospedeiro e de avirulência (*avr*) do patógeno (Morel and Dangl 1997; Mur et al. 2008; Wanderley-Nogueira et al. 2017). Os mecanismos de defesa são conservados no que diz respeito às modificações bioquímicas e metabólicas e, caso a infecção persista, vários genes de defesa serão induzidos, aumentando o nível de resistência e desencadeando a SAR (Dangl 1995; Bent 1996; Morel and Dangl 1997; Durrant and Dong 2004).

Inicialmente, a SAR envolve a geração de moléculas sinalizadoras nos tecidos infectados, com a finalidade de prover a percepção da invasão do patógeno pelos outros tecidos (Fu and Dong 2013; Kachroo and Robin 2013). Trata-se da forma de defesa mais duradoura (podendo chegar a alguns meses) e de largo espectro em plantas. Do ponto de vista molecular, ocorre aumento da expressão de uma vasta quantidade de genes *PR*, tanto no local da lesão, como em outros tecidos (Durrant and Dong 2004), um exemplo são as vias de sinalização por etileno que são ativadas pela liberação desse hormônio no local de infecção, fazendo com que ocorra um *feedback* positivo das proteínas PR (Kitajima and Sato 1999).

3.1.2. Genes R e PR

Dentre os mecanismos mais bem caracterizados de defesa de plantas e, portanto, amplamente utilizados em processos de transformação genética são aqueles mediados pelos genes *R* (Elvira et al. 2008; Pandolfi et al. 2017; Wanderley-Nogueira et al. 2017). Estes se apresentam com um alto grau de polimorfismo, um aparato de reconhecimento específico e

fazem parte de famílias gênicas que tendem a formar grupamentos (*clusters*) no genoma, os quais evoluíram por duplicação, recombinação e neofuncionalização (Ronald 1998; Kang et al. 2012). Só na cana-de-açúcar (*Saccharum officinarum*) e no eucalipto (*Eucalyptus* sp.), por exemplo, foram encontrados 280 e 210 genes *R*, respectivamente, com sítios de ligação de nucleotídeos (*Nucleotide Binding Site* - NBS) e repetições ricas em resíduos de leucina centrais (*Leucine Rich Repeat*, LRR) (Barbosa-da-Silva et al. 2005; Wanderley-Nogueira et al. 2007).

A resistência conferida pelos genes *R* parece seguir o padrão de interação *gene-a-gene* relatado por Flor (1971) e completado por Bourras et al. (2016), onde está relatado que a resistência seria mediada por interações envolvendo os genes *avr/svr* no patógeno e os genes *R* no hospedeiro. O mecanismo de resistência seria ativado quando o supressor (*svr*) estivesse inativo e os genes avr fossem produzidos em quantidade suficiente para serem detectados pelos genes *R*. Caso tais condições não ocorram, a planta pode sucumbir à infecção, o que geraria outra cascata de sinalização hormonal, dando início à SAR (Cameron et al. 1994; Vlot et al. 2008; Bourras et al. 2016).

A "*guard hypothesis"* proposta por Van Der Biezen e Jones (1998) por sua vez, sugere que as proteínas R (*guard*) estão constitutivamente associadas a proteínas celulares no hospedeiro (*guardee*), as quais são "exigidas" pelo patógeno para a infecção. Nesta situação, o patógeno causa modificações nas proteínas *guardee*, as quais são detectadas pela *guard.* Assim, qualquer modificação na estrutura quaternária das *guardee* pode resultar na detecção do patógeno e, portanto, induzir uma cascata de sinalização, resultando na resposta de resistência (Van Der Biezen and Jones 1998; Holt et al. 2003; Soosaar et al. 2005) (**Figura 1**).

Em ambas as hipóteses, as etapas subsequentes seguem um padrão, onde moléculas sinalizadoras (quinases, espécies reativas de oxigênio e alguns hormônios, por exemplo) atuam como indutores da expressão de proteínas PR (Soosaar et al. 2005; Sudisha et al. 2012).

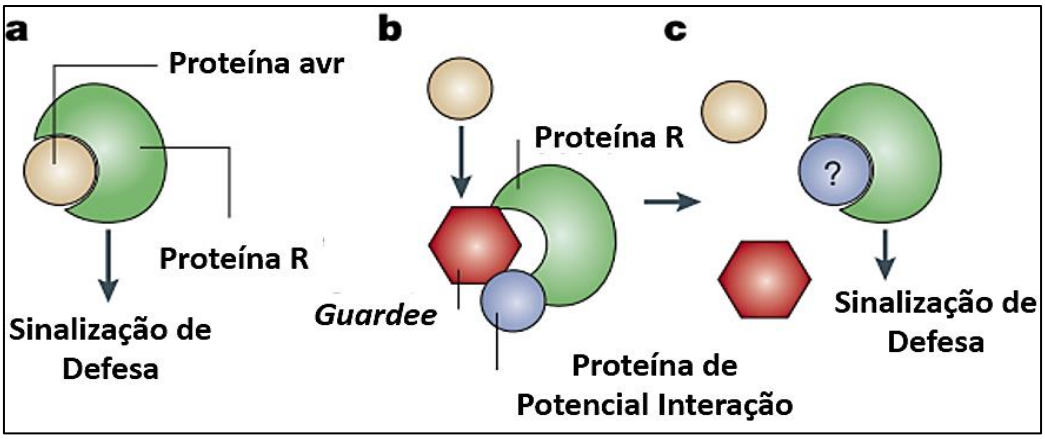


Figura 1. Hipótese do Receptor-Ligante *vs. guard hypothesis*. **(a)** Hipótese do Receptor-Ligante: proteínas de resistência (R) detectam o patógeno via interação direta com as proteínas de avirulência (avr), desencadeando a sinalização de defesa; **(b)** *Guard hypothesis:* os *guardees* são os alvos das proteínas avr, assim, as proteínas *guard* e *guardee* interagem dinamicamente podendo haver outras proteínas no complexo; **(c)** Situação onde ocorre a modificação das proteínas *guardee* (pelas proteínas avr), alterando sua interação e desencadeando os sinais de defesa (Soosaar et al. 2005).

Com base na sua estrutura e propriedades funcionais, as proteínas PR foram classificadas em 17 famílias (**Tabela 1**), tendo em vista suas homologias, similaridade, estrutura primária ou terciária e atividade biológica (Christensen et al. 2002; Benko-Iseppon et al. 2010; Sudisha et al. 2012). Outro parâmetro considerado nesta classificação foi o ponto isoelétrico (pI), prevalecendo a carga positiva para a maioria das PRs (van Loon et al. 1994; Christensen et al. 2002; Sudisha et al. 2012), uma vez que muitas delas tem a propriedade de aderir a membranas que possuem carga negativa (Zasloff 2002; Stec 2006). Dentre essas famílias, destacam-se as PRs 12, 13 e 14 (**Tabela 1**), que incluem as esnaquinas e as heveínas, bem como outros AMPs ricos em resíduos de cisteínas (Cys) (Silverstein et al. 2007).

Tabela 1. Classificação das proteínas relacionadas à Patogênese (proteínas PR)*.

Família	Membro tipo	Propriedades
PR1	PR-1a de tabaco	Antifúngica
PR2	PR-2 de tabaco	β-1,3-glucanase
PR3	Proteína P, Q de tabaco	Quitinase tipo: I, II, IV, V, VI e VII
PR4	Proteína R de tabaco	Quitinase tipo: I e II
PR5	Proteína S de tabaco	Taumatina, Antifúngica
PR6	Inibidor I de tomate	Inibidor de protease
PR7	P69 de tomate	Endoproteinase
PR8	Quitinase de pepino	Quitinase tipo III
PR9	Lignina peroxidase de tabaco	Peroxidase
PR10	*Petroselinum crispum* PR1	Ribonuclease
PR11	Quitinase classe V de tabaco	Quitinase tipo I
PR12	Rs-AFP3 de rabanete	Defensina
PR13	THI2.1 de Arabidopsis	Tionina
PR14	LTP4 de cevada	Proteína transferidora de lipídeos
PR15	OxOa de cevada	Oxalato oxidase
PR16	OxOLP de cevada	Oxalato oxidase-like
PR17	PRp27 de tabaco	Antifúngica e antiviral

*Fonte: Christensen et al. 2002; Sudisha et al. 2012.

Geralmente, as proteínas PR em plantas são expressas em decorrência do ataque de patógenos. No entanto, estes eventos não se referem apenas àqueles desencadeados pela HR ou SAR (onde as PRs são mais comumente observadas), mas em quase todos os tipos de situações de estresse causadas por patógenos, incluindo parasitismo e herbivoria (como nematoides e insetos, respectivamente). Em adição, proteínas PR podem estar presentes em situações de estresse abiótico, tanto em tecidos com sinais de clorose / necrose, como em tecidos intactos (van Loon 1999; Tuzun 2001; Sudisha et al. 2012). Embora frequentemente observada em tecidos foliares (van Loon et al. 2006), a expressão de PRs também tem sido relatada em outros tecidos/ órgãos, como tubérculos, flores, raízes e sementes (Belarmino et al. 2010; Sudisha et al. 2012) em uma grande variedades de espécies vegetais. Além disso, uma única espécie pode conter representantes de diferentes classes, como é o caso de *Arabidopsis thaliana*, *Medicago truncatula* e *Glycine max*, por exemplo, onde foram

identificados, respectivamente, 366, 585 e 310 transcritos de PRs das classes 1-9, 11, 12 e 14-16 (Wanderley-nogueira et al. 2012).

3.1.3 Peptídeos antimicrobianos de plantas ricos em cisteínas

De uma forma geral, peptídeos antimicrobianos são ubíquos e participam da defesa das plantas contra uma grande variedade de patógenos, incluindo microrganismos, fungos e animais (Egorov et al. 2005; Benko-Iseppon et al. 2010). Estruturalmente podem ser lineares (maioria) ou cíclicos (Tam et al. 2015; Lima et al. 2017), embora, em plantas se observe uma predominância da forma rica em resíduos de cisteínas (Cys; C) – fundamental para a formação das pontes de dissulfeto auxiliando na compactação da estrutura e servindo, portanto, de proteção contra eventual degradação proteolítica (Hammami et al. 2009b; Tam et al. 2015).

Em termos gerais, os AMPs de plantas compartilham algumas características com homólogos de organismos evolutivamente distantes (ex. mamíferos e insetos) como a carga positiva e natureza anfipática – propriedades associadas à atividade desestabilizadora da membrana dos microrganismos (Zasloff 2002; Padovan et al. 2010). Contudo, a característica que melhor separa as famílias de AMPs de plantas são os seus motivos ricos em cisteínas definidos, principalmente em relação aos demais resíduos de aminoácidos.

Baseado nessas propriedades, os AMPs de plantas foram classificados em diferentes famílias, tais como: defensinas, tioninas, LTPs, knotinas, heveínas e esnaquinas, entre outras, cuja quantidade de cisteínas pode variar de 4 a 12 resíduos (Manners 2007; Silverstein et al. 2007; Benko-Iseppon et al. 2010). Outras propriedades dos AMPs de plantas incluem: (i) tamanho pequeno (peso molecular de 2-6 kDa), básicos e com 2 - 6 pontes dissulfeto; (ii) similaridade de sequência e conservação estrutural secundária e terciária; (iii) múltiplas funções; (iv) todos são bioprocessados e, geralmente, possuem três domínios (peptídeo sinal, região variável além do peptídeo maduro) e (v) exibem alta estabilidade térmica, química e enzimática (Tam et al. 2015).

3.1.4. Esnaquinas (*Snakin*)

As esnaquina "*snakin'* (nomenclatura dada em função da similaridade de três aminoácidos com a "*desitegrin-like*" - uma peçonha da serpente *Agkistrodon rhodostoma*, embora também conhecidos por GSL - *gibberellin stimulated-like proteins*) são peptídeos

ricos em cisteínas, identificados pela primeira vez na batata (*Solanum tuberosum*) (Adler et al. 1991; Segura et al. 1999; Berrocal-lobo et al. 2002; Meiyalaghan et al. 2014).

Devido à sua similaridade com membros da família GASA (*Gibberellic Acid Stimulated,* de *Arabidopsis*), foram classificadas como membros integrantes da família "*snakin*/GASA" (Berrocal-lobo et al. 2002). Os genes que codificam esses peptídeos apresentam: (i) uma sequência sinal, de até 28 aminoácidos, servindo para seu direcionamento subcelular; (ii) uma região variável, e (iii) um peptídeo maduro, de ~60 aminoácidos, sendo 12 resíduos de cisteína (Cys) altamente conservados. Tais resíduos auxiliam na manutenção da estrutura tridimensional do peptídeo, através da formação de pontes dissulfeto, além de fornecer estabilidade à molécula – fato este de fundamental quando a planta está sob estresse (Segura et al. 1999; Berrocal-lobo et al. 2002; Mao et al. 2011; Yeung et al. 2016). A biossíntese desses peptídeos, embora ainda não tenha sido completamente eluncidada, relatos sugerem que, após sua síntese no núcleo, estes peptídeos são processados e transportados (atraves do retículo endoplasmático) para o ambiente extracelular, compondo a parede celular (Nahirñak et al. 2012).

São encontradas em diferentes tecidos vegetais: raíz, caule, folhas, flores e sementes (Zhang et al. 2009; Zimmermann et al. 2010; Almasia et al. 2010; Guzmán-rodríguez et al. 2013), com os mais variados níveis de expressão, desde constitutivo ao induzido por diferentes tipos de estresses (bióticos e/ou abióticos), especialmente agindo na desestabilização da membrana plasmática de fitopatógenos como fungos, bactérias e nematoides(Segura et al. 1999; Faccio et al. 2011; Mao et al. 2011; Herbel and Wink 2016). A expressão de homólogos de esnaquinas também tem sido relatada em eventos de divisão celular, elongação, crescimento, floração, embriogênese, além de várias vias de sinalização (Kotilainen et al. 1999; Furukawa et al. 2006; Roxrud et al. 2007; Lucau-Danila et al. 2010). Curiosamente, a atividade de um homólogo da família (*snakin*-Z - *Ziziphus jujuba* Mill.) foi associada ao tratamento de Alzheimer em humanos, tendo em vista sua ação antioxidante, aliada a uma significativa atividade como inibidor de colinesterase, fatores estes determinantes no tratamento dessa doença, embora tal propriedade ainda esteja em fase de estudo (Zare-zardini et al. 2013).

3.2 Bioinformática – Do armazenamento ao processamento de dados biológicos

Os recentes avanços nas tecnologias de sequenciamento têm favorecido, em grande escala, a disponibilidade de informações genômicas de interesse, principalmente na geração de marcadores moleculares, bem como na identificação de genes de interesse agronômico e para fins biotecnologicos. Assim, para que houvesse uma maior eficiência na caracterização dos genomas, fez-se necessário organizar a informação de forma "navegável" e desenvolver ferramentas que pudessem processá-la de forma eficiente. Neste sentido, bancos de dados inteiros vêm sendo dedicados ao armazenamento e tratamento deste tipo de informação (Edwards and Batley 2010; Batley 2015). Esses conjuntos de dados (*datasets*) representam uma fonte valiosa de informações derivadas das diferentes "ômicas" que têm proporcionando um melhor conhecimento/entendimento da genômica funcional às relações filogenéticas entre genes, proteínas e organismos (Soares-Cavalcanti et al. 2012; Porto et al. 2017).

A bioinformática, por sua vez, tem possibilitado a junção e interpretação do conhecimento biológico (Mochida and Shinozaki 2010), uma vez que todas as análises nessa área, necessariamente remetem à etapas *in silico*. Neste contexto, a bioinformática tem contribuido muito, principalmente pela constante evolução no desenvolvimento de ferramentas cada vez mais eficientes e aplicadas na área das ômicas.Dentre as principais contribuições da bioinformática no campo biológico destacam-se: (i) a obtenção da informação em bancos de dados biológicos (dedicados ou não a um objeto de estudo específico, organismo ou família gênica), (ii) alinhamento comparativo de sequências (seja ele local ou global), (iii) a análise funcional de domínios conservados e/ou motivos de sequências, (iv) a análise estrutural (gênica ou proteica), (v) a análise de elementos regulatórios, (vi) a identificação de polimorfismo de nucleotídeo único (*Single Nucleotide Polymorphism - SNP*) e/ou inserções e deleções (*Indels*), (vii) as análises fenéticas ou filogenéticas, (viii) a simulação de dinâmica molecular e (ix) a análise de transcritos, dentre outros (Rahman et al. 2016).

Assim, integrar essas informações e associá-las a um contexto biológico tem sido o grande desafio, uma vez que as interações entre os organismos e o meio em que estão inseridos são extremamente complexas (Sheth and Thaker 2014).

3.2.1. Bancos de dados biológicos

Bancos de dados biológicos (*databank* – DB) são coleções organizadas de dados de natureza diversa, que podem ser recuperados utilizando diferentes entradas (*inputs*). O gerenciamento dessa informação é feito através de vários *softwares* e *hardwares*, cuja recuperação e organização das informações são realizadas de forma mais rápida e eficiente quanto possível (Eltabakh et al. 2006). Em se tratando de dados biológicos, as informações podem ser classificadas em: (i) primária (sequências), (ii) secundária (de estrutura, de expressão, de vias metabólicas, tipos de drogas, etc..) e (iii) especializada (contendo informações sobre uma espécie ou classe de proteína) (Rahman et al. 2016).

Tendo em vista que a geração de dados biológicos tem ocorrido em uma escala exponencial, os repositórios dessas informações precisam acompanhar tal processo. Comparativamente, o desenvolvimento do primeiro banco de dados dedicado a sequências de proteínas, na década de 60, por Margaret Dayhoff (Strasser 2010), contava com apenas algumas dezenas de sequências. Nos últimos anos, somente no UniProt (*Universal Protein Resource*) foram depositados dezenas de milhões de sequências (Bairoch et al. 2004; Wasmuth and Lima 2016). Após uma busca simples por palavras-chave no NCBI (*National Center for Biotechnology Information*), por exemplo, é possível recuperar informações acerca de publicações (PubMed), sequências de proteínas e nucleotídeos, genomas completos, informações taxonômicas, entre muitas outras opções. Adicionalmente, uma ferramenta simples, com o uso de filtros, permite refinar a busca, auxiliando na identificação de novos genes e proteínas, bem como na elucidação de funções biológicas (Rahman et al. 2016; Porto et al. 2017).

Todas estas funcionalidades juntas possibilitaram a construção de outros tipos de bancos que tratam as informações de forma mais específica e dirigida. Neste sentido a construção de bases de dados destinados à AMPs foi um passo importante para organizar e disponibilizar dados provenientes deste tipo de peptídeo.

Ao longo da última década, várias bases de dados foram geradas com intuito de comportar a grande diversidade, deposição, consulta e mineração de AMPs (Tabela 2) (Zhou and Huang 2015). De forma geral, essas bases de dados são classificadas em dois grupos: gerais e específicas. Segundo Porto et al. (2017), as bases de dados específicas podem ser divididas em dois subgrupos - aquelas contendo apenas um grupo específico de AMPs (ex. defensinas ou ciclotídeos) e aquelas contendo dados de grandes grupos de peptídeos, incluindo os peptídeos vegetais, animais ou peptídeos cíclicos) (Tabela 2) (Mulvenna 2006;

Seebah et al. 2007; Hammami et al. 2009a). Ambos os bancos de dados compartilham de algumas características, como por exemplo, a forma como os dados estão disponíveis e/ou a forma de acesso às diferentes ferramentas de análise.

Tabela 2. Exemplos de bancos de dados de AMPs gerais e específicos

Banco de dados	**Tipo**	**Url**
CAMP	Geral	http://www.camp.bicnirrh.res.in/
APD	Geral	http://aps.unmc.edu/AP/main.php
LAMP	Geral	http://biotechlab.fudan.edu.cn/database/lamp
DBAASP	Geral	https://dbaasp.org/
Defensin Knowledgebase	Específico	http://defensins.bii.a-star.edu.sg/
Cybase	Específico	http://www.cybase.org.au/
Phytamp	Específico	http://phytamp.hammamilab.org/

Em linhas gerais, as bases de dados de AMPs podem reunir dados (peptídeos) advindos de diferentes fontes, de naturais a sintéticos, validados experimentalmente ou não, podendo possuir, também, sequências patenteadas (Wang 2004; Waghu et al. 2016). Com relação ao acesso às informações, essas podem ser obtidas por por palavras-chave ou através de alinhamentos locais. Algoritmos baseados em aprendizado de máquina que permitem a predição de AMPs, a partir de dados de sequências, tambem podem ser disponibilizados por alguns bancos de dados, a exemplo da *Collection of Antimicrobial Peptides* - CAMP (Waghu et al. 2016).

Consequentemente, os bancos de dadospassaram, ao longo dos anos, de repositórios de informação de sequências biológicas para servirem, tambem, como ferramentas na análise de dados em diversas instâncias, permitindo aos usuários, uma pesquisa bem direcionada e eficiente em termos de custo financeiro e/ou de tempo (Bry and Kruger 2003; Rahman et al. 2016).

3.2.2. Recuperação e anotação de sequências

O primeiro passo para a inferência da função de uma determinada sequência (anotação) é a sua recuperação nos bancos de dados. Para tal, duas metodologias apresentam grande utilidade: o alinhamento local (*Basic Local Alignment Search Tool* BLAST; Altschul

et al. 1990; FASTA – Pearson 1990) e a busca por padrões, a qual envolve duas estratégias principais: expressão regular (*Regular Expression* – REGEX) ou busca por padrões usando modelos de Markov (*Hidden Markov Model*- HMM) (Porto et al. 2017).

Alinhar sequências é a principal forma de compará-las, uma vez que a maior parte da informação disponível nos bancos encontra-se sob esta forma. O modo mais simples de conduzir esta abordagem é o alinhamento local (Polyanovsky et al. 2011), sendo o BLAST (*Basic Local Alignment Search Tool*) a ferramenta mais utilizada para este fim. O princípio desta ferramenta é dividir a sequência em tamanhos menores (*words*), comparando-os com o banco de dados de interesse. Tal abordagem, no entanto, apresenta uma limitação: pequenos motivos podem não ser encontrados, uma vez que muitos destes compreendem porções menores que 20 % do tamanho total para determinadas sequências (Tam et al. 2015; Porto et al. 2017). Visando diminuir os efeitos das limitações do alinhamento local, estratégias alternativas baseadas na busca por padrões são também utilizadas, como REGEX (Thompson 1968) ou HMM (Eddy 1998).

A ferramenta REGEX é uma forma precisa e sucinta de descrever um determinado padrão em uma sequência de caracteres (*string*) onde cada posição da REGEX deve ser fixada; embora caracteres ambíguos ou "curingas" também possam ser usados. Por exemplo, se quisermos encontrar uma sequência que tenha *match* com os fonemas da palavra "VALESCA", podemos utilizar a seguinte expressão: [VW]AL{1,2}ES[CK]A, ou seja, esta expressão encontraria uma palavra que iniciasse com "V" ou "W", seguida por um "A", um ou dois "L", "ES", "C" ou "K" e terminando com um "A". No caso dos perfis de HMM, há um perfil probabilístico inserido no modelo, o qual é calculado a partir de um alinhamento de sequências e uma pontuação é determinada sítio a sítio. Por não haver perfis estatísticos associados à REGEX, esta é menos restritiva que o HMM (Porto et al. 2017).

3.2.3. Alinhamento Múltiplo

O alinhamento múltiplo envolve a comparação entre três ou mais sequências, sejam elas de natureza nucleotídica ou peptídica. Estes alinhamentos podem também ser construídos comparando todas as sequências par a par, chamado de alinhamento pareado. Qualquer que seja a abordagem utilizada pelo algoritmo, a comparação é construída através da combinação e pontuação aplicada a *gaps*, *mismatches* e *matches*. Os programas costumam incluir ainda uma matriz de distância que funciona como medida qualitativa do conjunto de sequências a ser comparado e uma etapa de otimização, onde, dependendo da metodologia,

pode ser incluída a construção de uma árvore filogenética (árvore guia) que agrupa os entes mais similares e rodadas de iteração que refazem as comparações, assim possibilitando que os alinhamentos obtidos sejam aqueles com maior *score* (Figura 2) (Do and Katoh 2008; Pirovani and Heringa 2008; Pais et al. 2014).

Esta abordagem pode ser utilizada para se observar diferenças entre motivos proteicos, inserções e deleções em sequências nucleotídicas bem como auxiliar na construção de análises fenéticas e filogenéticas.

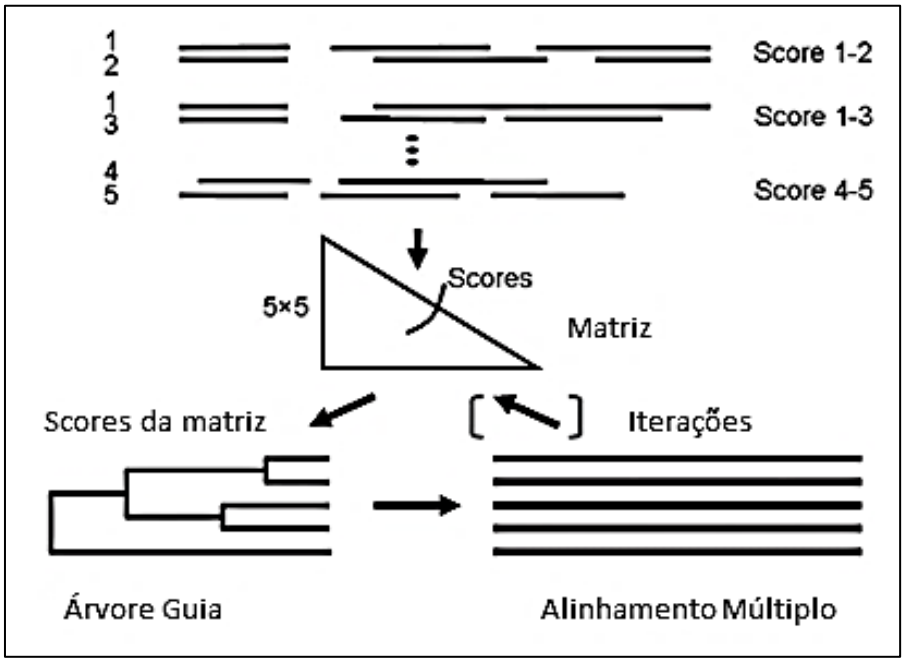


Figura 2. Representação de um alinhamento progressivo. Os alinhamentos construídos par a par passam por etapas qualitativas e de refinamento até o alinhamento final (Pirovani and Heringa 2008).

3.2.4. Domínios e motivos conservados

O "domínio conservado" (*conserved domain*) compreende uma região da estrutura proteica frequentemente (porém não sempre) composta por uma sequência contínua de aminoácidos e com uma função distinta do esqueleto proteico (Buljan and Bateman 2009). Já o termo "motivo" (*motif*) é utilizado em dois sentidos na biologia estrutural: primeiro, uma sequência de aminoácidos característica de uma função bioquímica específica. A sequência: "CXX(XX)CXXXXXXXXXXXXHXXXH", chamada de *zinc finger*, encontrada em muitas famílias proteicas com função de ligação ao DNA (Aitken 1999) é um

exemplo de motivo conservado; O segundo sentido envolve segmentos contínuos da estrutura secundária com significante importância funcional, por exemplo, o motivo hélice-volta-hélice presente em diversos peptídeos, como nas esnaquinas (Yeung et al. 2016), de proteínas com grampos (*hairpins*) (Nolde et al. 2011) e das tioninas (Romagnoli et al. 2000).

Como nem todos os domínios consistem de cadeias polipeptídicas contínuas, em alguns casos os domínios podem ser interrompidos por uma cadeia que se dobra em um domínio distinto, depois do qual a sequência original continua (como ocorre, por exemplo, na enzima alanina racemase) (Petsko and Ringe 2004a). Os domínios também podem variar em tamanho, embora não sejam maiores que 250 aminoácidos (a maioria varia entre 51-150 aminoácidos) (Petsko and Ringe 2004b). O conhecimento desses aspectos é fundamental quando a proposta inclui o estudo e a caracterização estrutural de um gene.

Existem várias ferramentas atualmente disponíveis em servidores *online* para identificação de domínios proteicos. Dentre as ferramentas mais utilizadas destacam-se oSMART (Letunic et al. 2002), o InterProScan (Jones et al. 2014) e o CD-Search (Marchler-Bauer and Bryant 2004). Essas ferramentas são baseadas em algoritmos que alinham as sequências em questão com bancos de domínios conhecidos (similar ao BLAST) ou, ainda, por comparação de posições específicas, sendo o domínio inferido. A busca por possíveis assinaturas contidas na sequência submetida à análise (*query*) também pode ser uma estratégia na busca de domínios e, neste caso, um dos programas mais utilizados é o InterProScan (Jones et al. 2014).

3.2.5. Predição *in silico* de estruturas proteicas

Um dos maiores desafios para a bioinformática estrutural consiste na predição da estrutura tridimensional das proteínas a partir de sua sequência unidimensional (sequência de aminoácidos). O objetivo é determinar que conformação determinados resíduos irão assumir em conjunto, gerando o desafio de determinar se a estrutura que a sequência vai adotar é nova ou possui semelhança com outras já conhecidas e depositadas em bancos de dados especializados. Desta forma, três principais abordagens são geralmente aplicadas para resolução do modelo: *ab initio* (ou *de novo*), *threading* e modelagem comparativa (Huzefa and Karypis 2010; Pantazes et al. 2011).

Quando os *templates* para a construção de um modelo não estão disponíveis, sua construção é guiada a partir dos princípios das leis físico-quimicas, baseadas na hipótese termodinâmica a qual tem como princípio que "*o estado nativo de uma proteína corresponde*

ao seu nível mínimo de energia livre" (Anfinsen 1973). Os métodos que se baseiam em tais ideias são: *ab initio* (Wu et al. 2007), também conhecidos como modelagem *de novo* (Bradley et al. 2005) ou modelagem livre (*free modeling*) (Jauch et al. 2008), cuja abordagem pode auxiliar a elucidar questões sobre como e por que as proteínas adotam certos padrões de dobragem.

Geralmente as técnicas de modelagem *ab initio* fazem uma busca conformacional orientada por uma dada função de energia. Este procedimento produz um número de possíveis conformações a partir das quais os modelos finais são selecionados. Assim, uma modelagem *ab initio* depende de três fatores: (1) uma função energética precisa com a qual a estrutura nativa de uma proteína corresponda ao estado mais termodinamicamente estável, em comparação com todas as estruturas possíveis; (2) um método de busca eficiente que possa identificar os estados de mínima energia através da busca conformacional; (3) seleção de modelos nativos a partir de um conjunto estruturas (Hardin et al. 2002; Lee et al. 2009).

Em alguns casos, uma *query* poderá ter um padrão de dobragem nativo semelhante a outro já conhecido e depositado em um banco de dados. Contudo, suas sequências podem não compartilhar muitas semelhanças. Uma estrutura que pode servir de exemplo é o barril-TIM, encontrado em diversas proteínas não relacionadas diretamente (Brändén 1991; Dorn et al. 2014).

À medida que os métodos de alinhamento ganham mais refinamento para encontrar similaridade entre sequências, como consequência do aumento na demanda causada em vista da quantidade crescente de sequências disponíveis em bancos de dados, o número de estruturas proteicas em questão não segue o mesmo ritmo (Figura 3). As técnicas para resolver as estruturas destas sequências residem em montar a *query* comparativamente a uma série de estruturas modelares (*templates*). A esta técnica dá-se o nome de *threading* (Eisenberg et al. 1991; Jones et al. 1992; Jones 1999).

A modelagem comparativa (ou modelagem por homologia) é utilizada quando existe uma relação clara entre uma *query* e a sequência de uma estrutura conhecida, geralmente quando a identidade está em torno de 30 % ou mais (Xiang 2006). A abordagem mais simples para a predição da estrutura de uma determinada proteína é realizada com um alinhamento de sequências pareado comparando a *query* contra as sequências das estruturas conhecidas, gerando-se, então, o alinhamento a partir do qual pode-se construir uma estrutura com base no melhor *template* reportado. Alguns algoritmos (como por exemplo, o MODELLER; Webb and Sali 2016) realizam ainda uma etapa de refinamento dos modelos

gerados, fornecendo modelos com o mínimo de energia livre (Schwede et al. 2008; Huzefa and Karypis 2010).

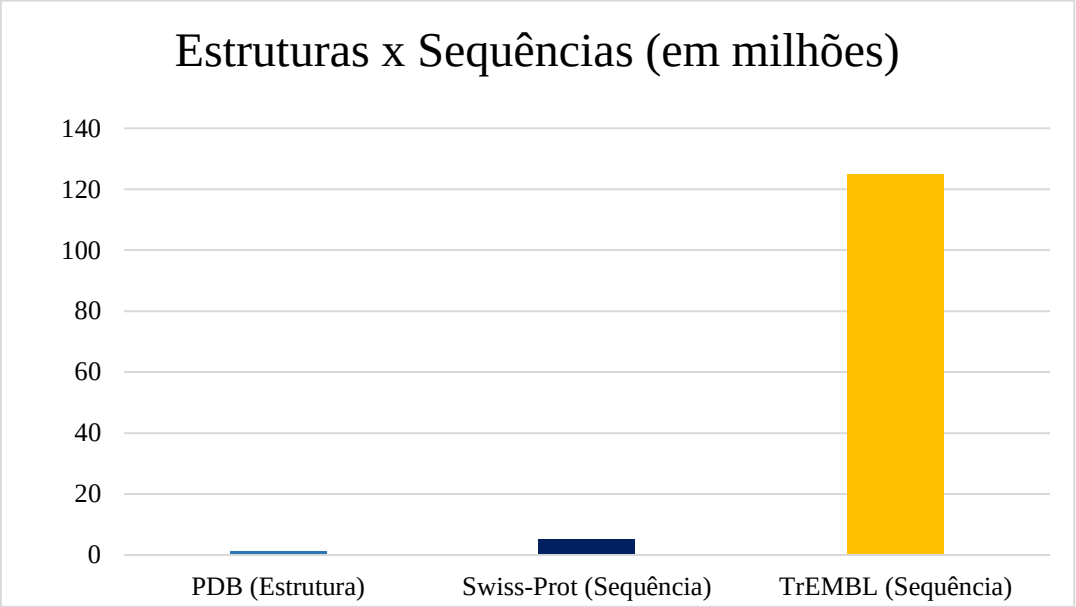


Figura 3. Quantidade de estruturas depositadas no PDB (azul claro) em comparação com o número de sequências manualmente anotadas (Swiss-Prot – Azul escuro) e automaticamente anotadas (TREMBL - Amarelo) até Junho de 2018.

3.2.6. Análises Filogenéticas e Fenéticas

Em biologia evolutiva, os principais métodos utilizados para resolver os problemas de classificação das espécies são a fenética e a filogenética, as quais diferem entre si em conceitos e abordagens (Zander 2013). Na fenética, duas espécies que compartilham mais caracteres similares, são alocadas em posições mais próximas no dendrograma. Este agrupamento é realizado com base em níveis hierárquicos, onde membros mais próximos nos ramos terminais são os que compartilham maior similaridade (Hall 1988; M. Williams and Ebach 2008). Por exemplo, ao analisarmos morfologicamente um crocodilo e um lagarto, observa-se que estes compartilham maior similaridade considerando seus caracteres morfológicos, do que quando se comparam estes táxons com um pássaro. No entanto, sabemos que filogeneticamente isso não ocorre. Assim, no fenograma teríamos um agrupamento entre os dois primeiros, com a ave em um ramo mais externo. Porém, na filogenia “crocodilo e pássaro” estariam mais próximos, uma vez que estes últimos possuem um ancestral comum mais próximo. Contudo, é importante ressaltar que filogenética e

fenética podem concordar entre si, a depender das características comparadas (Ridley 2004; Rahman et al. 2016) (Figura 4).

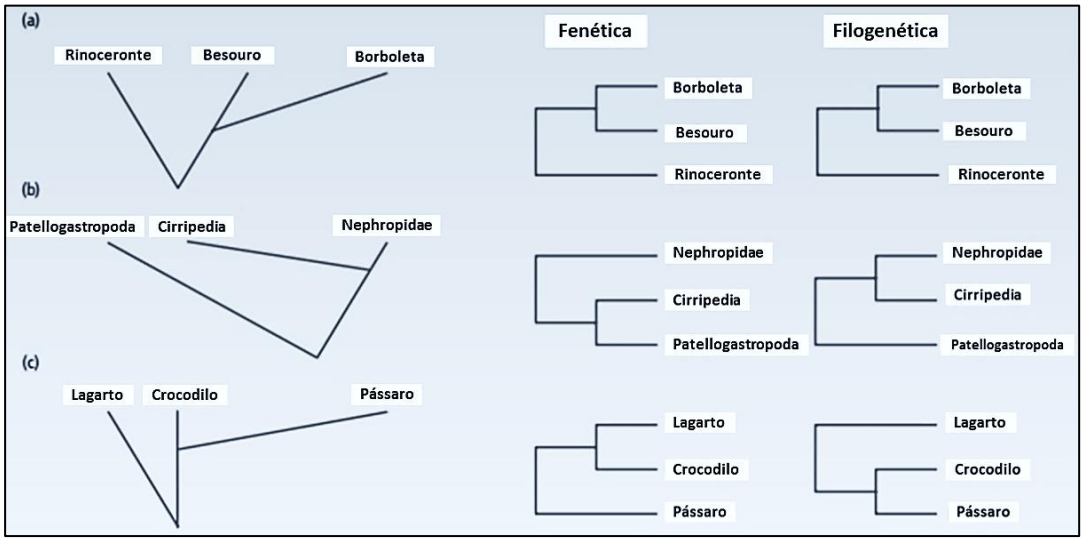


Figura 4. Os princípios de classificação fenética e filogenética podem convergir (a) ou divergir (b e c) entre si (Ridley 2004).

A fenética utiliza algoritmos baseados em matrizes de distância para construir suas representações, onde o *Unweighted Pair Group Method with Aritimetic Mean* (UPGMA) e o *Neighbor Joining* (NJ) recebem maior destaque (Lemey et al. 2009).

O UPGMA é provavelmente o método mais antigo e mais simples usado para reconstruir árvores filogenéticas a partir de dados de distância. O agrupamento é construído utilizando o menor valor fornecido pelas matrizes de distância pareadas. Esta abordagem constrói uma árvore identificando a menor distância na matriz transformando dois taxa em uma única OTU (Unidade Taxonômica Operacional – *Operational Taxonomic Unit*) que será submetida a todos os cálculos subsequentes, calculando uma nova matriz de distância e repetindo essas etapas. Em UPGMA, a distância do cluster recém-formado é a média das distâncias das OTUs que o originaram. O cálculo das médias assume que a taxa evolutiva do nó das duas OTUs agrupadas é idêntica para cada uma das entidades individualmente (Peng 2007).

Por sua vez, o método NJ constrói uma árvore alocando as OTUs vizinhas em único nó interno. O método de clusterização usado por esse algoritmo é bem diferente do UPGMA, porque ele não agrupa as OTUs com as médias mais relacionadas, mas minimiza o tamanho de todos os ramos internos e, portanto, o comprimento da árvore. O algoritmo NJ começa

assumindo uma árvore em formato de estrela que não possui ramificações internas. Na primeira etapa o algoritmo introduz a primeira ramificação interna e calcula o comprimento da árvore resultante. O algoritmo conecta sequencialmente todos os pares OTU possíveis e, finalmente, mantém os pares de OTUs que produzem a árvore mais curta (Saitou and Nei 1987; Lemey et al. 2009).

Por outro lado, a análise filogenética demanda a inclusão de um grupo externo ou grupo irmão, sendo baseada em parâmetros evolutivos, ou seja, leva em consideração a ancestralidade comum das OTUs (Ridley 2004). Em uma árvore filogenética entidades mais relacionadas serão agrupadas mais próximas nos ramos terminais, em contrapartida, à medida que o ancestral comum se torna cada vez mais distante, os agrupamentos se mostrarão cada vez mais separados na classificação (Gregory 2008). Para tal, as matrizes e métodos estatísticos possuem um arcabouço mais complexo do que aqueles utilizados nas análises fenéticas, pois analisam cada carácter das entidades taxonômicas (Peng 2007). Em uma filogenia de proteínas, cada aminoácido é tomado como um carácter a ser analisado e, então, a depender da matriz utilizada, outras variáveis são adicionadas à comparação, tais como: posição, composição química ou mutações. Por sua vez, na análise fenética apenas as diferenças e/ou semelhanças entre as entidades são levadas em consideração (Nei 1996; Lemey et al. 2009).

Entre os métodos aplicados na reconstrução filogenética incluem: máxima parcimônia (*maximum parcimony*), máxima verossimilhança (*maximum likelihood*) e inferência bayesiana (*bayesian inference*)

O método de máxima parcimônia é baseado no principio da navalha de Occam, preconizado pelo filósofo William de Occam (1285-1347), onde sugere que entre duas ou mais hipóteses que possam explicar um mesmo fenômeno, provavelmente a hipóotese mais simples seja a mais adequada (Thorburn 1915; Lemey et al. 2009). Os algoritmos deste tipo funcionam selecionando árvores construídas com o menor número de passos. Para cada *sítio* no alinhamento, todas as árvores possíveis são avaliadas - procedimento que difere daquele adotado em inferências fenéticas. Em comparação com outros métodos filogenéticos, este é menos dependente de dados pré-estabelecidos; no entanto, possui a desvantagem de ser pouco eficiente quando se tem dados muito heterogêneos (Rizzo and Rouchka 2007; Rahman et al. 2016).

A máxima verossimilhança se assemelha ao de máxima parcimônia no que diz respeito à atribuição de um *score* às diferentes topologias a serem comparadas, porém trata-se de um método probabilístico. Originalmente desenvolvido como uma metodologia

estatística para estimar parâmetros desconhecidos em um dado modelo, sendo sua função definida como a probabilidade dos dados em vista dos parâmetros (Yang and Rannala 2012). Para a construção da árvore, calculam-se as probabilidades associadas a diferentes topologias e cada uma delas com as variações nos tamanhos dos ramos, considerando o modelo evolutivo escolhido. , Trata-se do método mais amplamente utilizado na atualidade, devido à melhora do poder computacional (embora ainda oneroso) somado à implementação de programas e modelos estatísticos voltados, exclusivamente, para a análise de sequências. Além do custo computacional, apresenta a desvantagem de não apresentar a correção do modelo estatístico, que pode resultar - a depender dos parâmetros estabelecidos, na obtenção de dados discrepantes (Yang 1996; Lemey et al. 2009; Yang and Rannala 2012).

Por último, a inferência bayesiana - uma abordagem também estatística, que difere da máxima verossimilhança no modo em que tratam os parâmetros (na máxima verossimilhança são constantes fixas e na bayesiana são variáveis aleatórias), tendo como base o modelo criado por Thomas Bayes, o qual foi posteriormente adaptado para aplicação em análises filogenéticas (Yang and Rannala 1997). Antes da análise dos dados, os parâmetros devem ser atribuídos em uma distribuição anterior que é então combinada com os dados, gerando assim a distribuição (ou probabilidade) posterior (Bellhouse 2004; Yang and Rannala 2012). Assim, sua principal desvantagem seria a especificação desses parâmetros, e, apesar de demandar menor custo computacional que a máxima verossimilhança, continua sendo computacionalmente custoso (Holder and Lewis 2003).

Analogamente, outra medida qualitativa aplicada na construção de árvores, utilizada nas outras abordagens é o método de *bootstrapping.* Nesta abordagem, as colunas do alinhamento original são reamostradas aleatoriamente, repetidas vezes. e novas bases de dados (réplicas) são geradas, sendo cada uma delas usada para gerar uma árvore. O algoritmo de construção da árvore é então aplicado a esse novo conjunto de dados, onde todo o procedimento de seleção das colunas e construção da árvore é repetido em uma quantidade de vezes pré-estabelecida (geralmente 1000 vezes). Com isso, tem-se na árvore final a frequência com que um determinado conjunto de dados foi amostrado (Peng 2007).

Em conjunto, todas estas metodologias *in silico* podem ser aplicadas para uma maior compreensão dos diversos aspectos relacionados ao objeto de estudo (a exemplo dos peptídeos antimicrobianos), envolvendo informações a cerca da estrutura primária, diversidade e função, até a sua possível história evolutiva. Além disso, análises com ferramentas computacionais são também essenciais para a manipulação e tratamento de dados, os quais possam, posteriormente, ser validados através de experimentos laboratoriais,

implicando em inferências biotecnológicas. Neste contexto, a escolha do tipo de dado e metodologia de análise dependem da natureza da hipótese evolutiva que se deseja testar.

CAPÍTULO 1

Capítulo publicado: ***Current Protein and Peptide Science***

Current Protein and Peptide Science, 2017, *18*, 1-7 1

REVIEW ARTICLE

Snakin: Structure, Roles and Applications of a Plant Antimicrobial Peptide

Marx Oliveira-Lima[1], Ana Maria Benko-Iseppon[1], José Ribamar Costa Ferreira Neto[1], Susana Rodríguez-Decuadro[2], Ederson Akio Kido[1], Sergio Crovella[1,3] and Valesca Pandolfi[1,*]

[1]*Federal University of Pernambuco, Center of Biological Sciences, Department of Genetics, Laboratory of Genetics and Vegetal Biotechnology, Av. Prof. Moraes Rêgo 1235, CEP 50670-420, Recife, PE, Brazil;* [2]*Universidad de la República, Facultad de Agronomía, Departamento de Biología Vegetal, Garzón 780, Montevideo, Uruguay;* [3]*Genetic Service, IRCCS Burlo Garofolo and Department of Developmental and Reproductive Sciences, University of Trieste -Via dell'Istria, 65/1- 34137 Trieste, Italy*

ARTICLE HISTORY

Received: June 24, 2015
Revised: October 20, 2015
Accepted: November 12, 2015

DOI: 10.2174/1389203717666160619183140

Abstract: Snakins are plant antimicrobial peptides (AMPs) of the Snakin/GASA family, formed by three distinct regions: an N-terminal signal peptide; a variable site; and the GASA domain in the C-terminal region composed by twelve conserved cysteine residues that contribute to the biochemical stability of the molecule. These peptides are known to play different roles in response to a variety of biotic (i.e., induced by bacteria, fungi and nematode pathogens) and abiotic (salinity, drought and ROS) stressors, as well as in crosstalk promoted by plant hormones, with emphasis on abscisic and salicylic acid (ABA and SA, respectively). Such properties make snakin/GASA members promising biotechnological sources for potential therapeutic and agricultural applications. However, information regarding their tertiary structure, mode of action and function are not yet completely elucidated. The present review presents aspects of snakin structure, expression, functional studies and perspectives about the potential applications for agricultural and medical purposes.

Keywords: Snakin/GASA domain, cysteine-bridges, stress response, crosstalk.

1. INTRODUCTION

Along evolution plants developed complex defense mechanisms against a broad spectrum of pathogens. However, for many pathogenic microorganisms defense mechanisms are still inefficient, leading to disease development [1]. A common feature in defense strategies of plants and animals is the existence of a wide variety of small antimicrobial peptides (AMPs) that act as effector molecules of non-specific defense [2]. These peptides figure among most effective plant weapons, since they can be both constitutively expressed and stress induced by a wide range of biotic and abiotic stressors [3, 4]. Normally encoded by multigenic families, in which some genes are developmentally regulated, AMPs may act as potent broad-spectrum antibiotics, with potential as novel therapeutic agents.

Usually formed by 20 to 120 amino acid residues, these peptides can be rich in proline, glycine, histidine, arginine, tryptophan or cysteine [5]. Most of the cysteine-rich peptides have a globular structure, which is stabilized by disulfide bridges [6] and represent the major class of antimicrobial proteins found in plants. Among plant cysteine-rich AMPs, snakins are the most recently described peptides, isolated initially from potato and pepper, showing a broad spectrum activity against plant pathogens [7-9]. Studies based on sequence analysis, phenotypic characterization and expression pattern analysis have suggested that these peptides can be involved in diverse processes, both in development and in response to external factors [8, 10].

The present review focuses on the most significant features of this new family of antimicrobial peptides: structure, function, expression and potential applications, exploring its importance based on the current knowledge.

2. SNAKIN/ GASA FAMILY

2.1. Molecular Structure and Characterization

Snakin genes encode small proteins that comprise three distinct domains: (1) a signal peptide with 18-29 residues, (2) a variable region that is highly divergent between family members, both in amino acid composition and sequence length, and (3) the GASA domain, a conserved C-terminal region with approximately 60 amino acids, 12 of them are cysteine residues in conserved positions [10] (Fig. **1**).

*Address correspondence to this author at the Federal University of Pernambuco, Center of Biological Sciences, Department of Genetics, Laboratory of Genetics and Vegetal Biotechnology, Av. Prof. Moraes Rêgo 1235, CEP 50670-420, Recife, PE, Brazil; E-mail: valescapandolfi@gmail.com

1389-2037/17 $58.00+.00

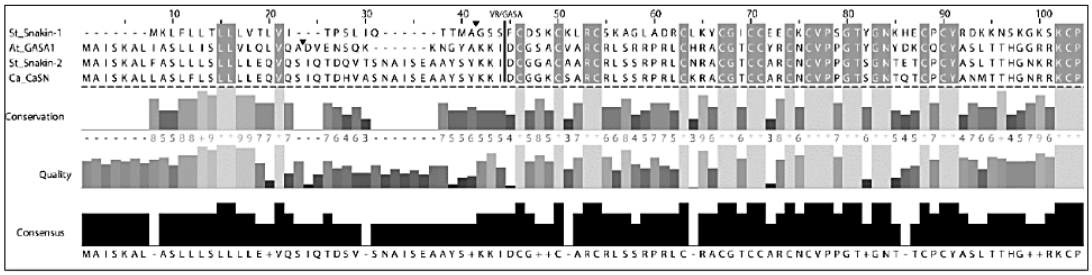


Fig. (1). Multiple sequence alignment of snakin/GASA peptides. The black triangles show de cleavage site for the signal peptide in *Solanum tuberosum* Snakin-1 (St-Snakin-1), *Arabidopsis thaliana* GASA1 (At_GASA1), *Solanum tuberosum* Snakin-2 (St_Snakin-2) and *Capsicum annuum* Snakin (Ca_CaSN), which is in the same position for At_Snakin-1, Ca_Casn and St_Snakin-2. The vertical bar indicates the limit between the variable region (VR) and the start of the GASA domain that exhibits twelve conserved cysteines and other nine conserved residues (dark gray in the alignment). Bars below the alignment indicate the levels of conservation, quality and, at the bottom, the consensus sequence.

Snakin/GASA (Gibberellin acid-stimulated from Arabidopsis) family comprises a group of widely distributed peptides among higher plants (Table **1**) sharing common structural features with many known antimicrobial peptides from eukaryotes, including the N-terminal putative signal sequence and a highly divergent acidic intermediate region. However, the main structural aspect is a cysteine-rich C-terminal signature, which contains 12 cysteine amino acids, in highly conserved positions (XCX3CX3CX8CX3CX2C CX2CXCX11CXCX12CX), forming a globular conformation, through six disulfide bonds (Fig. **1**). This C-Terminal region was described in *Arabidopsis thaliana* by Herzog *et al.* [10] and is called GASA domain. This arrangement is considered important for structure maintenance (specially the 3D conformation) and function of these peptides [11-15].

Studies have revealed that genes from this family may be regulated by plant hormones and participate in hormonal signaling pathways modulating the responses of the plant through their involvement in redox and hormone homeostasis. It is known that the presence of hormones such as Gibberellic Acid (GA) and SA has been associated to an increase in the expression level of snakin/GASA genes of *A. thaliana*, *Fagus sylvatica, Fragaria ananassa, Gerbera hybrida*, *Petunia hybrida*, *Oryza sativa* and *Zea mays* [10].

Genes of this family often encode a basic peptide with a molecular size below 11 kDa. After post-translational processes, the mature peptide basically consists of a structure with approximately 60 amino acid residues, with a molecular weight of approximately 7 KDa, for all snakin isolated so far. The basic isoelectric point (pI < 9.4) was described for Snakin-2 from potato and other GASA homologues, as well as for other antimicrobial peptides [7, 8, 16, 17], revealing that this feature is very conserved among AMPs and their homologues. Their basic character provides an advantage in interacting with negatively charged microbial membrane components [2]. All these features were also observed in a recently isolated snakin from a nematode-resistant pepper cultivar (*Capsicum annuum* cv. Santaka). According to Mao *et al.* [11] the CaSn (*C. annuum* snakin) gene encodes a protein of 104 amino acids residues (11 kDa) and a mature peptide of 66 amino acid residues that corresponds to a molecular mass of 7.03 kDa. Another member of this family was isolated from alfalfa (*Medicago sativa*), sharing many features (as a signal peptide comprising 25 residues) with the peptides of this group, representing a new member of the family [18].

Although there are three different genomic structures for Snakin/GASA members, only two structures were observed in snakins: (1) consisting of two exons and (2) consisting of three exons [11, 19]. Genes with two exons have been commonly described among antimicrobial peptides [6], including Snakin-1 from *Solanum tuberosum*, being considered a component of the constitutive defense against pathogens in both storage and reproductive plant organs [9, 20]. The second structure, also found in the Snakin-2, from potato [7] and in other AMPs [6] was associated with developmental stages and in response to specific pathogens, being considered an important component of both constitutive and induced defense [8, 21, 22]. The number of copies of Snakin-1 and Snakin-2 in the potato genome was estimated by southern blot hybridization, therefore the patterns observed indicated the existence of only one or two copies per genome [7, 9], whereas no CNV (Copy Number Variation) has been reported up to date.

Snakin/GASA proteins are expressed in different plant organs, however little is known about their 3D structure and mode of action [23]. To our knowledge, there is just one theoretical model for Snakin-1, whose putative structure is very similar to two other AMPs: thionins and α-helical hairpins [24, 25], formed by one short helix and two long α-helices. These findings strongly suggest that some classes of cysteine-rich AMPs share a common ancestor [6]. Furthermore, although the cysteine residues may be conserved in sequence and position, the disulfide bonds may be not structurally conserved, showing different patterns among snakin members [15, 26].

2.2. Synthesis, Developmental Process and Relationship to Stress Tolerance

2.2.1. Spatiotemporal Expression

When considering subcellular localization, there is a tendency among different snakin/GASA members to vary,

Table 1. Plant Snakin/GASA peptides previously described.

Sequence name	Species	Family	Reference
GEG	*Gerbera hybrida*	Asteraceae	[52]
GASA1-14	*Arabidopsis thaliana*	Brassicaceae	[10, 11, 12, 36]
GmSN-like	*Glycine max*	Fabaceae	[Oliveira-Lima *et al.*]*
GsGASA	*Glycine soja*		[35]
MsSN1	*Medicago sativa*		[18]
Snakin-like	*Phaseolus vulgaris*		[28]
VuSN-like	*Vigna unguiculata*		[Oliveira-Lima *et al.*]*
FsGASA	*Fagus sylvatica*	Fagaceae	[53]
OsGASR1-2; OsGSR1	*Oryza sativa*	Poaceae	[54, 55]
ZmGSL1-10	*Zea mays*		[31]
FaGAST1	*Fragaria ananassa*	Rosaceae	[30]
CaSN	*Capsicum annuum*	Solanaceae	[8]
Gip1-5	*Petunia hybrida*		[56, 57]
RSI-1; GAST1	*Solanum lycopersicum*		[29, 34]
Snakin1-2	*Solanum tuberosum*		[10, 22]

* Author's data, paper *in prep*.

whereas some representatives (i.e., Snakin-1 from an agroinfiltrated *Nicotiana benthamiana*) had their localization experimentally observed in the plasmatic membrane. Moreover, depending on their structural features (for example, when a signal peptide is present) not all snakin/GASA proteins are secreted to the extracellular matrix possibly due to post-translational modifications. Other interactions may lead their localization as shown by Nahirñak *et al*. [27] reporting that Snakin-1 self-interacts *in vivo*. Another exemplified interaction regarded SN2-like from common bean (*Phaseolus vulgaris*) which associated with a proline-rich protein, resulting in a protein complex named FBCBP (French bean chitin-binding protein) [28].

Currently it is known that the expression of these peptides may vary for each gene and organ or may still be tissue specific. Most information about the function of some Snakin/GASA members was based on phenotypic characterization of mutants and transgenic plants or, still, on expression profiling. Some candidates seem to be involved in processes of arrest or promotion of cell elongation and division in adult plants, root formation (in early stages), flowering time and also fruit ripening [27, 29-33].

Among snakin genes, *Snakin-1* and *2* have different levels of expression along the age of the plant and developmental stage in potato plants. *Snakin-1* had its expression level decreasing along the plant development, being more active in young stages in tissues like carpels, petals, floral buds and tubers [9]. In turn, *Snakin-2* exhibited a pattern varying for both, development and tissues, being found in stems, stamens, flower buds, leaves, shoot apices, carpels, petals and tubers, being also induced by pathogen infection [7].

2.2.2. Hormone Responsiveness

Many Snakin/GASA genes have been reported to have their transcription regulated by hormones [32, 34-36]. The first members of this family showing positive induction by GA and auxin hormones were the tomato *GAST1* (*Gibberellic Acid-Stimulated Transcript* 1) and *RSI-1* genes, respectively [29, 34]. Nevertheless, the positive induction by hormones (especially GA) was not considered a decisive factor for all members. For example, the transcription of potato *Snakin-2* was inhibited by GA, whereas the expression of *Snakin-1* of the same species had no alterations [7, 9]. On the other hand, under abscisic acid (ABA) treatment, the *Snakin-2* was induced, confirming the importance of crosstalk between snakin and this hormone, since ABA is an important wound signaling in Solanaceae species [7].

2.2.3. Redox Homeostasis

All identified members of the Snakin/GASA family presented a signature with twelve cysteine residues in the GASA domain, which is considered to exhibit a significant oxidative power due to their putative catalytic disulfide bonds (redox active cysteines), indicating a participation in redox regulation [15, 37, 38]. Besides, many metabolic pathways are activated by biotic stress which, in turn, trigger crosstalk responses, such as Reactive Oxygen Species (ROS) [10, 14, 39]. It is known that ROS is also involved in wounding where snakin may play a role in the redox regulation [10,

21]. Additionally, Nahirñak *et al.* [28] observed that metabolites that mitigate ROS effects were significantly reduced in *Snakin-1* silenced strains from *S. tuberosum* in comparison to wild type. These lines also exhibited affected cell division, since the silenced lines presented smaller leaf sizes, what might reflect an inhibition of the cell division or growth, suggesting that Snakin-1 might be involved in cell division by modulating some metabolites [27].

2.2.4. Biotic and Abiotic Stress

As previously mentioned, snakins have been associated with a wide range of biotic and abiotic stresses [7, 14, 20, 21, 40, 41]. Interestingly, *Snakin-1* was not induced in potato leaves by either biotic or abiotic stresses, indicating this peptide as a component of the plant constitutive defense [9]. In contrast to *Snakin-1*, *Snakin-2* was positively regulated in potato leaves by wounding and presented a weak response to salinity stress, while drought stress caused no effect [7].

The following section details the existing evidence that *Snakin* genes are induced when plants are challenged by a wide range of microorganisms, whereas their overexpression in several angiosperms resulted in increased resistance against several microbial diseases. These findings, together with their reported *in vitro* antimicrobial activities may indicate a potential role of snakins to act as resistance traits against the main phytopathogens.

3. ROLE IN DEFENSE RESPONSE: *IN VITRO* AND *IN VIVO* BIOLOGICAL ACTIVITIES

Several *in vitro* and *in vivo* antimicrobial activities have been attributed to plant snakins. It has been reported that Snakin-1 (StSN1) and Snakin-2 (StSN2), isolated from *S. tuberosum* cv. Desireé and cv. Caerla, respectively, exhibit *in vitro* antimicrobial activity against important pathogens of potato and other plant species. StSN2 and StSN1 have been effective against the bacteria *Clavivacter michiganensis* subsp. s*epedonicus* and against fungal pathogens such as *Fusarium solani*, *Colletotrichum lagenarium*, *Colletotrichum graminicola*, *Bipolaris maydis* and *Botrytis cinerea,* among others, but not against *Ralstonia solanacearum*, an important potato bacterial pathogen [7, 9]. It was further observed that StSN1 presented an intermediate inhibitory effect on *Listeria monocytogenes*, a foodborne bacteria [42].

Additionally, StSN1 was heterologously expressed in *Escherichia coli,* followed by peptide purification and evaluation of its activity. This recombinant protein showed antibacterial and antifungal activity, inhibiting completely the growth of *Clavibacter miciganensis* subsp. *sepedonicus* and spore germination of *Botrytis cinerea* and *Colletotrichum coccoides* at a concentration of 14 mM [43], also inhibiting the growth of the mycelium of *Gaeumannomyces graminis* var. *tritici* - an important wheat pathogen [44]. A peptide of *Capsicum annuum* (CaSN), with high homology to StSN2 was also produced in *E. coli.* Free-living nematodes, like *Caenorhabditis elegans* were fed with *E. coli* strains expressing CaSN, leading to an inhibition of larval development, reduced growth and mobility, as well as death of most nematodes. In addition, plant-parasitic nematode intoxication was observed in toxicity assays using the supernatant containing the CaSN protein [8].

A possible role *in planta* has been suggested by overexpressing a *snakin* gene resulting in plants with increased resistance against several diseases. Transgenic lines of *S. tuberosum* subsp. *tuberosum* cv. Kennebec that accumulated high levels of SN1 mRNA from a potato wild relative (*Solanum chacoense*) showed heightened resistance against important potato pathogens, such as the fungus *Rhizoctonia solani* and the gram negative bacteria *Erwinia carotovora* [40] and also against *Pectobacterium carotovorum* subsp. *carotovorum* and *Blumeria graminis* [20]. *Snakin-1* also decreased the virulence of the bacteria *Dickeya chrysanthemi* in sensitive tomato mutants [45], while the overexpression of a gene encoding SN2 increased tolerance to *Clavivacter michiganensis* subsp. *michiganensis* [21]. Transgenic wheat expressing StSN1 showed prominent resistance to 'take-all root rot', a wheat root disease caused by *Gaeumannomyces graminis* var. *tritici* [44], indicating a potential *in vivo* antifungal and antibacterial activity.

Plant snakins are induced by a wide range of biotic stresses and by plant hormones involved in stress signaling, also supporting their role in plant defense. In potato *StSN1* gene activation was not observed when young leaves were subjected to biotic and abiotic stimuli [9]. However, a down-regulation of the gene encoding *StSN2* was observed during *Ralstonia solanacearum* and *Erwinia chrysanthemi* infection, whereas an up-regulation was detected after infection with the compatible fungus *Botrytis cinerea* [7]. In *Nicotiana benthamiana* the silencing of *Snakin-2* homologs increased susceptibility to *Clavibacter michiganensis* subsp. *michiganensis* [41], while their over-expression in transgenic tomato plants limited the invasiveness of this tomato pathogenic bacterium [21]. Variation regarding resistance and/or susceptibility to different pathogens indicates possible pathogen-specific associations or still induction mediated by other signaling molecules.

4. SNAKIN BIOTECHNOLOGY: STATE OF ART, APPLICABILITY AND PERSPECTIVES

Due to the identified antimicrobial activities [9] snakins have been considered interesting biotechnological targets. Thus, their biotechnological significance was indeed highlighted after the recognition of their involvement in the regulation of developmental processes in plants and also by the possibility to obtain considerable amounts of this peptide by expression in *E. coli* [10, 46]. Although their limited exploration in relation to other AMPs, the potential of snakins has been analyzed mainly by methods of transgeny and cisgeny (artificial gene transfer between organisms that could otherwise be conventionally bred).

In this regard, Almasia *et al.* [40] generated transgenic potato plants (*S. tuberosum* subsp. *tuberosum* cv. Kennebec) transformed via *Agrobacterium tumefaciens* with a construct encoding the gene SN1 of *S. chacoense* under the control of the CaMV 35S promoter. The resulting transgenic lines were evaluated under inoculation with *Erwinia carotovora* and *Rhizoctonia solani,* both relevant bacterial and fungal potato pathogens, respectively, whereas the transformed lines showed high rates of survival when compared to wild-type controls, indicating that StSN1 is involved in the resistance process against these pathogens.

In turn, Balaji & Smart [43] obtained genetically engineered tomato (*Solanum lycopersicum*) plants overexpressing *StSN2* or *ELP* (*Extensin like protein*) genes of the own species. The generated OGMs were evaluated under infection with the bacteria *Clavibacter michiganensis* subsp. *michiganensis* (*Cmm*), causal agent of bacterial wilt and canker, one of the most important tomato diseases throughout the world. In the transformed plants, the bacterial population was significantly reduced (100–10,000 fold) when compared to the wild type, indicating that StSN2 and ELP products limit the action of this pathogen, suggesting a potential *in vivo* antibacterial activity.

Additionally, Mohan *et al.* [22] obtained transgenic lines of potato (cv. Iwa) overexpressing the gene *Gibberellin Stimulated-Like 2* (*GSL2*; also known as *Snakin-2*) from the same species. Three trials performed independently showed that the modified strains exhibited a significant induction of *GSL2* transcriptional activity resulting in increased resistance to *Pectobacterium atrosepticum* (previously known as *Erwinia carotovora* subsp. *atroseptica*) the causal agent of potato blackleg, strengthening the role of this AMP in the defense against pathogens in potato.

Evaluations addressed to other environmental factors have shown that snakins may have a wide range of functions, highlighting their biotechnological potential. For example, Almasia *et al.* [47] characterized the promoter region (*PStSN1*) of the potato gene *StSN1*. Using bioinformatic approaches the authors identified 55 possible target regulation motifs associated to tissue specificity, abiotic stresses (high and low temperatures), defense and response to hormones, among others. To confirm the indicated functional dynamism the authors evaluated transgenic lines of *A. thaliana* expressing a reporter gene under control of the identified promoter (PStSN1:GUS). Contrary to expectations, there was no correlation between bioinformatic simulations (which indicated their potential functions) and the activation by the expected stressors, including phytohormones, light, dark, UV radiation or infection by *Pseudomonas syringae*. However, plants transformed with the identified promoter presented induction of responses associated to high or low temperatures, and mechanical damage (wounding), providing interesting targets for development of plants tolerant to extreme temperatures, insectivory [48], attack by pathogenic fungi [49] and viruses [50].

In addition to the antimicrobial role in biotic and abiotic stresses, recent data on snakins have increased the range of their potential biotechnological applications. Zare-Zardini *et al.* [51] observed that a protein named Snakin-Z presented, in addition to the already known antioxidant role, inhibitory effects over acetylcholinesterase (AChE) and butyrylcolinesterase (BChE) enzymes. This was considered a very important finding, since there are evidences that the inhibition of these enzymes may be applicable for the treatment of Alzheimer's disease, indicating that snakins may be a potential target for pharmacological trials in animal models to confirm these properties by *in vivo* essays [51].

The range of functions regarding snakins may still be underestimated, since several proteins bearing the Snakin/GASA domain play important roles in response to various types of stress. Due to the observed properties, snakins may have broad applications in different sectors: from agriculture, by obtaining tolerant/resistant transgenic plants to a range of stressful conditions; pharmaceutical industry, since their antifungal and antibacterial effects can be exploited for the development of antibiotics or antimicrobial substances, as well as a possible role in medicine, in the field of neurological diseases, due to its proposed ability to inhibit target enzymes involved in Alzheimer's disease [51]. Therefore, additional efforts are necessary to identify additional representatives of these interesting AMPs and their possible biological roles, especially considering their functional plurality.

CONCLUDING REMARKS

Among the diverse families of plant antimicrobial peptides, snakins deserve a special attention, considering that, at one side, they share many features with other AMPs and, by the other side, they possess peculiar features, like the number of cysteines (twelve) in the C-terminal region, which is the most important GASA domain signature, an important player of redox homeostasis.

Snakins seem to represent different roles in diverse plant developmental stages and tissues, from young to differentiated tissues, some induced by stress, exhibiting activity against a variety of bacteria, fungi and also nematodes. It has been also associated to hormones specially the GA, auxin and SA, being also constitutively expressed with participation in key events like cell division. Based on all these evidences, the broad potential of snakins is still to be explored, expanding the knowledge of their diversity, evolution, mode of action and its correlation with the structure, which constitute its main knowledge gap.

LIST OF ABBREVIATIONS

ABA	=	Abscisic acid
AMP	=	Antimicrobial peptide
CaMV	=	Cauliflower mosaic virus
CaSN	=	*Capsicum annuum* snakin
ELP	=	Extensin-like protein
FaGASA	=	Fragaria ananassa GASA
FBCBP	=	French bean chitin-binding protein
FsGASA	=	Fagus sylvatica GASA
GA	=	Gibberellic acid
GASA	=	Gibberellin acid-stimulated Arabidopsis
GAST	=	Gibberellic acid-stimulated transcript
GEG	=	Gerbera homolog of *GAST1* gene
Gip	=	GA-induced proteins
GmSN	=	*Glycine max* snakin
GsGASA	=	*Glycine soja* GASA
GSL	=	Gibberellin stimulated-like
MsSN	=	Medicago sativa snakin

OsGASR = *Oryza sativa* GA-stimulated transcript-related gene

ROS = Reactive oxygen species

RSI-1 = Root system inducible-1

SA = Salicilic acid

StSN = Solanum tuberosum snakin

VuSN = Vigna unguiculata snakin

ZmGSL = *Zea mays* gibberellin-stimulated like

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We thank Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brasília, Brazil), Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE, Recife, Brazil) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brasília, Brazil) for fellowships and financial support.

REFERENCES

[1] Mysore, K.S.; Ryu, C.M. Nonhost resistance: How much do we know? *Trends Plant Sci.,* **2004**, *9*, 97-104.

[2] Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature,* **2002**, *415*, 389-395.

[3] García-Olmedo, F.; Molina, A.; Alamillo, J.M.; Rodríguez-Palenzuéla, P. Plant defense peptides. *Biopolymers,* **1998**, *47*, 479-491.

[4] Benko-Iseppon, A.M.; Galdino, S.L.; Calsa, T.; Kido, E.A.; Tossi, A.; Belarmino, L.C.; Crovella, S. Overview on plant antimicrobial peptides. *Curr. Protein Pept. Sci.,* **2010**, *11*, 181-188.

[5] Bulet, P.; Stöcklin, R.; Menin, L. Antimicrobial peptides: From invertebrates to vertebrates. *Immunol. Rev.,* **2004**, *198*, 169-184.

[6] Silverstein, K.T.; Moskal, W.; Wu, H.C.; Underwood, B.; Graham, M.; Town, C.D.; VandenBosch, K. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.,* **2007**, *51*, 262-280.

[7] Berrocal-Lobo, M.; Segura, A.; Moreno, M.; López, G.; García-Olmedo, F.; Molina, A. Snakin-2 , an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection. *Plant Physiol.,* **2002**, *128*, 951-961.

[8] Mao, Z.; Zheng, J.; Wang, Y.; Chen, G.; Yang, Y.; Feng, D.; Xie, B. The new casn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper. *Phytoparasitica,* **2011**, *39*, 151-164.

[9] Segura, A.; Moreno, M.; Madueño, F.; Molina, A.; García-Olmedo, F. Snakin-1 , a peptide from potato that is active against plant pathogens. *Mol. Plant-Microbe Interact.,* **1999**, *12*, 16-23.

[10] Nahirñak, V.; Almasia, N.I.; Hopp, H.E.; Vazquez-Rovere, C. Involvement in hormone crosstalk and redox homeostasis Snakin / GASA proteins. *Plant Signal. Behav.,* **2012**, *7*, 1004-1008.

[11] Herzog, M.; Dorne, M.; Grellet, F. GASA, a gibberellin-regulated gene family from *Arabidopsis thaliana* related to the tomato GAST1 gene. *Plant Mol. Biol.,* **1995**, *27*, 743-752.

[12] Aubert, D.; Chevillard, M.; Dorne, M.; Arlaud, G.; Herzog, M. Expression patterns of GASA genes in *Arabidopsis thaliana*: The GASA4 gene is up-regulated by gibberellins in meristematic regions. *Plant Mol. Biol.,* **1998**, *36*, 871-883.

[13] Jianzong, P.; Liujing, L.A.I.; Xiaojing, W. PRGL : A cell wall proline-rich protein containning GASA domain in *Gerbera hybrida*. *Sci. China C Life Sci.,* **2008**, *51*, 520-525.

[14] Sun, S.; Wang, H.; Yu, H.; Zhong, C.; Zhang, X.; Peng, J.; Wang, X. GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation. *J. Exp. Bot.,* **2013**, *64*, 1637-1647.

[15] Porto, W.F.; Franco, O.L. Peptides theoretical structural insights into the Snakin / GASA family. *Peptides,* **2013**, *44*, 163-167.

[16] Ko, C.B.; Woo, Y.M.; Lee, D.J.; Lee, M.C.; Kim, C.S. Enhanced tolerance to heat stress in transgenic plants expressing the GASA4 gene. *Plant Physiol. Biochem.,* **2007**, *45*, 722-728.

[17] Ferreira, R.B.; Monteiro, S.; Regina, F.; Santos, C.N.; Chen, Z.; Batista, L.M.; Duarte, J.; Borges, A.; Teixeira, A.R. The role of plant defence proteins in fungal pathogenesis. *Mol. Plant Pathol.,* **2007**, *8*, 677-700.

[18] García, A.; Ayub, N.; Fox, A.; Gómez, M.; Diéguez, M.; Pagano, E.; Berini, C.; Muschietti, J.; Soto, G. Alfalfa Snakin-1 prevents fungal colonization and probably coevolved with rhizobia. *BMC Plant Biol.,* **2014**, *14*, 248.

[19] Meiyalaghan, S.; Thomson, S.J.; Fiers, M.W.; Barrell, P.J.; Latimer, J.M.; Mohan, S.; Jones, E.E.; Conner, A.J.; Jacobs, J.M. Structure and expression of GSL1 and GSL2 genes encoding gibberellin stimulated-like proteins in diploid and highly heterozygous tetraploid potato reveals their highly conserved and essential status. *BMC Genomics,* **2014**, *15*, 2.

[20] Faccio, P.; Hopp, E.; Favret, E.; Franzone, P. Increased tolerance to wheat powdery mildew by heterologous constitutive expression of the *Solanum chacoense* Snakin-1 gene. *Czech J. Genet. Plant Breed.,* **2011**, *47*, 135-141.

[21] Balaji, V.; Smart, C.D. Over-expression of Snakin-2 and extensin-like protein genes restricts pathogen invasiveness and enhances tolerance to *Clavibacter michiganensis* Subsp. *michiganensis* in transgenic tomato (*Solanum lycopersicum*). *Transgenic Res.,* **2012**, *21*, 23-37.

[22] Mohan, S.; Meiyalaghan, S.; Latimer, J.M.; Gatehouse, M.L.; Monaghan, K.S.; Vanga, B.R.; Pitman, A.R.; Jones, E.E.; Conner, A.J.; Jacobs, J.M.E. GSL2 over-expression confers resistance to *Pectobacterium atrosepticum* in potato. *TAG. Theor. Appl. Genet.* **2013**, *127*, 677-689.

[23] Nawrot, R.; Barylski, J.; Nowicki, G.; Broniarczyk, J.; Buchwald, W.; Goździcka-Józefiak, A. Plant antimicrobial peptides. *Folia Microbiol. (Praha).,* **2014**, *59*, 181-196.

[24] Nolde, S.B.; Vassilevski, A.; Rogozhin, E.; Barinov, N.; Balashova, T.; Samsonova, O.V.; Baranov, Y.V.; Feofanov, A.V.; Egorov, T.; Arseniev, A.S.; Grishin, E.V. Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (Echinochloa *crus-galli*). *J. Biol. Chem.,* **2011**, *286*, 25145-25153.

[25] Oparin, P.B.; Mineev, K.S.; Dunaevsky, Y.E.; Arseniev, A.S.; Belozersky, M.; Grishin, E.V; Egorov, T.; Vassilevski, A. Buckwheat trypsin inhibitor with helical hairpin structure belongs to a new family of plant defence peptides. *Biochem. J.,* **2012**, *446*, 69-77.

[26] Padovan, L.; Scocchi, M.; Tossi, A. Structural aspects of plant antimicrobial peptides. *Curr. Protein Pept. Sci.,* **2010**, *11*, 210-219.

[27] Nahirñak, V.; Almasia, N.I.; Fernandez, P.V.; Hopp, H.E.; Estevez, J.M.; Carrari, F.; Vazquez-Rovere, C. Potato Snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. *Plant Physiol.,* **2012**, *158*, 252-263.

[28] Bindschedler, L.V; Whitelegge, J.P.; Millar, D.J.; Bolwell, G.P. A two component chitin-binding protein from french bean-Association of a proline-rich protein with a cysteine-rich polypeptide. *FEBS Lett.,* **2006**, *580*, 1541-1546.

[29] Taylor, B.H.; Scheuring, C.F. A molecular marker for lateral root initiation : The RSI-1 gene of tomato (*Lycopersicon esculentum* Mill) is activated in early lateral root primordia. *Mol. Genet. Genomics,* **1994**, *243*, 148-157.

[30] De la Fuente, J.I.; Amaya, I.; Castillejo, C.; Sánchez-Sevilla, J.F.; Quesada, M.; Botella, M.; Valpuesta, V. The strawberry gene FaGAST affects plant growth through inhibition of cell elongation. *J. Exp. Bot.,* **2006**, *57*, 2401-2411.

[31] Zimmermann, R.; Sakai, H.; Hochholdinger, F. The gibberellic acid stimulated-like gene family in maize and its role in lateral root development. *Plant Physiol.,* **2010**, *152*, 356-365.

[32] Zhang, S.; Yang, C.; Peng, J.; Sun, S.; Wang, X. GASA5, a regulator of flowering time and stem growth in *Arabidopsis thaliana*. *Plant Mol. Biol.,* **2009**, *69*, 745-759.

[33] Guzmán-Rodríguez, J.J.; Ibarra-Laclette, E.; Herrera-Estrella, L.; Ochoa-Zarzosa, A.; Suárez-Rodríguez, L.M.; Rodríguez-Zapata, L.C.; Salgado-Garciglia, R.; Jimenez-Moraila, B.; López-Meza,

J.E.; López-Gómez, R. Analysis of expressed sequence tags (ESTs) from avocado seed (*Persea americana* Var . *drymifolia*) reveals abundant expression of the gene encoding the antimicrobial peptide Snakin. *Plant Physiol. Biochem.,* **2013**, *70*, 318-324.

[34] Shi, L.; Gast, R.T.; Gopalraj, M.; Olszewski, N.E. Characterization of a shoot-specific, GA3- and ABA-regulated gene from tomato. *Plant J.,* **1992**, *2*, 153-159.

[35] Li, K.L.; Bai, X.; Li, Y.; Cai, H.; Ji, W.; Tang, L.L.; Wen, Y.D.; Zhu, Y.M. GsGASA1 mediated root growth inhibition in response to chronic cold stress is marked by the accumulation of DELLAs. *J. Plant Physiol.,* **2011**, *168*, 2153-2160.

[36] Roxrud, I.; Lid, S.E.; Fletcher, J.C.; Schmidt, E.D.L.; Opsahl-Sorteberg, H.G. GASA4, one of the 14-member Arabidopsis GASA family of small polypeptides, regulates flowering and seed development. *Plant Cell Physiol.,* **2007**, *48*, 471-483.

[37] Harris, P.W.R.; Yang, S.H.; Molina, A.; López, G.; Middleditch, M.; Brimble, M. Plant antimicrobial peptides Snakin-1 and Snakin-2: Chemical synthesis and insights into the disulfide connectivity. *Chemistry,* **2014**, 1-10.

[38] Wigoda, N.; Ben-Nissan, G.; Granot, D.; Schwartz, A.; Weiss, D. The gibberellin-induced, cysteine-rich protein GIP2 from *Petunia hybrida* exhibits in planta antioxidant activity. *Plant J.,* **2006**, *48*, 796-805.

[39] Torres, M.A. ROS in biotic interactions. *Physiol. Plant.,* **2010**, *138*, 414-429.

[40] Almasia, N.I.; Bazzini, A.A.; Hopp, H.E.; Vazquez-Rovere, C. Overexpression of Snakin-1 gene enhances resistance to *Rhizoctonia solani* and *Erwinia carotovora* in transgenic potato plants. *Mol. Plant Pathol.,* **2008**, *9*, 329-338.

[41] Balaji, V.; Sessa, G.; Smart, C.D. Silencing of host basal defense response-related gene expression increases susceptibility of *Nicotiana benthamiana* to *Clavibacter michiganensis* subsp. *michiganensis*. *Phytopathology,* **2011**, *101*, 349-357.

[42] López-Solanilla, E.; González-Zorn, B.; Novella, S.; Vázquez-Boland, J.A.; Rodríguez-Palenzuela, P. Susceptibility of *Listeria monocytogenes* to antimicrobial peptides. *FEMS Microbiol. Lett.,* **2003**, *226*, 101-105.

[43] Kovalskaya, N.; Hammond, R.W. Expression and functional characterization of the plant antimicrobial Snakin-1 and defensin recombinant proteins. *Protein Expr. Purif.,* **2009**, *63*, 12-17.

[44] Rong, W.; Qi, L.; Wang, J.; Du, L. Expression of a potato antimicrobial peptide SN1 increases resistance to take-all pathogen *Gaeumannomyces graminis* Var . *tritici* in transgenic wheat. *Funct. Integr. Genomics,* **2013**, *3*, 403-409.

[45] López-Solanilla, E.; García-Olmedo, F.; Rodríguez-Palenzuela, P. Inactivation of the sapA to sapF locus of *Erwinia chrysanthemi* reveals common features in plant and animal bacterial pathogenesis. *Plant Cell,* **1998**, *10*, 917-924.

[46] Meiyalaghan, S.; Latimer, J.M.; Kralicek, A.V; Shaw, M.L.; Lewis, J.G.; Conner, A.J.; Barrell, P.J. Expression and purification of the antimicrobial peptide GSL1 in bacteria for raising antibodies. *BMC Res. Notes,* **2014**, *7*, 777.

[47] Almasia, N.I.; Narhirñak, V.; Hopp, H.E.; Vazquez-Rovere, C. Isolation and characterization of the tissue and development-specific potato Snakin-1 promoter inducible by temperature and wounding. *Electron. J. Biotechnol.,* **2010**, *13*, 1-21.

[48] Gulbitti-Onarici, S.; Zaidi, M.A.; Taga, I.; Ozcan, S.; Altosaar, I. Expression of Cry1Ac in transgenic tobacco plants under the control of a wound-inducible promoter (AoPR1) isolated from *Asparagus officinalis* to control *Heliothis virescens* and *Manduca sexta*. *Mol. Biotechnol.,* **2009**, *42*, 341-349.

[49] Corrado, G.; Bovi, P.D.; Ciliento, R.; Gaudio, L.; Di Maro, A.; Aceto, S.; Lorito, M.; Rao, R. Inducible expression of a *Phytolacca heterotepala* ribosome-inactivating protein leads to enhanced resistance against major fungal pathogens in tobacco. *Phytopathology,* **2005**, *95*, 206-215.

[50] Corrado, G.; Scarpetta, M.; Alioto, D.; Di Maro, A.; Polito, L.; Parente, A.; Rao, R. Inducible antiviral activity and rapid production of the ribosome-inactivating protein I from *Phytolacca heterotepala* in tobacco. *Plant Sci.,* **2008**, *174*, 467-474.

[51] Zare-Zardini, H.; Tolueinia, B.; Hashemi, A.; Ebrahimi, L.; Fesahat, F. Antioxidant and cholinesterase inhibitory activity of a new peptide from *Ziziphus jujuba* fruits. *Am. J. Alzheimers. Dis. Other Demen.,* **2013**, *28*, 702-709.

[52] Kotilainen, M.; Helariutta, Y.; Mehto, M.; Pollanen, E.; Albert, V.; Elomaa, P.; Teeri, T. GEG participates in the regulation of cell and organ shape during corolla and carpel development in *Gerbera hybrida*. *Plant Cell,* **1999**, *11*, 1093-1104.

[53] Alonso-Ramírez, A.; Rodríguez, D.; Reyes, D.; Jiménez, J.A.; Nicolás, G.; López-Climent, M.; Gómez-Cadenas, A.; Nicolás, C. Evidence for a role of gibberellins in salicylic acid-modulated early plant responses to abiotic stress in Arabidopsis seeds. *Plant Physiol.,* **2009**, *150*, 1335-1344.

[54] Furukawa, T.; Sakaguchi, N.; Shimada, H. Two OsGASR Genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles. *Genes Genet. Syst.,* **2006**, *81*, 171-180.

[55] Wang, L.; Wang, Z.; Xu, Y.; Joo, S.H.; Kim, S.K.; Xue, Z.; Xu, Z.; Wang, Z.; Chong, K. OsGSR1 is involved in crosstalk between gibberellins and brassinosteroids in rice. *Plant J.,* **2009**, *57*, 498-510.

[56] Ben-Nissan, G.; Weiss, D. The petunia homologue of tomato GAST1: Transcript accumulation coincides with gibberellin-induced corolla cell elongation. *Plant Mol. Biol.,* **1996**, *32*, 1067-1074.

[57] Ben-Nissan, G.; Lee, J.Y.; Borohov, A.; Weiss, D. GIP, a *Petunia hybrida* GA-induced cysteine-rich protein: A possible role in shoot elongation and transition to flowering. *Plant J.,* **2004**, *37*, 229-238.

CAPÍTULO 2

Artigo a ser submetido à revista: ***Peptides***
(Online ISSN: 0196-9781)

Structural and functional analysis of Snakin/GASA members in soybean (*Glycine max*) and description of a new member (*Gm*SN2)

Marx Oliveira-Lima[1]; Valesca Pandolfi[1]; William Farias Porto[2]; Jose Ribamar Costa Ferreira Neto[1]; Octávio Luiz Franco[2,3]; Lidiane Lindinalva Barbosa Amorim[1]; Francismar Correa Marcelino Guimarães[4]; Ricardo Vilela Abdelnoor[4]; Ederson Akio Kido[1]; Ana Maria Benko-Iseppon[1*]

[1] Departamento de Genética, Centro de Biociências, Universidade Federal de Pernambuco, Av. Prof. Morais Rego 1235, CEP 50.670-420, Recife, PE, Brazil.

[2] Centro de Análises Proteômicas e Bioquímicas, Programa de Pós-graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, 70719-100 Brasília, DF, Brazil.

[3] S-Inova, Programa de Pós-Graduação em Biotecnologia, Universidade Católica Dom Bosco, 79117-900 Campo Grande, MS, Brazil.

[4] Empresa Brasileira de Pesquisa Agropecuária, Embrapa Soja, Rodovia Carlos João Strass, Caixa Postal 231, 86001-970 - Londrina, PR, Brazil.

Corresponding author[*]: E-mail: ana.benko.iseppon@pq.cnpq.br

Abstract

Several members of Snakin/GASA family have been described in plant species, all of them with a C-terminal conserved motif of 12 cysteines and six disulfide bonds. They are reported to be expressed in different tissues, developmental stages and under given environmental conditions - sometimes constitutively or induced by pathogens. This work focuses on the characterization of Snakin/GASA (SNK) representatives in soybean based on transcriptomic data, available at GENOSOJA (Brazilian Soybean Genome Consortium). Thirty-three non-redundant snakin sequences were identified in the soybean transcriptome of which 20 exhibited the complete GASA-domain. They share many characteristics with other (SNK) members, like isoelectric point ranging from 5.53 to 9.32, low molecular weight (6.88-18.04 kDa) and putative antimicrobial activity showed by *in silico* approaches. All sequences were addressed to the extracellular environment. A new snakin member (GmSN2) is here described for the first time, including gene isolation, genome sequencing, expression validation by RT-qPCR and analysis of the peptide structure and stability. This new SNK (GmSN2) presented an α-hairpin structure that is common in other AMPs, and it is also stabilized by disulfide bonds (six) with the pattern: Cys1-7, Cys2-5, Cys3-4, Cys6-12, Cys8-11, and Cys9-10. The validation by RT-qPCR showed that GmSN2 is constitutively expressed in leaves, maintaining the same expression level also after inoculation with the rust fungus *Phakopsora pachyrhizi*. The genome-wide analyses revealed that GmSN2 has 40 homologs distributed along 16 of 20 soybean chromosomes and also in six chromosomes of *P. vulgaris* and *M. truncatula*. This study provides an important source of information to map potential defense genes of biotechnological importance.

Keywords: antimicrobial peptides, molecular modeling, molecular dynamics, annotation.

Running title: Snakin/GASA in soybean

1. Introduction

Cysteine-rich small peptides (CRP) are effective components of eukaryotic genomes. In plants, they are constitutively expressed or induced under biotic and abiotic stresses and have emerged as an efficient alternative to mitigate the effects of stress factors [1,2]. These peptides can be classified into different families, such as thionins, defensins, cyclotides and nakins (SNKs) (Benko-Iseppon et al. 2010; Nawrot et al., 2014).

SNK proteins are members of the Snakin/GASA protein family [5] which exhibit three distinct domains: (i) a signal peptide with 18-29 residues, (ii) a variable region that is highly divergent between family members, both in amino acid composition and sequence length, and (iii) a conserved C-terminal region with approximately 60 residues, of which 12 are cysteine residues in conserved positions connected by six disulfide bonds which stabilize the mature peptide - the GASA domain [6–8]. Studies based on sequence analysis, phenotypic characterization and expression pattern have suggested that these peptides may be involved in plant growth and development, in response to abiotic or biotic stresses [9–12]. Concerning biotic stresses, a recent study (Oliveira-Lima et al. 2017) showed that Snakin-2 could cause membrane pores in fungi and bacteria, also participating in crosstalk promoted by phytohormone signaling.

SNK proteins have been identified in several plant species, including tomato (*Solanum lycopersicum*) [13], petunia (*Petunia hybrida*) [14], *Arabidopsis thaliana* [15], potato (*Solanum tuberosum*) [6,7], gerbera (*Gerbera hybrida*) [16], strawberry (*Fragaria ananassa*) [17], common bean (*Phaseolus vulgaris*) [18], rice (*Oryza sativa*) [19], beechnut (*Fagus sylvatica*) [20], maize (*Zea mays*) [21], pepper (*Capsicum annum*) [10], alfalfa (*Medicago sativa* L.) [22], rape (*Brassica napus*) [23], in wild soja bean (*Glycine soja;* Li et al. 2011) and, most recently, in soybean (*Glycine max*; He et al. 2017). The listed works report the expression of SNK coding genes under abiotic or biotic stresses.

In view of the importance of soybean and the limiting factors to the increase in its production, such as long periods of drought and susceptibility to pathogens, the GENOSOJA consortium was created aiming at the search for transcripts involved in biotic and abiotic stresses useful to increase soybean productivity (Benko-Iseppon et al. 2012). The GENOSOJA database was constructed with millions of soybean transcripts identified via RNAseq, DeepSuperSAGE, miRNA, and ESTs, under different biotic and abiotic stress

situations. Detailed information about available libraries, accessions, and experimental conditions are described in Benko-Iseppon et al. (2012) and do Nascimento et al. (2012).

The present work intends to answer the following questions: (1) How many SNKs exist in the soybean genome, what is their structure and how many are expressed? (2) Expression of these SNKs is tissue or specific stage or both, depending on the SNK analyzed? (3) Is the distribution of SNKs in the genome of soybeans and other legumes clustered or dispersed by several chromosomes? SNK expression is preferably constitutive or induced after stress? Are SNKs also induced under abiotic stress? The genomic and transcriptomic data available for soybean (with emphasis on those generated by the GENOSOJA consortium, which includes 29 libraries of different tissues and situations) allow us to shed light on these issues, helping to understand the function and range of action of SNKs.

2. Material and Methods

2.1. Screening for snakin seed sequences and soybean homologs

Complete sequences of experimentally validated SNKs (StSN1 and StSN2), previously identified in potato (*Solanum tuberosum* cv. Desireé; Segura et al. 1999; Berrocal-lobo et al. 2002) were used to retrieve homologs by BLASTp search [28] on PhytAMP (http://phytamp.pfba-lab-tun.org/about.php; Hammami et al. 2009) database. UniProt identifiers (IDs) of the retrieved homologs and both potato SNKs were used as seed sequences to retrieve orthologs with the complete GASA domain using SeedServer (http://biodados.icb.ufmg.br/seedserver/; Guedes et al., unpublished). The Batch CD-Search [30] was used in order to identify the conserved domain. Next, a tBLASTn (cut-off e-value $\leq e^{-04}$) was performed on the GENOSOJA database (http://lge.ibi.unicamp.br/soybean; do Nascimento et al. 2012) to identify soybean homologs. Redundant matches were eliminated, and the non-redundant were annotated to their putative homology performing a BLASTx against the NCBI nr database. The sequences identified were translated with ORF-Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), and the conserved domains were annotated via Batch CD-Search. The predicted subcellular localization, the putative antimicrobial activity, the signal peptide, the molecular weight (MW) and the isoelectric

point (pI) (for the mature peptide) were analyzed using Protcomp V.9 [31], $CAMP_{R3}$ algorithms [32] SignalP [33] and JVirgel [34], respectively.

2.2. In silicoprofiling of soybean snakins

The expression profile of putative SNKs was traced using transcripts downloaded from GENOSOJA databank (http://lge.ibi.unicamp.br/soybean; Benko-Iseppon et al. 2012; do Nascimento et al. 2012). After counting on different EST libraries, the evaluation of reads frequencies that compose each contig was submitted to the Hierarchical Clustering Method [35], which creates a matrix that compares the set of genes pair to pair to build a Neighbor-Joining tree. The dendrograms were visualized using Java TreeView [36]. Additionally, these contigs were used for anchoring the three comparisons (case x control) of the six DeepSuperSAGE libraries.

*2.3. Gm*SN2 *isolation and Neighbor-Joining analysis*

Soybean genomic DNA was extracted using CTAB method (Weising et al. 1995). The genomic sequence of *Gm*SN2 was obtained by PCR, using specific primers (F) AGCTTGTCTTTGGCACCCTA and (R) ACATTTGCACTCAGCACAGC, designed based on the genomic homology of the transcript Contig14826 soybean genomic data obtained from Phytozome v.9 [37] to amplify the complete GASA domain.

PCR amplification was performed with 10 ng DNA, 10X PCR buffer, 1,25 µL of 25 mM $MgCl_2$, 1.5 µL of 10 mM dNTPs, 1.0 µL of each primer (10 mM), 1.0 U of Taq DNA Polymerase, in a total reaction volume of 25 µL. PCR reactions were performed using the following cycling conditions: 95°C for 10 min, 35 cycles of 95°C for 1 min, 58°C for 1 min, 72°C for 1 min, followed by an extension step at 72°C for 10 min. The reactions were performed on a Techne TC-412 Thermocycler (Barloworld Scientific). The PCR products were separated in 1.5% (w/v) agarose gel, quantified by fluorometry (Qubit-Invitrogen) and its identity was confirmed by sequencing (from both directions: forward and reverse, in triplicate) using the BigDye Terminator v3.1 Cycle Sequence Kit (ThermoFisher) in a Genetic Analyzer 3500 (Applied Biosystems) at the Sequencing Platform LABBE (Universidade Federal de Pernambuco, Recife, Brazil).

The *Gm*SN2 sequences (in triplicate) were assembled using BioEdit, v.7 (Hall, 1999) and the contig was annotated using BLASTx against the NCBI nr database and translated

using ORF-Finder (https://www.ncbi.nlm.nih.gov/orffinder/). The conserved domain, signal peptide, pI, and MW were predicted using CD-Search [30], SignalP v. 4.1 [33] and JVirgel [34], respectively. SNK protein sequences described in previous works, the trancripts retrieved from GENOSOJA databank and the *Gm*SN2 sequence was aligned using ClustalW [38] of the MEGA6 package [39], followed by a Neighbor-Joining tree construction with bootstrap (1,000 replicates) and Dayhoff model [40].

2.4. Genomic distribution and gene structure of snakin homologs

The peptide sequence from *Gm*SN2 was aligned against soybean, *Medicago truncatula* and *Phaseolus vulgaris* genomes at Phytozome v.10.3 [37] and NCBI using tBLASTn. Alignment matches that shared the GASA domain were evaluated to identify their relative position in the respective virtual chromosomes, the gene structure, as well as their abundance. The graphic representation of the Snakin candidates in the virtual chromosomes was generated with Circos software package [41].

2.5. Molecular modeling

In order to select the best template to perform the comparative modeling for GmSN2 and the soybean transcripts, a BLASTp was performed against the Protein Data Bank [42] using the respective sequences as seeds. Using MODELLER 9.15 [43] one hundred of molecular models were constructed with snakin-1 structure as template (PDB ID: 5E5Q), whose final model was selected according to the discrete optimized protein energy (DOPE) score (a measure which evaluates the energy of the model and indicates the best one). Additionally, the final models were evaluated through ProSA II [44] which accesses to the folding quality of the protein and evaluated by ProFunc [45] to check the stereochemical quality of the model, using Ramachandran plot. Reliable models are expected to have more than 90% of the amino acid residues in the most favored and allowed regions. Structure visualization was carried out using PyMol v.1.6 (The PyMol Molecular Graphics System, Schrödinger, LLC).

2.6. Validation of GmSN2 gene expression by RT-qPCR

Soybean plants inoculated with the fungus *Phakopsora pachyrhizi* were used for expression analysis of *Gm*SN2 gene. Details on the soybean accession studied as well as growing conditions are available in Kulcheski et al. 2010. Three biological replicates were evaluated for controls and treatments. Leaf tissues from each plant were collected at 1, 12, 24, 48, 96 and 192 hai (hours after inoculation), frozen in liquid nitrogen and used for RNA extraction. Samples from 48 hai (hours after inoculation), including mock treatments, were chosen for expression analysis of *Gm*SN2 gene. Total RNA of leaf tissue was extracted using Trizol (Invitrogen) according to the manufacturer's instructions. The first-strand cDNA synthesis reaction was done with 1 µg of total RNA (DNA-free) using the ImPro-II Reverse Transcription System (Promega) and stored at -20°C until expression analysis by RT-qPCR.

The expression profile of *GmSN2* was analyzed by RT-qPCR using soybean cDNA samples (48 dai, including inoculated and mock treatments, with biological triplicates). A set of seven genes (previously identified as potential reference genes) were included in this approach (Supplementary Table S1). Firstly, the amplification efficiency (E) of each primer was calculated according to the equation: $E\ (\%) = (10^{-1/slope} - 1) \times 100$ (Rasmussen, 2001), using five serially diluted cDNA (concentrated, 1:10, 1:100, 1:1000 and 1:10,000).

All RT-qPCR amplifications were performed on the LineGene 9660 model (Bioer), using SYBR Green detection chemistry (USB). Each reaction mixture comprised 1 µL of template cDNA, 5 µL of HotStart-IT SYBR Green qPCR Master Mix 2x (USB), 0.05 µL of ROX, 1.95 µL of water, and 1 µL primer (500 nM each) to a final volume of 10 µL. The reactions were denatured at 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, 58°C for 30 s, and 72°C for 15 s in 96-well reaction plates, with the detection of the fluorescence signal at the end of each extension step. After amplification, dissociation curves were produced (60°C to 95°C at a heating rate of 0.1°C/sec and acquiring fluorescence data every 0.3°C) to discriminate the main reaction products from other nonspecific ones or primer-dimers. Each PCR always included an NTC (no template control) and reverse transcription negative controls.

The relative expression analysis was performed using the REST 2009 software package [47]. The expression stability of the reference genes was evaluated using the geNorm v. 3.5 software [48] which is based on the assumption that expression of two ideal reference genes will always have the same ratio among samples regardless of the experimental conditions before the qPCR.

3. Results

3.1. Snakin seed-sequences and homologs in soybean

Using the potato SN1 (ID: Q948Z4.1) and SN2 (ID:Q93X17.1) as seeds sequences on PhytAMP database, the BLASTp recovered sequences from four plant families as best matches: Orchidaceae (*Gymnadenia conopsea*, ID: A3F8U7), Brassicaceae (*Arabidopsis thaliana*, ID: Q1G2Y4), Rosaceae (*Fragaria ananassa*, ID: O49134) and Asteraceae (*Gerbera hybrida*, ID: Q9XGJ3). Afterward, these UniProt IDs were used as input in the SeedServer. This analysis retrieved 52 SNK candidate sequences, 47 of which showed complete GASA domain, distributed on 10 plant families: Brassicaceae (11 genes); Fabaceae (7 genes); Solanaceae (6 genes); Selaginellaceae (5 genes), Pinaceae and Poaceae (both 4 genes); Vitaceae, Salicaceae and Euphorbiaceae (both 3 genes); Orchidaceae (1 gene) (Supplementary Fig.1)**.**

The tBLASTn performed on GENOSOJA database (using the 47 SNK candidates as queries) recovered 33 non-redundant sequences within the established parameters (cut-off e^{-4}) (data not shown), of which, 20 sequences exhibited the complete GASA-domain. A subsequent BLASTx search on nr NCBI revealed that 14 sequences described as "unknown" were all from *G. max*. Sequence lengths ranged from 481 to 965 bp and from 65 to 138 amino acid residues (**Table 1**).

The virtual 2D-electrophoresis evaluation revealed signal peptides with isoelectric point (pI) varying between 4.30 (Contig15060) and 9.58 (Contig7785), while for the mature peptides, pI values ranged from 5.53 (Contig14720) to 9.32 (Contig9321) (**Table 1**). Regarding the molecular weight (MW), there was a discrete variation between the values in the signal peptide (from 2.37 to 2.99 kDa in the contigs Contig22342 and SJ01-E1-C06-026-A02-UC.F, respectively). For the mature peptides, this variation was between 6.88 (Contig15060) and 18.04 (Contig9293) kDa. For contig14826, only the MW value for the mature peptide is presented, due to the absence of the signal peptide (**Table 1**).

The subcellular localization analysis showed that all sequences possessed the signal peptide addressing to the extracellular environment, a result supported by optimal scores (between 8.9 and 9.6; data not shown), regarding the antimicrobial activity prediction all candidates were classified as AMPs.

Table 1 - Soybean clusters identified through tBLASTn tool in the GENOSOJA database. A search conducted with seed sequences (*e-value* ≤ e^{-4}) reported by the SeedServer (only those with complete GASA-domain) and annotated via BLASTx tool (GenBank).

Reference ID	Cluster	nr NCBI BLASTx best matches		Size		MW	pI
		Accession	Description	nt	aa		
GmSN-like1	**Contig1367**	ACU15001.1	Unknown	749	99	8.07	8.53
GmSN-like2	**Contig1373**	XP_003549639.1	Snakin-2-like	965	115	10.36	7.93
GmSN-like3	**Contig14720**	XP_003528161.1	Unc Prot	804	119	10.77	5.53
GmSN-like4	**Contig14826**	XP_003528161.1	Unc Prot	802	65	6.92	8.39
GmSN-like5	**Contig15060**	XP_003525918.1	Snakin-1	696	88	6.88	8.10
GmSN-like6	**Contig16419**	ACU14488.1	Unknown	608	106	9.00	8.50
GmSN-like7	**Contig18282**	ACU14567.1	Unknown	792	107	9.33	8.75
GmSN-like8	**Contig2059**	ACU14224.1	Unknown	481	106	9.11	7.72
GmSN-like9	**Contig20912**	ACU15584.1	Unknown	809	138	12.33	8.78
GmSN-like10	**Contig22342**	ACU14458.1	Unknown	637	115	10.31	7.93
GmSN-like11	**Contig23713**	ACU16624.1	Unknown	619	90	7.18	8.37
GmSN-like12	**Contig26289**	XP_003536595.1	Snakin-1-like	554	90	7.24	8.50
GmSN-like13	**Contig3375**	ACU13692.1	Unknown	735	99	8.10	8.80
GmSN-like14	**Contig3731**	ACU14995.1	Unknown	791	88	6.92	9.28
GmSN-like15	**Contig627**	ACU14995.1	Unknown	783	88	6.87	9.09
GmSN-like16	**Contig7139**	XP_003523021.1	GRP	798	106	9.01	8.90
GmSN-like17	**Contig7785**	ACU13226.1	Unknown	847	117	10.59	6.89
GmSN-like18	**Contig9293**	ACU16056.1	Unknown	935	191	18.04	9.29
GmSN-like19	**Contig9321**	ACU13162.1	Unknown	701	110	9.38	9.32
GmSN-like20	**SJ01-E1-C06-026-A02-UC.F**	ACU15001.1	Unknown	542	94	7.46	8.66

UncProt: Uncharacterized protein; MW: molecular weight; pI: isoelectric point (pI).

3.2. Analyses of snakin based on ESTs and DeepSuperSAGE data

Hierarchical clustering of the EST libraries was carried out, including 29 soybean libraries from different tissues and developmental stages. Among the organs and tissues evaluated, the most representative number of libraries regarded seed tissues (six libraries, with 1,988 reads, representing 82% of total 2,415). The remaining reads were present in tissues as leaves (129 reads), hypocotyls (61 reads), flowers (50 reads), roots (31 reads) and epicotyls (8 reads). Moreover, read frequency obtained from normalized data revealed the prevalence of transcripts in the early stages of development (**Figure 1**).

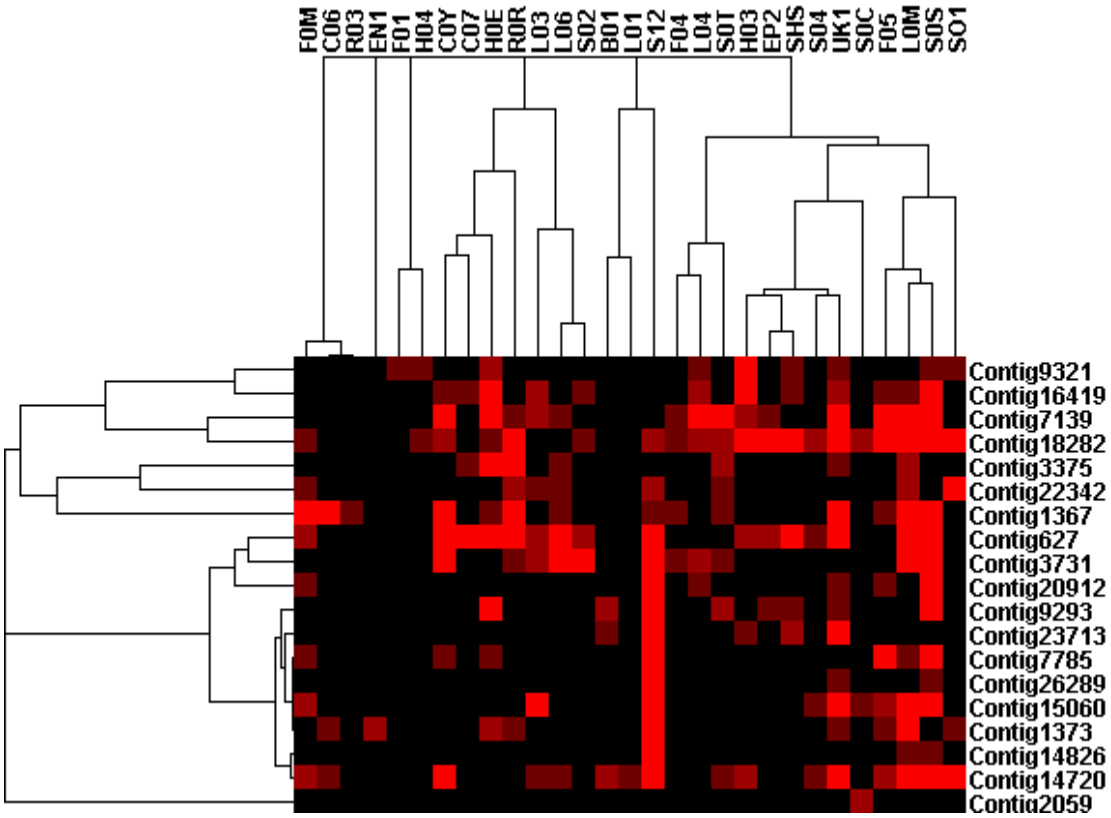


Figure 1 - Hierarchical clustering of transcriptional profile in soybean performed using the CLUSTER software. The red color indicates higher prevalence; the dark red indicates intermediate prevalence and black means absence of SNKs in the respective treatment. Libraries: B01 (vegetable buds; field-plants); C0Y (young cotyledons); C06 (wounded cotyledons); C07 (degenerating cotyledons; 10 days old, etiolated seedlings); EN1 (endosperm - developing seeds); EP2 (seedling epicotyl); F01 (floral meristem); F0M (mature flowers); F04 (floral meristem mRNA); F05 (immature flowers; field-plants); H03 (hypocotyl/plumule; germinating seeds); H04 (*P. sojae;* infected hypocotyl); H0E (etiolated hypocotyls); L01 (senescing leaf; greenhouse-plants); L03 (fully expanded leaves; greenhouse-plants); L04 (immature leaves; greenhouse-plants); L0M (mature leaves); L06 (leave; drought stressed); R0R (roots); R03 (seedling roots); S02 (mature seed pods; greenhouse-plants); S04 (young seeds); S0C (seed coats); S0S (seedlings); S12 (seeds; globular stage embryos); SHS (shoots); SO1 (somatic embryos; cultured on MSD 20); S0T (stem) and UK1 (unknown tissue).

The analysis of the DeepSuperSAGE data allowed the identification of 81 tags (out of 4,000,000 in the database), of which, 20 regarded the root dehydration comparison in the drought-tolerant contrast (DT *vs.* TC; **Figure 2A**), 16 were associated with the same

treatment in the drought-sensitive contrast (DS *vs*. SC; **Figure 2B**), whereas 45 tags occurred in the leave tissue contrast (RR *vs*. RC; i.e., rust-resistant *vs*. non-inoculated control; **Figure 2C**). Regarding exclusive tags in each stress condition, 22 tags were exclusive in the comparisons regarding root tissues (comparisons 12 TD *vs*. TC; 9 SD *vs*. SC; and one tag in both), while 34 tags were exclusive to the leave tissue contrasts (RR *vs*. RC). Among 81 tags obtained, 31 (38 %) showed up/down-regulation (UR/DR, respectively): 10 tags in the contrast TD *vs*. TC (1 UR and 9 DR); 1 DR tag in the contrast SD *vs*. SC and 20 tags (6 UR and 14 DR) in the contrast RR *vs*. RC.

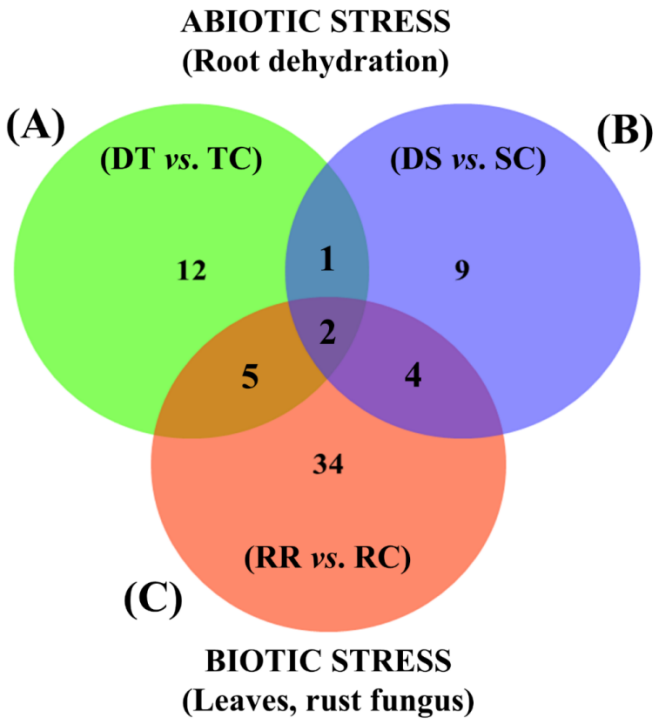


Figure 2 - Venn diagram showing the number of DeepSuperSAGE UR tags matching soybean SNK exclusive in each contrast. (**A**) DT *vs*. TC, roots of drought-tolerant accession (Embrapa 48); stressed *vs*. non-stressed control; (**B**) DS *vs*. SC, roots of drought-susceptible accession (BR16); stressed *vs*. non-stressed control and (**C**) RR *vs*. RC, leaves of rust-resistant accession (PI561356); inoculated *vs*. non-inoculated - Mock.

*3.3. Analysis of the snakin Gm*SN2

The *Gm*SN2 genomic sequence contains an ORF (215 bp), which encodes a peptide with a variable region (1-11 residues), followed by a 60-residue sequence corresponding to

the conserved GASA domain (**Figure 3**). This mature peptide is basic (pI = 8.52) with a predicted molecular mass of 6.32 kDa.

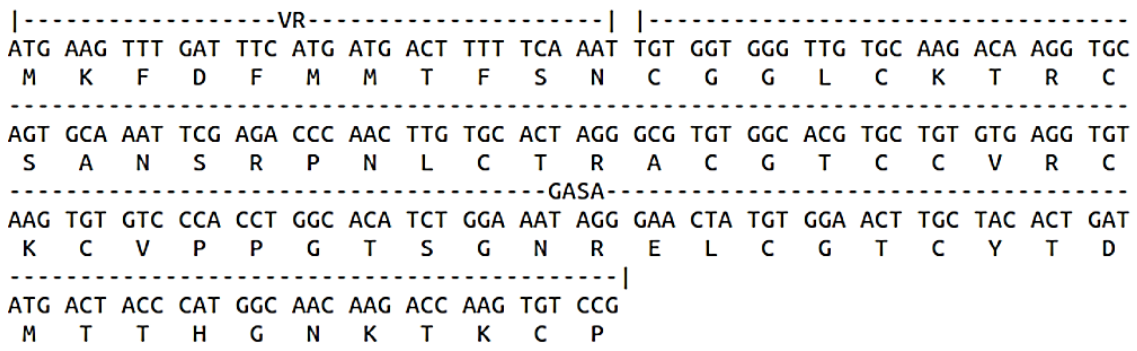


Figure 3. *Gm*SN2 gene sequence and the putative amino acid sequence. VR: Variable region; GASA: GASA domain.

3.4. Neighbor-Joining analysis

The alignment showed that GmSN2 shares a similarity of 75% in the amino acid sequence with StSN2 (snakin-2 of *S. tuberosum*). Additionally, 21 high conserved residues were observed, containing 12 cysteine motifs and other eight residues (R, V, P, G, G, Y, K, P), performing the general motif: CX3CX2RCX8CX3CX2CCX2CXCVPXGX2GX4CXCYX10/13KCP. This pattern was observed in all sequences (**Figure 4**) and regards the typical conserved structure of the Snakin/GASA family.

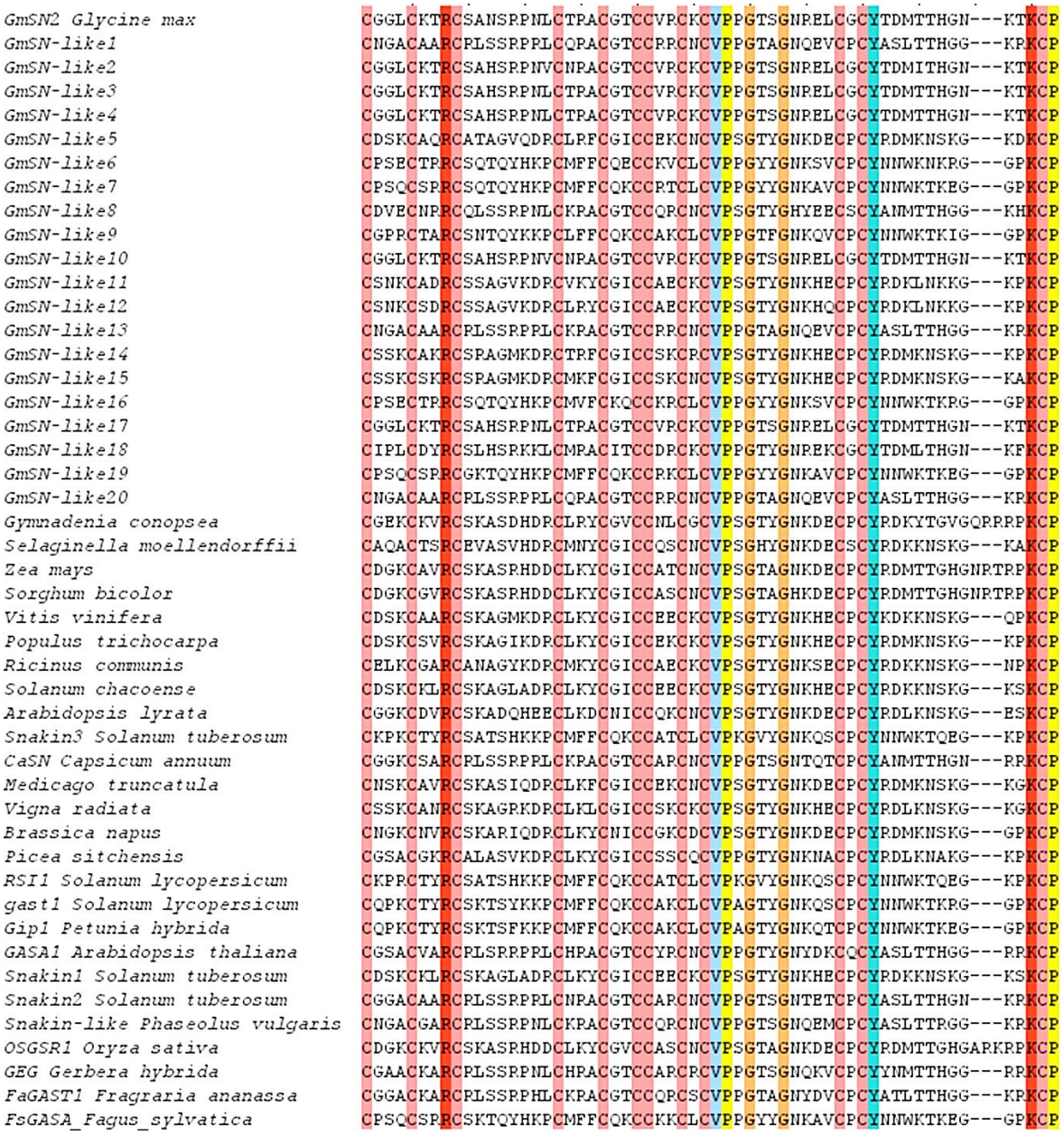


Figure 4 - Multiple sequence alignment performed with peptide sequences of Snakin/GASA family members, the *Gm*SN2 and the other soybean peptides (GmSN-like) identified showing the main conserved residues represented by the 12 cysteine motif (blue) and other eight residues, Arginine (R) – yellow; Lysine (K): pink; Valine (V): light green; Proline (P): red; Glycine (G): green and Tyrosine (Y): grey, that compose the general motif "CX3CX2RCX8CX3CX2CCX2CXCVPXGX2GX4CXCYX10/13KCP".

The Neighbor-Joining tree showed that Snakin/GASA family can be divided into three subfamilies (I-III). The subfamily I comprises the GmSN-like5, 11, 12, 14 and 15 with 15 other homologues, the general pattern of the group is: CX2A/KCX2RCX2AX3D/EXCX3CG/NI/VCCX2CXCVPP/SGH/TA/YGH/NKX2CP/SCYR/KDX2N/TX2GX/5KCP, where X is any of the 20 proteinogenic amino acids, the subfamily II comprises the GmSN2 and 10 other soybean transcripts (GmSN-like1-4, 6, 16-20) and the general pattern of the subfamily is: CX3CX2RCX8CX3CX2CCX2CXCVPP/SGT/YXGX4CXCYX4T/NX5KCP, finally the subfamily III which is the smallest group with 10 sequences with the transcripts GmSN-like6, 7, 9, 16 and 19 with the general pattern: CX3CT/SXRCSXTS/QXK/HKPCM/LF/VFCQ/KXCCX2CLCVPXGXF/YGNKX2CPCYNNWKN/TK/QXGK/GPKCP (**Figure 5).**

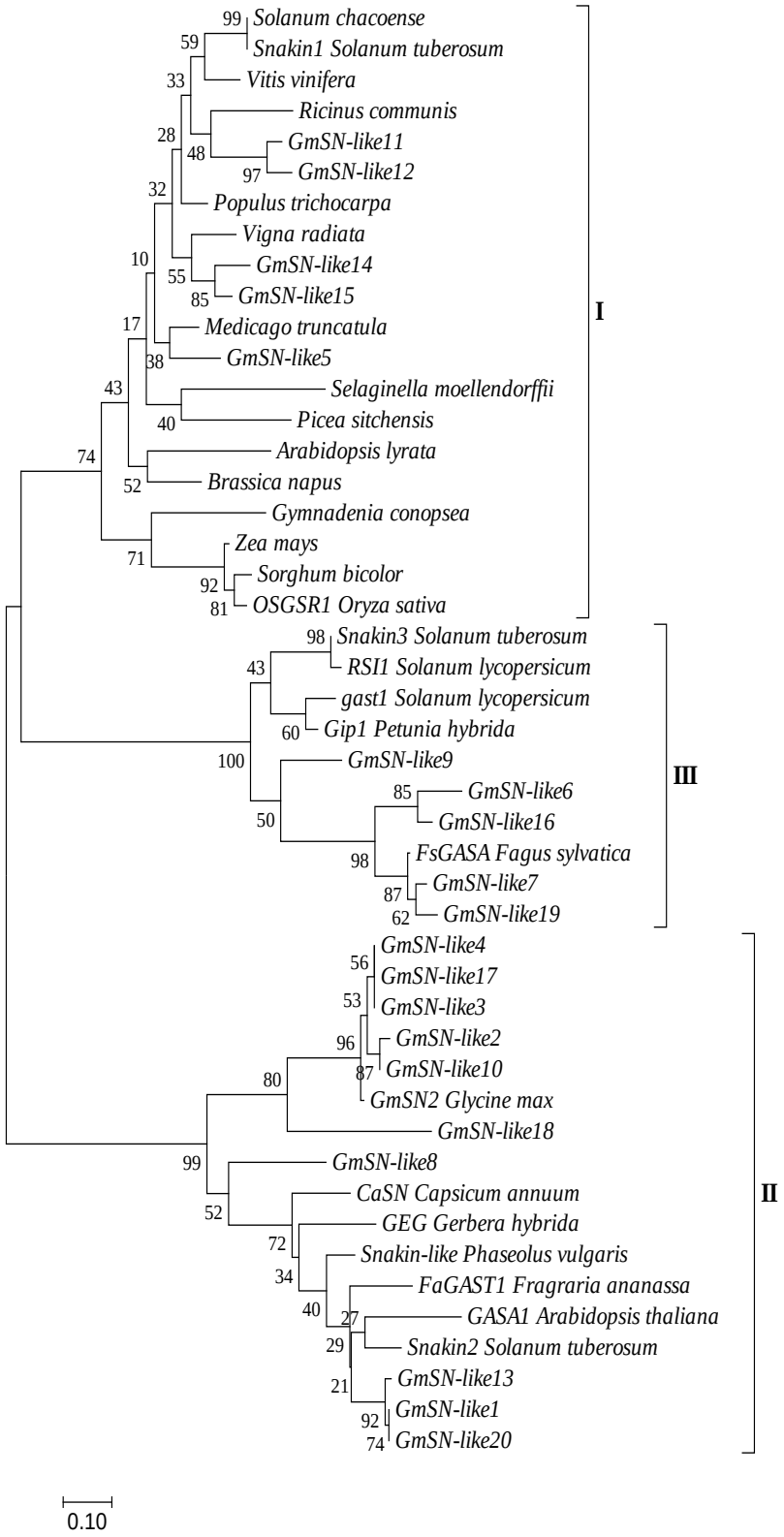


Figure 5 - Phenetic analysis of Snakin/GASA proteins including *Gm*SN2 and the GmSN-like peptides. The branches are supported by the bootstrap values showed in percent (%). The three subfamilies are evidenced by the bracket, at the right side. The scale length (0.1) indicates distance.

3.5. Genomic distribution and gene structure of snakin candidates

A tBLASTn search was carried out to investigate the distribution and abundance of *Gm*SN2 homologous in soybean virtual chromosomes and their homologs in other legumes. The chromosomal localization was plotted on a circular genome map with their predict gene structure which it was shown that possesses two, three or four exons (**Figure 6**).

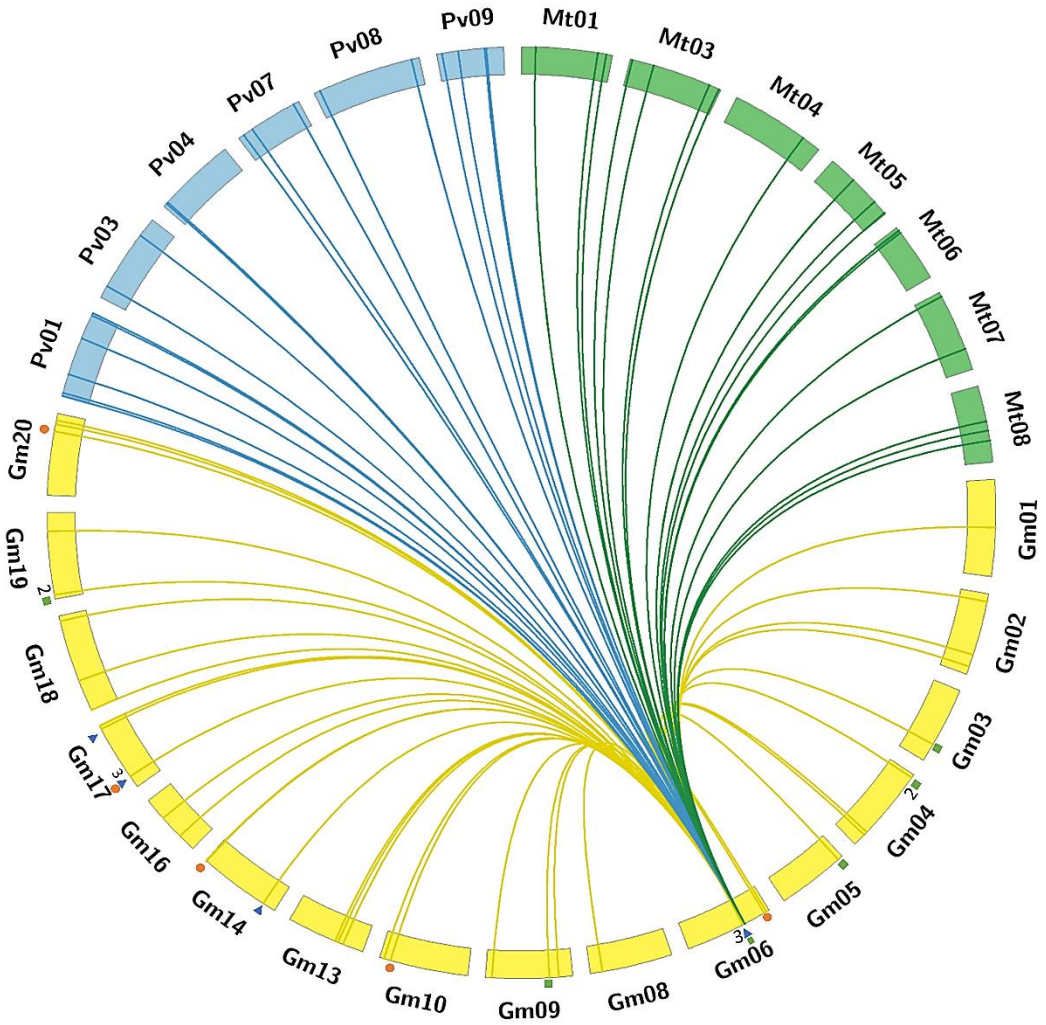


Figure 6 - Genomic distribution of *Gm*SN2 homologs in soybean (GM), *M. truncatula*, (Mt) and *P. vulgaris* (Pv). The orange circles, blue triangles and green squares indicate the soybean transcripts mapped and their gene structures with two, three and four exons respectively.

The anchoring based on identity and e-value showed that *GmSN2* has three exons and it is in the chromosome 6 with three links to mapped transcripts. Some transcripts may have the same position in the chromosomes due to alternative splicing variants. The

chromosomes Gm06 and Gm17 were the most representatives for the mapped transcripts with five links each, in the Gm06 all the gene structures were found, in total were mapped five genes with two exons and eight genes with three and four exons each (Supplementary Table S2). Counting all the links 41 homologues sequences had a heterogeneous distribution among 16 of the 20 chromosomes of soybean haploid set, 28 homologs in six of the 11 chromosomes in *P. vulgaris* and 26 distributed in seven of eight of the *M. truncatula* haploid chromosome set. In general, the distribution was diversified to all compared organisms with few clusters in soybean (e.g., Gm13 and Gm20), *P. vulgaris* (Pv01) and *M. truncatula* (Mt06).

3.6. Molecular Modelling

For molecular modeling, the snakin-1 (PDB ID: 5E5Q) was chosen as template. It shares 47% of identity with GmSN2 and from 50-74 % with the soybean transcripts (Supplementary Table S3). The three-dimensional modeling of GmSN2 showed an overall structure composed by two α-helix (residues ^{2}Gly to Ser^{10} and from ^{15}Pro to Cys^{29}), like an α-hairpin, a 3_{10} helix ^{40}REL^{42} and another α-helix in the C-terminal region composed by residues ^{43}GTCYTD^{48}. The disulfide bridges (**Figure 7A**, in yellow) showed a general pattern as follows: Cys1-7, Cys2-5, Cys3-4, Cys6-12, Cys8-11, and Cys9-10. In the Ramachandran plot for GmSN2, 91.5 % of the residues occurred in favored regions, 6.4 % in the additional allowed regions and 2.1 % in the generously allowed regions for GmSN2 (**Figure 7B**), the Z-score on ProSA (-3.79) indicated a valid model (**Figure 7C**). The variation of the Ramachandran plot to the different regions was from 91.3-96 % to the favored regions, 4-8.7 % to additional allowed regions and 0-2.1 % to generously allowed regions. Only the structure Contig3731 (GmSN2) had one amino acid in a disallowed region and the Z-Score varied from -5.64 to -3.72 (Supplementary Table S2).

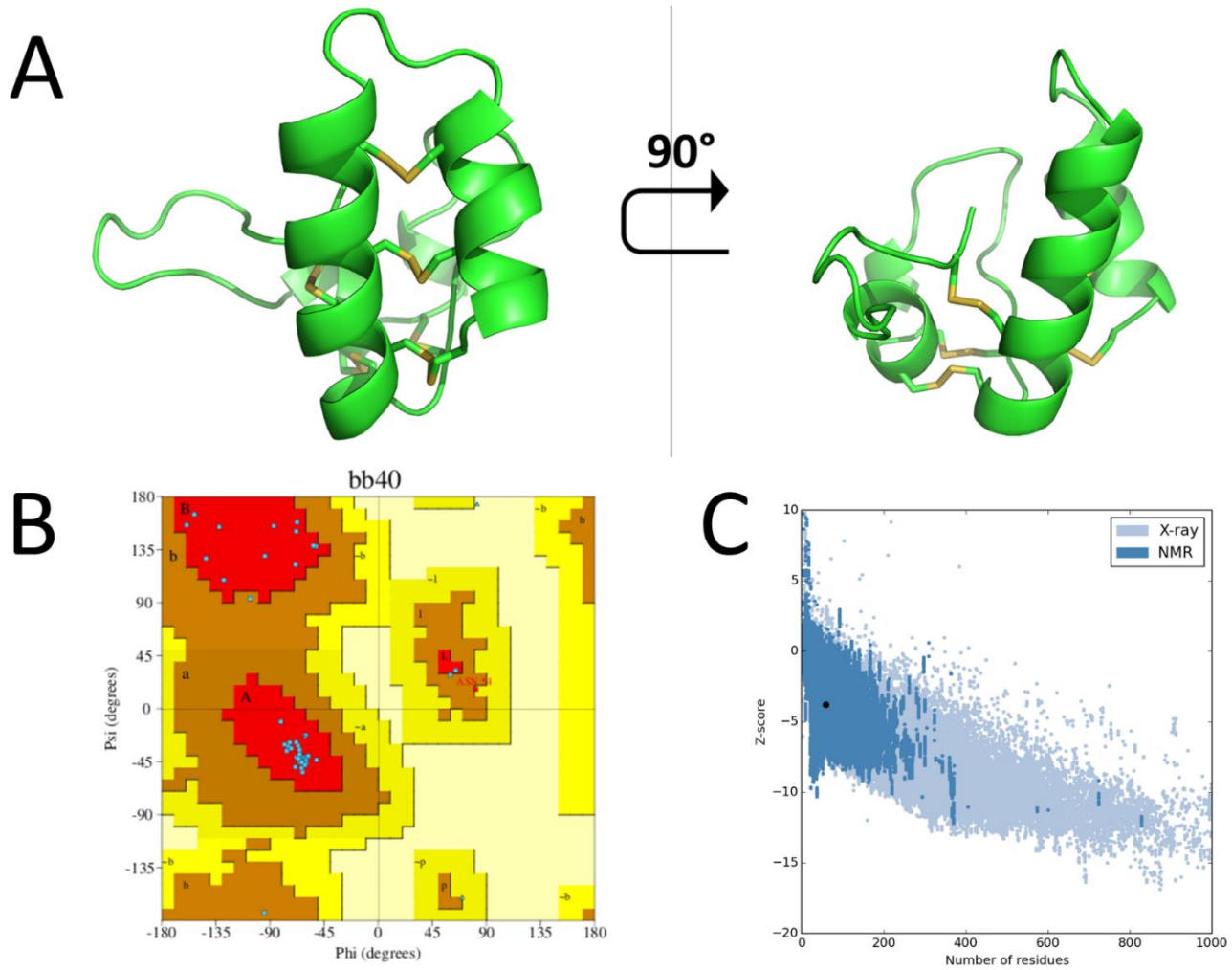


Figure 7 – GmSN2 structure showing the disulfide bonds (in yellow) (A). (B) Ramachandran plot where the blue dots are the aminoacids, the red regions are the most favored regions, in brown the additional allowed regions and in yellow the generously allowed regions and (C) Quality of folding showed by the Z-Score where the black spot is the GmSN2 model.

3.7. Expression analysis of GmSN2 in response to P. pachyrhizi

The RT-qPCR method was employed to analyze the expression profile changes of *Gm*SN2 in response to biotic stress (soybean, 48 h after inoculation with *P. pachyrhizi*). The primer pairs of the target and reference genes tested by melting curve and agarose gel electrophoresis revealed product specificity (data not shown). All primers (target and reference genes) showed amplification efficiency (E) higher than 90%, ranging from 90.25 to 96.06% for CYP2, CHLP, CWC15 and *Gm*SN2; and from 100.08 to 107.23% for ACT, TUA, ZMET, and 26P. The detection limit of the analyzed primer set (represented by the *y-intercept value*) ranged from 33.17 (CWC15) to 38.01 (CHLP) for the reference genes and 36.09 for *Gm*SN2 gene. The linear regression (R^2) analysis, in turn, revealed value ≥ 0.990, for all genes (see Supplementary Table S1).

In order to choose the best reference genes, geNorm software [48] was used. The most stable genes were CYP2 and 26P. The M value obtained for these two genes was 1.307 and 1.431, respectively (Supplementary Fig. 2A). Similarly, the pairwise variation ($V_{2/3}$) was 0.103 (Supplementary Fig. 2B)**,** that is, below the cut-off (0.15), suggesting the combination of CYP2 and 26P was adequate to normalize gene expression. However, in accordance with Vandesompele et al. (2002), we consider three reference genes as a minimal number for expression analysis. So, a third gene (ZMT) was used in order to normalize the relative expression of *Gm*SN2 **(**Supplementary Fig. 2).

After confirmation of the genomic localization and molecular characterization, the *Gm*SN2 gene was analyzed by RT-qPCR to explore the expression pattern of this antimicrobial peptide in soybean leaves (PI561356 accession) 48 hai (hours after inoculation) with *P. pachyrhizi*, in comparison with those in the mock-inoculated plants (control). The real-time PCR analysis revealed no change in the expression of *Gm*SN2 expression in the soybean leaves in response to biotic stimulus caused by the *P. pachyrhizi* (Supplementary Fig. 3)**.**

4. Discussion

The use of six Uniprot SNK homologs (from *S. tuberosum*, *A. thaliana*, *Ge. hybrida*, *Gy. conopsea* and *F. ananassa*) as seeds in the SeedServer allowed the identification of SNK candidates in different taxa, most of which (90%) with the expected structural features (Snakin/GASA domain). The search for homologs in the SeedServer

also permitted inferences about the possible origin of the SNK family in the Lycopodiophyta clade (Selaginellaceae), where the most basal SNK sequence was found (Supplementary Fig.1). It is likely that this family of antimicrobial peptides underwent events of duplication during evolution, being subsequently settled in upper clades, as it has been observed for other peptides [49,50], although no functional annotation to this gene was available in the searched databases.

The seeds allowed the identification of 20 putative SNKs in the soybean transcriptome, considering the adopted cut-off e^{-4}, plus the GmSN2 isolated. All sequences presented the complete GASA domain, shared by all SNK members, characterized by a cysteine-rich signature (C-terminal, 12 cysteine residues) (Figure 4), in highly conserved positions, which is essential for their biochemical activity, being also responsible for their protein structure [15,51–53], in addition 41 putative snakins were mapped in the genome, while *M. truncatula* and *P. vulgaris* showed 26 and 28 genes mapped respectively, experimental data of Nahirñaki et al 2016 showed 16 snakin members in potato, this data confirms that soybean and the other legumes are important sources to study this gene, being soybean, so far, the organism that has more predict snakins mapped in the genome than any other.

…

Regarding the subcellular localization, our data showed that all putative soybean SNKs were addressed to the extracellular environment, a role similar to *GASA5*, that acts as a regulator of flowering time and stem growth [54,55]. Moreover, it is known that SNK genes are also involved in other signaling pathways, i.e., hormonal [9] and stress response, as already reported for other AMPs [56]. Considering this evidence, a putative divergence of roles may be recognized in these peptides, depending on plant physiological state and/or stress condition imposed. However, all the peptides retrieved reported putative antimicrobial activity as it was already reported for other SNKs. Thus, the soybean putative SNK member emerged as an important candidate to study the antimicrobial activity of these peptides like He et al. 2017 showed to the GmSN1.

Although a high similarity among these peptides with “unknown” and “uncharacterized” sequences from soybean has been recognized, they exhibited conservation in structure, size, molecular weight, and pI very close to those found in other members of the family [10,57]. Besides the most conserved region described for these peptides, the C-terminal domain (GASA) [58] demonstrates strict conservation, which

can be a strong evidence of functional specialization, also highlighting their importance to the maintenance of the plant homeostasis.

Soybean SNK candidates were found in all the reads libraries, prevailing in tissues at early developmental stages (leaves, roots, seeds, and cotyledons) counting 29 different situations in total, this is the most complete *in silico* profile for this gene so far, where the cellular activity and the action of effectors that regulate the cell division (as gibberellin) are stronger. This pattern was also observed for other SNK members [59]. This fact may indicate that soybean SNK in some cases presented here follows the same model, as it was observed for homologs involved in shoot elongation [13], corolla and carpel [16], as well as in seed and root development [21,60]. Besides, it should be highlighted that these tissues are also prime targets of many pathogens and nematodes [7,10]. Therefore, the expression of antimicrobial peptides, like SNKs, have been associated as part of the permanent and inducible defense against microbial assault [61–63], in addition we also found that they may be involved in responses related to wounded tissues, from early from to late stages.

The analysis of DeepSuperSAGE libraries revealed a significant number of tags (81) with homology to SNKs. This may indicate the critical role of SNK in response to both biotic and abiotic stresses in soybean. As expected, most DeepSuperSAGE unitags (45 unitags) were identified in the biotic stress condition (after rust fungus inoculation). It is known that rust fungus often infects leaves, a tissue which has a ubiquitous expression of SNKs, both constitutive and induced (Balaji and Smart 2012; Nahirñak et al. 2012). However, to our knowledge, there are no reports of SNK transcripts associated with this pathogen. In concordance with the constitutive expression observed of *St*SN1 [6,64] in potato, as well as, *Ms*SN1 in alfalfa (Garcia et al, 2014), the expression pattern of *Gm*SN2, obtained by RT-qPCR, shows to be a component of the constitutive defense barrier like Snakin-1 from *S. tuberosum* [6].

Regarding libraries submitted to abiotic stress (root dehydration), the considerable number of tags may indicate constitutive expression and/or crosstalk responses, since water deficit causes osmotic and oxidative disturbs, triggering responses via reactive oxygen species (ROS), reflecting in the induced expression of SNKs, as reported for *A. thaliana*, where the GASA14 modulates the accumulation of ROS. It was also known that members of this family, such as StSN2, GIP4, and GIP5, respond to effectors responsive to this stress as the hormone ABA [9]. Nevertheless, most of these transcripts are downregulated, indicating that the factors involved in growth tend to be

suppressed in this situation [65]. In soybean the snakin play roles in many different situations, the reads counting data showed a diverse pattern of expression for stages of development, treatments and tissues, however the DeepSuperSAGE evidenced that they are related more to biotic stresses by fungus even, they are related as well to abiotic stresses.

The isolated *Gm*SN2 exhibits a structure like other typical SNK members: the cysteine motifs, molecular weight, isoelectric point, subcellular localization and the structure of the peptide. Thus, it is plausible to affirm that it is a new cysteine-rich antimicrobial peptide of *G. max*. Based on the amino acid conservation, the phenetic tree confirmed the division of SNK family into three subfamilies, as proposed by Berrocal-lobo et al. (2002). According to this classification, *Gm*SN2 belongs to subfamily II, which includes GEG (*G. hybrida*), FaGAST1 (*F. ananassa*), GASA1 (*A. thaliana*) and Snakin-2 (*S. tuberosum*). This C-terminal conservation (especially for the cysteine residues) might be required not only for the maintenance of the 3D structure, as it was shown in our data, but also to the interaction with other proteins [66]

In comparison to the genome-wide analysis performed in potato (where 16 SNK family members were identified in nine of the 12 chromosomes by Nahirñak et al. (2016) or rice (which has nine homologs distributed over six of the 12 chromosomes according to Zimmermann et al., 2010), our study identified 40 SNK homologs to *Gm*SN2 in the genome using *in silico* approaches, 26 in *M. truncatula* and 28 in *P. vulgaris* the pattern of distribution is very distributed and there is no clear relation between the distribution of the subfamilies or the gene structure once one chromosome like Gm06 has all the structures found for this gene (two, three or four exons) and has snakin genes form different subfamiles. The higher amount of genes can be justified by the paleotetraploid condition of the soybean genome (2n = 40 chromosomes), also identified previously for other gene families [68,69].

Nevertheless, the genomic distribution of the SNK members follows a similar pattern among different organisms, being present in an isolated form or with few neighboring SNK members, presenting, therefore, a distinct distribution as compared to other AMPs that are usually present in gene clusters [70]. However, the positions in the chromosomes may suggest that they are conserved among syntenic blocks in the compared genomes, and since the positions of the links are similar among them it is likely to conjecture that duplication and genomic rearrangement events lead the distribution of this family, similar events are known in the R genes [69,71,72].

The identified soybean candidates exhibited three gene structures related to the SNK family: two, three and four exons [67], regarding probably neofunctionalized candidates (paralogs). This structural variation reflects their functional scope, and this is shown by the broad spectrum of activities like development, biotic and abiotic stress [62,73]. Thus, soybean arises as a promising organism to study these genes, since its genome comprises significant structural and functional variations.

The overall structure of *Gm*SN2 and the other putative soybean peptides show similarities with plant thionins and α-helical hairpin [74,75], peptides that share a hairpin stabilized by disulfide bridges. This reinforces the Silverstein's hypothesis which proposes that some cysteine-rich AMPs have a common ancestor, based on structural features, especially the cysteine pattern and connectivity [2].

5. Conclusions

The results from this investigation were able to indicate 20 transcripts retrieved from Genosoja and also 40 links to snakin in the soybean genome, this amount indicates the highest number of putative genes from this family tagged in a single organism, in addition we demonstrated that their distribution is heterogeneous not following the pattern for defense genes, gene structure that has two, three or four exons or subfamily divisions that are divided by different sequence motifs. The snakin/GASA peptides in *G. max* are related to biotic and abiotic stresses, being induced or constitutively expressed, moreover they are also related to the basal metabolism in both early and late stages of the plant

The isolated GmSN2 is a new member of this subfamily in soybean, this gene is constitutively expressed against the phytopathogen *P. pachyrhizi*. The gene was mapped in the chromosome 6 and has three exons in its genic structure, in addition we mapped potential paralogues in *M. truncatula* and *P. vulgaris.* The peptide structure is composed by an alpha-hairpin and two small helixes all stabilized by six disulfide bridges, this may corroborate the hypothesis of common ancestrality among the cysteine rich AMPs.

The SNK candidates retrieved from GENOSOJA database are related to biotic and abiotic stresses, but also to the developmental stages, furthermore, the comparative distribution of the homologs in soybean, *M. truncatula,* and *P. vulgaris* evidenced that even for a defense related gene they are not distributed like them. Regarding the new

GmSN2, our study provided significant elucidations from the structural point of view to the functional, since it possesses structural and potential functional aspects that mayshares structural and potential functional aspects with other AMPs. This study is an important source to map potential defense genes in soybean with biotechnological importance.

References

[1] F. García-Olmedo, a Molina, J.M. Alamillo, P. Rodríguez-Palenzuéla, Plant defense peptides., Biopolymers. 47 (1998) 479–91. doi:10.1002/(SICI)1097-0282(1998)47:6<479::AID-BIP6>3.0.CO;2-K.

[2] K. a T. Silverstein, W. a Moskal, H.C. Wu, B. a Underwood, M. a Graham, C.D. Town, K. a VandenBosch, Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants., Plant J. 51 (2007) 262–80. doi:10.1111/j.1365-313X.2007.03136.x.

[3] A.M. Benko-Iseppon, S.L. Galdino, T. Calsa, E.A. Kido, A. Tossi, L.C. Belarmino, S. Crovella, Overview on plant antimicrobial peptides., Curr. Protein Pept. Sci. 11 (2010) 181–8. http://www.ncbi.nlm.nih.gov/pubmed/20088772.

[4] R. Nawrot, J. Barylski, G. Nowicki, J. Broniarczyk, W. Buchwald, A. Goździcka-Józefiak, Plant antimicrobial peptides., Folia Microbiol. (Praha). 59 (2014) 181–96. doi:10.1007/s12223-013-0280-4.

[5] M. Berrocal-Lobo, A. Segura, M. Moreno, G. Lopez, F. Garcia-Olmedo, A. Molina, Snakin-2, an Antimicrobial Peptide from Potato Whose Gene Is Locally Induced by Wounding and Responds to Pathogen Infection, PLANT Physiol. 128 (2002) 951–961. doi:10.1104/pp.010685.

[6] A. Segura, M. Moreno, F. Madueño, A. Molina, F. García-olmedo, Snakin-1 , a peptide from potato that is active against plant pathogens, Mol. Plant-Microbe Interact. 12 (1999) 16–23.

[7] M. Berrocal-lobo, A. Segura, M. Moreno, G. López, F. Garcı, A. Molina, Snakin-2 , an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection, Plant Physiol. 128 (2002) 951–961. doi:10.1104/pp.010685.1.

[8] H. Yeung, C.J. Squire, Y. Yosaatmadja, S. Panjikar, G. López, A. Molina, E.N. Baker, P.W.R. Harris, M.A. Brimble, Radiation damage and racemic protein crystallography reveal the unique structure of the GASA/Snakin protein superfamily, Angew. Chemie Int. Ed. (2016) 1–5. doi:10.1002/anie.201602719.

[9] V. Nahirñak, N.I. Almasia, H.E. Hopp, C. Vazquez-rovere, Involvement in hormone crosstalk and redox homeostasis Snakin / GASA proteins, Plant Signal. Behav. 7 (2012) 1004–1008.

[10] Z. Mao, J. Zheng, Y. Wang, G. Chen, Y. Yang, D. Feng, B. Xie, The new CaSn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper, Phytoparasitica. 39 (2011) 151–164. doi:10.1007/s12600-011-0149-5.

[11] M. Oliveira-Lima, A.M. Benko-Iseppon, J.R.C.F. Neto, S. Rodríguez-Decuadro, E.A. Kido, S. Crovella, V. Pandolfi, Snakin: Structure, roles and applications of a plant antimicrobial peptide., Curr. Protein Pept. Sci. 18 (2017) 1–7. doi:10.2174/1389203717666160619183140.

[12] V. Herbel, M. Wink, Mode of action and membrane specificity of the antimicrobial peptide snakin-2, PeerJ. 4 (2016) 1–13. doi:10.7717/peerj.1987.

[13] L. Shi, R.T. Gast, M. Gopalraj, N.E. Olszewski, Characterization of a shoot-specific, GA3- and ABA-regulated gene from tomato, Plant J. 2 (1992) 153–159.

[14] G. Ben-Nissan, D. Weiss, The petunia homologue of tomato gast1: transcript accumulation coincides with gibberellin-induced corolla cell elongation., Plant Mol. Biol. 32 (1996) 1067–74. http://www.ncbi.nlm.nih.gov/pubmed/9002605.

[15] D. Aubert, M. Chevillard, A.M. Dorne, G. Arlaud, M. Herzog, Expression patterns of GASA genes in Arabidopsis thaliana: the GASA4 gene is up-

regulated by gibberellins in meristematic regions., Plant Mol. Biol. 36 (1998) 871–83. http://www.ncbi.nlm.nih.gov/pubmed/9520278.

[16] M. Kotilainen, Y. Helariutta, M. Mehto, E. Pollanen, V. Albert, P. Elomaa, T. Teeri, GEG participates in the regulation of cell and organ shape during corolla and carpel development in gerbera hybrida, Plant Cell. 11 (1999) 1093–104. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=144246&tool=pmcentrez&rendertype=abstract.

[17] J.I. de la Fuente, I. Amaya, C. Castillejo, J.F. Sánchez-Sevilla, M. a Quesada, M. a Botella, V. Valpuesta, The strawberry gene FaGAST affects plant growth through inhibition of cell elongation., J. Exp. Bot. 57 (2006) 2401–11. doi:10.1093/jxb/erj213.

[18] L. V Bindschedler, J.P. Whitelegge, D.J. Millar, G.P. Bolwell, A two component chitin-binding protein from French bean -- association of a proline-rich protein with a cysteine-rich polypeptide., FEBS Lett. 580 (2006) 1541–6. doi:10.1016/j.febslet.2006.01.079.

[19] L. Wang, Z. Wang, Y. Xu, S.-H. Joo, S.-K. Kim, Z. Xue, Z. Xu, Z. Wang, K. Chong, OsGSR1 is involved in crosstalk between gibberellins and brassinosteroids in rice., Plant J. 57 (2009) 498–510. doi:10.1111/j.1365-313X.2008.03707.x.

[20] A. Alonso-Ramírez, D. Rodríguez, D. Reyes, J.A. Jiménez, G. Nicolás, M. López-Climent, A. Gómez-Cadenas, C. Nicolás, Evidence for a role of Gibberellins in Salicylic Acid-modulated early plant responses to abiotic stress in Arabidopsis Seeds, Plant Physiol. 150 (2009) 1335–1344. doi:10.1104/pp.109.139352.

[21] R. Zimmermann, H. Sakai, F. Hochholdinger, The Gibberellic Acid Stimulated-Like gene family in maize and its role in lateral root development., Plant Physiol. 152 (2010) 356–65. doi:10.1104/pp.109.149054.

[22] A. García, N. Ayub, A. Fox, M. Gómez, M. Diéguez, E. Pagano, C. Berini, J. Muschietti, G. Soto, Alfalfa snakin-1 prevents fungal colonization and probably coevolved with rhizobia., BMC Plant Biol. 14 (2014) 248. doi:10.1186/s12870-014-0248-9.

[23] T. Ke, H. Cao, J. Huang, F. Hu, J. Huang, C. Dong, X. Ma, J. Yu, H. Mao, X. Wang, Q. Niu, F. Hui, S. Liu, EST-based in silico identification and in vitro test of antimicrobial peptides in Brassica napus, BMC Genomics. 16 (2015) 653. doi:10.1186/s12864-015-1849-x.

[24] K.-L. Li, X. Bai, Y. Li, H. Cai, W. Ji, L.-L. Tang, Y.-D. Wen, Y.-M. Zhu, GsGASA1 mediated root growth inhibition in response to chronic cold stress is marked by the accumulation of DELLAs., J. Plant Physiol. 168 (2011) 2153–60. doi:10.1016/j.jplph.2011.07.006.

[25] H. He, X. Yang, H. Xun, X. Lou, S. Li, Z. Zhang, L. Jiang, Y. Dong, S. Wang, J. Pang, B. Liu, Over-expression of GmSN1 enhances virus resistance in Arabidopsis and soybean, Plant Cell Rep. (2017). doi:10.1007/s00299-017-2167-3.

[26] A.M. Benko-Iseppon, A.L. Nepomuceno, R. V Abdelnoor, GENOSOJA - The Brazilian Soybean Genome Consortium: High throughput omics and beyond., Genet. Mol. Biol. 35 (2012) i–iv. doi:10.1590/S1415-47572012000200001.

[27] L.C. do Nascimento, G.G.L. Costa, E. Binneck, G.A.G. Pereira, M.F. Carazzolle, A web-based bioinformatics interface applied to the GENOSOJA Project: Databases and pipelines., Genet. Mol. Biol. 35 (2012) 203–11. doi:10.1590/S1415-47572012000200002.

[28] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs., Nucleic Acids Res. 25 (1997) 3389–402. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=146917&tool=pmcentrez&rendertype=abstract.

[29] R. Hammami, J. Ben Hamida, G. Vergoten, I. Fliss, PhytAMP: a database dedicated to antimicrobial plant peptides., Nucleic Acids Res. 37 (2009) D963-8. doi:10.1093/nar/gkn655.

[30] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M. V Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R. a Yamashita, D. Zhang, N. Zhang, C. Zheng, S.H. Bryant, CDD: a Conserved Domain Database for the functional annotation of proteins., Nucleic Acids Res. 39 (2011) D225-9. doi:10.1093/nar/gkq1189.

[31] E.W. Klee, L.B.M. Ellis, Evaluating eukaryotic secreted protein prediction., BMC Bioinformatics. 6 (2005) 256. doi:10.1186/1471-2105-6-256.

[32] F.H. Waghu, R.S. Barai, P. Gurung, S. Idicula-Thomas, CAMPR3: A database on sequences, structures and signatures of antimicrobial peptides, Nucleic Acids Res. 44 (2016) D1094–D1097. doi:10.1093/nar/gkv1051.

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

[34] K. Hiller, A. Grote, M. Maneck, R. Münch, D. Jahn, JVirGel 2.0: computational prediction of proteomes separated via two-dimensional gel electrophoresis under consideration of membrane and secreted proteins., Bioinformatics. 22 (2006) 2441–3. doi:10.1093/bioinformatics/btl409.

[35] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns., Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 14863–8. doi:10.1073/pnas.95.25.14863.

[36] A.J. Saldanha, Java Treeview--extensible visualization of microarray data., Bioinformatics. 20 (2004) 3246–8. doi:10.1093/bioinformatics/bth349.

[37] D.M. Goodstein, S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D.S. Rokhsar, Phytozome: a comparative platform for green plant genomics., Nucleic Acids Res. 40 (2012) D1178-86. doi:10.1093/nar/gkr944.

[38] M. a Larkin, G. Blackshields, N.P. Brown, R. Chenna, P. a McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, a Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0., Bioinformatics. 23 (2007) 2947–8. doi:10.1093/bioinformatics/btm404.

[39] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729. doi:10.1093/molbev/mst197.

[40] M. Dayhoff, R. Schwartz, A Model of evolutionary change in proteins, Atlas Protein Seq. Struct. (1978) 345–352. doi:10.1.1.145.4315.

[41] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, S.J. Jones, M.A. Marra, Circos : an Information Aesthetic for Comparative Genomics, Genome Res. 9 (2009) 1639–1645.

[42] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank., Nucleic Acids Res. 28 (2000)

235–242. doi:10.1093/nar/28.1.235.
[43] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using MODELLER, Curr. Protoc. Protein Sci. (2007) 1–31. doi:10.1002/0471140864.ps0209s50.
[44] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) 407–410. doi:10.1093/nar/gkm290.
[45] R.A. Laskowski, J.D. Watson, J.M. Thornton, ProFunc: A server for predicting protein function from 3D structure, Nucleic Acids Res. 33 (2005) 89–93. doi:10.1093/nar/gki414.
[46] F.R. Kulcheski, F.C. Marcelino-Guimaraes, A.L. Nepomuceno, R.V. Abdelnoor, R. Margis, The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean, Anal. Biochem. 406 (2010) 185–192. doi:10.1016/j.ab.2010.07.020.
[47] M.W. Pfaffl, G.W. Horgan, L. Dempfle, Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR., Nucleic Acids Res. 30 (2002) e36. doi:10.1093/nar/30.9.e36.
[48] J. Vandesompele, K. De Preter, ilip Pattyn, B. Poppe, N. Van Roy, A. De Paepe, rank Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002) 34–1. doi:10.1186/gb-2002-3-7-research0034.
[49] O.L. Franco, Peptide promiscuity: an evolutionary concept for plant defense., FEBS Lett. 585 (2011) 995–1000. doi:10.1016/j.febslet.2011.03.008.
[50] D.A. Phoenix, S.R. Dennison, F. Harris, Antimicrobial Peptides : Their History , Evolution , and Functional Promiscuity, in: Antimicrob. Pept., 1st ed., Weinheim, 2013: pp. 1–37.
[51] M. Herzog, a M. Dorne, F. Grellet, GASA, a gibberellin-regulated gene family from Arabidopsis thaliana related to the tomato GAST1 gene., Plant Mol. Biol. 27 (1995) 743–52. http://www.ncbi.nlm.nih.gov/pubmed/7727751.
[52] P. Jianzong, L.A.I. Liujing, W. Xiaojing, PRGL : A cell wall proline-rich protein containning GASA domain in Gerbera hybrida, 51 (2008) 520–525. doi:10.1007/s11427-008-0067-z.
[53] S. Sun, H. Wang, H. Yu, C. Zhong, X. Zhang, J. Peng, X. Wang, GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation., J. Exp. Bot. 64 (2013) 1637–1647. doi:10.1093/jxb/ert021.
[54] S. Zhang, C. Yang, J. Peng, S. Sun, X. Wang, GASA5, a regulator of flowering time and stem growth in Arabidopsis thaliana., Plant Mol. Biol. 69 (2009) 745–59. doi:10.1007/s11103-009-9452-7.
[55] E. Moyano-Cañete, M.L. Bellido, N. García-Caparrós, L. Medina-Puche, F. Amil-Ruiz, J. a González-Reyes, J.L. Caballero, J. Muñoz-Blanco, R. Blanco-Portales, FaGAST2, a strawberry ripening-related gene, acts together with FaGAST1 to determine cell size of the fruit receptacle., Plant Cell Physiol. 54 (2013) 218–36. doi:10.1093/pcp/pcs167.
[56] S. V Oard, F.M. Enright, Expression of the antimicrobial peptides in plants to control phytopathogenic bacteria and fungi., Plant Cell Rep. 25 (2006) 561–72. doi:10.1007/s00299-005-0102-5.
[57] C.-B. Ko, Y.-M. Woo, D.J. Lee, M.-C. Lee, C.S. Kim, Enhanced tolerance to

heat stress in transgenic plants expressing the GASA4 gene., Plant Physiol. Biochem. 45 (2007) 722–8. doi:10.1016/j.plaphy.2007.07.010.

[58] L. Padovan, M. Scocchi, A. Tossi, Structural aspects of plant antimicrobial peptides., Curr. Protein Pept. Sci. 11 (2010) 210–9. http://www.ncbi.nlm.nih.gov/pubmed/20088769.

[59] V. Nahirñak, N.I. Almasia, P.V. Fernandez, H.E. Hopp, J.M. Estevez, F. Carrari, C. Vazquez-Rovere, Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition., Plant Physiol. 158 (2012) 252–63. doi:10.1104/pp.111.186544.

[60] I. Roxrud, S.E. Lid, J.C. Fletcher, E.D.L. Schmidt, H.-G. Opsahl-Sorteberg, GASA4, one of the 14-member Arabidopsis GASA family of small polypeptides, regulates flowering and seed development., Plant Cell Physiol. 48 (2007) 471–83. doi:10.1093/pcp/pcm016.

[61] P. Veronese, M.T. Ruiz, M.A. Coca, A. Hernandez-lopez, H. Lee, I. Ibeas, B. Damsz, M. Pardo, P.M. Hasegawa, R.A. Bressan, M.L. Narasimhan, In Defense against Pathogens . Both Plant Sentinels and Foot Soldiers Need to Know the Enemy, Plant Physiol. 131 (2003) 1580–1590. doi:10.1104/pp.102.013417.1580.

[62] N.I. Almasia, A.A. Bazzini, H.E. Hopp, C. Vazquez-Rovere, Overexpression of snakin-1 gene enhances resistance to Rhizoctonia solani and Erwinia carotovora in transgenic potato plants, Mol. Plant Pathol. 9 (2008) 329–338. doi:10.1111/J.1364-3703.2008.00469.X.

[63] V. Balaji, C.D. Smart, Over-expression of snakin-2 and extensin-like protein genes restricts pathogen invasiveness and enhances tolerance to Clavibacter michiganensis subsp. michiganensis in transgenic tomato (Solanum lycopersicum)., Transgenic Res. 21 (2012) 23–37. doi:10.1007/s11248-011-9506-x.

[64] S. Meiyalaghan, S.J. Thomson, M.W. Fiers, P.J. Barrell, J.M. Latimer, S. Mohan, E.E. Jones, A.J. Conner, J.M. Jacobs, Structure and expression of GSL1 and GSL2 genes encoding gibberellin stimulated-like proteins in diploid and highly heterozygous tetraploid potato reveals their highly conserved and essential status., BMC Genomics. 15 (2014) 1–16. doi:10.1186/1471-2164-15-2.

[65] A. Blum, Plant Water Relations, Plant Stress and Plant Production, in: Plant Breed. Water-Limited Environ., Springer New York, New York, NY, 2011: pp. 11–52. doi:10.1007/978-1-4419-7491-4.

[66] G. Ben-Nissan, J.-Y. Lee, A. Borohov, D. Weiss, GIP, a Petunia hybrida GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering, Plant J. 37 (2004) 229–238. doi:10.1046/j.1365-313X.2003.01950.x.

[67] V. Nahirñak, M. Rivarola, M. Gonzalez de Urreta, N. Paniego, H.E. Hopp, N.I. Almasia, C. Vazquez-Rovere, Genome-wide Analysis of the Snakin/GASA Gene Family in Solanum tuberosum cv. Kennebec, Am. J. Potato Res. (2016) 172–188. doi:10.1007/s12230-016-9494-8.

[68] N.M. Soares-Cavalcanti, L.C. Belarmino, E. a Kido, V. Pandolfi, F.C. Marcelino-Guimarães, F. a Rodrigues, G. a G. Pereira, A.M. Benko-Iseppon, Overall picture of expressed Heat Shock Factors in Glycine max, Lotus japonicus and Medicago truncatula., Genet. Mol. Biol. 35 (2012) 247–59. doi:10.1590/S1415-47572012000200006.

[69] A.C. Wanderley-nogueira, L.C. Belarmino, N.M. Soares-cavalcanti, J.P. Bezerra-neto, A.M. Benko-iseppon, An overall evaluation of the Resistance (R) and Pathogenesis-Related (PR) superfamilies in soybean , as compared with Medicago and Arabidopsis, Genet. Mol. Biol. 35 (2012) 260–271.

[70] E. Marshall, L.M. Costa, J. Gutierrez-Marcos, Cysteine-Rich Peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development, J. Exp. Bot. 62 (2011) 1677–1686. doi:10.1093/jxb/err002.
[71] H. Zhu, S.B. Cannon, N.D. Young, D.R. Cook, Phylogeny and genomic organization of the TIR and non-tIR NBS-LRR resistance gene family in Medicago truncatula., Mol. Plant. Microbe. Interact. 15 (2002) 529–539. doi:10.1094/MPMI.2002.15.6.529.
[72] Y.J. Kang, K.H. Kim, S. Shim, M.Y. Yoon, S. Sun, M.Y. Kim, K. Van, S.-H. Lee, Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean., BMC Plant Biol. 12 (2012) 139–152. doi:10.1186/1471-2229-12-139.
[73] S. Lee, S. Han, S. Kim, Salt- and ABA-inducible OsGASR1 is Involved in Salt Tolerance, (2015) 96–101. doi:10.1007/s12374-014-0497-z.
[74] S. Romagnoli, R. Ugolini, F. Fogolari, G. Schaller, K. Urech, M. Giannattasio, L. Ragona, H. Molinari, NMR structural determination of viscotoxin A3 from Viscum album L., Biochem. J. 350 Pt 2 (2000) 569–77. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1221286&tool=pmcentrez&rendertype=abstract.
[75] M. Fujimura, M. Ideguchi, Y. Miname, K. Watanabe, K. Tadera, Purification, Characterization, and Sequencing of Novel Antimicrobial Peptides, Tu -AMP 1 and Tu -AMP 2, from Bulbs of Tulip (Tulipa gesneriana L.), Biosci. Biotechnol. Biochem. 68 (2004) 571–577. doi:10.1271/bbb.68.571.

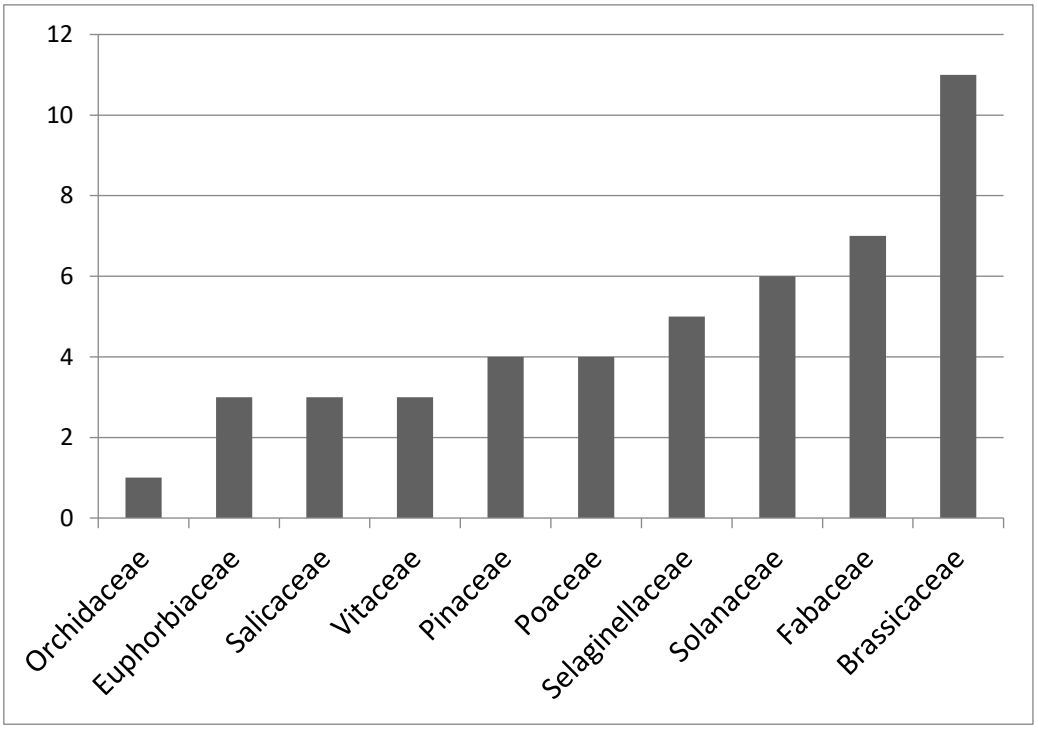
SUPPLEMENTARY MATERIAL

Table S1. Primer sequences of the seven candidate reference genes and of the gene of interest (*Gm*SN2) used in the RT- qPCR analysis

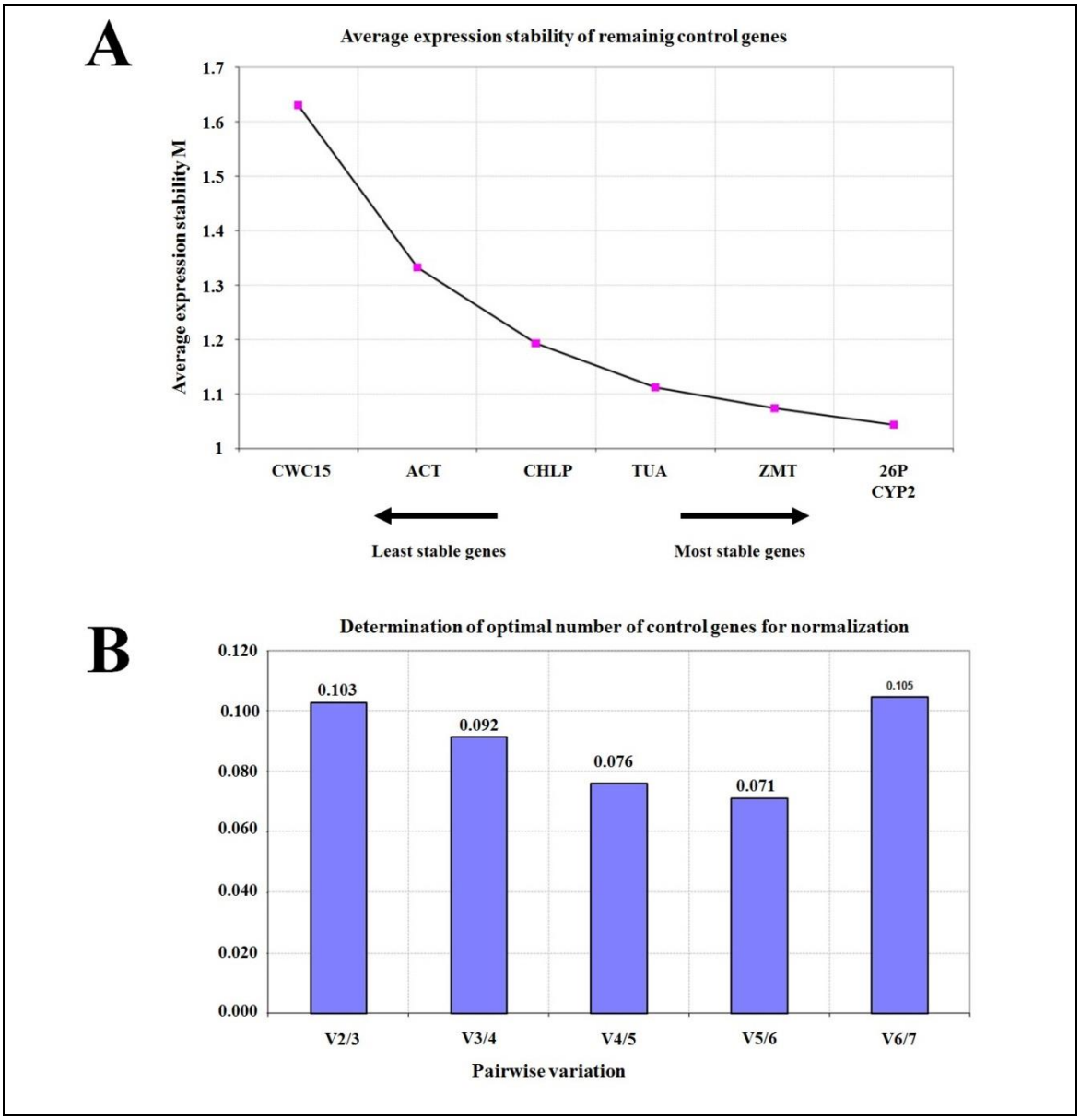
Gene Name (Acronym)	Accession Number	Unigene ID	Gene (Function)	Primers Sequence (5'-3')	Amplicon (bp)	Efficiency (%)	R^2
Actin (*ACT 11)*	BW652479	Gma.32186	Cytoskeletal structural protein	(F) CGGTGGTTCTATCTTGGCATC (R) GTCTTTCGCTTCAATAACCCTA	144	100.08	0.989
Alpha Tubulin *(TUA5)*	CA801144.1	Gma.13580	Structural constituent of cytoskeleton	(F) AGGTCGGAAACTCCTGCTGG (R) AAGGTGTTGAAGGCGTCGTG	159	107.23	0.999
Zinc metallopeptidase (*ZMET)*	XM_003521127.3	Gma.7635	Unknown	(F) GCAACCAACCTTTCATCAGC (R) GCCTCGACCCTTTGCTCAAT	150	102.65	0.998
26S proteasome *(26P)*	XM_003530886.3	Gma.59029	Protein degradation	(F) ATGGCTGTGGATGAGGAACC (R) TCAAGTGGGCAACAGAGCAG	169	107.23	0.999
Geranylgeranyl diphosphate reductase *(CHLP)*	XM_003524728.3	Gma.1279	Reduction of geranylgeranyl diphosphate to phytyl diphosphate	(F) GAGTCGCCAAGTCTATTGAT (R) CCGTAGAAATCTGGAGAAAC	145	96.06	0.998
Spliceosome-associated protein *(CWC15)*	XM_003549867.3 XP_003549915.1	Glyma.17G141500	Spliceosome-associated protein CWC15	(F) GCTCAAAGTAAAGGAGGCAGAGC (R) CCACGGGCTTGGTTCTTAAAC	120	96.06	0.999
cytochromes P4502 gene family *(CYP2)*	CF806591	Gma.31599	Biosynthetic and detox pathways	(F) CGGGACCAGTGTGCTTCTTCA (R) CCCCTCCACTACAAAGGCTCG	154	90.25	0.999
GmSN2			*G. max* Snakin-2	(F) AGACGCGAACAGAAGGCTAA (R) TGCACCTTGTCTTGCACAAT	215	96.06	0.999

R^2, Regression coefficient

Supplementary Fig.1 Taxonomic distribution of representative gene-candidates of Snakin/GASA family identified using SeedServer. The y-axis represents the number of genes, and the x-axis represents the plant family.



Supplementary Fig. 2 - geNorm analysis, indicating the average expression stability (M value) (A) and pairwise variation (B), calculated by pair-wise variation analysis between normalization factors (NF_n and NF_{n+1}).



Supplementary Fig. 3 - Expression profile of GmSN2 gene from soybean plants (PI561356 accession) in response to *Phakopsora pachyrhizi*, 48 hai (hours after inoculation) in comparison with those in the mock-inoculated plants (control). Soybean genes encoding CYP2 (cytochromes P4502 gene family, 26P (26S proteasome) and ZMT (Zinc metallopeptidase) were used as reference genes. The error bars represent the set of three independent biological replicates and two technical repeats.

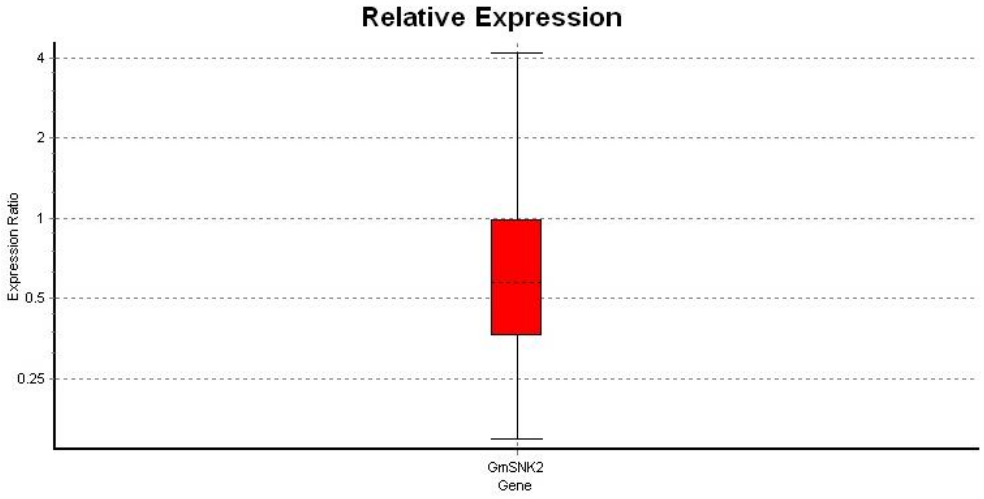


Table S2. Soybean transcripts anchored in the virtual chromosomes

Transcript	Chromosome	Identity (%)	Start	End	Evalue	Exons
GmSN2	6	95.361	16099119	16098926	1.08E-82	3
GmSN-like1	17	100.000	41223285	41222787	0.00E+00	3
GmSN-like2	17	100.000	7253935	7253516	0.00E+00	3
GmSN-like3	6	99.789	16099107	16098634	0.00E+00	3
GmSN-like4	6	99.859	16099369	16098660	0.00E+00	3
GmSN-like5	6	98.622	3382377	3381874	0.00E+00	2
GmSN-like6	6	99.699	17279227	17279558	5.77E-168	4
GmSN-like7	9	99.078	46048084	46047651	0.00E+00	4
GmSN-like8	17	100.000	41213214	41213527	6.25E-160	3
GmSN-like9	3	99.002	34663864	34664364	0.00E+00	4
GmSN-like10	5	100.000	3022907	3023335	0.00E+00	4
GmSN-like11	20	99.502	37179214	37178813	0.00E+00	2
GmSN-like12	10	93.484	48562512	48562864	6.80E-154	2
GmSN-like13	14	98.936	48414359	48413797	0.00E+00	3
GmSN-like14	14	99.824	7800129	7799562	0.00E+00	2
GmSN-like15	17	99.445	39246408	39246948	0.00E+00	2
GmSN-like16	4	99.838	42390534	42389919	0.00E+00	4
GmSN-like17	4	100.000	44455449	44455978	0.00E+00	4
GmSN-like18	19	100.000	1247143	1247613	0.00E+00	4
GmSN-like19	19	100.000	2576710	2577165	0.00E+00	4
GmSN-like20	17	100.000	41223285	41222960	2.17E-166	3

Table S3. Qualitative aspects of the soybean homolog structures, the Z-Score and the Ramachandran plot in % comparing with snakin-1 model (PDB ID 5e5q)

Reference ID	Z-Score	Favored regions (%)	Additional allowed regions (%)	Generously allowed regions (%)	Disallowed regions (%)
GmSN-like1	-4.88	96	4	0	0
GmSN-like2	-4.4	95.7	4.3	0	0
GmSN-like3	-3.73	91.5	8.5	0	0
GmSN-like4	-4.36	96	4	0	0
GmSN-like5	-4.28	91.5	8.5	0	0
GmSN-like6	-5.14	91.8	6.1	0	2
GmSN-like7	-3.95	91.5	8.5	0	0
GmSN-like8	-4.62	95.9	4.1	0	0
GmSN-like9	-3.77	91.5	6.4	2.1	0
GmSN-like10	-4.59	92	8	0	0
GmSN-like11	-4.12	91.3	8.7	0	0
GmSN-like12	-3.8	93.6	6.4	0	0
GmSN-like13	-3.8	94	6	0	0
GmSN-like14	-5.64	91.5	8.5	0	0
GmSN-like15	-3.96	91.5	8.5	0	0
GmSN-like16	-3.85	91.5	8.5	0	0
GmSN-like17	-4.04	91.3	8.7	0	0
GmSN-like18	-3.72	91.5	6.4	2.1	0
GmSN-like19	-4.92	93.9	6.1	0	0
GmSN-like20	-4.4	95.7	4.3	0	0

6. CONCLUSÕES GERAIS

- ✓ Nesta análise foi possível identificar 20 esnaquinas-candidatas no transcriptoma da soja e 40 homólogos no genoma distribuídos em 16 dos 20 cromossomos. Todas apresentaram 12 cisteínas, conservadas em posição, independente de sua subfamília (I, II ou III);

- ✓ As esnaquinas identificadas em soja estão relacionadas ao metabolismo de desenvolvimento, bem como a estresses bióticos e abióticos;

- ✓ Sua distribuição genômica é heterogênea, não dependendo da estrutura do gene que pode possuir dois, três ou quatro éxons ou subfamília.

- ✓ O isolado GmSN2 é um novo membro que é expresso constitutivamente contra o fungo *P. pachyrhizi*, este gene possui três éxons, está localizado no cromossomo 6 da soja e possui homólogos em P. vulgaris e M. truncatula;

- ✓ A estrutura do peptídeo isolado bem como dos transcritos identificados está relacionada a outros AMPs previamente descritos, como alfa-hairpins e tioninas

- ✓ Este estudo provê uma importante fonte para o estudo da função e mapeamento destes genes em leguminosas

REFERÊNCIAS

Abraham MJ, Murtola T, Schulz R, et al (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25.

Adler M, Lazarus RA, Dennis MS, Wagner G (1991) Solution structure of kistrin, a potent platelet aggregation inhibitor and GP IIb-IIIa antagonist. Science. 253:445–448.

Aitken A (1999) Protein consensus sequence motifs. Mol Biotechnol 12:241–53.

Almasia NI, Narhirñak V, Hopp HE, Vazquez-Rovere C (2010) Isolation and characterization of the tissue and development-specific potato snakin-1 promoter inducible by temperature and wounding. Electron J Biotechnol 13:1–21.

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–10.

Andrew JM, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590.

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230.

Asensio JL, Cañada FJ, Bruix M, et al (1998) NMR investigations of protein – carbohydrate interactions : refined three-dimensional structure of the complex between hevein and methyl β -chitobioside. Glycobiology 8:569–577.

Atkinson NJ, Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. J Exp Bot 63:3523–3544.

Bairoch A, Boeckmann B, Ferro S, Gasteiger E (2004) Swiss-Prot: juggling between evolution and stability. Brief Bioinform 5:39–55.

Barbosa-da-silva A, Wanderley-nogueira AC, Silva RRM, et al (2005) *In silico* survey of resistance (R) genes in *Eucalyptus* transcriptome. Genet Mol Biol 28:562–574.

Batley, J. (2015). Molecular Marker Databases. In J. Batley (Ed.), Methods in Molecular Biology (pp. 29–49). Crawley: Humana Press.

Beintema JJ (1994) Structural features of plant chitinases and chitin-binding proteins. FEBS Lett 350:159–163.

Belarmino LC, Capriles PVSZ, Crovella S, et al (2010) EST-Database Search of Plant Defensins – An Example Using Sugarcane , a Large and Complex Genome. Curr Protein Pept Sci 11:1–7.

Bellhouse DR (2004) The reverend Thomas Bayes, FRS: A biography to celebrate the tercentenary of his birth. Stat Sci 19:3–43.

Benko-Iseppon AM, Galdino SL, Calsa T, et al (2010) Overview on plant antimicrobial peptides. Curr Protein Pept Sci 11:181–8.

Bent AF (1996) Plant disease resistance genes: Function meets structure. Plant Cell 8:1757–1771.

Berrocal-lobo M, Segura A, Moreno M, et al (2002) Snakin-2 , an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection. Plant Physiol 128:951–961.

Broekaert W, Franky W, Prost P (1992) Antimicrobial peptides from *Amaranthus caudatus* seeds with sequence homology to the Cystein/ Glycine rich Domain of chiting-biding proteins. Biochemistry 31:4308–4314.

Bry, F., & Kruger, P. (2003). A computational biology database digest: Data , data analysis, and data management. Distributed and Parallel Databases, 13(1), 7–42.

Buljan M, Bateman A (2009) The evolution of protein domain families. Biochem Soc Trans 37:751–755.

Cameron RK, Dixon RA, Lamb CJ (1994) Biologically induced systemic acquired resistance in *Arabidopsis thaliana*. Plant J 5:715–725.

Christensen AB, Cho BHO, Naesby M, et al (2002) The molecular characterization of two barley proteins establishes the novel PR-17 family of pathogenesis-related proteins. Mol Plant Pathol 3:135–144.

Cristobal S, Zemla a, Fischer D, et al (2001) A study of quality measures for protein threading models. BMC Bioinformatics 2:5.

Dangl JL (1995) Piece de Resistance: novel classes of plant disease resistance genes. Cell 80:363–366.

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833.

Diaz-Perales A, Collada C, Blanco C, et al (1999) Cross-reactions in the latex-fruit syndrome: A relevant role of chitinases but not of complex asparagine-linked glycans. J Allergy Clin Immunol 104:681–687.

Dixon, R A, Harrison, M J (1994) Early events in the activation of plant defense responses. Annu Rev Phytopathol 32:479–501.

Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S (2005) ProbCons:Probabilistic consistency-based multiple sequence alignment. Genome Res 15:330–340.

Dorantes-Acosta AE, Sánchez-Hernández C V., Arteaga-Vázquez MA (2012) Biotic stress in plants: Life lessons from your parents and grandparents. Front Genet 3:1–2.

Droit A, Poirier GG, Hunter JM (2005) Experimental and bioinformatic approaches for interrogating protein-protein interactions to determine protein function. J Mol Endocrinol 34:263–80.

Durrant WE, Dong X (2004) Systemic Acquired Resistance. Phytopathology 185–209.

Eddy S (1998) Profile hidden Markov models. Bioinformatics 14:755–763.

Edwards D, Batley J (2004) Plant bioinformatics: from genome to phenome. Trends Biotechnol 22:232–7.

Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnol J 8:2–9.

Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet 126:1–11.

Egorov TA, Odintsova TI, Pukhalsky VA, Grishin E V. (2005) Diversity of wheat anti-microbial peptides. Peptides 26:2064–2073.

Eltabakh MY, Ouzzani M, Aref WG (2006) bdbms - A database management system for biological data. In: https://arxiv.org/abs/cs/0612127. http://arxiv.org/abs/cs/0612127.

Elvira MI, Galdeano MM, Gilardi P, et al (2008) Proteomic analysis of pathogenesis-related proteins (PRs) induced by compatible and incompatible interactions of pepper mild mottle virus (PMMoV) in *Capsicum chinense* L3 plants. J Exp Bot 59:1253–1265.

Eswar N, Webb B, Marti-Renom MA, et al (2007) Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci 1–31.

Faccio P, Hopp E, Favret E, Franzone P (2011) Increased tolerance to wheat powdery mildew by heterologous constitutive expression of the *Solanum chacoense* snakin-1 gene. Czech J Genet Plant Breed 47:135–141.

Flor HH (1971) Current status of the gene-fob-gene concept. Annu Rev Phytopathol 275–296.

Fu ZQ, Dong X (2013) Systemic Acquired Resistance: turning local infection into global defense. Annu Rev Plant Biol 64:839–63.

Furukawa T, Sakaguchi N, Shimada H (2006) Two OsGASR genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles. Genes Genet Syst 81:171–80.

Goodstein DM, Shu S, Howson R, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178-86.

Gregory TR (2008) Understanding evolutionary trees. Evol Educ Outreach 1:121–137.

Guzmán-rodríguez JJ, Ibarra-laclette E, Herrera-estrella L, et al (2013) Analysis of expressed sequence tags (ESTs) from avocado seed (*Persea americana* var . *drymifolia*) reveals abundant expression of the gene encoding the antimicrobial peptide snakin. Plant Physiol Biochem 70:318–324.

Hall A V. (1988) A joint phenetic and cladistic approach for systematics. Biol J Linn Soc 33:367–382.

Hammami R, Ben Hamida J, Vergoten G, Fliss I (2009) PhytAMP: a database dedicated to antimicrobial plant peptides. Nucleic Acids Res 37:D963-8.

Herbel V, Wink M (2016) Mode of action and membrane specificity of the antimicrobial peptide snakin-2. PeerJ 4:1–13.

Holder M, Lewis PO (2003) Phylogeny estimation: Traditional and bayesian approaches. Nat Rev Genet 4:275–284.

Holt BF, Hubert DA, Dangl JL (2003) Resistance gene signaling in plants - Complex similarities to animal innate immunity. Curr Opin Immunol 15:20–25.

Jauch R, Yeo HC, Kolatkar PR, Clarke ND (2008) Assessment of CASP7 structure predictions for template free targets. Proteins 70:311–319.

Jiménez-Barbero J, Cañada JF, Asensio JL, et al (2006) Hevein domains: An attractive model to study carbohydrate-protein interactions at atomic resolution. In: Advances in Carbohydrate Chemistry and Biochemistry. pp 303–354

Jones P, Binns D, Chang HY, et al (2014) InterProScan 5: Genome-scale protein function classification. Bioinformatics 30:1236–1240.

Kachroo A, Robin GP (2013) Systemic signaling during plant defense. Curr Opin Plant Biol 16:527–533.

Kang YJ, Kim KH, Shim S, et al (2012) Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol 12:139–152.

Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. Proc Natl Acad Sci U S A 102:6679–6685.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772–780.

Kini SG, Nguyen PQT, Weissbach S, et al (2015) Studies on the chitin binding property of novel Cysteine-Rich peptides from *Alternanthera sessilis*. Biochemistry 54:6639–

6649.

Koo JC, Lee SY, Chun HJ, et al (1998) Two hevein homologs isolated from the seed of *Pharbitis nil* L. exhibit potent antifungal activity. Biochim Biophys Acta - Protein Struct Mol Enzymol 1382:80–90.

Kopec W, Telenius J, Khandelia H (2013) Molecular dynamics simulations of the interactions of medicinal plant extracts and drugs with lipid bilayer membranes. FEBS J 280:2785–2805.

Kotilainen M, Helariutta Y, Mehto M, et al (1999) GEG participates in the regulation of cell and organ shape during corolla and carpel development in *Gerbera hybrida*. Plant Cell 11:1093–104.

Lemey P, Salemi M, Vandamme A-M (2009) The Phylogenetic Handbook.

Letunic I, Goodstadt L, Dickens NJ, et al (2002) Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res 30:242–244.

Lima SCB da S, Benko-Iseppon AM, Bezerra-neto JP, et al (2017) Plants defense-related cyclic peptides: Diversity, structure and applications. Curr Protein Pept Sci 18:1–6.

Lipkin A, Anisimova V, Nikonorova A, et al (2005) An antimicrobial peptide Ar-AMP from amaranth (*Amaranthus retroflexus* L.) seeds. Phytochemistry 66:2426–2431.

Liu J-J, Ekramoddoullah AKM (2006) The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. Physiol Mol Plant Pathol 68:3–13.

Lucau-Danila A, Laborde L, Legrand S, et al (2010) Identification of novel genes potentially involved in somatic embryogenesis in chicory (*Cichorium intybus* L.). BMC Plant Biol 10:122–137.

Mackerell AD (2004) Empirical force fields for biological macromolecules: Overview and issues. J Comput Chem 25:1584–1604.

Madej T, Gibrat J-F, Bryant SH (1995) Threading a dabase of protein cores. Proteins Struct Funct Genet 23:356–369.

Manners JM (2007) Hidden weapons of microbial destruction in plant genomes. Genome Biol 8:225.

Mao Z, Zheng J, Wang Y, et al (2011) The new CaSn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper. Phytoparasitica 39:151–164.

Marchler-Bauer A, Bryant SH (2004) CD-Search: Protein domain annotations on the fly. Nucleic Acids Res 32:327–331.

Maróti G, Kereszt A, Kondorosi E, Mergaert P (2011) Natural roles of antimicrobial peptides in microbes, plants and animals. Res Microbiol 162:363–74.

Meiyalaghan S, Latimer JM, Kralicek A V, et al (2014) Expression and purification of the antimicrobial peptide GSL1 in bacteria for raising antibodies. BMC Res Notes 7:777.

Mochida K, Shinozaki K (2010) Genomics and bioinformatics resources for crop improvement. Plant Cell Physiol 51:497–523.

Mochida K, Shinozaki K (2017) Advances in omics and bioinformatics tools for systems analyses of plant functions. Plant Cell Physiol 52:2017–2038.

Morel JB, Dangl JL (1997) The hypersensitive response and the induction of cell death in plants. Cell Death Differ 4:671–83.

Mur LAJ, Kenton P, Lloyd AJ, et al (2008) The hypersensitive response; The centenary is upon us but how much do we know? J Exp Bot 59:501–520.

Nahirñak V, Almasia NI, Fernandez PV, et al (2012) Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. Plant Physiol 158:252–63.

Nawrot R, Barylski J, Nowicki G, et al (2014) Plant antimicrobial peptides. Folia Microbiol (Praha) 59:181–96.

Nei M (1996) Phylogenetic analysis in molecular evolutionary genetics. Annu Rev Genet 30:371–403.

Nolde SB, Vassilevski AA, Rogozhin EA, et al (2011) Disulfide-stabilized helical Hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (*Echinochloa crus-galli*). J Biol Chem 286:25145–25153.

Oliveira-Lima M, Benko-Iseppon AM, Neto JRCF, et al (2017) Snakin: Structure, roles and applications of a plant antimicrobial peptide. Curr Protein Pept Sci 18:1–7.

Padovan L, Scocchi M, Tossi A (2010) Structural aspects of plant antimicrobial peptides. Curr Protein Pept Sci 11:210–9.

Pais FS, Ruy PDC, Oliveira G, Coimbra RS (2014) Assessing the efficiency of multiple sequence alignment programs. Algorithms Mol Biol 9:1–8.

Pandey D, Rajendran SRCK, Gaur M, et al (2016) Plant defense signaling and responses against necrotrophic fungal pathogens. J Plant Growth Regul 1–16.

Pandolfi V, Neto JRCF, Silva MD, et al (2017) Resistance (R) genes: Applications and prospects for plant biotechnology and breeding. Curr Protein Pept Sci 18:1–12.

Pantazes RJ, Grisewood MJ, Maranas CD (2011) Recent advances in computational

protein design. Curr Opin Struct Biol 21:467–472.

Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 183:63–98.

Peng C (2007) Distance based methods in phylogenetic tree construction. Neural Parallel Sci Comput 1–11.

Petsko, G. A., & Ringe, D. (2004). From sequence to structure. In Protein structure and function (p. 195). New Science Press Ltd.

Pinto M dos ST, Ribeiro JM, Oliveira EAG de (2011) O estudo de genes e proteínas de defesa em plantas. Brazilian J Biosci 9:241–248.

Polyanovsky VO, Roytberg MA, Tumanyan VG (2011) Comparative analysis of the quality of a global algorithm and a local algorithm for alignment of two sequences. Algorithms Mol Biol 6:25.

Porto WF, Pires ÁS, Franco OL (2017) Computational tools for exploring sequence databases as a resource for antimicrobial peptides. Biotechnol Adv 1–13.

Rahman M-, Shaheen T, Rahman M-, Iqbal MA (2016) Bioinformatics : A way forward to explore "Plant Omics." In: Abdurakhmonov I y (ed) Bioinformatics - Updated Features and Applications. InTech, pp 203–230

Ridley, M. (2004). Evolution. Evolution (3rd ed.). Malden: Blackwell Publishing.

Rizhsky L, Liang H, Shuman J, et al (2004) When defense pathways collide. The response of *Arabidopsis* to a combination of drought and heat stress. Plant Physiol 134:1683–1696.

Rizzo J, Rouchka EC (2007) Review of phylogenetic tree construction. Univ Louisv Bioinforma Lab Tech Rep Ser 1–8.

Robert-Seilaniantz A, Grant M, Jones JDG (2011) Hormone crosstalk in plant disease and defense: More than just Jasmonate-Salycate antagonism. Annu Rev Phytopathol 49:317–343.

Romagnoli S, Ugolini R, Fogolari F, et al (2000) NMR structural determination of viscotoxin A3 from *Viscum album* L. Biochem J 350 Pt 2:569–77.

Ronald PC (1998) Resistance gene evolution. Curr Opin Plant Biol 1:294–298.

Roxrud I, Lid SE, Fletcher JC, et al (2007) GASA4, one of the 14-member *Arabidopsis* GASA family of small polypeptides, regulates flowering and seed development. Plant Cell Physiol 48:471–83.

Roy, A., & Zhang, Y. (2012). Protein Structure Prediction. eLS, (8), 1–11.

Saboki E, Usha K, Singh B (2011) Pathogenesis Related (PR) proteins in plant defense

mechanism age-related pathogen resistance. Curr Res Technol Adv 1043–1054.

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Segura A, Moreno M, Madueño F, et al (1999) Snakin-1 , a peptide from potato that is active against plant pathogens. Mol Plant-Microbe Interact 12:16–23.

Sels J, Mathys J, De Coninck BM a, et al (2008) Plant pathogenesis-related (PR) proteins: a focus on PR peptides. Plant Physiol Biochem 46:941–950.

Serrano M, Coluccia F, Torres M, et al (2014) The cuticle and plant defense to pathogens. Front Plant Sci 5:1–8.

Sheth BP, Thaker VS (2014) Plant systems biology: Insights, advances and challenges. Planta 240:33–54.

Sievers F, Wilm A, Dineen D, et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

Silverstein K a T, Moskal W a, Wu HC, et al (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J 51:262–80.

Slavokhotova, A. A., Naumann, T. A., Price, N. P. J., Rogozhin, E. A., Andreev, Y. A., Vassilevski, A. A., & Odintsova, T. I. (2014). Novel mode of action in plant defense peptides: hevein-like antimicrobial peptides from wheat inhibit fungal metalloproteases. The FEBS Journal.

Soares-Cavalcanti NM, Belarmino LC, Kido E a, et al (2012) Overall picture of expressed Heat Shock Factors in *Glycine max*, *Lotus japonicus* and *Medicago truncatula*. Genet Mol Biol 35:247–59.

Soosaar JLM, Burch-Smith TM, Dinesh-Kumar SP (2005) Mechanisms of plant resistance to viruses. Nat Rev Microbiol 3:789–98.

Strasser BJ (2010) Collecting, comparing, and computing sequences: The making of , Margaret O. Dayhoff's atlas of protein sequence and structure, 1954-1965. J Hist Biol 43:623–660.

Sudisha J, Sharathchandra RG, Amruthesh KN, et al (2012) Plant Defence: Biological Control. In: Plant Defense: Biological controll. Springer Netherlands, pp 379–403

Tam JP, Wang S, Wong KH, Tan WL (2015) Antimicrobial peptides from plants. Pharmaceuticals 8:711–757.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815.

Thompson K (1968) Programming Techniques: Regular expression search algorithm. Commun ACM 11:419–422.

Thorburn, W. M. (1915). Occam's razor. Mind, XXIV(2), 287–288.

Tuzun S (2001) The relationship between pathogen-induced systemic resistance (ISR) and multigenic (horizontal) resistance in plants. Eur J Plant Pathol 107:85–93.

Van Der Biezen EA, Jones JDG (1998) Plant disease-resistance proteins and the gene-for-gene concept. Trends Biochem Sci 23:454–456.

van Loon LC (1999) Occurrence and properties of plant Pathogenesis-Related proteins. In: Datta SK, Muthukrishman S (eds) Pathogenesis-Related Proteins in Plants. CRC Press, Boca Raton,

van Loon LC, Pierpoint WS, Boller T, Conejero V (1994) Recommendations for naming plant pathogenesis-related proteins. Plant Mol Biol Report 12:245–264.

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol 44:135–62.

van Verk MC, Gatz C, Linthorst HJM (2009) Transcriptional regulation of plant defense responses. In: Advances in Botanical Research, 1st edn. Elsevier Ltd., pp 397–438

Vanetten HD, Mansfield JW, Bailey JA, Farmer EE (1994) Two classes of plant antibiotics: Phytoalexins versus "Phytoanticipins." Plant Cell 101:106–109.

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530.

Vlot AC, Klessig DF, Park SW (2008) Systemic acquired resistance: the elusive signal(s). Curr Opin Plant Biol 11:436–442.

Waghu FH, Gopi L, Barai RS, et al (2014) CAMP: Collection of sequences and structures of antimicrobial peptides. Nucleic Acids Res 42:1154–1158.

Wanderley-nogueira AC, Belarmino LC, Soares-cavalcanti NM, et al (2012) An overall evaluation of the Resistance (R) and Pathogenesis-Related (PR) superfamilies in soybean , as compared with *Medicago* and *Arabidopsis*. Genet Mol Biol 35:260–271.

Wanderley-Nogueira AC, Bezerra-Neto JP, Kido EA, et al (2017) Plant elite squad: First defense line and resistance genes - identification, diversity and functional roles. Curr Protein Pept Sci 18:1–17.

Wanderley-Nogueira AC, Soares-Cavalcanti N., Morais DA., et al (2007) Abundance and diversity of resistance genes in the sugarcane transcriptome revealed by *in silico* analysis. Genet Mol Res 6:866–889.

Wasmuth E V, Lima CD (2016) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:1–12.

Wu S, Zhang Y (2007) LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Res 35:3375–3382.

Xu D, Zhang Y (2012) *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins Struct Funct Bioinforma 80:1715–1735.

Yan J, Yuan S su, Jiang L luan, et al (2015) Plant antifungal proteins and their applications in agriculture. Appl Microbiol Biotechnol 99:4961–4981.

Yang Z (1996) Phylogenetic analysis using parsimony and likelihood methods. J Mol Evol 42:294–307.

Yang Z, Rannala B (1997) Bayesian phylogentic inference using DNA sequences: A Markov chain monte carlo method. Mol Biol Evol 14:717–724.

Yang Z, Rannala B (2012) Molecular phylogenetics: principles and practice. Nat Rev Genet 13:303–314.

Yeung H, Squire CJ, Yosaatmadja Y, et al (2016) Radiation damage and racemic protein crystallography reveal the unique structure of the GASA/Snakin protein superfamily. Angew Chemie Int Ed 1–5.

Zare-zardini H, Tolueinia B, Hashemi A, Fesahat F (2013) Antioxidant and Cholinesterase From *Ziziphus jujuba* Fruits. Am J Alzheimer's Dis Other Dementias 4–12.

Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature 415:389–395.

Zhang S, Yang C, Peng J, et al (2009) GASA5, a regulator of flowering time and stem growth in *Arabidopsis thaliana*. Plant Mol Biol 69:745–59.

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40.

Zimmermann R, Sakai H, Hochholdinger F (2010) The Gibberellic Acid Stimulated-Like gene family in maize and its role in lateral root development. Plant Physiol 152:356–65.

Zurbriggen MD, Carrillo N, Hajirezaei M-R (2010) ROS signaling in the hypersensitive response. Plant Signal Behav 5:393–396.

ANEXOS

Anexo I. Instruções para autores – Revista: *Peptides*

Guide for Authors
All journal information and instructions compiled in one document (PDF) in just one mouse-click Author information pack

• Your Paper Your Way
INTRODUCTION
• Submission checklist
BEFORE YOU BEGIN
• Ethics in publishing
• Declaration of interest
• Submission declaration and verification
• Use of inclusive language
• Contributors
• Changes to authorship
• Copyright
• Role of the funding source
• Open access
• Submission
PREPARATION
• NEW SUBMISSIONS
• Peer review
• REVISED SUBMISSIONS
• Article structure
• Essential title page information
• Keywords
• Artwork
• Tables
• References
• Data visualization
• Supplementary material
• Research data
AFTER ACCEPTANCE
• Online proof correction
• Offprints
AUTHOR INQUIRIES
Your Paper Your Way

We now differentiate between the requirements for new and revised submissions. You may choose to submit your manuscript as a single Word or PDF file to be used in the refereeing process. Only when your paper is at the revision stage, will you be requested to put your paper in to a 'correct format' for acceptance and provide the items required for the publication of your article.
To find out more, please visit the Preparation section below.

Peptides is an international journal presenting original contributions on the biochemistry, physiology and pharmacology of biological active peptides, as well as their functions that relate to gastroenterology, endocrinology, and behavioral effects.

Peptides emphasizes all aspects of high profile peptide research in mammals and non-mammalian vertebrates. Special consideration can be given to plants and invertebrates. Submission of articles with clinical relevance is particularly encouraged.

Types of paper

Research articles
Letter to the Editor.
Review articles
Submission checklist

You can use this list to carry out a final check of your submission before you send it to the journal for review. Please check the relevant section in this Guide for Authors for more details.

Ensure that the following items are present:

One author has been designated as the corresponding author with contact details:
• E-mail address
• Full postal address

All necessary files have been uploaded:
Manuscript:
• Include keywords
• All figures (include relevant captions)
• All tables (including titles, description, footnotes)
• Ensure all figure and table citations in the text match the files provided
• Indicate clearly if color should be used for any figures in print
Graphical Abstracts / Highlights files (where applicable)
Supplemental files (where applicable)

Further considerations
• Manuscript has been 'spell checked' and 'grammar checked'
• All references mentioned in the Reference List are cited in the text, and vice versa
• Permission has been obtained for use of copyrighted material from other sources (including the Internet)
• A competing interests statement is provided, even if the authors have no competing interests to declare
• Journal policies detailed in this guide have been reviewed
• Referee suggestions and contact details provided, based on journal requirements

For further information, visit our Support Center.

Ethics in publishing

Please see our information pages on Ethics in publishing and Ethical guidelines for journal publication.

Declaration of interest

All authors must disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work. Examples of potential competing interests include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding. Authors must disclose any interests in two places: 1. A summary declaration of interest statement in the title page file (if double-blind) or the manuscript file (if single-blind). If there are no interests to declare then please state this: 'Declarations of interest: none'. This summary statement will be ultimately published if the article is accepted. 2. Detailed disclosures as part of a separate Declaration of Interest form, which forms part of the journal's official records. It is important for potential interests to be declared in both places and that the information matches. More information.

Submission declaration and verification

Submission of an article implies that the work described has not been published previously (except in the form of an abstract, a published lecture or academic thesis, see 'Multiple, redundant or concurrent publication' for more information), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. To verify originality, your article may be checked by the originality detection service Crossref Similarity Check.

Preprints
Please note that preprints can be shared anywhere at any time, in line with Elsevier's sharing policy. Sharing your preprints e.g. on a preprint server will not count as prior publication (see 'Multiple, redundant or concurrent publication' for more information).

Use of inclusive language

Inclusive language acknowledges diversity, conveys respect to all people, is sensitive to differences, and promotes equal opportunities. Articles should make no assumptions about the beliefs or commitments of any reader, should contain nothing which might imply that one individual is superior to another on the grounds of race, sex, culture or any other characteristic, and should use inclusive language throughout. Authors should ensure that writing is free from bias, for instance by using 'he or she', 'his/her' instead of 'he' or 'his', and by making use of job titles that are free of stereotyping (e.g. 'chairperson' instead of 'chairman' and 'flight attendant' instead of 'stewardess').
Contributors

Each author is required to declare his or her individual contribution to the article: all authors must have materially participated in the research and/or article preparation, so roles for all authors should be described. The statement that all authors have approved the final article should be true and included in the disclosure.

Addition, deletion, or rearrangement of author names in the authorship of accepted manuscripts
Before the accepted manuscript is published in an online issue
Requests to add or remove an author, or to rearrange the author names, must be sent to the Journal Manager from the corresponding author of the accepted manuscript and must include:

The reason the name should be added or removed or the author names rearranged.
Written confirmation (email, fax, letter) from all authors that they agree with the addition, removal or rearrangement. In the case of addition or removal of authors, this includes confirmation from the author being added or removed.
Requests that are not sent by the corresponding author will be forwarded by the Journal Manager to the corresponding author, who must follow the procedure as described above.
Note that:
Journal Managers will inform the Journal Editors of any such requests.
Publication of the accepted manuscript in an online issue is suspended until authorship has been agreed.
After the accepted manuscript is published in an online issue Any requests to add, delete, or rearrange author names in an article published in an online issue will follow the same policies as noted above and result in a corrigendum.

Changes to authorship

Authors are expected to consider carefully the list and order of authors before submitting their manuscript and provide the definitive list of authors at the time of the original submission. Any addition, deletion or rearrangement of author names in the authorship list should be made only before the manuscript has been accepted and only if approved by the journal Editor. To request such a change, the Editor must receive the following from the corresponding author: (a) the reason for the change in author list and (b) written confirmation (e-mail, letter) from all authors that they agree with the addition, removal or rearrangement. In the case of addition or removal of authors, this includes confirmation from the author being added or removed.
Only in exceptional circumstances will the Editor consider the addition, deletion or rearrangement of authors after the manuscript has been accepted. While the Editor considers the request, publication of the manuscript will be suspended. If the manuscript has already been published in an online issue, any requests approved by the Editor will result in a corrigendum.

Copyright

Upon acceptance of an article, authors will be asked to complete a 'Journal Publishing Agreement' (see more information on this). An e-mail will be sent to the corresponding author confirming receipt of the manuscript together with a 'Journal Publishing Agreement' form or a link to the online version of this agreement.

Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution and for all other derivative works, including compilations and translations. If excerpts from other copyrighted works are

included, the author(s) must obtain written permission from the copyright owners and credit the source(s) in the article. Elsevier has preprinted forms for use by authors in these cases.

For gold open access articles: Upon acceptance of an article, authors will be asked to complete an 'Exclusive License Agreement' (more information). Permitted third party reuse of gold open access articles is determined by the author's choice of user license.

Author rights
As an author you (or your employer or institution) have certain rights to reuse your work. More information.

Elsevier supports responsible sharing
Find out how you can share your research published in Elsevier journals.

Role of the funding source

You are requested to identify who provided financial support for the conduct of the research and/or preparation of the article and to briefly describe the role of the sponsor(s), if any, in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. If the funding source(s) had no such involvement then this should be stated.

Funding body agreements and policies
Elsevier has established a number of agreements with funding bodies which allow authors to comply with their funder's open access policies. Some funding bodies will reimburse the author for the gold open access publication fee. Details of existing agreements are available online.

Open access

This journal offers authors a choice in publishing their research:

Subscription
• Articles are made available to subscribers as well as developing countries and patient groups through our universal access programs.
• No open access publication fee payable by authors.
• The Author is entitled to post the accepted manuscript in their institution's repository and make this public after an embargo period (known as green Open Access). The published journal article cannot be shared publicly, for example on ResearchGate or Academia.edu, to ensure the sustainability of peer-reviewed research in journal publications. The embargo period for this journal can be found below.
Gold open access
• Articles are freely available to both subscribers and the wider public with permitted reuse.
• A gold open access publication fee is payable by authors or on their behalf, e.g. by their research funder or institution.

Regardless of how you choose to publish your article, the journal will apply the same peer review criteria and acceptance standards.

For gold open access articles, permitted third party (re)use is defined by the following Creative Commons user licenses:

Creative Commons Attribution (CC BY)
Lets others distribute and copy the article, create extracts, abstracts, and other revised versions, adaptations or derivative works of or from an article (such as a translation), include in a collective work (such as an anthology), text or data mine the article, even for commercial purposes, as long as they credit the author(s), do not represent the author as endorsing their adaptation of the article, and do not modify the article in such a way as to damage the author's honor or reputation.

Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)
For non-commercial purposes, lets others distribute and copy the article, and to include in a collective work (such as an anthology), as long as they credit the author(s) and provided they do not alter or modify the article.

The gold open access publication fee for this journal is USD 3000, excluding taxes. Learn more about Elsevier's pricing policy: https://www.elsevier.com/openaccesspricing.

Green open access
Authors can share their research in a variety of different ways and Elsevier has a number of green open access options available. We recommend authors see our green open access page for further information. Authors can also self-archive their manuscripts immediately and enable public access from their institution's repository after an embargo period. This is the version that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and in editor-author communications. Embargo period: For subscription articles, an appropriate amount of time is needed for journals to deliver value to subscribing customers before an article becomes freely available to the public. This is the embargo period and it begins from the date the article is formally published online in its final and fully citable form. Find out more.

This journal has an embargo period of 12 months.

Elsevier Researcher Academy
Researcher Academy is a free e-learning platform designed to support early and mid-career researchers throughout their research journey. The "Learn" environment at Researcher Academy offers several interactive modules, webinars, downloadable guides and resources to guide you through the process of writing for research and going through peer review. Feel free to use these free resources to improve your submission and navigate the publication process with ease.

language and language services
Please write your text in good English. Only American usage is accepted, e.g., utilize, not utilise; color, not colour; while, not whilst.

Authors who require information about language editing and copyediting services pre- and post-submission please visit http://www.elsevier.com/languageediting or our customer support site at http://service.elsevier.com for more information.

Submission

Our online submission system guides you stepwise through the process of entering your article details and uploading your files. The system converts your article files to a single PDF file used in the peer-review process. Editable files (e.g., Word, LaTeX) are required to typeset your article for final publication. All correspondence, including notification of the Editor's decision and requests for revision, is sent by e-mail.

NEW SUBMISSIONS

Submission to this journal proceeds totally online and you will be guided stepwise through the creation and uploading of your files. The system automatically converts your files to a single PDF file, which is used in the peer-review process.
As part of the Your Paper Your Way service, you may choose to submit your manuscript as a single file to be used in the refereeing process. This can be a PDF file or a Word document, in any format or lay-out that can be used by referees to evaluate your manuscript. It should contain high enough quality figures for refereeing. If you prefer to do so, you may still provide all or some of the source files at the initial submission. Please note that individual figure files larger than 10 MB must be uploaded separately.

References

There are no strict requirements on reference formatting at submission. References can be in any style or format as long as the style is consistent. Where applicable, author(s) name(s), journal title/book title, chapter title/article title, year of publication, volume number/book chapter and the article number or pagination must be present. Use of DOI is highly encouraged. The reference style used by the journal will be applied to the accepted article by Elsevier at the proof stage. Note that missing data will be highlighted at proof stage for the author to correct.

Formatting requirements

There are no strict formatting requirements but all manuscripts must contain the essential elements needed to convey your manuscript, for example Abstract, Keywords, Introduction, Materials and Methods, Results, Conclusions, Artwork and Tables with Captions.
If your article includes any Videos and/or other Supplementary material, this should be included in your initial submission for peer review purposes.
Divide the article into clearly defined sections.

Peer review

This journal operates a single blind review process. All contributions will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then typically sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper. The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final. More information on types of peer review.

REVISED SUBMISSIONS

Use of word processing software
Regardless of the file format of the original submission, at revision you must provide us with an editable file of the entire article. Keep the layout of the text as simple as possible. Most formatting codes will be removed and replaced on processing the article. The electronic text should be prepared in a way very similar to that of conventional manuscripts (see also the Guide to Publishing with Elsevier). See also the section on Electronic artwork.
To avoid unnecessary errors you are strongly advised to use the 'spell-check' and 'grammar-check' functions of your word processor.

Article structure

Subdivision - numbered sections
Divide your article into clearly defined and numbered sections. Subsections should be numbered 1.1 (then 1.1.1, 1.1.2, ...), 1.2, etc. (the abstract is not included in section numbering). Use this numbering also for internal cross-referencing: do not just refer to 'the text'. Any subsection may be given a brief heading. Each heading should appear on its own separate line.

Introduction
State the objectives of the work and provide an adequate background, avoiding a detailed literature survey or a summary of the results.

Material and methods
Provide sufficient details to allow the work to be reproduced by an independent researcher. Methods that are already published should be summarized, and indicated by a reference. If quoting directly from a previously published method, use quotation marks and also cite the source. Any modifications to existing methods should also be described.

Results
Results should be clear and concise. Results and Discussion sections should be separate, even for papers submitted as Brief Communications.

Discussion
This should explore the significance of the results of the work, not repeat them. Avoid extensive citations and discussion of published literature.

Conclusion
The main conclusions of the study may be presented in a short Conclusions section, which may stand alone or form a subsection of a Discussion section.

Glossary
Please supply, as a separate list, the definitions of field-specific terms used in your article. Italics are not to be used for expressions of Latin origin, for example, in vivo, et al., per se.

Appendices. If there is more than one appendix, they should be identified as A, B, etc. Formulae and equations in appendices should be given separate numbering: (Eq. A.1), (Eq. A.2), etc.; in a subsequent appendix, (Eq. B.1) and so forth.

Essential title page information

• Title. Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible.
• Author names and affiliations. Please clearly indicate the given name(s) and family name(s) of each author and check that all names are accurately spelled. You can add your name between parentheses in your own script behind the English transliteration. Present the authors' affiliation addresses (where the actual work was done) below the names. Indicate all affiliations with a lower-case superscript letter immediately after the author's name and in front of the appropriate address. Provide the full postal address of each affiliation, including the country name and, if available, the e-mail address of each author.
• Corresponding author. Clearly indicate who will handle correspondence at all stages of refereeing and publication, also post-publication. This responsibility includes answering any future queries about Methodology and Materials. Ensure that the e-mail address is given and that contact details are kept up to date by the corresponding author.
• Present/permanent address. If an author has moved since the work described in the article was done, or was visiting at the time, a 'Present address' (or 'Permanent address') may be indicated as a footnote to that author's name. The address at which the author actually did the work must be retained as the main, affiliation address. Superscript Arabic numerals are used for such footnotes.

Abstract
A concise and factual single paragraph abstract without headings is required. The abstract should state briefly the purpose of the research, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided. Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself.

We will not evaluate any abstracts or submissions outside the EES system. Editorials and letter to the editor do not have an abstract.

Graphical abstract
Although a graphical abstract is optional, its use is encouraged as it draws more attention to the online article. The graphical abstract should summarize the contents of the article in a concise, pictorial form designed to capture the attention of a wide readership. Graphical abstracts should be submitted as a separate file in the online submission system. Image size: Please provide an image with a minimum of 531 × 1328 pixels (h × w) or proportionally more. The image should be readable at a size of 5 × 13 cm using a regular screen resolution of 96 dpi. Preferred file types: TIFF, EPS, PDF or MS Office files. You can view Example Graphical Abstracts on our information site.
Authors can make use of Elsevier's Illustration Services to ensure the best presentation of their images and in accordance with all technical requirements.

Highlights

Highlights are mandatory for this journal. They consist of a short collection of bullet points that convey the core findings of the article and should be submitted in a separate editable file in the online submission system. Please use 'Highlights' in the file name and

include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point). You can view example Highlights on our information site.

Keywords

Immediately after the abstract, provide a maximum of 6 keywords, using American spelling and avoiding general and plural terms and multiple concepts (avoid, for example, 'and', 'of'). Be sparing with abbreviations: only abbreviations firmly established in the field may be eligible. These keywords will be used for indexing purposes.

Abbreviations
Define abbreviations that are not standard in this field in a footnote to be placed on the first page of the article. Such abbreviations that are unavoidable in the abstract must be defined at their first mention there, as well as in the footnote. Ensure consistency of abbreviations throughout the article.

Acknowledgements
Acknowledgements. Place acknowledgements, including information on grants received, before the references, in a separate section, and not as a footnote on the title page.

Formatting of funding sources
List funding sources in this standard way to facilitate compliance to funder's requirements:

Funding: This work was supported by the National Institutes of Health [grant numbers xxxx, yyyy]; the Bill & Melinda Gates Foundation, Seattle, WA [grant number zzzz]; and the United States Institutes of Peace [grant number aaaa].

It is not necessary to include detailed descriptions on the program or type of grants and awards. When funding is from a block grant or other resources available to a university, college, or other research institution, submit the name of the institute or organization that provided the funding.

If no funding has been provided for the research, please include the following sentence:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Units
Follow internationally accepted rules and conventions: use the international system of units (SI). If other units are mentioned, please give their equivalent in SI. For numbers, use decimal points (not commas); use a space for thousands (10 000 and above).

Drugs
Proprietary (trademarked) names should be capitalized. The chemical name should precede the trade, popular name, or abbreviation of a drug the first time it occurs.

Amino Acids
The first letter of the 3-letter abbreviations for amino acids should be capitalized.

Anesthesia
In describing surgical procedures on animals, the type and dosage of the anesthetic agent should be specified. Curarizing agents are not anesthetics; if these were used, evidence must be provided that anesthesia of suitable grade and duration was employed

Footnotes
Footnotes should be used sparingly. Number them consecutively throughout the article. Many word processors build footnotes into the text, and this feature may be used. Should this not be the case, indicate the position of footnotes in the text and present the footnotes themselves separately at the end of the article.

Artwork

Electronic artwork
General points
• Make sure you use uniform lettering and sizing of your original artwork.
• Preferred fonts: Arial (or Helvetica), Times New Roman (or Times), Symbol, Courier.
• Number the illustrations according to their sequence in the text.
• Use a logical naming convention for your artwork files.
• Indicate per figure if it is a single, 1.5 or 2-column fitting image.
• For Word submissions only, you may still provide figures and their captions, and tables within a single file at the revision stage.
• Please note that individual figure files larger than 10 MB must be provided in separate source files.
A detailed guide on electronic artwork is available.
You are urged to visit this site; some excerpts from the detailed information are given here.
Formats
Regardless of the application used, when your electronic artwork is finalized, please 'save as' or convert the images to one of the following formats (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below):
EPS (or PDF): Vector drawings. Embed the font or save the text as 'graphics'.
TIFF (or JPG): Color or grayscale photographs (halftones): always use a minimum of 300 dpi.
TIFF (or JPG): Bitmapped line drawings: use a minimum of 1000 dpi.
TIFF (or JPG): Combinations bitmapped line/half-tone (color or grayscale): a minimum of 500 dpi is required.
Please do not:
• Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); the resolution is too low.
• Supply files that are too low in resolution.
• Submit graphics that are disproportionately large for the content.

Color artwork
Please make sure that artwork files are in an acceptable format (TIFF (or JPEG), EPS (or PDF) or MS Office files) and with the correct resolution. If, together with your accepted article, you submit usable color figures then Elsevier will ensure, at no additional charge, that these figures will appear in color online (e.g., ScienceDirect and other sites) in addition to color reproduction in print. Further information on the preparation of electronic artwork.

Figure captions
Ensure that each illustration has a caption. A caption should comprise a brief title (not on the figure itself) and a description of the illustration. Keep text in the illustrations themselves to a minimum but explain all symbols and abbreviations used.

Tables

Please submit tables as editable text and not as images. Tables can be placed either next to the relevant text in the article, or on separate page(s) at the end. Number tables consecutively in accordance with their appearance in the text and place any table notes below the table body. Be sparing in the use of tables and ensure that the data presented in them do not duplicate results described elsewhere in the article. Please avoid using vertical rules and shading in table cells.

References

Citation in text
Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be given in full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either 'Unpublished results' or 'Personal communication'. Citation of a reference as 'in press' implies that the item has been accepted for publication.

Web references
As a minimum, the full URL should be given and the date when the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references can be listed separately (e.g., after the reference list) under a different heading if desired, or can be included in the reference list.

Data references
This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.

Reference management software

Most Elsevier journals have their reference template available in many of the most popular reference management software products. These include all products that support Citation Style Language styles, such as Mendeley and Zotero, as well as EndNote. Using the word processor plug-ins from these products, authors only need to select the appropriate journal template when preparing their article, after which citations and bibliographies will be automatically formatted in the journal's style. If no template is yet

available for this journal, please follow the format of the sample references and citations as shown in this Guide. If you use reference management software, please ensure that you remove all field codes before submitting the electronic manuscript. More information on how to remove field codes.

Users of Mendeley Desktop can easily install the reference style for this journal by clicking the following link:
http://open.mendeley.com/use-citation-style/peptides
When preparing your manuscript, you will then be able to select this style using the Mendeley plug-ins for Microsoft Word or LibreOffice.

Reference formatting

There are no strict requirements on reference formatting at submission. References can be in any style or format as long as the style is consistent. Where applicable, author(s) name(s), journal title/book title, chapter title/article title, year of publication, volume number/book chapter and the article number or pagination must be present. Use of DOI is highly encouraged. The reference style used by the journal will be applied to the accepted article by Elsevier at the proof stage. Note that missing data will be highlighted at proof stage for the author to correct. If you do wish to format the references yourself they should be arranged according to the following examples:

Reference style
Text: Indicate references by number(s) in square brackets in line with the text. The actual authors can be referred to, but the reference number(s) must always be given. All references should be in English - native language publications other than English are not accepted. References for normal research articles should be less than 50.
List: The list of references is arranged alphabetically and then numbered (numbers in square brackets).
Examples:
Reference to a journal publication:
[1] Van der Geer J, Hanraads JAJ, Lupton RA. The art of writing a scientific article. J Sci Commun 2000;163:51–9.
Reference to a book:
[2] Strunk Jr W, White EB. The elements of style. 3rd ed. New York: Macmillan; 1979.
Reference to a chapter in an edited book:
[3] Mettam GR, Adams LB. How to prepare an electronic version of your article. In: Jones BS, Smith RZ, editors. Introduction to the electronic age, New York: E-Publishing Inc; 1999, p. 281–304.
[4] M. Oguro, S. Imahiro, S. Saito, T. Nakashizuka, Mortality data for Japanese oak wilt disease and surrounding forest compositions, Mendeley Data, v1, 2015. http://dx.doi.org/10.17632/xwj98nb39r.1.
Note shortened form for last page number. e.g., 51–9, and that for more than 6 authors the first 6 should be listed followed by "et al." For further details you are referred to "Uniform Requirements for Manuscripts submitted to Biomedical Journals" (J Am Med Assoc 1997;277:927–934) (see also http://www.nlm.nih.gov/bsd/uniform_requirements.html).

Journal abbreviations source
Journal names should be abbreviated according to the List of Title Word Abbreviations.

Data visualization

Include interactive data visualizations in your publication and let your readers interact and engage more closely with your research. Follow the instructions here to find out about available data visualization options and how to include them with your article.

Supplementary material

Supplementary material such as applications, images and sound clips, can be published with your article to enhance it. Submitted supplementary items are published exactly as they are received (Excel or PowerPoint files will appear as such online). Please submit your material together with the article and supply a concise, descriptive caption for each supplementary file. If you wish to make changes to supplementary material during any stage of the process, please make sure to provide an updated file. Do not annotate any corrections on a previous version. Please switch off the 'Track Changes' option in Microsoft Office files as these will appear in the published version.

Research data

This journal encourages and enables you to share data that supports your research publication where appropriate, and enables you to interlink the data with your published articles. Research data refers to the results of observations or experimentation that validate research findings. To facilitate reproducibility and data reuse, this journal also encourages you to share your software, code, models, algorithms, protocols, methods and other useful materials related to the project.

Below are a number of ways in which you can associate data with your article or make a statement about the availability of your data when submitting your manuscript. If you are sharing data in one of these ways, you are encouraged to cite the data in your manuscript and reference list. Please refer to the "References" section for more information about data citation. For more information on depositing, sharing and using research data and other relevant research materials, visit the research data page.

Data linking
If you have made your research data available in a data repository, you can link your article directly to the dataset. Elsevier collaborates with a number of repositories to link articles on ScienceDirect with relevant repositories, giving readers access to underlying data that gives them a better understanding of the research described.

There are different ways to link your datasets to your article. When available, you can directly link your dataset to your article by providing the relevant information in the submission system. For more information, visit the database linking page.

For supported data repositories a repository banner will automatically appear next to your published article on ScienceDirect.

In addition, you can link to relevant data or entities through identifiers within the text of your manuscript, using the following format: Database: xxxx (e.g., TAIR: AT1G01020; CCDC: 734053; PDB: 1XFN).

Mendeley Data
This journal supports Mendeley Data, enabling you to deposit any research data (including raw and processed data, video, code, software, algorithms, protocols, and methods) associated with your manuscript in a free-to-use, open access repository. During the submission process, after uploading your manuscript, you will have the opportunity to upload your relevant datasets directly to Mendeley Data. The datasets will be listed and directly accessible to readers next to your published article online.

For more information, visit the Mendeley Data for journals page.

Data in Brief
You have the option of converting any or all parts of your supplementary or additional raw data into one or multiple data articles, a new kind of article that houses and describes your data. Data articles ensure that your data is actively reviewed, curated, formatted, indexed, given a DOI and publicly available to all upon publication. You are encouraged to submit your article for Data in Brief as an additional item directly alongside the revised version of your manuscript. If your research article is accepted, your data article will automatically be transferred over to Data in Brief where it will be editorially reviewed and published in the open access data journal, Data in Brief. Please note an open access fee of 500 USD is payable for publication in Data in Brief. Full details can be found on the Data in Brief website. Please use this template to write your Data in Brief.

Data statement
To foster transparency, we encourage you to state the availability of your data in your submission. This may be a requirement of your funding body or institution. If your data is unavailable to access or unsuitable to post, you will have the opportunity to indicate why during the submission process, for example by stating that the research data is confidential. The statement will appear with your published article on ScienceDirect. For more information, visit the Data Statement page.

Online proof correction

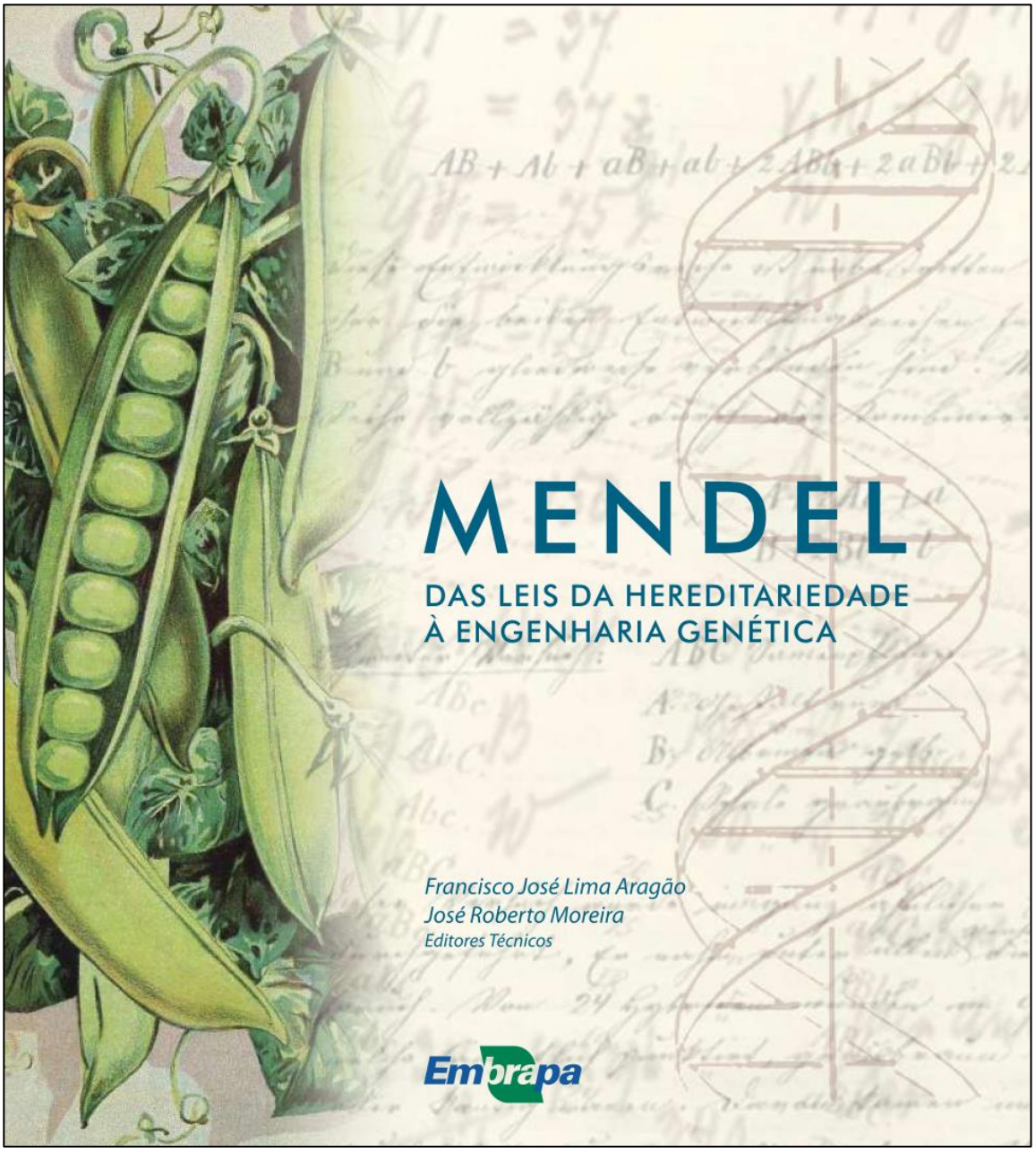
Corresponding authors will receive an e-mail with a link to our online proofing system, allowing annotation and correction of proofs online. The environment is similar to MS Word: in addition to editing text, you can also comment on figures/tables and answer questions from the Copy Editor. Web-based proofing provides a faster and less error-prone process by allowing you to directly type your corrections, eliminating the potential introduction of errors.
If preferred, you can still choose to annotate and upload your edits on the PDF version. All instructions for proofing will be given in the e-mail we send to authors, including alternative methods to the online version and PDF.
We will do everything possible to get your article published quickly and accurately. Please use this proof only for checking the typesetting, editing, completeness and correctness of the text, tables and figures. Significant changes to the article as accepted for publication will only be considered at this stage with permission from the Editor. It is important to ensure that all corrections are sent back to us in one communication. Please check carefully before replying, as inclusion of any subsequent corrections cannot be guaranteed. Proofreading is solely your responsibility.

Offprints

The corresponding author will, at no cost, receive a customized Share Link providing 50 days free access to the final published version of the article on ScienceDirect. The Share Link can be used for sharing the article via any communication channel, including email and social media. For an extra charge, paper offprints can be ordered via the offprint order form which is sent once the article is accepted for publication. Both corresponding and co-authors may order offprints at any time via Elsevier's Webshop. Corresponding authors who have published their article gold open access do not receive a Share Link as their final published version of the article is available open access on ScienceDirect and can be shared through the article DOI link.

Visit the Elsevier Support Center to find the answers you need. Here you will find everything from Frequently Asked Questions to ways to get in touch.
You can also check the status of your submitted article or find out when your accepted article will be published.

Anexo II. Capítulo de livro publicado: **"Mendel: Das leis de Hereditariedade à engenharia genética"**



Empresa Brasileira de Pesquisa Agropecuária
Embrapa Recursos Genéticos e Biotecnologia
Ministério da Agricultura, Pecuária e Abastecimento

MENDEL

DAS LEIS DA HEREDITARIEDADE À ENGENHARIA GENÉTICA

Francisco José Lima Aragão
José Roberto Moreira
Editores Técnicos

Embrapa
Brasília, DF
2017

Exemplares desta publicação podem ser adquiridos na:

Embrapa Recursos Genéticos e Biotecnologia
Parque Estação Biológica (PqEB)
Av. W5 Norte (final)
CEP 70770-917 Brasília, DF
Fone: (61) 3448-4700
Fax: (61) 3340-3624
www.embrapa.br
www.embrapa.br/fale-conosco/sac

Unidade responsável pelo conteúdo
Embrapa Recursos Genéticos e Biotecnologia

Embrapa Informação Tecnológica
Parque Estação Biológica (PqEB)
Av. W3 Norte (Final)
CEP 70770-901 Brasília, DF
Fone: (61) 3448-4236
Fax: (61) 3448-2494
www.embrapa.br/livraria
livraria@embrapa.br

Unidade responsável pela edição
Embrapa Informação Tecnológica

Coordenação editorial
Selma Lúcia Lira Beltrão
Lucilene Maria de Andrade
Nilda Maria da Cunha Sette

Supervisão editorial
Erika do Carmo Lima Ferreira

Revisão de texto
Letícia Ludwig Loder

Normalização bibliográfica
Márcia Maria Pereira de Souza

Projeto gráfico e capa
Carlos Eduardo Felice Barbeiro

1ª edição
1ª impressão (2017): 1.000 exemplares

Dados Internacionais de Catalogação na Publicação (CIP)
Embrapa Informação Tecnológica

Mendel: das leis da hereditariedade à engenharia genética / Francisco José Lima Aragão, José Roberto Moreira, editores técnicos. – Brasília, DF : Embrapa, 2017.
502 p. : il. color. ; 21 cm x 23 cm.

ISBN 978-85-7035-705-2

1. Pesquisa agrícola. 2. Cientista. 3. Genética. 4. Engenharia Genética. I. Aragão, Francisco José Lima. II. Moreira, José Roberto. III. Embrapa Recursos Genéticos e Biotecnologia.

CDD 630.5

Autores

Alexandre Rodrigues Caetano
Zootecnista, Ph.D. em Genética Animal, pesquisador da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF

Ana C. Wanderley-Nogueira
Bióloga, doutora em Ciências Biológicas, bolsista de pós-doutorado da Universidade Federal de Pernambuco, Recife, PE

Ana Christina Brasileiro-Vidal
Engenheira-agrônoma, doutora em Ciências Biológicas, professora da Universidade Federal de Pernambuco, Recife, PE

Ana M. Benko-Iseppon
Bióloga, Ph.D. em Ciências Naturais, professora da Universidade Federal de Pernambuco, Recife, PE

Ângela de Fátima Barbosa Abreu
Engenheira-agrônoma, doutora em Genética e Melhoramento de Plantas, pesquisadora da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO

Augusto Lima Diniz
Biólogo, Ph.D. em Genética e Melhoramento de Plantas, Esalq, Universidade de São Paulo, Piracicaba, SP

Bruna Line Carvalho
Engenheira-agrônoma, doutora em Genética e Melhoramento de Plantas, pesquisadora da Monsanto do Brasil, Rolândia, PR

Bruna Piereck Moura
Bióloga, mestre em Genética, doutoranda da Universidade Federal de Pernambuco, Recife, PE

Elibio L. Rech
Engenheiro-agrônomo, Ph.D. em Biologia Molecular e Celular, pesquisador da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF

Fabiana A. S. Brandão
Biomédica, Ph.D. em Biologia Molecular e Celular, pesquisadora da Universidade de Brasília, Brasília, DF

Fernando Flores Cardoso
Médico-veterinário, Ph.D. em Ciências Animais, pesquisador da Embrapa Pecuária Sul, Bagé, RS

Francisco G. Nóbrega
Médico, Ph.D. em Bioquímica, professor aposentado da Universidade de São Paulo, São Paulo, SP

Francisco José Lima Aragão
Engenheiro-agrônomo, Ph.D. em Biologia Molecular e Celular, pesquisador da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF

Gianna Carvalheira
Bióloga, Ph.D em Ciências Biológicas, Universidade Federal de São Paulo, São Paulo, SP

Glaucia Barbosa Cabral
Engenheira-agrônoma, Ph.D. em Biologia, pesquisadora da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF

Hayana M. de A. Azevedo
Bióloga, doutora em Ciências Biológicas pela Universidade Federal de Pernambuco, Recife, PE

Helaine Carrer
Engenheira-agrônoma, Ph.D. em Biologia Molecular de Plantas, professora da Esalq, Universidade de São Paulo, Piracicaba, SP

Jeferson Gross
Geneticista, Ph.D. em Genética Molecular de Plantas, pesquisador do Instituto de Pesquisas em Bioenergia, Unesp, Rio Claro, SP

João P. B. Neto
Biólogo, doutor em Ciências Biológicas, bolsista de pós-doutorado da Universidade Federal de Pernambuco, Recife, PE

José Roberto Moreira
Engenheiro-agrônomo, Ph.D. em Ecologia, pesquisador da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF

Larissa Arrais Guimarães
Bióloga, Ph.D em Biologia Molecular, pesquisadora da Universidade da Georgia, Georgia, Estados Unidos

Lidiane L. B. Amorim
Bióloga, doutora em Ciências Biológicas, professora do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Oeiras, PI

Luciana Magalhães Melo
Médica-veterinária, doutora em Biologia Molecular, professora da Universidade Estadual do Ceará e da Faculdade Metropolitana da Grande Fortaleza, Fortaleza, CE

Luiz Ricardo Hanai
Bacharel em Ciência Molecular, Ph.D. em Genética e Melhoramento de Plantas, pesquisador da Coodetec, filiada da Dow AgroSciences Sementes & Biotecnologia Brasil Ltda., Cascavel, PR

Luiz Sérgio de Almeida Camargo
Médico-veterinário, doutor em Reprodução Animal, pesquisador da Embrapa Gado de Leite, Juiz de Fora, MG

Magno Antonio Patto Ramalho
Engenheiro-agrônomo, doutor em Genética e Melhoramento de Plantas, professor da Universidade Federal de Lavras, Lavras, MG

Marcio José Poças-Fonseca
Biólogo, doutor em Biologia Molecular, professor da Universidade de Brasília, Brasília, DF

Maria Lúcia Carneiro Vieira
Bióloga, Ph.D. em Genética e Melhoramento de Plantas, Professora Titular do Departamento de Genética da Esalq, Universidade de São Paulo, Piracicaba, SP

Mariana Moysés-Oliveira
Biomédica, doutoranda em Genética Humana da Universidade Federal de São Paulo, São Paulo, SP

Marx O. de Lima
Biólogo, mestre em Ciências Biológicas pela Universidade Federal de Pernambuco, Recife, PE

Mirella Pupo Santos
Bióloga, doutora em Genética e Melhoramento de Plantas, professora da Universidade Federal do Rio de Janeiro, Núcleo em Ecologia e Desenvolvimento Socioambiental de Macaé, Macaé, RJ

Pablo Rodrigues Gonçalves
Biólogo, doutor em Zoologia, professor da Universidade Federal do Rio de Janeiro, Núcleo em Ecologia e Desenvolvimento Socioambiental de Macaé, Macaé, RJ

Patrícia Biegelmeyer
Médica-veterinária, doutora em Zootecnia, pós-doutoranda na Embrapa Pecuária Sul, Bagé, RS

Ronyere Olegário de Araújo
Biólogo, doutor em Ciências Animais, professor do Instituto Tocantinense Presidente Antônio Carlos, Porto Nacional, TO

Samuel Rezende Paiva
Biólogo, doutor em Genética e Melhoramento Animal, pesquisador da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF

Santelmo Vasconcelos
Biólogo, doutor em Biologia Vegetal, pesquisador do Instituto Tecnológico Vale, Belém, PA

Valesca Pandolfi
Bióloga, doutora em Ciências pela Universidade de São Paulo, Piracicaba, SP

Vera Tavares de Campos Carneiro
Bióloga, doutora em Biologia Celular e Molecular Vegetal, pesquisadora do Laboratório de Reprodução Vegetal da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF

Vicente José de Figueirêdo Freitas
Médico-veterinário, doutor em Reprodução Animal, professor da Universidade Estadual do Ceará, Fortaleza, CE

Werner Arber
Químico e físico, Ph.D., prêmio Nobel de Fisiologia e Medicina (1978), professor da Biozentrum der Universität, Basiléia, Suíça.

Sumário

Capítulo 1 **A tradução do artigo seminal de Gregor Mendel** 17

Capítulo 2 **Legado e controvérsias de Gregor Mendel** 43

Capítulo 3 **Mendel e a fundação da genética** 79

Capítulo 4 **Mendel e a produção de grãos no Brasil** 85

Capítulo 5 **Citogenética: da teoria cromossômica da hereditariedade à citogenética molecular** 101

Capítulo 6 **Para além das leis de Mendel: genomas das organelas, plastídio e mitocôndria** 139

Capítulo 7 **Apomixia: a resposta ao enigma de Mendel** 177

Capítulo 8 **Aplicações práticas de dados genômicos no melhoramento genético animal: progressos recentes e perspectivas futuras** 193

219 Capítulo 9 **De Mendel a Hennig: a incorporação de dados genéticos na sistemática filogenética**

239 Capítulo 10 **Marcadores moleculares: de Gregor Mendel à era genômica**

281 Capítulo 11 **Mendel e a herança epigenética: a pré-programação parental**

301 Capítulo 12 **Mendel e suas exceções à luz das ômicas e da biologia de sistemas**

373 Capítulo 13 **Biodiversidade como fonte para a domesticação sintética de características genéticas**

385 Capítulo 14 **A importância dos fatores de Mendel: aspectos genéticos, moleculares e fisiológicos relacionados ao estresse abiótico**

445 Capítulo 15 **Engenharia genética aplicada à agricultura**

473 Capítulo 16 **Engenharia genética em animais**

Capítulo 12

Mendel e suas exceções à luz das ômicas e da biologia de sistemas

Ana M. Benko-Iseppon
Ana C. Wanderley-Nogueira
Santelmo Vasconcelos
Lidiane L. B. Amorim
Bruna Piereck Moura
João P. B. Neto
Marx O. de Lima
Hayana M. de A. Azevedo
Valesca Pandolfi

Introdução

Em vida, Gregor Mendel nunca recebeu reconhecimento por seus estudos, nem mesmo prêmios ou homenagens. Apesar disso, hoje é praticamente impossível estudar até mesmo os aspectos mais básicos da biologia sem encontrar seus – agora famosos e polêmicos – experimentos com plantas de ervilha (espécie botânica *Pisum sativum* L. da família Fabaceae).

Um dos aspectos que distinguiram Mendel de outros naturalistas de sua época foi o seu excelente conhecimento de matemática. Embora a matemática envolvida na genética mendeliana básica não seja profunda ou mesmo remotamente difícil, merece destaque o fato de que Mendel guardava meticulosas anotações sobre suas experiências [estima-se que tenha cultivado mais de 33.500 plantas de ervilhas no período de estudos mais intensos, entre 1856 e 1863 (HARTL; JONES, 1998)]. Ele documentava experimentos que outros possivelmente considerariam rotineiros, mas que lhe permitiram reconhecer padrões que o levaram à sua descoberta. Assim, pode-se dizer que os experimentos de Mendel estão entre os mais bem desenhados, acuradamente executados e elegantemente interpretados da história da ciência experimental e que sua redescoberta lançou as bases para uma moderna ciência: a genética.

Entre suas principais conclusões, Mendel mostrou que, em cruzamento de linhagens puras de ervilhas (uma com característica dominante e outra, recessiva), o híbrido era idêntico ao parental dominante (não havia mistura das características). Da mesma forma, quando cruzados os híbridos, a descendência era constituída tanto por indivíduos dominantes quanto por recessivos.

Mendel não encontrou evidência de que os fatores (genes) hereditários fossem modificados por sua associação nos híbridos parentais. Cada caráter seria condicionado por um par de fatores (genes), que segregavam e recombinavam-se

Anexo III - Aceito Para Publicação

Antimicrobial peptides as plant defense weapons: characterization and functional genomics

Chapter 7 of the book entitled:
Applied Plant Biotechnology for Improving Resistance to Biotic Stress. Editors: Poltronieri, P. & Hong, Y. Elsevier, 1600 John F. Kennedy Boulevard Suite 1800 Philadelphia, USA.

Submitted: 20th August 2018

Authors:

Carlos André dos Santos Silva[1]; **Marx Oliveira de Lima[1]**; Lívia Maria Batista Vilela[1]; João Pacifico Bezerra Neto[1]; José Ribamar Ferreira Neto[1,2]; José Diogo Cavalcanti[1]; Roberta Lane de Oliveira Silva[1]; Carolline de Jesús Pires[1]; Flavia Figueira Aburjaile[1]; Marianne Firmino de Oliveira[1]; Ederson Akio Kido[2] and Ana Maria Benko-Iseppon[1]

Authors Affiliation

[1]Universidade Federal de Pernambuco, Departamento de Genética, Laboratório de Genética e Biotecnologia Vegetal, Recife, Pernambuco, Brazil.

[2]Universidade Federal de Pernambuco, Departamento de Genética, Laboratório de Genética Molecular, Recife, Pernambuco, Brazil.

Abstract

Even before the perception or the contact with pathogens, plants rely on guardian molecules that may be expressed constitutively, in a tissue/stage specific manner or, still, induced after pathogen expression. They are antimicrobial peptides (AMPs), small molecules, generally rich in cysteines, capable of preventing the establishment of invading pathogens. Some of these peptides are shared with other eukaryotes such as defensins and cyclotides, or, still, restricted to the plant kingdom (as snakins). In turn, other AMP classes are specific to some plant taxonomic groups (such as heveins). Even when shared with other groups of organisms, a given AMP class presents a much higher number of isoforms in plants, due to gene duplications or genomic redundancy, an occurrence possibly also associated to the sessile habit of plants, which prevents them from evading biotic or environmental stresses. Therefore, plants are the resource of new AMP molecules. This chapter compiles information on their structural features at genomic, transcriptomic and protein level. Since AMPs are often difficult to recognize, based on simple BLAST alignments, a description of database resources and bioinformatics tools available for their identification is provided. Finally, we highlight the still almost unexplored biotechnological potential of AMPs in the generation of both transgenic plants resistant to pathogens, and new drugs or bioactive compounds for treatment of human and animal diseases.

Keywords: Defensin, Lipid Transfer Protein, Hevein, Cyclotide, Snakin, Knotin, Macadamia β-barrelins, Impatiens-Like, Puroindoline, Thaumatin.

Introduction

Proteins and peptides are formed by small subunits composed of amino acids (aa), forming a chain that can range from a few tens to thousands aa. Peptides are conventionally understood as having less than 50 aa (Marmiroli and Maestri 2014). Proteins, on the other hand, would be any molecule presenting higher amino acid content, being widely studied in plants. A data-mining at the PubMed database (https://www.ncbi.nlm.nih.gov/pubmed) was carried out in April/2018 using the keywords "Plants AND Proteomics", returned about 5900 articles. However, when using the keywords "Plants AND Peptidomics" only 30 publications returned. This difference evidences the discrepancy between systematic research on plant proteins and peptides, despite the relevant roles of both categories.

It is well known that the biochemical machinery necessary for the synthesis and metabolism of peptides is present in every living organism. From this machinery, a wide diversity of peptides is generated, justifying the growing interest in their study. In animals, peptides are prevalent in intercellular communication, performing as hormones, growth factors and neuropeptides (Germain, Chevalier, and Matton 2006).

In several pathogenic microorganisms, peptides can serve as classical virulence factors which disrupt the epithelial barrier, damage cells and activate or modulate host immune responses. Example of this performance is verified with Candidalysin (Moyes et al. 2016), a fungal cytolytic peptide toxin found in the pathogenic fungus *Candida albicans*. This secreted toxin directly damages epithelial membranes, triggers a danger response signaling pathway and activates epithelial immunity. There are also reports of fungal peptides helping in the organism defense. For example, the Copsin, a peptide-based fungal antibiotic recently identified in the fungus *Coprinopsis cinerea* (Essig et al. 2014) acts killing bacteria by inhibiting their cell wall synthesis. Regarding some bacterial peptides, certain species from the gastrointestinal microbial community can release low-molecular-weight peptides, which trigger immune responses (Singh et al. 2009). Another example includes probiotic bacteria-derived proteases that can degrade cow milk casein and thereby generate peptides with suppressive effects on the lymphocyte proliferation in healthy individuals (Singh et al. 2009). There are additionally peptides that function as bacterial "hormones" that allow bacterial communities to organize multicellular behavior such as biofilm formation (Flaherty, Freed, and Lee 2014).

Plant peptides, in turn, can also be multifunctional, and according to Farrokhi, Whitelegge, and Brusslan (2008) they could be classified into two main categories (**Figure 7.1**):

- Peptides with no bioactivity, primarily resulting from the degradation of proteins by proteolytic enzymes, aiming at their recycling;
- Bioactive peptides (BP), which are encrypted in the structure of the parent proteins and are released mainly by enzymatic processes.

The first group is innocuous regarding signaling, regulatory functions, and bioactivity. So far, it has been known that some of them may play an important role in nitrogen mobilization across cellular membranes (Higgins and Payne 1982). The second group (BP) has a substantial impact on the plant cell physiology, and some peptides can act in the plant growth regulation (through cell-to-cell signaling), endurance against pests and pathogens by acting as toxins or elicitors, or even detoxification of heavy metals by ion-sequestration.

Comprising BPs, additional subcategorizations have been proposed. Tavormina et al. (2015), based on the type of precursor, divided BPs into three groups (**Figure 7.1**):

- Derived from functional precursors: originated from a functional precursor protein;
- Derived from nonfunctional precursors: originated from a longer precursor that has no known biological function (as a preprotein, proprotein, or preproprotein);
- Not derived from a precursor protein: some sORFs (small Open Read Frames; usually <100 codons) are considered to represent a potential new source of functional peptides (known as ‘short peptides encoded by sORFs’);

In turn Farrokhi, Whitelegge and Brusslan (2008) present a more intuitive classification of BPs, according to their intracellular role (**Figure 7.1**):

- Phytohormone peptides: the characteristic feature of these peptides is the regulation of fundamental plant physiological processes. They can be classified into those with signaling roles in non-defense functions or those with signaling roles in plant defense. Concerning the first group (**Figure 7.1**), the peptide CLE25 (CLAVATA3/EMBRYO-SURROUNDING REGION-RELATED 25) is one of the representatives. This peptide transmits water-deficiency signals through vascular tissues in *Arabidopsis thaliana*, affecting abscisic acid biosynthesis and stomatal control of transpiration in association with BAM (BARELY ANY MERISTEM) receptors in leaves (Takahashi et al. 2018). Another example is the PLS (POLARIS)

peptide that acts during early embryogenesis but later activates auxin synthesis, also affecting the cytokines synthesis and the ethylene response (J. Liu et al. 2013). Regarding the second group, it includes peptides with signaling roles in plant defense, comprising at least four subgroups, including SYST (systemin) (**Figure 1-1**). SYST peptides were identified in Solanaceae members, as tomato and potato (Pearce, Bhattacharya, and Chen 2008) acting on the signaling response to herbivory. SYST leads to the production of a plant protease inhibitor that suppresses insect´s proteases (Ryan 1990). Stratmann (2003) suggested that in plants SYSTs act to stimulate the jasmonic acid signaling cascade within vascular tissues to induce a systemic wound response.

- Defense peptides or antimicrobials peptides (AMPs): to be fitted into this class, a plant peptide must fulfill some specific biochemical and genetic prerequisites. Regarding biochemical feature, an *in vitro* antimicrobial activity is required. Concerning the genetic condition, the gene encoding the peptide should be inducted in the presence of infectious agents (García-Olmedo et al. 1998). In practice this last requirement is not ever fulfilled since some AMPs are tissue-specific and are considered as part of the plant innate immunity, while other isoforms of the same class appear induced after pathogen inoculation (Benko-Iseppon et al. 2010).

Plant AMPs are the central focus of this chapter, comprising information on their structural features (at genomic, gene and protein levels), resources and bioinformatics tools available. Their biotechnological potential is also highlighted in the generation of both transgenic plants resistant to pathogens, and new drugs or bioactive compounds.

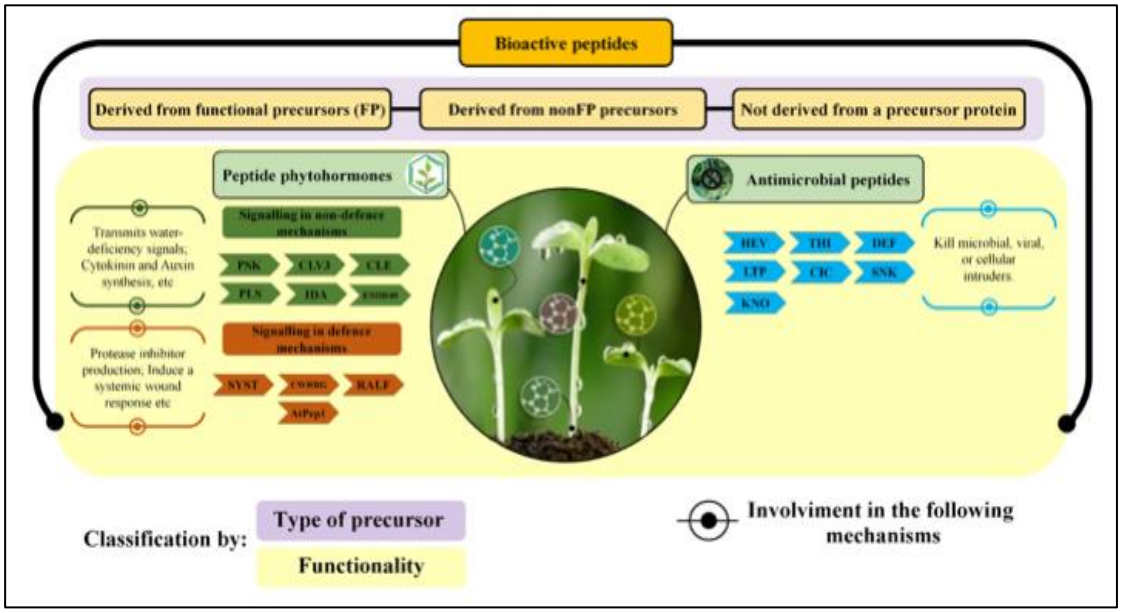


Figure 1-1. Plant bioactive peptides (BP): classification system and some biological processes involved. Tavormina et al. (2015), considering the BP precursors, classified them into: (1) derived from functional precursors or (2) nonfunctional precursors, and (3) not derived from a precursor protein. Farrokhi, Whitelegge, and Brusslan (2008), based on functional groups, classified them into peptide phytohormones (with signaling in defense or non-defense mechanisms) and antimicrobial peptides.

Overall Features of Plant AMPs

AMPs are ubiquitous host defense weapons against microbial pathogens. The overall plant AMP characterization regards the following variables (**Figure 1-2**): electrical charge, hydrophilicity, secondary and three-dimensional structures, and the abundance or spatial pattern of cysteine residues (Benko-Iseppon et al. 2010). These features are primarily related to their defensive role(s) as membrane-active antifungal, antibacterial or antiviral peptide(s).

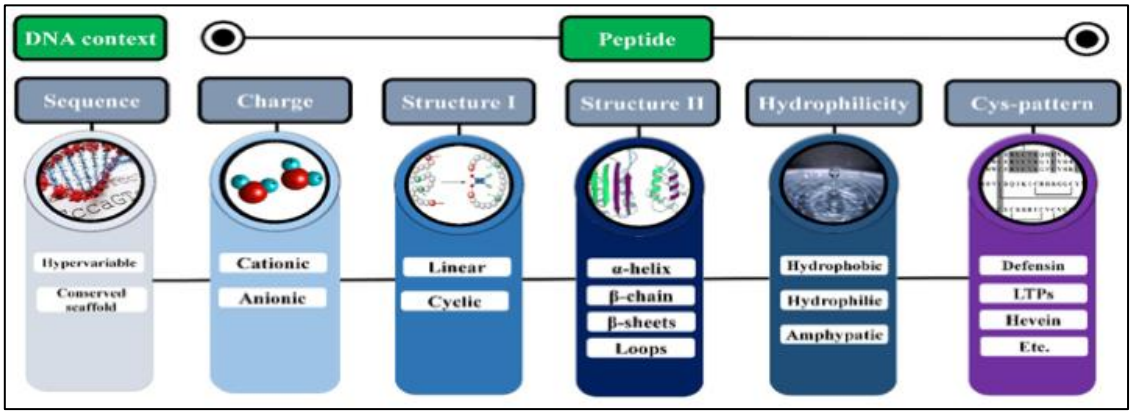


Figure 1-2. Plant antimicrobial peptides features considering DNA sequence level, protein structure and physicochemical properties.

Regarding the nucleotide sequence, plant AMPs are hypervariable (**Figure 1-2**). This genetic variability provides diversity and the ability to recognize different targets. Regarding their charges, AMPs can be classified as cationic or anionic (**Figure 1-2**), but most of the plant AMPs have positive charges, which is a fundamental feature for the interaction with membrane lipid head groups of pathogens (Kaas, Westermann and Craik 2010). Concerning hydrophilicity, AMPs are generally amphipathic, i.e., they exhibit molecular conformation with both hydrophilic and hydrophobic domains (Yu et al. 2013). In regard to their tridimensional structure, AMPs can be either linear or cyclic (**Figure 1-2**). Some linear AMPs adopt an amphipathic α-helical conformation whereas non-α-helical linear peptides generally show one or two predominant amino acids (Zanetti 2004). In turn, cyclic AMPs, including cysteine-containing peptides, can be divided into two subgroups, based on the presence of single or multiple disulfide bonds. A usual feature of these peptides is a cationic and amphipathic character, which accounts for their functioning as membrane- permeabilizing agents (Zanetti 2004).

Considering the secondary structures, AMPs may exhibit α-helices, β-chains, β-pleated sheets and loops (**Figure 1-2**). Wang (2010) classified plant AMPs into four families (α, β, αβ, and non-αβ), based on the protein classification of Murzin et al. (1995), with some modifications. AMPs of the 'α' family present α-helical structures (Wang, Wacklin and Craik 2012), whereas AMPs from the 'β' family contains β-sheet structures usually stabilized by disulfide bonds (Willem F. Broekaert et al. 1997; Wang, Wacklin and Craik 2012). Some plant AMPs showing a α-hairpinin motif formed by antiparallel α-helices are stabilized by two disulfide bridges (Terras et al. 1992). Such AMPs present a higher resistance to enzymatic, chemical or thermal degradation (Vriens, Cammue and

Thevissen 2014). AMPs from the ‘αβ’ family having ‘α’ and ‘β’ structures are also stabilized by disulfide bridges. An example of AMP presenting ‘αβ’ structures are defensins, usually composed of a cysteine-stabilized αβ motif (CSαβ), an α-helix and a triple-stranded antiparallel β sheet stabilized mostly by four disulfide bonds (Bruix et al. 1995). Finally, AMPs that do not belong to the ‘αβ’ group exhibit no clearly defined ‘α’ or ‘β’ structures (Wang, Wacklin and Craik 2012).

Plant AMPs are also classified into families, considering protein sequence similarity, cysteine motifs and distinctive patterns of disulfide bonds, which determine the folding of the tertiary structure (Tam et al. 2015). Therefore, plant AMPs are commonly grouped as thionins, defensins, heveins, knottins (linear and cyclic), lipid transport proteins, snakins and cyclotides (Broekaert et al. 1997; Tam et al. 2015). Some of these AMP groups will be detailed in the next sections, with comments on their functions, tissue-specificity, and scientific data availability.

Defensin

Both plants and animals produce defensins in response to microbial challenges (Wong, Xia and Ng 2007), being present also in microorganisms. Such peptides comprise a well-known group of plant AMPs showing membranolytic function (Tam et al. 2015). They are abundant and known for their antimicrobial activities even at low concentrations against bacteria (gram-positive and negative), fungi, viruses and parasitic protozoa (Brogden 2005). Usually, defensins are located in plant parts often exposed to pathogens (Holly, Diaz and Smith 2017), such as leaves, roots, barks, pods, tubers, fruits and floral tissues (Lay 2003).

A data mining performed in the PubMed database, with the keywords "plant AND defensins" presented 790 scientific publications covering diverse approaches, from genetic diversity, gene identification/functional analysis, and protein isolation/purification methods. The current knowledge of plant defensins reflects the importance of this AMP family and it’s potential for biotechnological manipulation considering new therapeutic properties or the host fighting against microbial pathogens.

Lipid Transfer Protein (LTP)

A data-mining performed in the PubMed database using the keywords "Plant AND Lipid Transfer Proteins" revealed more than 750 scientific manuscripts.

LTPs are encoded by all land plants, being expressed in the majority of plant tissues (Salminen, Blomqvist and Edqvist 2016). LTPs are considered important proteins also for the plant establishment, colonization and survival on the soil. Generally, LTPs are present in extracellular spaces; however, many LTPs are also found in the cell wall (Tsuboi et al. 1992; Carvalho et al. 2004). Additionally, they can be localized in cellular organelles, such as the glyoxysomes (Tsuboi et al. 1992) and vacuoles (Carvalho et al. 2004). The role of LTPs is not entirely understood. However, published data suggest their importance in the assembly of physical barriers against water by depositing monomers that form waxes (such as cutin and suberin) in plant cell walls (Domínguez, Heredia-Guerrero and Heredia 2015). Scientific data also suggest the association of LTPs with signaling against infection by fungi, bacteria, and viruses (Tsuboi et al. 1992; Maldonado et al. 2002; Carvalho and Gomes 2007). Recently, the LTP participation in tolerance mechanisms to abiotic stresses was also reported (Edqvist et al. 2018).

Hevein

As suggested by the nomenclature, heveins were initially isolated from the rubber tree *Hevea brasiliensis* (Archer 1960). To date, they have been identified in more than 20 different plant species (Slavokhotova et al. 2017). Heveins are commonly expressed in plant tissues such as leaves (Games et al. 2016), latex (Archer 1960), and seeds (Gijzen et al. 2001), among others. There are no reports of their identification in other organisms, except plants.

The hevein action is mainly reported against plant pathogens that are rich in chitin, such as fungi. These microorganisms may bind to the structural components of the chitin-binding site present on heveins (Slavokhotova et al. 2017). It is worth noting that heveins may additionally affect microorganisms devoid of these polysaccharides. Hevein action against oomycetes and bacteria has been reported (Cândido et al. 2014). Besides, heveins have attracted the interest of scientists, with a growing number of researchers in recent years; their mechanism of action, however, remains unsolved. A PubMed data mining performed with the terms "Plant AND Heveins" revealed 215 scientific manuscripts, covering several aspects of heveins.

Cyclotide

Cyclotides are peptides found in all kingdoms (Park et al. 2017). A fundamental characteristic regards their cyclic structure, stabilized by three disulfide bridges (Craik et al. 1999). These features make them functional peptides that tolerate high temperatures, and also support some chemical degradation (Colgrave and Craik 2004). Their involvement in defense functions is also well known, since cyclotides exhibit diverse biological activities, including antifungal, insecticidal (Jennings et al. 2001), antibacterial (Pränting et al. 2010), anthelmintic (Colgrave et al. 2008), molluscicidal (Plan et al. 2008), and anti-HIV (Gustafson et al. 1994).

The cyclotide "[T20K] Kalata B1", isolated from *Oldenlandia affinis* (Rubiaceae) leaves showed to be active during *in vivo* assays against multiple sclerosis, preventing the disease progression without adverse effects (Henriques et al. 2015). These results indicate plant cyclotides as potential candidates for drug development (Thell et al. 2016). Additionally, they are relatively small, what facilitates their production by chemical synthesis and heterologous expression in bacterial, yeast or animal cells. All these features turn them into ideal candidates for new candidates and research tools based on peptides (Camarero 2017). Besides their importance, the mechanisms of action of cyclotides remain not yet fully elucidated (Park et al. 2017). However, more than 400 cyclotides have been already described from several plant species, which are available in the CyBase (http://cybase.org.au/), a database dedicated to cyclic peptides. A data-mining in the PubMed database applying the keywords “cyclotides AND plants” revealed 286 scientific articles, covering different subjects, from their chemical characterization (Hashempour et al. 2013), involvement in the dynamics of cell membranes (Wang 2012), and other issues.

Snakin

Snakins are also plant-exclusive AMPs. They were first described in potato (*Solanum tuberosum*), showing 12 conserved cysteine residues, forming six disulfide bridges (Segura et al. 1999; Nahirnak et al. 2012). These AMPs were also described in several crops such as rice (Furukawa, Sakaguchi and Shimada 2006), tobacco (Kovalskaya, Zhao and Hammond 2011), wheat (Rong et al. 2013), tomato (Herbel, Schäfer and Wink 2015), and in the medicinal plant *Peltophorum dubium* (Fabaceae) (Rodríguez-Decuadro et al. 2018).

They were observed expressing in diverse plant tissues as inflorescences (Boonpa et al. 2018), tubers (Segura et al. 1999), leaves, stems and roots (García et al. 2014) and seedlings (Rodríguez-Decuadro et al. 2018). Regarding their subcellular location, the presence of snakins in potato cell walls (Nahirnak et al. 2012) suggested a role in the plant cell defense as a physical obstacle, preventing the invasion of some pathogens (Segura et al. 1999). According to Nahirnak et al. (2012), the silencing of potato snakin gene (StSN1) affected the plant metabolism and the cell wall composition. This result together with those from the scientific literature suggests that snakins may play other functions in plant growth and development, and also in hormonal crosstalk (Nahirnak et al. 2012), besides response to biotic stress (García et al. 2014). They can act against some fungal and bacterial pathogens at micromolar concentrations (EC 50 <10 μM), as observed by snakin SN1 performance in potato plants (Segura et al. 1999). Only a few studies covering the snakin mechanism of action against plant pathogens were published so far, regarding only 36 articles after a search at the PubMed database using the terms "plant AND snakin".

Structural features of AMPs at gene and protein levels

Defensin

The first plant defensins were isolated from wheat (Colilla, Rocher and Mendez 1990) and barley grains (Mendez et al. 1990), initially called γ-hordothionins. Due to some similarities as cysteine content and molecular weight they were classified as γ-thionins. Later, the term ‘γ-thionin’ was replaced by ‘defensin’ based on the higher number of primary and tertiary structures of these proteins and also to their antifungal activities more related to insect and mammalian defensins than to plant thionins (Broekaert et al. 1995).

Plant defensins belong to a diverse protein superfamily called cis-defensin (Parisi et al. 2018) and exhibit cationic charge, consisting of 45 to 54 aa with two to four disulfide bonds (W. F. Broekaert et al. 1995; Janssen et al. 2003; Pelegrini and Franco 2005). The defensins share similar tertiary structures and typically exhibit a triple-stranded antiparallel β sheet, enveloped by an α-helix and confined by intramolecular disulfide bonds (Shafee et al. 2017) (**Figure 1-3**).

Defensins are known for their antimicrobial activity at low micromolar concentrations against Gram-positive and -negative bacteria (Kraszewska et al. 2016),

fungi (Oddepally and Guruprasad 2015), parasitic viruses and protozoa (Brogden 2005). Additionally, they present inhibitory, insecticidal, and antiproliferative activity, acting as an ion-channel blocker, being also associated with protein synthesis (Carvalho and Gomes 2009).

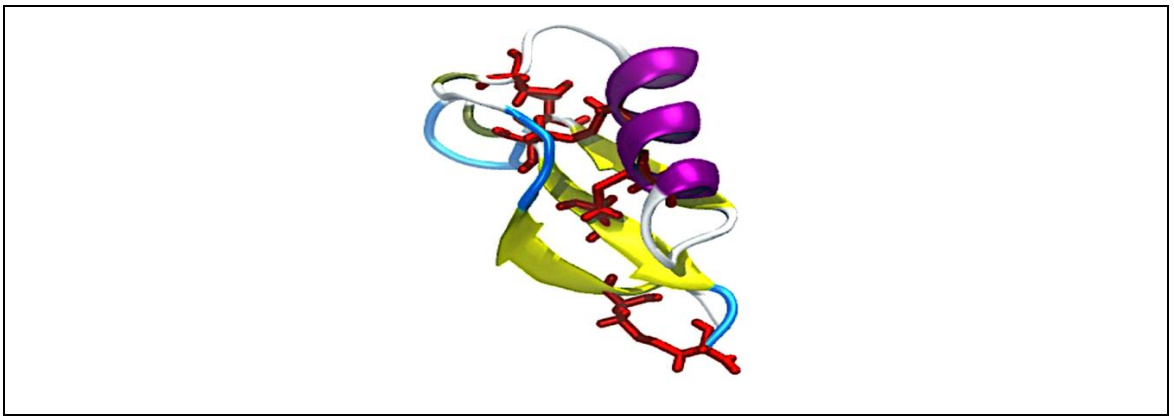


Figure 1-3. Three-dimensional structure NaD1 defensin from *Nicotiana alata* (Solanaceae) (1mr4.pdb available in https://www.rcsb.org/structure/1mr4). In yellow the three β-sheets; in purple the α-helix and in red the three disulfide bonds.

Lipid Transfer Protein (LTP)

Non-specific Lipid Transfer Proteins (nsLTPs) were first isolated from potato tubers (Mazliak et al. 1975) and are actually identified in diverse terrestrial plant species. They concern a large gene family, are abundantly expressed in most tissues, and are absent in most basal plant groups as chlorophyte and charophyte green algae (Edstam et al. 2011). They generally include an N-terminal signal peptide that directs the protein to the apoplastic space (Salminen, Blomqvist and Edqvist 2016). Some LTPs have a C-terminal sequence that allows their post-translational modification with a glycosylphosphatidylinositol molecule, facilitating the integration of LTP on the extracellular side of the plasma membrane.

nsLTPs regard small proteins which were thus named because of their function of transferring lipids between the different membranes carrying lipids (unspecifically, the list includes phospholipids, fatty acids, their acylCoAs or sterols). They have approximately 100 aa and are relatively larger in size than other AMPs, as defensins.

Depending on their molecular mass LTPs may be classified into two subfamilies: LTP1 and LTP2 with relative molecular weight of 9 kDa and 7 kDa, respectively (Kader 1996; Castro et al. 2003). The limited sequence conservation turned this classification

inadequate. Thus, a modified and expanded classification system was proposed, producing five main types (LTP1, LTP2, LTPc, LTPd, and LTPg) and four additional types with a smaller number of members (LTPe, LTPf, LTPh, LTPj, and LTPk) (Edstam et al. 2011). The new classification system is not based on molecular size but rather on (i) the position of a conserved intron; (ii) the identity of the amino acid sequence and (iii) the spacing between the cysteine (Cys) residues (**Figure 1-4)**. Although this latter classification system is the most recent, the conventional classification of types LTP1 and LTP2 has been maintained by most working groups.

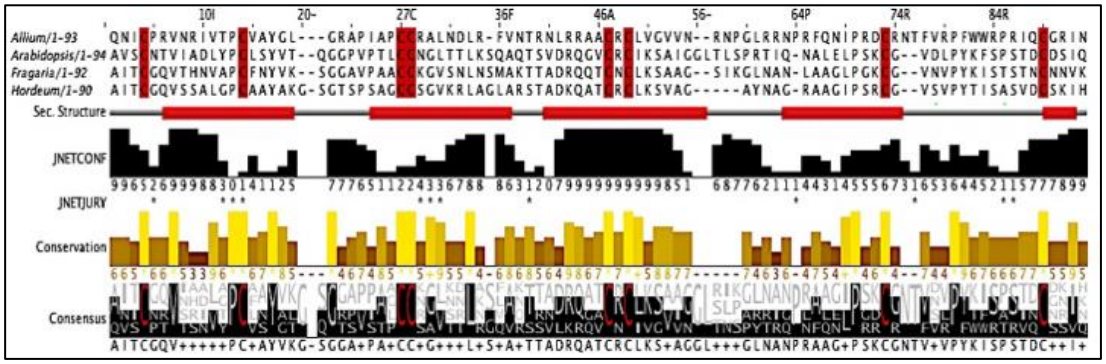


Figure 1-4. Alignment of plant LTP1 sequences from onion, arabidopsis, strawberry and barley. Cysteine residues are highlighted in red in the alignment. Below the alignment, the secondary structure prediction generated by JalView. Bars in yellow/brown show sequence conservation. Amino acid consensus sequence is shown at the bottom of the figure.

LTPs nomenclature has been confusing and without consistent guidelines or standards. There are several examples where specific LTPs receive different names in separate articles. The lack of a robust naming system sometimes made it quite difficult, extremely time-consuming and sometimes frustrating to compare LTPs with different roles (Salminen, Blomqvist and Edqvist 2016).

An additional nomenclature was also proposed by Salminen et al. (2016), that named the LTPs as follows: AtLTP1.3, OsLTP2.4, HvLTPc6, PpLTPd5 and TaLTPg7, with the first two letters indicating the species of plants (At = *Arabidopsis thaliana*, Pp = *Physcomitrella patens*, etc.), whereas LTP1, LTP2, LTPc indicate the type, while the last digit (here 3-7) regard the specific number given to each gene or protein within a given LTP type. For the sake of clarity, they recommend the inclusion of a point between the type specification and the gene number in LTP1 and LTP2. For LTPc, LTPd, LTPg and

other types of LTP defined with a letter; the punctuation mark was not recommended. This latter classification system is currently recommended since it comprises several features of LTPs and is more robust than previous classification systems.

LTPs are small cysteine-rich proteins, having four to five helices in their tertiary structure (**Figure 1-5**), which is stabilized by several hydrogen bonds. Such a folding gives LTPs a hydrophobic cavity to bind the lipids through hydrophobic interactions. This structure is stabilized by four disulfide bridges formed by eight conserved cysteines, similar to defensins, although bound by cysteines in different positions. The disulfide bridges promote LTP folding into a very compact structure, which is extremely stable at different temperatures and denaturing agents (Lindorff-Larsen and Winther 2001; Berecz et al. 2010; Edstam and Edqvist 2014). These foldings provide a different specificity of lipid binding at the LTP binding site, where the LTP2 structure is relatively more flexible and present a lower lipid specificity when compared to LTP1 (Goyal and Mattoo 2016).

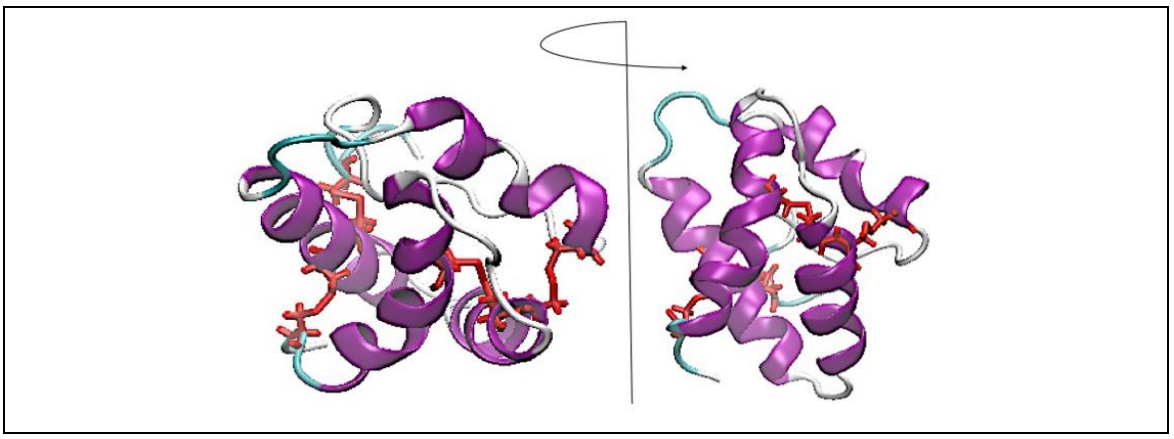


Figure 1-5. Three-dimensional structure of TaLTP1.1 purified from wheat seeds. Model available at: Protein Data Bank (identifier PDB ID: 1GH1). Four helices of TaLTP1.1 are colored pink, while the four conserved disulfide bridges are colored red.

The first three-dimensional structure of an LTP was established for TaLTP1.1 (**Figure 1-5**) based on 2D and 3D data of 1H-NMR, purified from wheat (*Triticum aestivum*) seeds in aqueous solution (Simorre et al. 1991; Gincel et al. 1994). Currently, several three-dimensional structures of LTPs have been determined, either by RMN or X-ray crystallography, either in their free, unbound form or in a complex with ligands.

Heveins

The heveins were first identified in 1960 in *H. brasiliensis*, but its sequence was determined later, whereas a similarity was detected to the chitin-binding domain of an agglutinin isolated from *Urtica dioica* L. (Peumans, De Ley and Broekaert 1984) with eight cysteine residues forming a typical Cys motif (Slavokhotova et al. 2017).

The structure of the hevein consists of 29 to 45 aa, positively charged, with abundant glycine and cysteine residues ranging from 6, 8 to 10 C (Slavokhotova et al. 2017) and aromatic residues (Asensio et al. 2000; Tam et al. 2015). The chitin-binding domain is a determinant component in the identification of hevein-like peptides whose binding site is represented by the amino acid sequence SXFGY / SXYGY, where X regards any amino acid (Wong et al. 2016; Slavokhotova et al. 2017). Most heveins have a coil-β1-β2-coil-β3 structure that occurs by variations with the secondary structural motif in the presence of turns in two long coils in the β3 chain (Tam et al. 2015). Antiparallel β chains form the central β sheet of the hevein motif with two long coils stabilized by disulfide bonds (**Figure 1-6**).

Although the presence of chitin has not been identified in plants, there are chitin-like structures present in proteins that exhibit strong affinity to this polysaccharide isolated from different plant sources (Raikhel, Lee and Broekaert 1993). The presence of three aromatic amino acids in the chitin-binding domain favors chitin binding by providing stability to the hydrophobic group C-H and the π electron system through van der Waals forces, as well as the hydrogen bonds between serine and N-acetylglucosamine (GlcNAc) present in the chitin structure (Asensio et al. 2000; Slavokhotova et al. 2017). This domain is commonly found in chitinases of classes I to V, in addition to other plant antimicrobial proteins, such as lectins and PR-4 (pathogenesis-related protein 4) members (Beintema 1994; Rogozhin et al. 2015). It may also occur in other proteins that bind to polysaccharide chitin (Beintema 1994), such as the antimicrobial AC-AMP1 and AC-AMP2 proteins of *Amaranthus caudatus* (Amaranthaceae) seeds which are homologous to hevein but lack the C-terminal glycosylated region (Broekaert et al. 1992). Among the several classes of proteins mentioned, the proteins with a high degree of similarity to hevein are classes I and IV of chitinases (Slavokhotova et al. 2017).

Chitinases are known to play an essential role in plant defense against pathogens (Iseli, Boller and Neuhaus 1993), also inhibiting fungal growth *in vitro* (Broekaert et al. 1989), especially when combined with β-1,3-glucanases (Leah et al. 1991) and interferes

with the growth of hyphae, resulting in abnormal ramification, delay and swelling in their stretching (Rogozhin et al. 2015). However, it has been shown that heveins have a higher inhibitory potential than chitinases and that their antifungal effect is not only related to the presence of chitinases (Van Parijs et al. 1991). Pn-AMP1 and Pn-AMP2 antimicrobial peptides with hevein domains have potent antifungal activities against a broad spectrum of fungi, including those without chitin in their cell walls (Koo et al. 1998), evidencing the statement raised by Van Parijs et al. (1991). The described modes of action of chitinases usually result in degradation and disruption of the fungal cell walls and plasma membranes due to the hydrolytic action of the enzyme, causing the extravasation of plasma particles (Koo et al. 1998; Kaas et al. 2010). Therefore, heveins have good antifungal activity, and only a few are active against bacteria, most of them with low activity (Lipkin et al. 2005; Rogozhin et al. 2015).

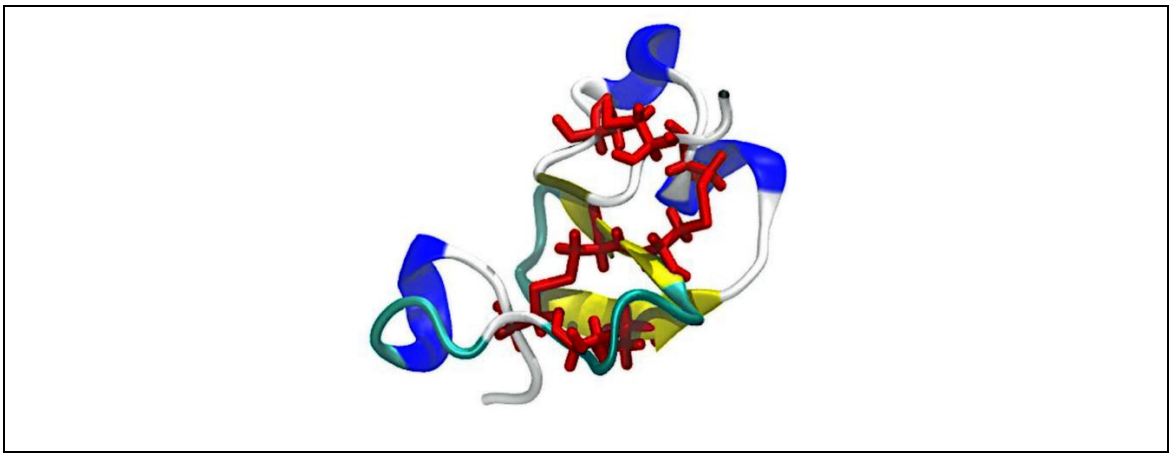


Figure 1-6. Three-dimensional structure of hevein peptide from *Hevea brasiliensis* latex (1hev.pdb available in http://www.rcsb.org/structure/1HEV). In yellow the three β-sheets; in blue three α-helix and in red the three disulfide bonds.

Knottins (Cystine-knot peptides)

Knottins are part of the superfamily of cystine-rich peptides (CRPs), share the cystine-knot motif and therefore resemble other families as defensins, heveins, and cyclotides (Molesini et al. 2017). Their structure was initially identified by crystallography of carboxypeptides isolated from potato, showing the cystine-knot motif, 39 aa and six cysteine residues (Rees and Lipscomb 1982). They are also called cystine-knot peptides, inhibitor cystine-knot peptides or even cystine-knot miniproteins because

their mature peptide presents less than 50 aa, forming three interconnected disulfide bonds in the cystine-knot motif, characterizing a particular scaffold (Molesini et al. 2017). This conformation confers thermal stability at high temperatures. For example, the Cystine Stabilized β-sheet (CSB) motif derived from knottins presents stability at approximately 100ºC with only 2 disulfide bonds (Chiche et al. 2004). The knottins may have linear or cyclic conformation, however, both exhibit connectivity between the cysteines at positions 1-4C, 2-5C, 3-6C, forming a ring at the last bridge (Molesini et al. 2017) (**Figure 1-7**).

In addition, the knottins have different functions, such as signaling molecules (Murphy, Smith and De Smet 2012), response against biotic and abiotic stresses (Li and Asiegbu 2004), root growth (Iyer and Acharya 2011), symbiotic interactions as well as antimicrobial activity against bacteria (Aboye et al. 2015), fungi (Göransson et al. 2012), virus (Gustafson et al. 1994), and insecticidal activity (Jennings et al. 2005), among others. The antimicrobial activity has been attributed to the action on the functional components of the plasma membrane, leading to alterations of lipids, ion flux, and exposed charge (Göransson et al. 2012). The accumulation of peptides on the surface of the membrane results in the weakening of the pathogen membrane (Burman et al. 2011) resulting in transient and toroidal perforations (Göransson et al. 2012).

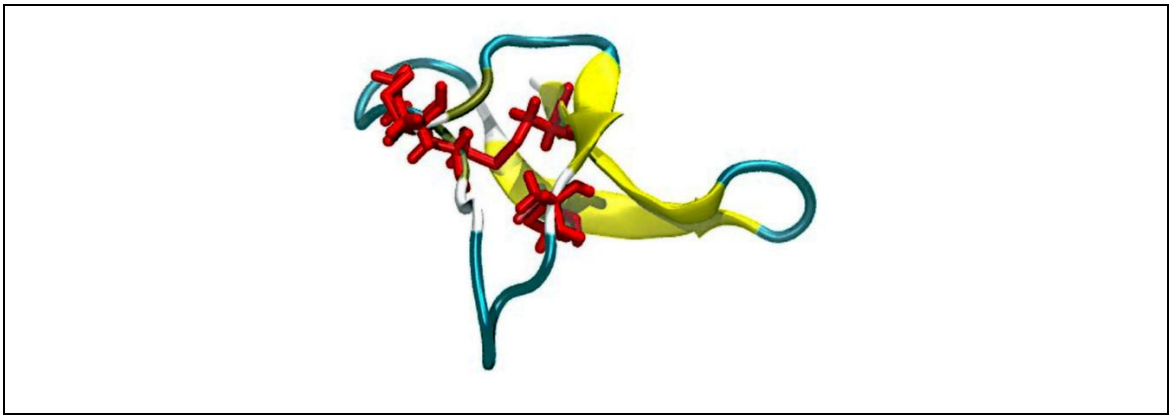


Figure 1-7. Three-dimensional structure of cysteine-knot (or knottin) peptide (Ep-AMP1) from *Echinopsis pachanoi* (Cactaceae; source: 2mfs.pdb available in https://www.rcsb.org/structure/2mfs). In yellow the three β-sheets and in red the three disulfide bonds.

Macadamia (β-barrelins)

In the course of a large-scale survey to identify novel antimicrobial peptides from Australian plants (Marcus et al. 1997; McManus et al. 1999), an antimicrobial peptide with no sequence homology was purified. Its cDNA was cloned from *Macadamia integrifolia* (Proteaceae) seeds, containing the complete peptide coding region. The peptide was named MiAMP1, being highly basic with an estimated PI of 10 and a mass of 8 kDa.

MiAMP1 is 102 aa long, including a 26 aa signal peptide in the N-terminal region, bound to a 76 aa mature region with 6 cysteine residues. Its three-dimensional structure was determined using NMR spectroscopy (McManus et al. 1999) revealing a unique conformation amongst plant AMPs, with eight beta-strands arranged in two Greek key motifs associated to form a Greek key beta-barrel (**Figure 1-8**). Due to its particularities, MiAMP1 was classified as a new structural family of plant AMPs, and the name β-barrelins was proposed for this class (McManus et al. 1999). This structural fold resembles a superfamily of proteins called *γ-crystallin-like* characterized by the precursors *βγ- crystallin* (Ohno et al. 1998). This family includes AMPs from other organisms, for example, WmKT a toxin produced by the wild yeast species *Williopsis mraki* (Antuch, Güntert, and Wüthrich 1996).

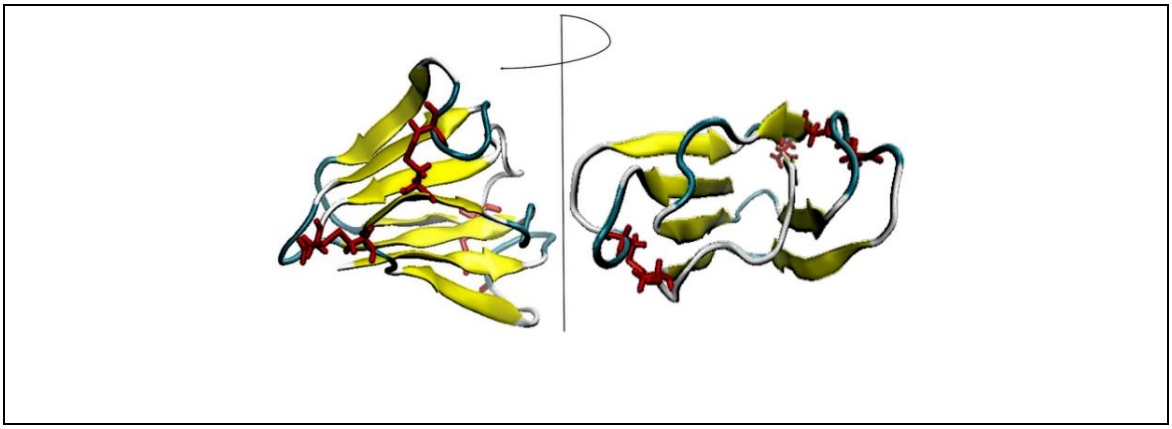


Figure 1-8. Three-dimensional structure of MiAMP1 purified from *Macadamia integrifolia* (source *Protein Data Bank*, PDB ID: 1C01). In yellow the eight beta leaves and in red the three disulfides conserved bridges.

MiAMP1 exhibited *in vitro* antimicrobial activity against various phytopathogenic fungi, oomycetes and gram-positive bacteria (Marcus et al. 1997) with a concentration range of 0.2 to 2 μM generally required for a 50% growth inhibition (IC_{50}). In addition, the transient expression of MiAMP1 in canola (*Brassica napus*) provided resistance against blackleg disease caused by the fungus *Leptosphaeria maculans* (Kazan et al. 2002), turning MiAMP1 potentially useful for genetic engineering aiming at disease resistance in crop plants.

There are few scientific publications with Macadamia-like peptides, maybe because they have been found only in primitive plant groups, being apparently absent or difficult to recognize in derived angiosperms, including most studied model and crop plants. In fact, peptides similar to MiAMP1 appear to play a role in the defense against pathogens in gymnosperms (Manners 2009) including species of economic importance, thus deserving attention for their biotechnological potential.

Impatiens

Four closely related AMPs (Ib-AMP1, Ib-AMP2, Ib-AMP3, and Ib-AMP4) were isolated from seeds of *Impatiens balsamina* (Balsaminaceae) with antimicrobial activity to a variety of fungi and bacteria, with low toxicity to human cells in culture. They present only 20 aa in length and are the smallest antimicrobial peptides isolated from plants to date. Ib-AMPs are highly basic and contain four cysteine residues that form two disulfide bonds, with no significant homology with AMPs available in public databases. Sequencing of cDNAs isolated from *I. balsamina* revealed that all four peptides are encoded within a single transcript. The predicted Ib-AMP precursor protein consists of a pre-peptide followed by six mature peptide domains, each flanked by propeptide domains ranging from 16 to 35 aa in length (**Figure 1-9**). This primary structure with repeated domains of alternating basic peptides and acid propeptide domains was reported before in plants (Tailor et al. 1997).

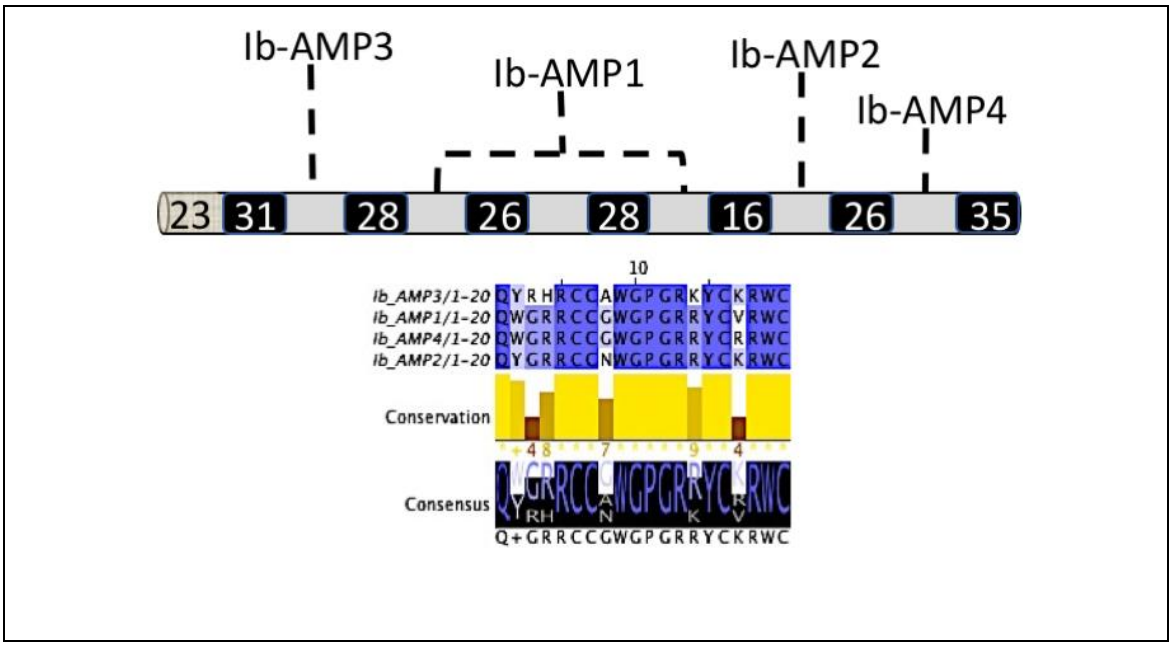


Figure 1-9. General gene structure and alignment of peptide sequences. Figure based on Tailor et al. 1997. (A) Structure of the predicted translation product of 333 amino acids with the individual 20 amino acid domains representing the mature Ib-AMP peptides (in gray), with the Ib-AMPs indicated above the dashed line. The regions of the propeptide are in black, the predetermined region of the pre-peptide (signal peptide) is textured at the beginning of the sequence, and the number of amino acids comprising each of those regions is indicated. (B) Alignment of the predicted translation amino acid sequence of each mature domain. In blue the conserved amino acid residues among Ib-AMPs. The yellow bars show conservation generally between the sequences and below the consensus sequence.

Circular dichroism (CD) and two-dimensional proton nuclear magnetic resonance (1H NMR) were performed in IbAMP1 by Patel et al. (1998) indicating that the peptide presents a turn but shows no evidence of helical or sheet structure over the analyzed temperature and pH range (**Figure 1-10**). NMR structural information was achieved in the form of proton-proton internuclear distances deduced from NOEs (Nuclear Overhauser Enhancement) and dihedral angle restraints from spin-spin coupling constants that were used for distance geometry calculations.

Due to the difficulty in obtaining the correct disulfide bridges by chemical methods, three separate calculations were performed: without disulfide bridges and with two bridge formation alternatives. Calculations have shown that although the peptide is

small, the cysteines restrict part of it to adopt a well-defined chain conformation. From residue 6 to 20, the main chain is well defined, whereas residues 1-5 in the N-terminal region present few restrictions and appear to be more flexible (Patel et al. 1998) .

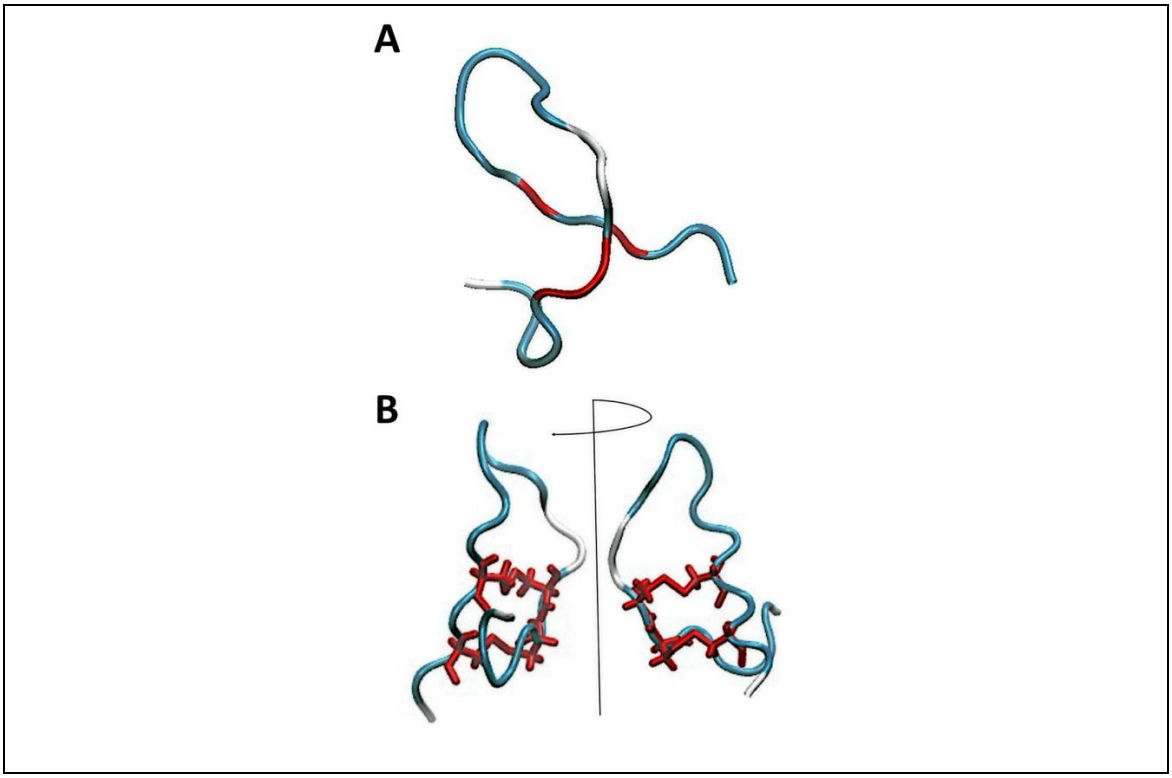


Figura 1-10. (**A**) Tridimensional structure of Ib-AMP1. Alternative model generated by *ab initio* modelling with Rosetta algorithm, evidencing the four conserved cysteines (in red). (**B**) An alternative ab initio model generated with the aid of the Rosetta algorithm, presenting two disulfide bridges.

Little is known about the mode of action of Impatiens-like AMPs. Lee et al. (1999) investigated the antifungal mechanism of Ib-AMP1, noting that when oxidized (bound by disulfide bridges) there was a fourfold increase in antifungal activity against *Aspergillus flavus* and *Candida albicans* as compared to reduced Ib-AMP1 (without disulfide bridges). Confocal microscopy analyzes have shown that Ib-AMP1 binds to the cell surface or penetrates cell membranes, indicating an antifungal activity by inhibiting a distinct cellular process, rather than ion channel or membrane pore formation. Modifications of these peptides have been carried out, indicating that the synthetic variant of Ib-AMP1 is totally active against yeasts and fungi. It has been proposed that the substitution of amino acid residues by arginine or tryptophan may improve by more than

twice the antifungal activity (Thevissen et al. 2005). Another study involving AMP modification generated a synthetic peptide without the disulfide bridges (i.e., a linear analog of Ib-AMP1) which showed an antimicrobial specificity 3.7-4.8 times higher than the wild-type Ib-AMP1 (P. Wang et al. 2009).

Puroindoline

Puroindolines (PINs) are small basic proteins that contain a single domain rich in tryptophan. These proteins were isolated from the wheat endosperm, have molecular masses around 13 kDa and a calculated isoelectric point higher than 10. At least two major isoforms (called PIN-a, and PIN-b) are known, which are encoded by *Pina-D1* and *Pinb-D1* genes, respectively. These genes share 70.2% identical coding regions, but exhibit only 53% identity in the 3' untranslated region (Gautier et al. 1994).

Both PIN-a and PIN-b contain a structure with ten conserved cysteine residues and a tertiary structure similar to LTPs, consisting of four α-helices separated by loops of varying lengths, with the tertiary structure joined by five disulfide bonds, four of which identical to ns-LTPs (Gautier et al. 1994) (**Figure 1-11**).

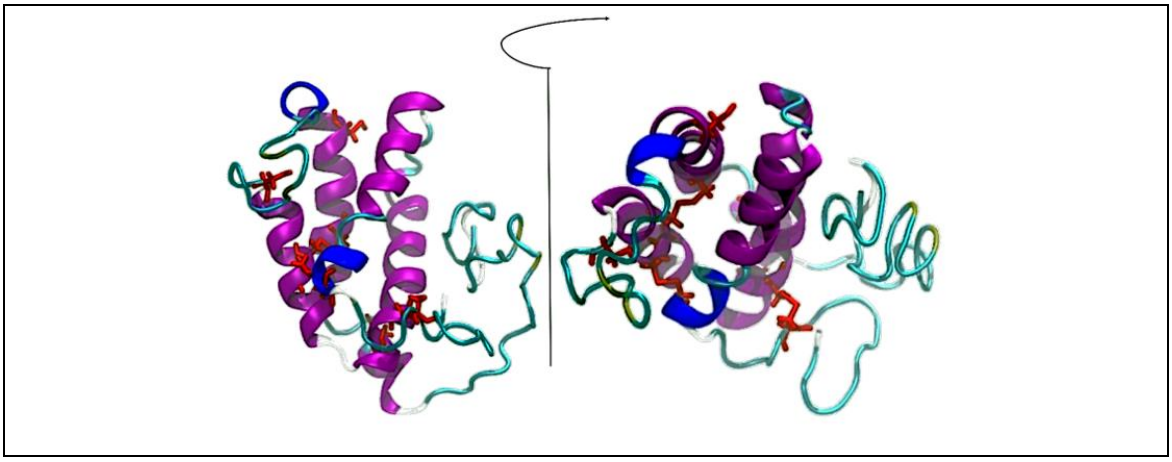


Figure 1-11. Tridimensional structure of PIN-a from *Aegilops kotschyi* (Poaceae). Alternative model generated by *ab initio* modeling using the Rosetta algorithm, showing in pink the four β-sheets and in red conserved cysteine bridges.

The conformation of the two PIN isoforms was studied by infrared and Raman spectroscopy. The results showed that PIN-a and PIN-b have similar secondary structures comprising approximately 30% helices, 30% β-sheets and 40% non-ordered structures at

pH 7. It has been proposed that the folding of both PINs is highly dependent on the pH of the medium. The reduction of the disulfide bridges results in a decrease of puroindolines solubility in water and to an increment of the ß-sheet content by about 15% at the expense of the α-helix content (Le Bihan et al. 1996). There is no high-resolution structure for any of the PIN isoforms, bringing challenges to understanding the function of their hydrophobic regions, with some evidence coming only from homolog peptides (Gautier et al. 1994).

PINs are proposed to be functional components of wheat grain hardness loci, control core texture, besides antifungal activity (Giroux et al. 2003; Bhave and Morris 2008; Dhatwalia et al. 2009; J. Zhang et al. 2011). Although the biological function of PINs is unknown, its involvement in lipid binding has been proposed. While LTPs bind to hydrophobic molecules in a large cavity, PINs interact only with lipid aggregates, i.e., micelles or liposomes, through a single stretch of tryptophan residues. This stretch of tryptophan residues is especially significant in the main form, PIN-a (WRWWKWWK), while it is truncated in the smaller form, PIN-b (WPTWWK) (Douliez et al. 2000; Morris 2002; Marion, Bakan, and Elmorjani 2007).

Puroindolines form protein aggregates in the presence of membrane lipids, and the organization of these aggregates is controlled by the lipid structure. In the absence of lipids, these proteins may aggregate, but there is no accurate information on the relationship between aggregation and interaction with lipids. The antimicrobial activity of PINs is targeted to cell membranes. Charnet et al. (2003) indicated that PIN is capable of forming ion channels in artificial and biological membranes that exhibit some selectivity over monovalent cations. The stress and Ca^{2+} ions modulate the formation and/or opening of channels. PINs may also be membranotoxins which may play a role in the plant defense mechanism against microbial pathogens.

Snakin

Snakins are cysteine-rich peptides first identified in potato (*Solanum tuberosum*) (Segura et al. 1999; Berrocal-Lobo et al. 2002). Due to their sequence similarity to GASA (Gibberellic Acid Stimulated in Arabidopsis) proteins, the snakins were classified as members of the snakin/GASA family (Oliveira-Lima et al. 2017). The genes that encode these peptides have: (i) a signal sequence of approximately 28 aa (ii) a variable region, and (iii) a mature peptide of approximately 60 residues, with 12 highly conserved cysteine

residues. These cysteine residues maintain the three-dimensional structure of the peptide through the disulfide bonds, besides providing stability to the molecule when the plant is under stress (Segura, et al. 1999; Berrocal-Lobo et al. 2002; Mao et al. 2011; Yeung et al. 2016) (Figure 1-12).

Snakins may be expressed in different parts of the plant, like stem, leaves, flowers, seeds and roots (S. Zhang et al. 2009; Almasia et al. 2010; R. Zimmermann, Sakai, and Hochholdinger 2010; Guzmán-Rodríguez et al. 2013), both constitutive or induced by biotic or abiotic stresses. *In vitro* activity was observed against a variety of fungi, bacteria and nematodes, acting as destabilizer of the plasma membrane (Segura et al. 1999; Faccio et al. 2011; Herbel, Schäfer, and Wink 2015). They were reported as important agents in biological processes such as cell division, elongation, cell growth, flowering, embryogenesis, and signaling pathways (Kotilainen et al. 1999; Furukawa, Sakaguchi, and Shimada 2006; Roxrud et al. 2007; Lucau-Danila et al. 2010).

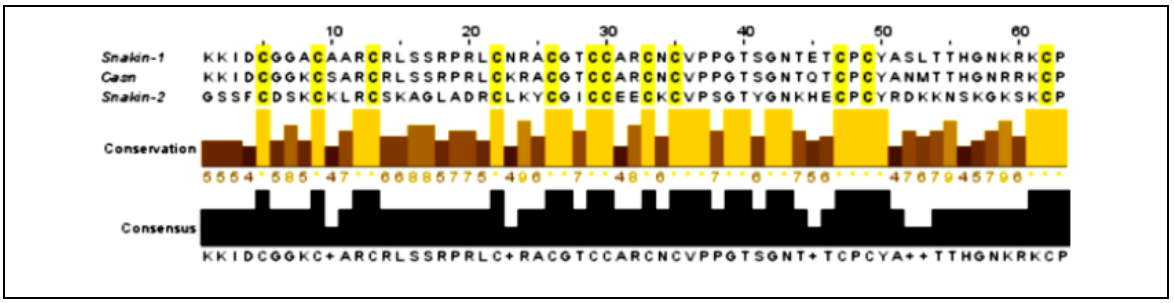


Figure 1-12. Comparative alignment of three Snakins from *Solanum tuberosum* (Snakin-1 and Snakin-2) and *Capsicum annuum* (Casn), evidencing the GASA domain with the conserved motif of twelve cysteines (yellow). The yellow bars show conservation between the sequences and below the consensus sequence is presented.

Cyclotide

The term cyclotide was created at the end of the past century to designate a family of plant peptides with approximately 30 aa in size and a structural motif called cyclic cystine knot (CCK) (Craik et al. 1999). This motif is composed by a head-to-tail cyclization that is stabilized by a knotted arrangement of disulfide bridges, with six

conserved cysteines, connected as follows: C1-2, C3-6, C4-5 (Lima et al. 2016). Cyclotides are generally divided into two subfamilies, Möbius, and Bracelets, based on structural aspects. In addition to CCKs, two loops (between C1-2 and C4-5) have high similarity between different subfamilies, while other two loops (between C2-3 and C3-4) exhibit some conservation within the subfamilies (Weidmann and Craik 2016; Park et al. 2017) (**Figure 1-13**).

To date, several cyclotides were identified in eudicot families such as, Rubiaceae (Gran 1973), Violaceae (Claeson et al. 1998), Fabaceae (Poth et al. 2011) and Solanaceae (Poth et al. 2012), in addition to some monocots of Poaceae family (Nguyen et al. 2011). In general, cyclotides may act in defense against a range of agents like insects, helminths or mollusks, in addition they can also act as ecbolic (Gran 1973), antibacterial (Pränting et al. 2010), anti-HIV (Gustafson et al. 1994) and anticancer (Lindholm et al. 2002). All these characteristics added to the stability conferred by the CCK motif turn these peptides into excellent candidates for drug development (Craik, Clark, and Daly 2007; Northfield et al. 2014).

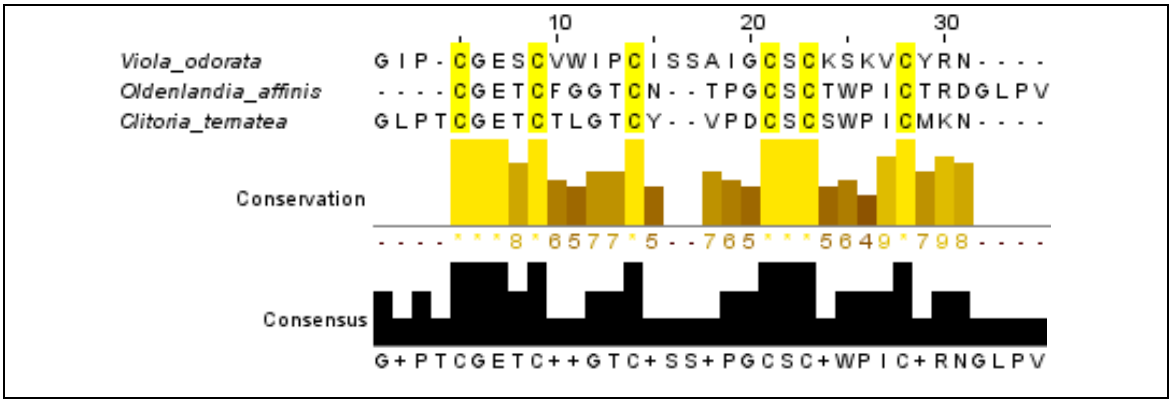


Figure 1-13. Alignment of cyclotides from three plant species: *Viola odorata* (Violaceae), *Oldenlandia affinis* (Rubiaceae) and *Clitoria ternatea* (Fabaceae) evidencing the conserved motif of eight cysteines (yellow). The yellow bars show conservation between the sequences. Consensus sequence is presented at the bottom of the figure.

Thaumatin-like protein (TLP)

Thaumatins or Thaumatin-like proteins (TLPs) belong to the PR-5 (Pathogen-related protein) family and received this name due to its first isolation from the fruit of *Thaumatococcus daniellii* (Maranthaceae) from West Africa (Liu, Sturrock and

Ekramoddoullah 2010). TLPs are abundant in the plant kingdom, being found in angiosperms, gymnosperms and bryophytes (Liu, Sturrock, and Ekramoddoullah 2010), being also identified in other organisms, including fungi (Grenier, Potvin, and Asselin 2000; Sakamoto et al. 2006), insects (Brandazza et al. 2004), and nematodes (Kitajima and Sato 1999).

TLPs are known for their antifungal activity, either by permeating fungal membranes (Batalia et al. 1996) or by binding and hydrolyzing β-1,3-glucans (Trudel et al. 1998; Grenier et al. 1999). In addition, they may act to inhibit fungal enzymes, such as xylanases (Fierens et al. 2007), α-amylases or trypsin (Schimoler-O'Rourke, Richardson and Selitrennikoff 2001). Besides, the expression of TLPs is regulated in response to some stress factors, such as drought (Jung et al. 2005), injuries (Ruperti et al. 2002), freezing (Hon et al. 1995), and infection by fungi (Kumar and Kirti 2011; Rudd et al. 2015), viruses and bacteria (Breiteneder and Radauer 2004).

As to the TLP structure, this protein presents characteristic thaumatin signature (PS00316): G-x-[GF]-x-C-x-T-[GA]-D-C-x(1,2)-[QG]-x(2,3)-C (Jami, Anuradha, and Guruprasad 2007; Tachi et al. 2009). Most of the TLPs have molecular mass ranging from 21 to 26 kDa (J.-J. Liu, Sturrock, and Ekramoddoullah 2010), possessing 16 conserved cysteine residues (**Figure 1-14**) involved in the formation of eight disulfide bonds (Breiteneder 2004), which help in the stability of the molecule, allowing a correct folding even under extreme conditions of temperature and pH (Fierens et al. 2009). TLPs also contain a signal peptide at the N-terminal which is responsible for targeting the mature protein to a particular secretory pathway (J.-J. Liu, Sturrock, and Ekramoddoullah 2010).

The tertiary structure, in turn, presents three distinct domains, which are conserved and form the central cleft, responsible for the enzymatic activity of the protein, being located between domains I and II (Leone et al. 2006). This central cleft may be of an acidic, neutral or basic nature depending on the binding of the different linkers / receptor. All TLPs from plants with antifungal activity have an acidic cleft known as motif REDDD due to five highly conserved amino acid residues (arginine, glutamic acid, and three aspartic acid; **Figure 1-14**), being very relevant for specific receptor binding of antifungal activity (Batalia et al. 1996; Koiwa et al. 1999; Min et al. 2004).

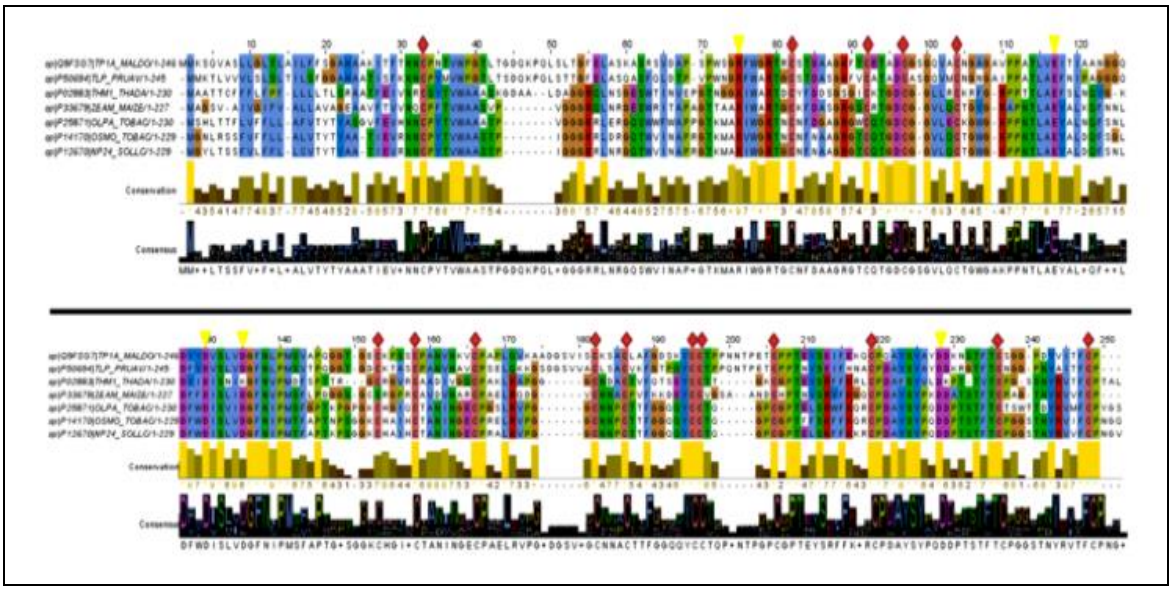


Figure 1-14. Alignment of thaumatin-like amino-acid sequences of six different species (*Malus domestica*, *Prunus avium*, *Thaumatococcus daniellii*, *Zea mays*, *Nicotiana tabacum*, and *Solanum lycopersicum*). Sequence comparison was carried out with Clustal Omega software with standard parameters. Colored bars show conserved residues and motifs. Pink diamonds indicate the cysteine residues that form the eight disulfide bonds; Yellow triangles indicate the position of the REDDD motif and yellow bars in the base show the level of conservation for each position.

Crystallized structures were determined for some plant TLPs, such as thaumatin (Ogata, Gordon, and Vos 1992; **Figure 1-15A)**, zeamatin (Batalia et al. 1996; **Figure 1-15B)**, tobacco PR-5d (Koiwa et al. 1999) and osmotin (Min et al. 2004; **Figure 1-15C)**, the cherry allergen PruAv2 (Dall'Antonia et al. 2005), banana allergen Ba-TLP (Leone et al. 2006; **Figure 1-15D)** and tomato NP24-I (Ghosh and Chakrabarti 2008; **Figure 1-15E)**, among other TLPs.

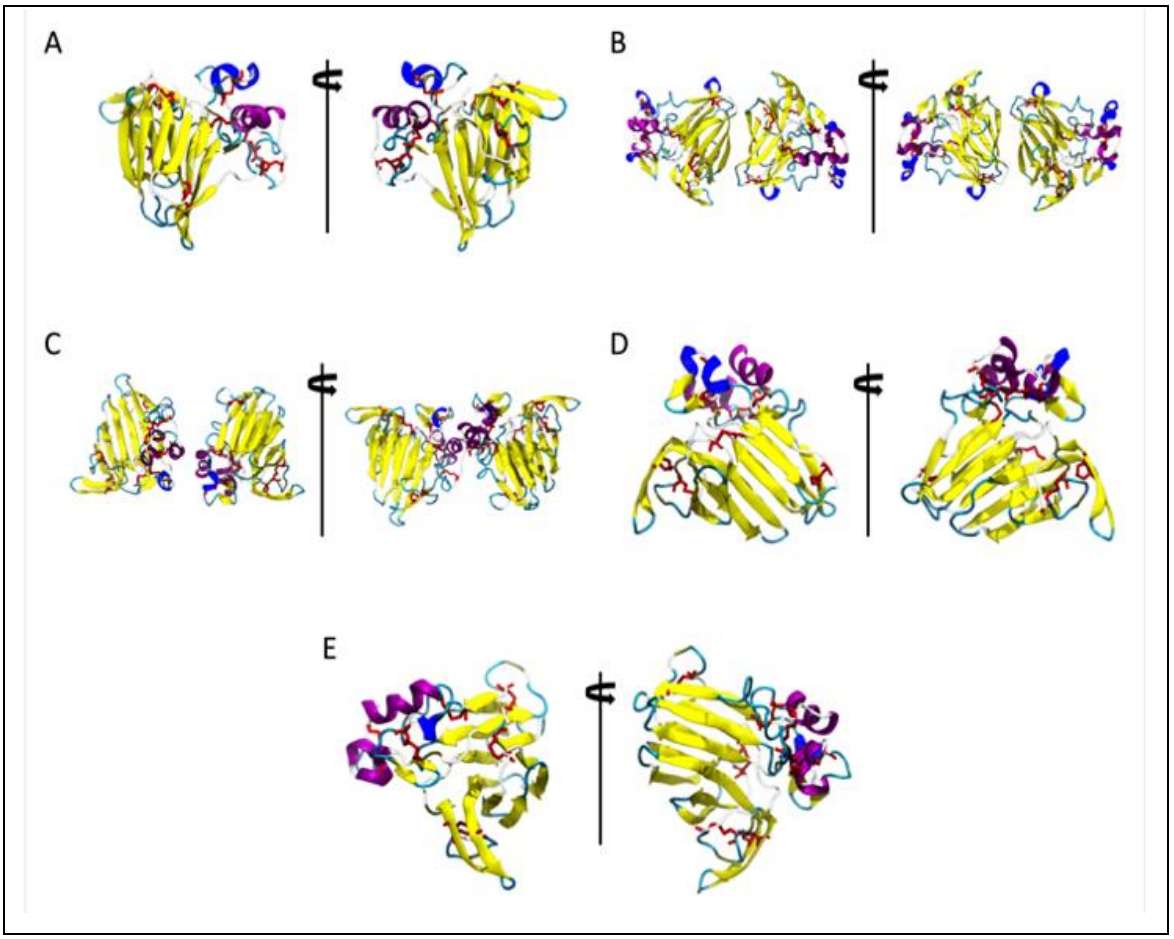


Figure 1-15. Crystal structure of representative plant TLPs available at PDB and visualized and viewed at two angles through the software VMD. **(a)** Thaumatin (1RQW), *Thaumatococcus daniellii*; **(b)** Zeamatin (1DU5), *Zea mays*; **(c)** Osmotin (1PCV), *Nicotiana tabacum*; **(d)** Allergen with antifungal properties (1Z3Q), *Musa acuminata*; **(e)** NP24-I (2I0W), *Solanum lycopersicum*.

Some TLPs are known as sTLPs (small TLPs) due to the deletion of peptides in one of their domains, culminating in the absence of the typical central cleft. These sTLPs exhibit only ten conserved cysteine residues, forming five disulfide bonds, resulting in a molecular weight of approximately 16-17 kDa. They have been described in monocots, conifers and fungi, so far (Osherov et al. 2002; Greenstein et al. 2006; Liu, Sturrock, and Ekramoddoullah 2010). Other TLPs exhibit an extracellular TLP domain and an intracellular kinase domain, being known as PR5K (PR5-like receptor kinases) (X. Wang et al. 1996) and are present in both monocots and dicots. For example, *Arabidopsis* contains three PR5K genes while rice has only one (Liu, Sturrock, and Ekramoddoullah 2010).

Bioinformatics tools and databases including plant AMPs

Databases

With the rapid growth in the number of available sequences, it is almost unfeasible to handle such amount of data manually. Thus, AMP sequences (as well as their biological information) have been deposited in large general databases, such as UniProt and TrEMBL, which contain sequences of multiple origins (Apweiler et al. 2004; Torrent et al. 2012). In this sense, the construction of databases that deal directly with AMPs was an important step to organize the data.

During the last decade, several databases were built to support the deposition, consultation, and mining of AMPs. Thus, these databases can be classified into two groups: general and specific, according to Porto, Pires, and Franco (2017). The specific databases can be divided into two subgroups: those containing only one specific group (defensins or cyclotides) and those containing data from a supergroup of peptides (plant, animal or cyclic peptides) (**Table 1-1**). In general, both types of databases share some characteristics such as the way that the data are available or the access to tools to analyze AMPs.

The Collection of Antimicrobial Peptides ($CAMP_{R3}$) is a database that comprises experimentally validated peptides, sequences experimentally deduced and still those with patent data, besides putative data based on similarity (Thomas et al. 2010; Waghu et al. 2014, 2016). The current version includes structures and signatures specific to families of prokaryotic and eukaryotic AMPs (Waghu et al. 2016). The platform also includes some tools for AMP prediction.

The antimicrobial peptide database (APD) (Z. Wang 2004; G. Wang, Li, and Wang 2016) collects mature AMPs from natural sources, ranging from protozoa to bacteria, archaea, fungi, plants, and animals, including humans. AMPs encoded by genes that undergo post-translational modifications are also part of the scope, besides some peptides synthesized by multienzyme systems. APD provides interactive interfaces for peptide research, prediction, and design, statistical data for a specific group, or for all peptides available in the database.

The LAMP (Database Linking Antimicrobial Peptides) comprises natural and synthetic AMPs, which can be separated into three groups: experimentally validated, predicted, and patented. Their data were primarily collected from the scientific literature, from UniProt and from other databases related to AMPs (Zhao et al. 2013).

DBAASP (Gogoladze et al. 2014) contains information about AMPs from different origins (synthetic or non-synthetic) and complexity levels (monomers and dimers) that were retrieved from PubMed using keywords: antimicrobial, antibacterial, antifungal, antiviral, antitumor, anticancer and antiparasitic peptides. This database is manually curated and provides information about peptides that have specific targets validated experimentally. This database also includes information on chemical structure, post-translational modifications, modifications in the N/C terminal amino acids, antimicrobial activities, cell target and experimental conditions in which a given activity was observed, besides information about the hemolytic and cytotoxic activities of the peptides (Gogoladze et al. 2014).

Due to the diversity of AMPs and the need to accommodate the most representative subclasses, several databases were established focusing on specific types, sources or features. There are several ways to classify AMPs, and they can range from biological sources such as bacterial AMPs (bacteriocins), plants, animals, etc.; biological activity: antibacterial, antiviral, antifungal and insecticide, and based on molecular properties, pattern of covalent bonds, 3D structure and molecular targets (Jorge, Lourenço and Pereira 2012; H.-T. Lee et al. 2015).

The ‘Defensins Knowledgebase’ is a database with manual curation and focused exclusively on defensins. This database contains information about sequence, structure, and activity, with a web-based interface providing access to information and enabling text-based search. In addition, the site presents information on patents, grants, laboratories, researchers, clinical studies and commercial entities (Seebah et al. 2007; S. Liu et al. 2017).

The CyBase is a database dedicated to the study of sequences and three-dimensional structures of cyclized proteins and their synthetic variants, including tools for the analysis of mass spectral fingerprints of cyclic peptides, also assisting in the discovery of new circular proteins (Mulvenna , Wang and Craik 2006).

Finally, the PhytAMP is a database designed solely dedicated to plant AMPs based on information collected from the UniProt database and from the scientific literature through PubMed (Hammami et al. 2009).

Table 1- General and specific antimicrobial peptide databases with their corresponding Uniform Resource Locator (URL).

Database	Type	Url
CAMP	General	http://www.camp.bicnirrh.res.in/
APD	General	http://aps.unmc.edu/AP/main.php
LAMP	General	http://biotechlab.fudan.edu.cn/database/lamp
DBAASP	General	https://dbaasp.org/
DEFENSIN KNOWLEDGEBASE	Specific	http://defensins.bii.a-star.edu.sg/
CYBASE	Specific	http://www.cybase.org.au/
PHYTAMP	Specific	http://phytamp.hammamilab.org/

Biological databanks (DB) are organized collections of data of diverse nature that can be retrieved using different inputs. The management of this information is done through various software and hardware, whose retrieval and organization is performed as quickly and efficiently as possible (Eltabakh, Ouzzani, and Aref 2006). Considering biological data, information can be classified into: (i) primary (sequences), (ii) secondary (structure, expression, metabolic pathways, types of drugs, etc.) and (iii) specialized, e.g. containing information on a species or class of protein (Mehboob-ur-Rahman, Mahmood-ur-Rahman and Zafar 2016). Within this third group some references to AMPs can be mentioned, such as $CAMP_{R3}$ (Waghu et al. 2014) and APD (G. Wang, Li, and Wang 2016) that compile sequence data and structure retrieved from diverse sources, and also the Defensin knowledgebase (Seebah et al. 2007) and the Cybase (Mulvenna , Wang and Craik 2006) which are dedicated to specific classes of peptides (defensins and cyclotides respectively), in addition to PhytAMP (R. Hammami et al. 2009), a specific database of plant antimicrobial peptides.

Retrieving and annotating sequences from databases

The first step to infer the function of a given sequence (annotation) is to retrieve it in databases. For this purpose, two methodologies have been used mostly: local alignments, especially by using BLAST (Altschul et al. 1990) and FASTA (Pearson 1990), but also by searching for specific patterns with Regular Expression (REGEX) or Hidden Markov Model (HMM) (Porto, Pires, and Franco 2017). Since most of the information is available in the databases as sequences, to align them is the best way to

compare. Thus the local alignment is the most commonly used approach whereas the BLAST is the primary tool for doing so (Polyanovsky, Roytberg, and Tumanyan 2011). This tool splits the sequence into small pieces (words), comparing it with the database. However, this approach has an important limitation. Small motifs may not be significantly aligned since they comprise small portions of the sequences that can be smaller than 20% of the total size (J. Tam et al. 2015; Porto, Pires, and Franco 2017).

In order to reduce the effects of local alignment limitations, other strategies based on the search for specific patterns were introduced, such as REGEX (Thompson 1968) and HMM (Eddy 1998). Regular expression (REGEX) is a precise way of describing a pattern in a string where each REGEX position must be set; although ambiguous characters (or wildcards) can also be used. For example, if we want to find a match for both amino acid sequences CAIESSK and WAIESK, we can use the following expression: [CW]AIES{1,2}K, this expression would find a sequence starting with the letter "C" or "W", followed by an "A", an "I" and an "E", one or two "S" and ending with a "K". By the other hand, for HMM, there is a statistic profile inserted in the model, which is calculated from a sequence alignment and a score that is determined site-to-site, with conserved and variable positions defined a priori (Yoon 2009; Porto, Pires, and Franco 2017).

Predicting antimicrobial activity

The design of new AMPs led to the development of methods for the discovery new peptides, thus allowing new experiments to be done by the researchers. In this way, the new challenge lies in the construction of new prediction models capable of discovering peptides with desired activities.

The APD databank has established a prediction interface based on some parameters defined by the entire set of peptides available in this database. These values are calculated from natural AMPs to consider features as length, net charge, hydrophobicity, amino acid composition, etc. If we take as an example the net load, the AMPs deposited in the APD range from -12 to +30. This is the first parameter incorporated into the prediction algorithm. However, most AMPs have a net load ranging from -5 to +10, which then becomes the alternative prediction condition. Therefore, the same method is applied to the remaining parameters. The prediction in APD is performed in three main steps. First, the sequence parameters will be calculated and compared. If defined as an AMP, the peptide can then be classified into three groups: (i) rich in given

amino acids, (ii) stabilized by disulfide and (iii) linear bridges. Finally, sequence alignments will be conducted to find five peptides of greater similarity (Zhou et al. 2013; Zhou and Huang 2015; Guangshun Wang, Li, and Wang 2016).

The advent of machine learning (ML) methods has promoted the field of drug discovery. In ML inferences, both, a positive and a negative dataset, are usually required to train the predictive models. The positive data, in this case, regard preferably experimentally validated AMPs that can be collected in databases, whereas negative data are randomly selected protein sequences that do not have AMP characteristics (Niarchou et al. 2013; Waghu et al. 2016). ML methods based on support vector machine (SVM), random forest (RF) and neural networks (NN) have been the most widely used. SVM is a specific type of supervised method of machine learning, aiming to classify data points by maximizing the margin between classes in a high-dimensional space. RF is a non-parametric tree-based approach that combines the ideas of adaptive neighbors with bagging for efficient adaptive data inference. NN is an information processing paradigm inspired by how a biological nerve system process information. It is composed of highly interconnected processing elements (neurons or nodes) working together to solve specific problems (Lin and Jeon 2006; Karasuyama and Takeuchi 2010; Ding et al. 2013).

Evaluating Proteomic Data

Regarding the use of AMPs in peptide therapeutics, as an alternative to antibacterial and antimicrobial treatment, new efficient and specific antimicrobials are demanded. As described previously in this review, AMPs are naturally occurring across all classes of life, presenting high active potential as therapeutic agents against various kinds of bacteria (Dobslaff et al. 2012). The identification of novel AMPs in databases is primarily dependent on knowing about specific AMPs together with a sufficient sequence similarity (Fedders and Leippe 2008). However, orthologs may be divergent in terms of sequence, mainly because they are under strong positive selection for variation in many taxa (Tennessen 2008), leading to extremely lower similarity, even in closely related species. In this scenario, where alignment tools present limited use, one strategy to identify AMPs is related to proteomic approaches.

Proteins and peptides are biomolecules responsible for various biochemical events in living organisms, from formation and composition to regulation and functioning. The search for the understanding of the expression, function, and regulation of the proteins encoded by an organism initiated the so-called ‘Proteomic Era’. The term "proteome" was

first used by Marc Wilkins in 1994 and represents the set of proteins encoded by the genome of a biological system (cell, tissue, organ, biological fluid or organism) at a specific time under certain conditions (Wilkins et al. 1996). Protein extraction, purification, and identification methods have significantly advanced our capacity to elucidate many biological questions using proteomic approaches (Wasinger and Corthals 2002; Sheoran et al. 2009). The wide diversity of proteomic analysis methods makes the choice of the correct methodology dependent mainly on the type of material and compounds that will be analyzed. Thus, proteomic analysis plays an important role in the discovery of new proteins and peptides, as well as providing several other tools to understand the mechanisms involved in the connection between proteins and its functions (J. D. Thompson, Schaeffer-Reiss and Ueffing 2008).

Two main tools are used to isolate proteins: (1) the two-dimensional electrophoresis (2-DE) associated with mass spectrometry (MS), and (2) liquid chromatography associated with MS, each one with its own limitations (Cho 2007; Baracat-Pereira et al. 2012). Obtaining native proteins is a challenge in proteomics or peptidomics, due to high protein complexity in samples, like the occurrence of post-translational modifications. Alternative strategies applied to extraction, purification, biochemical and functional analyses of these molecules have been proposed, favoring access to structural and functional information of hard-to-reach proteins and peptides (Kolodziejek and van der Hoorn 2010).

Based in 2D gel, Al Akeel et al. (2017) evaluated 14 spots obtained from seeds of *Foeniculum vulgare* (Apiaceae) aiming at proteomic analyses and isolation of small peptides. Extracted proteins were subjected to 3 kDa dialysis, and separation was carried out by DEAE-ion exchange chromatography while further proteins were identified by 2D gel electrophoresis. One of its spots showed highest antibacterial activity against *Pseudomonas aeruginosa*, pointing to promising antibacterial effects, but requiring further proteomic research to authenticate the role of the anticipated proteins.

For AMPs, 2-DE is challenging, due to the low concentration of the peptide molecules, their small sizes and their ionic features (strongly cationic). The low number of available specific databases turns their identification through limited proteolysis techniques and MALDI-MS difficult. In addition, the partial hydrophobicity characteristics and surface charges facilitate peptide molecular associations, making analysis difficult by any known proteomic approaches (Baracat-Pereira et al. 2012).

Additionally, peptides are most often cleaved from larger precursors by various releasing or processing enzymes (Menschaert et al. 2010).

However, profiles generated sometimes do not represent integral proteome, since this technique has limitations to detect proteins present in low concentration, with values of extreme molecular masses, pIs and hydrophobic proteins, including those of membranes (Nägele et al. 2004). Due to these limitations, multidimensional liquid chromatography (MDLC-HPLC) has been successfully employed as an alternative to two-dimensional gels. Techniques and equipment for the newly developed separation and detection of proteins and peptides, such as nano-HPLC and multidimensional HPLC, have allowed the proteomics improvement (Mitulovic and Mechtler 2006).

Molecular mass values obtained are used in computational searches in which they are compared with in silico digestion results of proteins in databases, using specific *in silico* approaches, usually by the action of trypsin as proteolytic agent, generating a set of unique peptides whose masses are determined by mass spectrometry (Blueggel, Chamrad, and Meyer 2004; McHugh and Arthur 2008). These methodologies are widely adopted for large-scale identification of peptide from MS/MS spectra (Hughes, Ma and Lajoie 2010). Theoretical spectra are generated using fragmentation patterns known for specific series of amino acids. The first two widely used search engines in database searching were SEQUEST (Eng et al. 2008) and MASCOT [Matrix Science, Boston, MA (www.-matrixscience.com)] (Perkins et al. 1999). They rank peptide matches based on a cross-correlation to match the hypothetical spectra to the experimental one.

MASCOT is widely used for peptidomics and proteomics analysis, including AMP identification in many organisms or to evaluate the antibacterial efficacy of novel antimicrobial peptides. Evaluating new AMP against multidrug-resistant (MDR) *Salmonella enterica*, Tsai et al. (2016) used two-dimensional gel electrophoresis and liquid chromatography-electrospray ionization-quadrupole- time-of-flight tandem mass spectrometry to determine the protein profiles. The protein identification was performed using the MASCOT with trypsin used as the cutting enzyme, whereas NCBI nr protein was set as a reference database. The methodology used in this study indicated that the novel AMP might serve as a potential candidate for antimicrobial drug development against MDR strains, confirming the usability of MASCOT.

In a similar way, Umadevi et al. (2018) described the AMP signature profile of black pepper (*Piper nigrum* L.) and their expression upon *Phytophthora* infection using label-free quantitative proteomics strategy. For protein/peptide identification, MS/MS

data were searched against the APD database (G. Wang, Li, and Wang 2016) using an in-house MASCOT server, established full tryptic peptides with a maximum of three missed cleavage sites and carbamidomethyl on cysteine, besides an oxidized methionine included as variable modifications. The APD database was used for AMP signature identification (G. Wang, Li, and Wang 2016), together with PhytAMP (R. Hammami et al. 2009) and $CAMP_{R3}$ (Waghu et al. 2016). To enrich the characterization parameters, isoelectric point, aliphatic index and grand average of hydropathy were also used (Gasteiger et al. 2005) (GRAVY) (using ProtParam tool) besides the net charge from PhytAMP database. Based on label-free proteomics strategy, they established for the first time the black pepper peptidomics associated with the innate immunity against *Phytophthora*, evidencing de usability of proteomics/peptidomics data for AMP characterization in any taxa, including plant AMPs, aiming the exploitation of these peptides as next-generation molecules against pathogens (Umadevi et al. 2018).

In the literature, other search tools are described using database searching algorithms, such as X!TANDEM (Craig and Beavis 2004), OMSSA (Geer et al. 2004), ProbID (N. Zhang, Aebersold and Schwikowski 2002), RADARS (Field, Fenyö and Beavis 2002), etc. The search engines are based on database search but use different scoring schemes to determine the top hit for a peptide match. More general information on database search engines, their algorithms and scoring schemes was reviewed by Nesvizhskii, Vitek and Aebersold (2007). Despite its efficient ability to identify peptides, the method presents several drawbacks, like false positive identifications because overly noisy spectra and lower quality peptides score related to short peptides. So, the identification is strongly influenced by the amount of protein in the sample, the degree of post-translational modification, the quality of automatic searches and the presence of the protein in the databases (Gazzana and Borlak 2007; Pestana-Calsa, Ribeiro and Calsa Jr. 2010). In this scenario, the knowledge about the genome from a specific organism has importance to allow the identification of the exact pattern of a given peptide. If an organism has not been sequenced, it is not searchable using these methods (Hughes, Ma and Lajoie 2010; Menschaert et al. 2010).

More recently a free interactive web software platform, MixProTool, was developed, aiming to process multigroup proteomics data sets. This tool is compiled in R (www.r-project.org), providing integrated data analysis workflow for quality control assessment, statistics, gene ontology enrichment, and other facilities. The MixProTool is compatible with identification and quantification results from other programs, such as

MaxQuant and MASCOT, where results may be visualized as vector graphs and tables for further analysis, in contrast to existing softwares, such as GiaPronto (Weiner et al. 2017). According to the authors, the web tool can be conveniently operated, even by users without bioinformatics training, and it is beneficial for mining the most relevant features among different samples (G. Wang 2010).

AMP modeling and simulation in molecular dynamics

The central tenet of structural biology is that structure determines function. For proteins it is often said the *"function follows form"* and *"form defines function"*. Therefore, to understand protein function in detail at the molecular level it is mandatory to know its tertiary structure (Jothi 2012). Experimental techniques for determining structures, such as X-ray crystallography, NMR, electron paramagnetic resonance and electron microscopy, require significant effort and investments (Kaufmann et al. 2010).

All methods mentioned have their own limitations, and the gap between the number of known proteins and the number of known structures is still substantial. Thus, there is a need for computational framework methods to predict protein structures based on the knowledge of the sequence (Jothi 2012). In addition, in recent years there has been impressive progress in the development of algorithms for protein folding that may aid in the prediction of protein structures from amino acid sequence information (Moult et al. 2018).

Historically, the prediction of a protein structure has been classified into three categories: comparative modeling, threading and *ab initio*. The first two approaches construct protein models by aligning the query sequences with already solved model structures. If the models are absent in the Protein Data Bank, the models must be constructed from scratch, i.e., by *ab initio* modeling, considered the most challenging way to predict protein structures.

In the case of comparative modeling methods, when inserting a target sequence, the programs identify evolutionarily related models of solved structures based on their sequence or profile comparison, thus constructing structure models supported by these previously resolved models (Källberg et al. 2012). This approach consists of four main steps: (i) fold assignment, which identifies similarity between the target and the structure of the solved model, (ii) alignment of the target sequence to the model, (iii) generation of a model based on alignment with the chosen template and (iv) analysis of errors considering the generated model (Martí-Renom et al. 2000).

There are several servers and computer models that automate the comparative modeling process, with SWISS-MODEL and MODELLER figuring as the most used (Biasini et al. 2014; Webb and Sali 2016). Although automation makes comparative modeling accessible to experts and beginners, some adjustments are still needed in most cases to maximize model accuracy, especially in the case of more complex proteins (Webb and Sali 2016). Therefore, some caution must be taken regarding the generated models, taking into account the resolution and quality of the model used, as well as homology between the model and the protein of interest.

Threading modeling methods are based on the observation that known protein structures appear to comprise a limited set of stable folds, and that similarity elements in evolutionarily distant or unrelated proteins are often found. The most used servers based on this approach are MUSTER (Wu and Zhang 2008), SPARKS-X (Yang et al. 2011), RaptorX (Källberg et al. 2012), ProSa-II (Wiederstein and Sippl 2007) and most notably the I-TASSER (Yang Zhang 2008). In some cases, the incorporation of structural information to combine the sequence used in the search with possible models allows the detection of similarity in the fold, even in the absence of an explicit evolutionary relation.

The prediction of structures from known protein models is, at first sight, a more straightforward task than the prediction of protein structures from available sequences. Therefore, when no solved model is available, another approach is recommended, namely the *ab initio* modeling. This method is intended to predict the structure only from the sequence information, without any direct assistance from previously known structures. The *ab initio* modeling aims to predict the best model, based on the minimum energy for a potential energy function by sampling the potential energy surface using various searchable information (Yang Zhang, Kolinski and Skolnick 2003; Wu, Skolnick and Zhang 2007). Such approaches turn it challenging to produce high-resolution modeling, essential for determining the native protein folding and its biochemical interpretation. On the other hand later resolved structures and comparisons with previously predicted proteins point to a higher successful modeling generated by *ab initio* methods, than those generated by pure energy minimization methods, classical or even pure methods (Jothi 2012).

Among the most used servers and programs for *ab initio* modeling, we highlight the ROSETTA (Kaufmann et al. 2010), and TOUCHSTONE II (Yang Zhang, Kolinski and Skolnick 2003). The accuracy of the models calculated by many of these methods is evaluated by CAMEO (Continuous Automated Model Evaluation) (Haas et al. 2018) and

by CASP (Critical Assessment of protein Structure Prediction) experiment (Moult et al. 2018). Probably the first reasonably accurate *ab initio* model was built in CASP4. Since then, sustained progress was achieved in *ab initio* prediction, but mainly for small proteins (120 residues or less). In CASP11, for the first time, a novel 256-residue protein with a sequence identity with known structures lower than 5% was constructed with high precision for sequences of this size (Moult et al. 2016).

Molecular Dynamics Simulation

Molecular dynamics (MD) is a computational simulation technique that predicts the changes of the positions and velocities of the constituent atoms of a system under a given time and condition. This calculation is done through a classical approximation of empirical parameters, called "force field" (Robustelli, Piana and Shaw 2018). If, on the one hand, this approximation makes the dynamics of a system containing thousands of atoms numerically accessible, it obviously limits the nature of the processes that can be observed during the simulations. No quantum effect is visualized in a molecular dynamics simulation, just as no chemical bond is broken, no interactions occur between orbitals, resonance, polarization or charge transfer effects (Freddolino et al. 2010). However, the molecules go beyond a static system. Thus, MD is a computational technique that can be used for predicting or refining structures, dynamics of molecular complexes, drug development and action of molecular biological systems (Karplus and McCammon 2002). MD simulation is widely used for protein research, aiming to extract information about the physical properties of individual proteins. The results of such simulations are then compared with experimental results. Since these experiments are generally carried out in solvents, it is necessary to simulate molecular systems of protein in water. These simulations are conducted for a variety of purposes, such as determining the folding of a structure to a native structure and analyzing the dynamic stability of this structure (Mitsutake and Takano 2018).

The use of MD to simulate protein folding processes is one of the most challenging applications and should be relatively long (in the order of microseconds to milliseconds) to allow observing a single fold event. In addition, the force field used must correctly describe the relative energies of a wide variety of shapes, including unfolding and poorly folded shapes that may occur during the simulation (Freddolino et al. 2010). The considerable application potential led to the implementation of MD simulation in many software packages, including GROMACS (Hess et al. 2008; Pronk et al. 2013; Abraham

et al. 2015), AMBER (Case et al. 2005), NAMD (Phillips et al. 2005), CHARMM (Brooks et al. 1983), LAMMPS (Plimpton 1995) and Desmond (Bowers et al. 2006). In addition to those mentioned, there are other types of simulation available, such as the Monte Carlo Method, Stochastic Dynamics and Brownian Dynamics (Abraham et al. 2015).

In the last decades, molecular dynamics simulation has become a standard tool in theoretical studies of large biomolecular systems, such as proteins or DNA, in environments with near realistic solvents. Historically the computational complexity of this type of computation has been extremely high, and much research has focused on algorithms to achieve unique simulations that are as long or large as possible (Hess et al. 2008).

Pathogen interaction studies and molecular docking

Currently, the control of bacterial pathogens is a great challenge, due to high microbial resistance to antibiotic treatment. Most market available antimicrobial drugs act under the synthesis of DNA, RNA and proteins, disruption of the membrane, and inhibition of pathogen-specific metabolic pathways (Neu and Gootz 1996; Hong, Zeng, and Xie 2014; Cheng et al. 2016). Despite different strategies, bacteria became adapted to these drug targets and developed mechanisms to overcome available drugs. Among the main mechanisms of resistance to antibacterial agents, we can mention antibiotic inactivation, target modification, resistance and plasmid efflux (Nikaido 1994; Sun, Deng, and Yan 2014; Munita and Arias 2016). The acquisition of these mechanisms has led to a worsening of clinical conditions, and new alternative methods have emerged to improve antibiotics efficacy (Tam et al. 2015). In this context, plant-derived AMPs have great potential for combating pathogens (Chandra et al. 2017), especially because plants have unique AMP classes, not present in other organisms (Benko-Iseppon et al., 2010). The differences of repertoire and composition in an amino acid of plant AMPs reveal structural and functional alterations of their protein families (Bolouri Moghaddam, Vilcinskas, and Rahnamaeian 2016) that allow a new perspective since they have a broad-spectrum antimicrobial target. Although AMPs have not yet obtained authorization for clinical use in humans, they are a promising future alternative (Nawrot et al. 2014).

As a perspective for this field, systems biology stands out because it is capable of correlating omics data and promoting studies of plant-pathogen interaction. The construction of plant-pathogen interaction network includes the reconstruction of

metabolic pathways of these organisms, identification of the degree of pathogenicity, besides expression of genes and proteins from both, plant and pathogen. The networks can be classified into five types: (i) regulatory, (ii) metabolic, (iii) protein-protein interaction, (iv) signaling and regulatory and (v) signaling, regulatory and metabolic (Botero et al. 2018). Each of these networks can be plotted according to computational approaches.

Also, further studies are required to contemplate the construction of evolutionary *in silico* models and the characterization of these molecular targets *in vitro* (Schneider and Collmer 2010; Peyraud et al. 2016). Studies of protein-protein interactions to understand the regulatory process are essential (Ramakrishnan, Chandra, and Srinivasan 2014) and new computational methods are necessary for this purpose with more optimized algorithms, also to remove the false positives detected nowadays in most softwares. Thus, in-depth studies on the orientation of molecules and their linkages to the formation of a stable complex are of great importance for understanding plant-pathogen studies and also to develop new drugs (R. Hammami and Fliss 2010).

Molecular docking

The understanding of the regulatory principles by which protein receptors recognize, interact and associate with molecular substrates or inhibitors is of paramount importance to generate new therapeutic strategies (Sousa, Fernandes and Ramos 2006). In modern drug discovery, docking plays an important role in predicting the orientation of the binder when it is attached to a protein receptor or enzyme, using forms and electrostatic interactions, van der Walls, Colombic and hydrogen bond as parameters to quantify or predict a given interaction (London, Raveh and Schueler-Furman 2013; Pagadala, Syed and Tuszynski 2017). Protein-protein interactions are essential for cellular and immune function and, in many cases, due to the absence of an experimentally determined structure of the complex, these interactions must be modeled to obtain an understanding about their molecular basis (Pierce et al. 2014).

Drug research based on structure is a powerful technique for the rapid identification of small molecules against the three-dimensional structure of available macromolecular targets, usually by X-ray crystallography, NMR structures or homology models. Due to abundant information on protein sequences and structures, the structural information on specific proteins and their interactions have become very important for current pharmacological research (Petrey and Honig 2014; van Zundert et al. 2016).

Even in the absence of knowledge about the binding site and limited backbone movements, a variety of algorithms have been developed for docking over the past two decades. Although the ZDOCK (Pierce et al. 2014), the rDOCK (Ruiz-Carmona et al. 2014) and the HEX (Ghoorah et al. 2013) have provided results with high coupling precision, the complexes provided are not very useful for designing inhibitors for protein interfaces due to constraints on rigid body docking (Pagadala, Syed and Tuszynski 2017). In this context, more flexible approaches have been developed which generally examine very limited conformations compared to rigid body methods. These docking methods predict that binding is more likely to occur in broad surface regions and then defines the sites in complex structures of high affinity (Andrusier et al. 2008).

The best example is the HADDOCK software (van Zundert et al. 2016), which has been successful in solving a large number of precise models for protein-protein complexes. A good example of its use is the study of the complex formed between plectasin, a member of the innate immune system, and a precursor lipid of bacterial cell wall II. The study identified the residues involved in the binding site between the two proteins, providing valuable information for planning new antibiotics (Schneider et al. 2010).

However, the absolute energies associated with intermolecular interaction are not estimated with satisfactory accuracy by the current algorithms. Some significant issues as solvent effects, entropic effects, and receptor flexibility still need to be addressed. However, some methods such as MOE-Dock (Corbeil, Williams and Labute 2012), GOLD (Jones et al. 1997), Glide (Friesner et al. 2006), FlexX (Sousa, Fernandes and Ramos 2006) and Surflex (Spitzer and Jain 2012), which deal with lateral chain flexibility, have proven to be effective and adequate in most cases. Realistic interactions between small molecules and receptors still depend on experimental wet-lab validation (Rentzsch and Renard 2015; Pagadala, Syed and Tuszynski 2017).

Despite the existing difficulties, there is a growing interest in the mechanisms and prediction of small molecules such as peptides, since they bind to proteins in a highly selective and conserved manner, being promising as new medicinal and biological agents (Audie and Swanson 2013). While both ‘small molecule docking methods’ and ‘custom protocols’ can be used, short peptides are difficult targets because of their high torsional flexibility (Rentzsch and Renard 2015). Protein-peptide docking is generally more challenging than those related to other small molecules, and a variety of methods have been applied so far. However, few of these approaches have been published in a way that

can be reproduced with ease (Audie and Swanson 2012; London, Raveh and Schueler-Furman 2013; Sousa et al. 2013). Although it is difficult to use peptide docking, a recent focus of basic and pharmacological research has used computational tools with modified peptides to predict the selective disruption of protein-protein interactions. These studies are based on the involvement of some critical amino acid residues that contribute most to the binding affinity of a given interaction, also called hot-spots (Wilson 2009; Unal, Gursoy and Erman 2010).

Despite the number of existing docking programs, the algorithms are not accurate and need to be improved. However, approaches are being developed to improve all issues related to punctuation, protein flexibility, interaction with plain water, among other issues (Lensink, Velankar and Wodak 2017). In this context, the CAPRI (Critical Assessment of Predicted Interactions) is a community that provides a quality assessment of different docking approaches. It started in 2001 and since then has aided the development and improvement of the methodologies applied for docking (Lensink and Wodak 2013).

An evaluation was carried out for CAPRI in 2016 resulting in an improvement in the integration of different modeling tools with docking procedures, as well as the use of more sophisticated evolutionary information to classify models. However, adequate modeling of conformational flexibility in interacting proteins remains an essential demand with a crucial need for improvement (Lensink, Velankar and Wodak 2017). Different docking programs are currently available, and new alternatives continue to appear. Some of these alternatives will disappear, just as others will become the top choices among field users.

5. Plant biotechnology involving AMPs

Several works reporting the overexpression of AMPs in transgenic plants have been published in the last decades mainly involving plant defensins (**Table 1-2**), but also regarding other plant AMPs (**Table 1-3**). The first transgenic plant expressing a defensin was generated by Gao et al. (2000) producing a transgenic potato carrying the alfalfa alfAFP defensin. The transformed plants did not exhibit phenotypic differences while the level of infection by *Verticillium dahliae* fungus was reduced by six-fold when compared to untransformed controls. In addition, the results of greenhouse experiments were consistent with the field analyzes, highlighting the *in vivo* activity of defensin.

Another experiment carried out by Kanzaki et al. (2002) generated transgenic rice plants expressing a defensin isolated from *Eutrema japonicum* (Brassicaceae) and

observed that the transformed plants presented quite variable levels of resistance to the fungus *Magnaporthe grisea* when compared to the controls. The authors suggested that this variation could be related to events of gene silencing or to post-transcriptional changes.

In another experiment, Chen, Liu, and Zou (2006) transformed tomato plants with three vector constructs using: (1) the *alfAFP* gene encoding an alfalfa defensin, (2) the *GLU* gene encoding a tobacco defensin, and (3) both genes *alfAFP* + *GLU*. Interestingly the levels of resistance to *Ralstonia solanacearum* were similar in transgenic plants carrying a single defensin gene, regardless of the type. However, transgenic plants carrying both genes exhibited a more pronounced resistance, evidencing the synergistic effect of their coexpression.

Anuradha et al. (2008) transformed tobacco and peanut plants with the mustard defensin gene *BjD,* generating individuals with high levels of resistance to fungal infections (*Phytophthora parasitica, Fusarium moniliforme, Pheaoisariopsis personata* and *Cercospora arachidicola*). In another assay, Abdallah et al. (2010) evaluated transgenic tomato plants expressing the *Msdef1* (*Medicago sativa* defensin) gene. High levels of resistance to *F. oxysporum* were observed. In the inoculation bioassays, 70% of the transgenic plants showed no fusarium wilt symptoms while 7% were asymptomatic.

The first transgenic cotton cv. Coker was generated by Gaspar et al. (2014), using class II defensin. Greenhouse tests and field bioassays confirmed the resistance level increase to *V. dahliae* and *F. oxysporum* in three times when compared to untransformed plants. In addition, the productivity of plants constitutively expressing defensins was similar to that of untransformed plants in the absence of the pathogen. This fact is possibly related to the storage of defensin in the vacuole, which promotes lower phytotoxicity.

Transgenic rice plants expressing the defensin gene *NmDef02* (from *Nicotiana megalosiphon*) acquired resistance to the fungus *Sarocladium oryzae* (Pérez-Bernal et al. 2017). This same defensin gene was tested in transgenic tobacco and potato plants that acquired resistance to *P. infestans, P. parasitica var. nicotianae, Alternaria solani, F. oxysporum* and *V. dahlia*. Interestingly, the antifungal activity raised no toxic effects for host plants, highlighting the efficacy of defensin in conferring resistance to different pathogens (Portieles et al. 2010).

Thao et al. (2017) transformed tobacco plants with a defensin gene from *Vigna radiata* (*VrDEF1*; Fabaceae) related to inhibition of weevil alpha-amylase. After an alpha-amylase enzymatic assay with total seed proteins of transgenic and control plants,

a reduction of 18% in enzyme activity was observed, demonstrating the potential of defensin as a candidate for biotechnological purposes related to insect resistance or tolerance. Overexpression of defensin J1-1 (from pepper) in transgenic tobacco was carried out using two constructs: (i) the constitutive promoter p35S and (ii) root-specific promoter pPRP3. The resistance to *P. parasitica* var. *nicotianae* was similar for both promoters tested. In addition, overexpression of exogenous J1-1 promoted increased expression of endogenous PR2 and PR10 genes, evidencing its potential in altering signaling pathways and increasing resistance to phytopathogens (H.-H. Lee et al. 2018).

Biotechnological approaches related to plant transformation with genes encoding other AMPs (such as snakins, LTPs, alpha-thionin, and heveins) are still poorly represented in the literature when compared to the use of defensin genes. The only exception involves thaumatins, which are extensively characterized and have been researched since the late 1990s (**Table 1-2**).

In regard to snakins, Almasia et al. (2008) developed a transgenic potato overexpressing snakin-1, which promoted increased resistance to fungal infection by *Rhizoctonia solani*, achieving survival rates of 75% in transgenic plants versus 17% in wild plants. In addition, the transformed plants were tested for resistance to *Erwinia carotovora* resulting in a reduction of 88% of the symptoms. Balaji and Smart (2012) used a snakin-2 gene to generate cisgenic tomato plants. Transformed plants infected with the bacterium *Clavibacter michiganensis* ssp. *michiganensis* presented a delay in the emergence of the disease symptoms, lesion size, and reduction of colonizing bacteria.

Scientific reports related to transformation with LTPs are still scarce. The study by Patkar and Chattoo (2006) is worth mentioning. The authors used the *Ace-AMP1*, an ns-LTP gene from onion to generate transgenic rice plants with a constitutive maize ubiquitin (UbI) promoter or by the pathogen-induced Phenylalanine Ammonia-lyase (PAL). The bioassays involved blast fungus (*Magnaporthe grisea*), sheath blight fungus (*R. solani*) and leaf blight bacterium (*Xanthomonas oryzae*). A significant reduction in the lesion size caused by the three phytopathogens was observed, as well as a delay in the appearance of the symptoms. However, no transgenic lineage acquired total resistance to any of the pathogens, even considering the different promoters used.

With respect to alpha-thionines, the first record of a transformed plant was carried out by Carmona et al. (1993) that overexpressed alpha-thionines (derived from barley and wheat) in tobacco. Transgenic plants bearing the barley gene increased resistance to *Pseudomonas syringae* pv. *syringae* and presented a drastic reduction in the lesion area.

In contrast, when wheat alpha-thionin was used, no difference in transformed plants was observed relative to controls, possibly due to the low level of gene expression in transgenic tobacco plants. Transgenic rice transformed with an oat alpha-thionin (*Asthi1*) was resistant against *Burkholderia plantarii* and *B. glumae*, with transformed plants showing similar growth to healthy controls (Iwai et al. 2002).

The alpha-thionin *Thi2.1* was used for transformation of tomato and resistance evaluation of symptoms after inoculation with *F. oxysporum* and *R. solanacearum* (Chan et al. 2005). Some transformed plants submitted to infection by *F. oxysporum* were as resistant as the resistant wild strain. In turn, the plants infected with *R. solanacearum* showed higher resistance than the untransformed ones. However, the tomato strain naturally resistant to *R. solanacearum* presented lower symptoms severity when compared to the transformed ones.

Muramoto et al. (2012) developed a transgenic sweet potato overexpressing barley alpha-hordothionin and found increased resistance to the fungus *Ceratocystis fimbriata* in leaves and roots.

The use of hevein in plant transformation was reported by Koo et al. (2002) that generated transgenic tobacco plants overexpressing the gene *Pn-AMP-h2* from *Pharbitis nil* (Convolvulaceae). Transgenic plants showed increased resistance to fungus *P. parasitica*, presenting a reduction in the symptoms and severity of the disease.

Later, the *in vivo* activity of the hevein *SmAMP1* gene from *Stellaria media* (Caryophyllaceae) was evaluated in transgenic tobacco and in Arabidopsis, both inoculated with the fungus *Bipolaris sorokiniana* (R. Shukurov et al. 2012). The untransformed plants presented smaller size and lower number of flowers when compared to the transgenic lines that grew similarly to the inoculated wild-type. In an additional trial, the authors evaluated the survival rate of transformed tobacco after inoculation with the fungus *Thielaviopsis basicola* and found that one of the transgenic lines had superior survival (89%) than the wild-type (48%). In contrast, another transformed strain showed high susceptibility to the pathogen, with a survival rate of only 15% compared to 48% in the control.

The first record of genetic transformation using thaumatin was performed by Datta et al. (1999) who developed transgenic rice by overexpressing an endogenous thaumatin. In a bioassay for sheath blight (*R. solani*) a reduction in the area and number of lesions caused by the phytopathogen was observed in the transformed plants. In the same year, (Chen et al. 1999) verified that transgenic wheat plants overexpressing a rice thaumatin

(*tlp*) showed retardation and decreased symptoms of fusarium ear blight caused by *F. graminearum*. However, Anand et al. (2003) reported opposite results in a field bioassay, with transformed individuals showing more intense symptoms than infected wild plants. In the following year, Velazhahan and Muthukrishnan (2004) demonstrated that transgenic tobacco plants carrying *Tlp* inoculated with the fungus *Alternaria alternata* showed a similar number of lesions to the wild type, even though the number of necrotic lesions was significantly lower in the transformed plants.

Recently, He et al. (2017) isolated and overexpressed a thaumatin from *Vitis amurensis* (VaTLP) in *V. vinifera* aiming resistance to Downy mildew caused by *Plasmopara viticola*, reporting a reduction of the lesion area and in the number of spores in leaves infected with the pathogen. Histological analyzes also confirmed disease resistance, demonstrating inhibition and malformation in the development of hyphae in leaf tissues of transgenic plants.

Aghazadeh et al. (2017) developed transgenic canola overexpressing a rice thaumatin (*Ostlp*) and found that the diameter of the lesions caused by *S. sclerotiorum* was reduced by half when compared to the wild plant. *Ostlp* was also used in the transformation of cassava (*Manihot esculenta*, Euphorbiaceae) infected with *Colletotrichum gloeosporioides*. The transformed plants had a delay in the onset of symptoms, as well as a reduction in the lesion diameter by more than 50% compared to controls (Ojola et al. 2018).

Gaspar et al. (2014) mentioned more than 100 publications reporting the expression of genes with antifungal properties in transgenic plants. However, few reports present validated greenhouse data on field bioassays. This fact must be related to the complexity of reproducing laboratory results in the field or to the difficulty of meeting the regulatory requirements necessary for this type of experiment in most countries. To date, no transgenic plant with antimicrobial peptide was introduced commercially, probably due to bureaucratic procedures for GMOs (Genetically Modified Organisms) testing or the cost of resistance to yield.

Zeller, Kalinina and Schmid (2013) comment that in 88 studies related to transgenic plants with resistance to pathogens, about half of the resistant plants had lower productivity than their respective controls. The authors evidenced these data in a study addressing resistance to pathogenic fungi in genetically modified wheat, in which four transformants carrying resistance genes to powdery mildew *Blumeria graminis* were evaluated. It was verified that even in the presence of the pathogen, three transformants

did not differ in the agronomic performance in relation to the untransformed plants of the same cultivar. The single accession that had higher resistance showed better agronomic performance only when the infection levels were extremely high. However, under natural conditions, it is known that infection levels generally vary yearly according to environmental factors.

It is worth noting that even considering the reported difficulties it is expected that in the near future biotech agricultural products based on AMPs will be essential for increasing agricultural production, accelerating the transition between biotechnological research and field bioproducts (Lacerda et al. 2014).

Table 1-1. Plant defensins used for transformation of various plant species, with their donator species, transformed plants and *in vivo* activities.

Name	Origin	Transformed plants	*in vivo* activity	Reference
BrD1	*Brassica rapa*	Rice	Insecticide against nymphs and adult females of the brown spittlebug (*Nilaparvata lugens*)	Choi et al. 2009
alfAFP (MsDef1)	*Medicago sativa*	Potato and tomato	Antifungal action against *Verticillium dahliae* (potato*), Fusarium oxysporum* (tomate) and antibacterial against *Ralstonia solanacearum* (tomato)	Gao et al. 2000; Chen et al. 2006; Abdallah et al. 2010
Wasabi defensin	*Wasabia japonica*	Rice, tabacco, eggplant, tomato, *Phalaenopsis*, melon 'Egusi' (*Colocynthis citrullus*)	Antifungal action against *Magnaporthe grisea* (rice), *Botrytis cinerea* (tabacco and tomato), *Alternaria solani* (eggplant and tomato), *Alternaria* leaf spot and *Fusarium* wilt (Melon 'Egusi'), *Fusarium oxysporum* and *Erysiphe lycopersici* (tomato) and antibacterial against *Erwinia carotovora* (*Phalaenopsis*)	Kanzaki et al. 2002; Khan et al. 2006; Sjahril et al. 2006; Ntui et al. 2010; Khan et al. 2011; Darwishet et al. 2014
BjD	*Brassica japonica*	Tabacco and peanut	Antifungal action against *Phytophthora parasitica* and *Fusarium moniliforme* (tabacco), *Phaeosariopsis personata* and *Cercospora arachidicola* (peanut)	Anuradha et al. 2008
NaD1	*Nicotiana alata*	Cotton	Antifungal action against *Fusarium oxysporum* and *Verticillium dahliae*	Gaspar et al. 2014
Sm-AMP-D1	*Stellaria media*	Banana	Antifungal action against *Fusarium oxysporum* f. sp. *cubense* (Foc)	Ghag et al. 2014
Rs-AFP2	*Raphanus sativus*	Tabacco, wheat and peanut	Antifungal action against *Alternaria longipes* (tabacco), *F. graminearum* and *Rhizoctonia cerealis* (wheat), *Phaeoisariopsis personata* (peanut)	Terras et al. 1995; Li et al. 2011; Vasavirama and Kirti 2012
BSD1	*Brassica campestris*	Tabacco	Antifungal action against *Phytophthora parasitica*	Park et al. 2002

Table 1-2 continued...

Name	Origin	Transformed plants	*in vivo* activity	Reference
PhDef1	*Petunia hybrida*	Banana	Antifungal action against *Fusarium oxysporum* f. sp. *cubense* (Foc)	Ghag et al. 2012
PhDef2	*Petunia hybrida*	Banana	Antifungal action against *Fusarium oxysporum* f. sp. *cubense* (Foc)	Ghag et al. 2012
DDR30	*Pisum sativum*	Canola	Antifungal action against *Leptosphaeria maculans*	Wang et al. 1999
DmAMP1	*Dahlia merckii*	Papaya and rice	Antifungal action against *Phytophthora palmivora* (papaya), *Magnaporthe oryzae* and *Rhizoctonia solani* (rice)	Zhu et al. 2007; Jha et al. 2009
J1-1	*Capsicum annuum*	Pepper and tobacco	Antifungal action against *Colletotrichum gloeosporioides* (Antracnose) in pepper and *Phytophthora parasítica* in tabacco	Seo et al. 2014; Lee et al. 2018
MtDef4.2	*Medicago truncatula*	*Arabidopsis thaliana*	Antifungal action against *Hyaloperonospora arabidopsidis* Noco2 and *Fusarium graminearum*	Kaur et al. 2012
NmDef02	*Nicotiana megalosiphon*	Rice, potato and tabacco	Antifungal action against *P. parasitica var. nicotianae* e *P. hyoscyami* f.sp. *tabacina* (tabaco), *A. solani* and *P. infestans* (potato), *Sarocladium oryzae* (rice)	Portieles et al. 2010; Pérez-Bernal et al. 2017
VrDEF1	*Vigna radiata*	Tabacco	Inhibitor insecticide of caruncho α-amylase	Thao et al. 2017

Table 1-2. Plant Snakins, LTPs, α-Thionins, Heveins and Thaumatins used for transformation of various plant species, with their donator species, transformed plants and *in vivo* activities.

Name	Origin	Transformed plants	*in vivo* activity	Reference
Snakin-1 (SN-1)	*Solanum chacoense*	Potato and Wheat	Antifungal action against *Rhizoctonia solani* (potato), *Blumeria graminis* (wheat) and Antibacterial against *Erwinia carotovora* (potato)	Almasia et al. 2008; Faccio et al. 2011
Snakin-2 (SN-2)	*Solanum lycopersicum*	Tomato	Antibacterial against *Clavibacter michiganensis*	Balaji and Smart 2012
Ace-AMP1 (ns-LTP)	*Allium cepa*	Rice	Antifungal action against *Magnaporthe grisea*, *Rhizoctonia solani* and antibacterial against *Xanthomonas oryzae*	Patkar and Chatoo 2006
α-Thionin	Cevada	Tabacco	Antibacterial against *Pseudomonas syringae*	Carmona et al. 1993
α-Thionin	Aveia	Rice	Antibacterial against *Burkholderia plantarii* and *B. glumae* (rice)	Iwai et al. 2002
α-Thionin (Thi2.1)	Arabidopsis	Arabidopsis and tomato	Antifungal action against *Ralstonia solanacearum* (tomato) and Antifungal action against *Fusarium oxysporum* (arabidopsis and tomato)	Epple et al. 1997; Chan et al. 2005
α-HordoThionin (α-HT)	Cevada	Sweet potato	Antifungal action against *Ceratocystis fimbriata*	Muramoto et al. 2012
Pn-AMP-h2 (Hevein)	*Pharbitis nil*	Tabacco	Antifungal action against *Phytophthora parasitica*	Koo et al. 2002
SmAMP1 (Hevein)	*Stellaria media*	Tabacco and Arabidopsis	Antifungal action against *Bipolaris sorokiniana* (arabidopsis) and *Thielaviopsis basicola* (tabacco)	Shukurov et al. 2012

Table 1-3, continued

Thaumatin	**Arroz**	**Rice, Wheat, Tabacco, Banana and Cassava**	**Antifungal action against *Rhizoctonia solani* (rice), *Fusarium graminearum* (wheat), Alternaria alternata (tabacco), *Fusarium oxysporum* (banana), *Sclerotinia sclerotiorum* (canola), *Colletotrichum gloeosporioides* (cassava)**	**Datta et al. 1999; Chen et al. 1999; Velazhahan and Muthukrishnan 2003; Mahdavi et al. 2012; Aghazadeh et al. 2017; Ojala et al. 2018**
Thaumatin	*Thaumatococcus daniellii*	Tabacco	Antifungal action against *Pythium aphanidermatum* and *Rhizoctonia solani* and increased tolerance to water and saline stress	Rajam et al. 2007
VaTLP (Thaumatin)	*Vitis amurensis*	Grape	Antifungal action against *Plasmopara viticola*	He et al. 2017
ObTLP1 (Thaumatin)	*Ocimum basilicum*	Arabidopsis	Antifungal action against *Scleretonia sclerotiorum* and *Botrytis cinerea* and increased tolerance to water and saline stress	Misra et al. 2016
CsTLP (Thaumatin)	*Camellia sinensis*	Potato	Antifungal action against *Macrophomina phaseolina* and *Phytophthora infestans*	Acharya et al. 2013

Plant AMPs in the development of new drugs and bioactive compounds

Plants play an essential role in animal and human survival as a source of food and oxygen, besides their use as therapeutic agents presenting action against diverse human pathologies. There is a rich and unexplored ethnobotanical diversity especially in tropical regions, with numerous plants traditionally used by human populations (Benko-Iseppon and Crovella 2010; Benko-Iseppon et al. 2012).

Some historical examples include *Cinchona pubescens* quinidine for the treatment of cardioarrhythmia and vinblastine of *Catharanthus roseus* for the treatment of various cancer types from a drug derived from a symbiotic organism associated with the plant (Craik et al. 2018). Another example is the paclitaxel (Taxol) a drug used to treat breast cancer, is derived from symbiotic fungi and from the gymnosperm tree *Taxos brevifolia* (Stierle, Strobel, and Stierle 1993).

Many new drugs discovered come from bioactive molecules that are responsible or participate in the therapeutic action of medicinal plants, as is the case of AMPs, which stand out for being active against a wide range of pathogenic microorganisms that attack plants and animals (Thevissen et al. 2007). Also called peptidic antibiotics, AMPs present a new generation of biocidal agents for plant protection, as well as for the treatment of microbial diseases in humans and animals (Gerwick 2013). Different applications have been reported on the action of these peptides, demonstrating broad-spectrum activity, such as (i) action against cancer cells (Hoskin and Ramamoorthy 2008; Felício et al. 2017), (ii) production and accumulation of immune cells, (iii) wound healing and (iv) angiogenesis stimulation (Elsbach 2003; Bowdish et al. 2005).

In general, the AMPs exhibit low cytotoxicity to mammalian cells and have a particular mode of action (Holaskova et al. 2015). However, the interaction of these proteins with pathogens has not yet been fully unveiled. The proposed modes of action for AMPs involve the interaction of peptides with microbial membranes and the formation of pores leading to one or more processes including micellisation, depolarization of membranes, leakage of cytoplasmic material, internalization of biocidal peptides or damage to intracellular macromolecules (Straus and Hancock 2006; Melo, Ferre and Castanho 2009). The action also depends on the peptide structural conformation (α-helices, β-sheets, disordered loops) and their interaction with the microbial membrane through hydrophobic or electrostatic forces (Holaskova et al. 2015).

As mentioned before, with the advent of bioinformatics, a number of tools and databases have provided relevant information on natural and synthetic AMPs and are

useful to screen and predict functional activities, helping to drive the process of discovery and design of AMPs with therapeutic potential or antimicrobial properties (Holaskova et al. 2015). Laverty, Gorman, and Gilmore (2011) estimate that more than 1700 natural AMPs were identified and that thousands of derivatives and analogs were designed or generated synthetically using these molecules as a model.

Over the years, modern medicine has been dealing with incidences of antimicrobial resistance, and the current and inadequate use of antibiotics are essential factors to generate the great crisis of drug resistance, considered worldwide as a public health problem recognized by the World Health Organization (WHO) and the United Nations (Silva et al. 2011; Tacconelli et al. 2017) that launched a campaign to encourage and prioritize research for the development of new drugs in the fight against resistant bacteria. In this scenario plant AMPs figure as promising candidates, offering a broad spectrum of activity against the most diverse microorganisms, as well as presenting low or no side effects (Hancock 2001; Galdino da Rocha Pitta, Galdino da Rocha Pitta and Lins Galdino 2010; Peters, Shirtliff and Jabra-Rizk 2010).

Limitations to the large-scale commercial production of peptide-derived drugs include the lack of suitable production platforms in terms of yield, cost, and product purity. Techniques such as heterologous expression and solid phase chemical synthesis are the most used. Therefore, the interest in producing large-scale AMPs using plants has increased significantly during the last decade (Nadal et al. 2012; Cabanos et al. 2013). Biotechnological advances in the last decade allowed the use of plant bioreactors as attractive platforms for the large-scale production of peptides, proteins, and drugs (Fukuzawa et al. 2010; Sathish et al. 2011; Rubio-Infante et al. 2012; Cabanos et al. 2013).

Concluding remarks and perspectives

As emphasized in this chapter, plant antimicrobial peptides (AMPs) show greater diversity and abundance, when compared to other kingdoms. It can be speculated that plants shelter yet undescribed AMP classes, given their huge abundance and isoform diversity.

The genomic and peptidic structure of AMPs can be variable, with few key residues conserved, what turns their identification, classification and comparison challenging also in the omics age. Nevertheless, advances in the generation of new

bioinformatics tools and specialized databases have led to new and more efficient approaches for both the identification of primary sequences and molecular modeling, besides the analysis of the stability of generated models.

Plants transformed with AMPs generally have greater resistance to pathogens, with an emphasis on fungi, generally not affecting productivity, although a low number of field assays are available. Despite this, no transgenic or cisgenic plants expressing AMPs have been commercially launched, possibly due to logistic difficulties or the lack of interest of large biotech companies that are also generally suppliers of pesticides.

The greatest potential has been recognized in the production of next generation antibiotics due to the bactericidal and fungicidal action of AMPs with low toxicity to mammals. In this particular, the greatest limitation has been the large-scale production of AMPs by heterologous expression systems (which generally use bacteria and yeasts), since the growth of these organisms is affected by AMPs, requiring immobilization techniques and posterior purification. Their potential as new drugs has been recognized not only individually, but as a coadjuvant in synergism with traditional antibiotic treatment.

References

Abdallah, N. A., Shah, D., Abbas, D., and Madkour, M. 2010. “Stable Integration and Expression of a Plant Defensin in Tomato Confers Resistance to Fusarium Wilt.” *GM Crops* 1 (5): 344–50. https://doi.org/10.4161/gmcr.1.5.15091.

Aboye, T. L., Strömstedt, A. A., Gunasekera, S., Bruhn, J. G., El-Seedi, H., Rosengren, K. J., and Göransson, U. 2015. “A Cactus-Derived Toxin-Like Cystine Knot Peptide with Selective Antimicrobial Activity.” *ChemBioChem* 16 (7): 1068–77. https://doi.org/10.1002/cbic.201402704.

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. 2015. “GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers.” *SoftwareX* 1–2 (September): 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

Acharya, K., Pal, A. K., Gulati, A., Kumar, S., Singh, A. K., and Ahuja, P. S. 2013. "Overexpression of *Camellia sinensis* thaumatin-like protein, CsTLP in potato confers enhanced resistance to *Macrophomina phaseolina* and P*hytophthora infestans infection*". *Molecular biotechnology* 54 (2): 609-622.

Aghazadeh, R., Zamani, M., Motallebi, M., and Moradyar, M. 2017. “*Agrobacterium*-Mediated Transformation of the *Oryza sativa* Thaumatin-Like Protein to Canola (R Line Hyola308) for Enhancing Resistance to *Sclerotinia sclerotiorum*.” *Iranian Journal of Biotechnology* 15 (3): 201–7. https://doi.org/10.15171/ijb.1585.

Al Akeel, R., Mateen, A., Syed, R., Alyousef, A. A., and Shaik, M. R. 2017. “Screening, Purification and Characterization of Anionic Antimicrobial Proteins from *Foeniculum vulgare*.” *Molecules* 22 (4). https://doi.org/10.3390/molecules22040602.

Almasia, N. I., Bazzini, A. A., Hopp, H. E., and Vazquez-Rovere, C. 2008. “Overexpression of Snakin-1 Gene Enhances Resistance to *Rhizoctonia solani* and *Erwinia carotovora* in Transgenic Potato Plants.” *Molecular Plant Pathology* 9 (3): 329–38. https://doi.org/10.1111/j.1364-3703.2008.00469.x.

Almasia, N. I., Narhirñak, V., Hopp, H. E., and Vazquez-Rovere, C. 2010. “Isolation and Characterization of the Tissue and Development-Specific Potato Snakin-1 Promoter Inducible by Temperature and Wounding.” *Electronic Journal of Biotechnology* 13 (5): 0–0. https://doi.org/10.2225/vol13-issue5-fulltext-12.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anand, A., Zhou, T., Trick, H. N., Gill, B. S., Bockus, W. W., and Muthukrishnan, S. 2003. “Greenhouse and Field Testing of Transgenic Wheat Plants Stably Expressing Genes for Thaumatin-like Protein, Chitinase and Glucanase against *Fusarium graminearum*.” *Journal of Experimental Botany* 54 (384): 1101–11. https://doi.org/10.1093/jxb/erg110.

Andrusier, N., Mashiach, E., Nussinov, R., and Wolfson, H. J. 2008. “Principles of Flexible Protein-Protein Docking.” *Proteins: Structure, Function, and Bioinformatics* 73 (2): 271–89. https://doi.org/10.1002/prot.22170.

Antuch, W., Güntert, P., and Wüthrich, K. 1996. “Ancestral Beta Gamma-Crystallin Precursor Structure in a Yeast Killer Toxin.” *Nature Structural Biology* 3 (8): 662–65.

Anuradha, T. S., Divya, K., Jami, S. K., and Kirti, P. B. 2008. “Transgenic Tobacco and Peanut Plants Expressing a Mustard Defensin Show Resistance to Fungal Pathogens.” *Plant Cell Reports* 27 (11): 1777–86. https://doi.org/10.1007/s00299-008-0596-8.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., ... and Martin, M. J. 2004. “UniProt: The Universal Protein Knowledgebase.” *Nucleic Acids Research* 32 (90001): 115D – 119. https://doi.org/10.1093/nar/gkh131.

Archer, B. L. 1960. “The Proteins of *Hevea brasiliensis* Latex. 4. Isolation and Characterization of Crystalline Hevein.” *The Biochemical Journal* 75 (May): 236–40.

Asensio, J. L., Siebert, H. C., Laynez, J., Poveda, A., Nieto, P. M., Soedjanaamadja, U. M., ... and Jiménez-Barbero, J. 2000. “Structural Basis for Chitin Recognition by Defense Proteins: GlcNAc Residues Are Bound in a Multivalent Fashion by Extended Binding Sites in Hevein Domains.” *Chemistry and Biology* 7 (7): 529–43. https://doi.org/10.1016/S1074-5521(00)00136-8.

Audie, J., and Swanson, J. 2012. “Recent Work in the Development and Application of Protein–Peptide Docking.” *Future Medicinal Chemistry* 4 (12): 1619–44. https://doi.org/10.4155/fmc.12.99.

Audie, J., and Swanson, J. 2013. “Advances in the Prediction of Protein-Peptide Binding Affinities: Implications for Peptide-Based Drug Discovery: Protein-Peptide Binding Affinities.” *Chemical Biology and Drug Design* 81 (1): 50–60. https://doi.org/10.1111/cbdd.12076.

Balaji, V., and Smart, C. D. 2012. “Over-Expression of Snakin-2 and Extensin-like Protein Genes Restricts Pathogen Invasiveness and Enhances Tolerance to *Clavibacter michiganensis* Subsp. Michiganensis in Transgenic Tomato (*Solanum lycopersicum*).” *Transgenic Research* 21 (1): 23–37. https://doi.org/10.1007/s11248-011-9506-x.

Baracat-Pereira, M. C., Barbosa, M. D. O., Magalhães Júnior, M. J., Carrijo, L. C., Games, P. D., Almeida, H. O., ... and Barros, E. G. D. 2012. “Separomics Applied to the Proteomics and Peptidomics of Low-Abundance

Proteins: Choice of Methods and Challenges - A Review." *Genetics and Molecular Biology* 35 (SUPPL.1): 283–291. https://doi.org/10.1590/S1415-47572012000200009.

Batalia, M. A., Monzingo, A. F., Ernst, S., Roberts, W., and Robertus, J. D. 1996. "The Crystal Structure of the Antifungal Protein Zeamatin, a Member of the Thaumatin-like, PR-5 Protein Family." *Nature Structural Biology* 3 (1): 19–23. https://doi.org/10.1038/nsb0196-19.

Beintema, J. J. 1994. "Structural Features of Plant Chitinases and Chitin-Binding Proteins." *FEBS Letters* 350 (2–3): 159–63.

Benko-Iseppon, A. M., and Crovella, S. 2010. "Ethnobotanical Bioprospection of Candidates for Potential Antimicrobial Drugs from Brazilian Plants: State of Art and Perspectives." *Current Protein and Peptide Science* 11 (3): 189–94.

Benko-Iseppon, A. M., Lins Galdino, S., Calsa Jr, T., Kido, E. A., Tossi, A., Carlos Belarmino, L., and Crovella, S. 2010. "Overview on Plant Antimicrobial Peptides." *Current Protein and Peptide Science* 11 (3): 181–88.

Benko-Iseppon, A. M., Pinangé, D. S. B., Chang, S. C., and Morawetz, W. 2012. "Ethnobotanical Uses of the Native Flora from Brazilian North-Eastern Region." *Medicinal Plants: Biodiversity and Drugs*, 84.

Berecz, B., Mills, E. C., Tamás, L., Lang, F., Shewry, P. R., and Mackie, A. R. 2010. "Structural stability and surface activity of sunflower 2S albumins and nonspecific lipid transfer protein". *Journal of agricultural and food chemistry* 58 (10): 6490-6497.

Berrocal-Lobo, M., Segura, A., Moreno, M., López, G., Garcıa-Olmedo, F., and Molina, A. 2002. "Snakin-2, an Antimicrobial Peptide from Potato Whose Gene Is Locally Induced by Wounding and Responds to Pathogen Infection." *Plant Physiology* 128 (3): 951–61. https://doi.org/10.1104/pp.010685.

Bhave, M., and Morris, C. F. 2008. "Molecular Genetics of Puroindolines and Related Genes: Allelic Diversity in Wheat and Other Grasses." *Plant Molecular Biology* 66 (3): 205–19. https://doi.org/10.1007/s11103-007-9263-7.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., ... and Schwede, T. 2014. "SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information." *Nucleic Acids Research* 42 (W1): W252–58. https://doi.org/10.1093/nar/gku340.

Blueggel, M., Chamrad, D., and Meyer, H. E. 2004. "Bioinformatics in Proteomics." *Current Pharmaceutical Biotechnology* 5 (1): 79–88.

Bolouri Moghaddam, M. R., Vilcinskas, A., and Rahnamaeian, M. 2016. "Cooperative Interaction of Antimicrobial Peptides with the Interrelated Immune Pathways in Plants: AMPs in Plant Immunity." *Molecular Plant Pathology* 17 (3): 464–71. https://doi.org/10.1111/mpp.12299.

Boonpa, K., Tantong, S., Weerawanich, K., Panpetch, P., Pringsulaka, O., Yingchutrakul, Y., ... and Sirikantaramas, S. 2018. "Heterologous Expression and Antimicrobial Activity of OsGASR3 from Rice (*Oryza sativa* L.)." *Journal of Plant Physiology* 224–225 (May): 95–102. https://doi.org/10.1016/j.jplph.2018.03.013.

Botero, D., Alvarado, C., Bernal, A., Danies, G., and Restrepo, S. 2018. "Network Analyses in Plant Pathogens." *Frontiers in Microbiology* 9 (January). https://doi.org/10.3389/fmicb.2018.00035.

Bowdish, D. M., Davidson, D. J., Lau, Y. E., Lee, K., Scott, M. G., and Hancock, R. E. 2005. "Impact of LL-37 on Anti-Infective Immunity." *Journal of Leukocyte Biology* 77 (4): 451–59. https://doi.org/10.1189/jlb.0704380.

Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., ... and Salmon, J. K. 2006. "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters." In, 43–43. IEEE. https://doi.org/10.1109/SC.2006.54.

Brandazza, A., Angeli, S., Tegoni, M., Cambillau, C., and Pelosi, P. 2004. "Plant Stress Proteins of the Thaumatin-like Family Discovered in Animals" 572: 3–7. https://doi.org/10.1016/j.febslet.2004.07.003.

Breiteneder, H. 2004. "Thaumatin-like Proteins – a New Family of Pollen and Fruit Allergens," no. 16: 479–481.

Breiteneder, H., and Radauer, C. 2004. "A Classification of Plant Food Allergens." *Journal of Allergy and Clinical Immunology* 113 (5): 821–30. https://doi.org/10.1016/j.jaci.2004.01.779.

Broekaert, W. F., Van Parijs, J. A. N., Leyns, F., Joos, H., and Peumans, W. J. 1989. "A Chitin-Binding Lectin from Stinging Nettle Rhizomes with Antifungal Properties." *Science* 245 (4922): 1100–1102. https://doi.org/10.1126/science.245.4922.1100.

Broekaert, W. F., Marien, W., Terras, F. R., De Bolle, M. F., Proost, P., Van Damme, J., ... and Rees, S. B. 1992. "Antimicrobial Peptides from *Amaranthus caudatus* Seeds with Sequence Homology to the Cysteine/Glycine-Rich Domain of Chitin-Binding Proteins." *Biochemistry* 31 (17): 4308–14. https://doi.org/10.1021/bi00132a023.

Broekaert, W. F., Terras, F. R., Cammue, B. P., and Osborn, R. W. 1995. "Plant Defensins: Novel Antimicrobial Peptides as Components of the Host Defense System." *Plant Physiology* 108 (4): 1353–58.

Broekaert, W. F., Cammue, B. P., De Bolle, M. F., Thevissen, K., De Samblanx, G. W., Osborn, R. W., and Nielson, K. 1997. "Antimicrobial Peptides from Plants." *Critical Reviews in Plant Sciences* 16 (3): 297–323. https://doi.org/10.1080/07352689709701952.

Brogden, K. A. 2005. "Antimicrobial Peptides: Pore Formers or Metabolic Inhibitors in Bacteria?" *Nature Reviews Microbiology* 3 (3): 238–50. https://doi.org/10.1038/nrmicro1098.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. A., and Karplus, M. 1983. "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations." *Journal of Computational Chemistry* 4 (2): 187–217. https://doi.org/10.1002/jcc.540040211.

Bruix, M., Gonzalez, C., Santoro, J., Soriano, F., Rocher, A., Mendez, E., and Rico, M. 1995. "^{1}H-NMR Studies on the Structure of a New Thionin from Barley Endosperm: Structure of a New Thionin." *Biopolymers* 36 (6): 751–63. https://doi.org/10.1002/bip.360360608.

Burman, R., Strömstedt, A. A., Malmsten, M., and Göransson, U. 2011. "Cyclotide–Membrane Interactions: Defining Factors of Membrane Binding, Depletion and Disruption." *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1808 (11): 2665–73. https://doi.org/10.1016/j.bbamem.2011.07.004.

Cabanos, C., Ekyo, A., Amari, Y., Kato, N., Kuroda, M., Nagaoka, S., ... and Maruyama, N. 2013. "High-Level Production of Lactostatin, a Hypocholesterolemic Peptide, in Transgenic Rice Using Soybean A1aB1b as Carrier." *Transgenic Research* 22 (3): 621–29.

Camarero, J. A. 2017. "Cyclotides, a Versatile Ultrastable Micro-Protein Scaffold for Biotechnological Applications." *Bioorganic and Medicinal Chemistry Letters* 27 (23): 5089–99. https://doi.org/10.1016/j.bmcl.2017.10.051.

Cândido, E. S., Cardoso, M. H., Sousa D. A., Viana, J. C., Oliveira-Júnior, N. G., Miranda, V., and Franco, O. L. 2014. "The Use of Versatile Plant Antimicrobial Peptides in Agribusiness and Human Health." *Peptides* 55 (May): 65–78. https://doi.org/10.1016/j.peptides.2014.02.003.

Carmona, M. J., Molina, A., Fernández, J. A., López-Fando, J. J., and García-Olmedo, F. 1993. "Expression of the Alpha-Thionin Gene from Barley in Tobacco Confers Enhanced Resistance to Bacterial Pathogens." *The Plant Journal: For Cell and Molecular Biology* 3 (3): 457–62.

Carvalho, A. O., Teodoro, C. E. S., Cunha, M., Okorokova-Façanha, A. L., Okorokov, L. A., Fernandes, K. V., and Gomes, V. M. 2004. "Intracellular Localization of a Lipid Transfer Protein in *Vigna unguiculata* Seeds." *Physiologia Plantarum* 122 (3): 328–36. https://doi.org/10.1111/j.1399-3054.2004.00413.x.

Carvalho, A. O., and Gomes V. M. 2007. "Role of Plant Lipid Transfer Proteins in Plant Cell Physiology-A Concise Review." *Peptides* 28 (5): 1144–53. https://doi.org/10.1016/j.peptides.2007.03.004.

Carvalho, A. O., and Gomes V. M. 2009. "Plant Defensins-Prospects for the Biological Functions and Biotechnological Properties." *Peptides* 30 (5): 1007–20. https://doi.org/10.1016/j.peptides.2009.01.018.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., ... and Woods, R. J. 2005. "The Amber Biomolecular Simulation Programs." *Journal of Computational Chemistry* 26 (16): 1668–88. https://doi.org/10.1002/jcc.20290.

Castro, M. S., Gerhardt, I. R., Orru, S., Pucci, P., and Bloch Jr, C. 2003. "Purification and Characterization of a Small (7.3 KDa) Putative Lipid Transfer Protein from Maize Seeds." *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* 794 (1): 109–14.

Chan, Y. L., Prasad, V., Chen, K. H., Liu, P. C., Chan, M. T., and Cheng, C. P. 2005. "Transgenic Tomato Plants Expressing an *Arabidopsis* Thionin (Thi2.1) Driven by Fruit-Inactive Promoter Battle against Phytopathogenic Attack." *Planta* 221 (3): 386–93. https://doi.org/10.1007/s00425-004-1459-3.

Chandra, H., Bishnoi, P., Yadav, A., Patni, B., Mishra, A. P., and Nautiyal, A. R. 2017. "Antimicrobial Resistance and the Alternative Resources with Special Emphasis on Plant-Based Antimicrobials - A Review." *Plants* 6 (2). https://doi.org/10.3390/plants6020016.

Charnet, P., Molle, G., Marion, D., Rousset, M., and Lullien-Pellerin, V. 2003. "Puroindolines Form Ion Channels in Biological Membranes." *Biophysical Journal* 84 (4): 2416–26. https://doi.org/10.1016/S0006-3495(03)75046-2.

Chen, W. P., Chen, P. D., Liu, D. J., Kynast, R., Friebe, B., Velazhahan, R., ... and Gill, B. S. 1999. "Development of Wheat Scab Symptoms Is Delayed in Transgenic Wheat Plants That Constitutively Express a Rice Thaumatin-like Protein Gene." *TAG Theoretical and Applied Genetics* 99 (5): 755–60. https://doi.org/10.1007/s001220051294.

Chen, S. C., Liu, A. R., and Zou, Z. R. 2006. "Overexpression of Glucanase Gene and Defensin Gene in Transgenic Tomato Enhances Resistance to *Ralstonia solanacearum*." *Russian Journal of Plant Physiology* 53 (5): 671–77. https://doi.org/10.1134/S1021443706050116.

Cheng, G., Dai, M., Ahmed, S., Hao, H., Wang, X., and Yuan, Z. 2016. "Antimicrobial Drugs in Fighting against Antimicrobial Resistance." *Frontiers in Microbiology* 7 (April). https://doi.org/10.3389/fmicb.2016.00470.

Chiche, L., Heitz, A., Gelly, J. C., Gracy, J., Chau, P. T., Ha, P. T., ... and Le-Nguyen, D. 2004. "Squash Inhibitors: From Structural Motifs to Macrocyclic Knottins." *Current Protein and Peptide Science* 5 (5): 341–49.

Cho, W. C. 2007. "Proteomics Technologies and Challenges." *Genomics, Proteomics and Bioinformatics* 5 (2): 77–85. https://doi.org/10.1016/S1672-0229(07)60018-7.

Choi, M. S., Kim, Y. H., Park, H. M., Seo, B. Y., Jung, J. K., Kim, S. T., ... and Kim, C. K. 2009. "Expression of BrD1, a plant defensin from *Brassica rapa*, confers resistance against brown planthopper (*Nilaparvata lugens*) in transgenic rices". *Molecules and cells* 28 (2): 131-137.

Claeson, P., Göransson, U., Johansson, S., Luijendijk, T., and Bohlin, L. 1998. "Fractionation Protocol for the Isolation of Polypeptides from Plant Biomass." *Journal of Natural Products* 61 (1): 77–81. https://doi.org/10.1021/np970342r.

Colgrave, M. L., and Craik, D. J. 2004. "Thermal, Chemical, and Enzymatic Stability of the Cyclotide Kalata B1: The Importance of the Cyclic Cystine Knot." *Biochemistry* 43 (20): 5965–75. https://doi.org/10.1021/bi049711q.

Colgrave, M. L., Kotze, A. C., Huang, Y. H., O'Grady, J., Simonsen, S. M., and Craik, D. J. 2008. "Cyclotides: Natural, Circular Plant Peptides That Possess Significant Activity against Gastrointestinal Nematode Parasites of Sheep †." *Biochemistry* 47 (20): 5581–89. https://doi.org/10.1021/bi800223y.

Colilla, F. J., Rocher, A., and Mendez, E. 1990. "γ-Purothionins: Amino Acid Sequence of Two Polypeptides of a New Family of Thionins from Wheat Endosperm." *FEBS Letters* 270 (1–2): 191–94. https://doi.org/10.1016/0014-5793(90)81265-P.

Corbeil, C. R., Williams, C. I., and Labute, P. 2012. "Variability in Docking Success Rates Due to Dataset Preparation." *Journal of Computer-Aided Molecular Design* 26 (6): 775–86. https://doi.org/10.1007/s10822-012-9570-1.

Craig, R., and Beavis, R. C. 2004. "TANDEM: Matching Proteins with Tandem Mass Spectra." *Bioinformatics* 20 (9): 1466–1467. https://doi.org/10.1093/bioinformatics/bth092.

Craik, D. J., Daly, N. L., Bond, T., and Waine, C. 1999. "Plant Cyclotides: A Unique Family of Cyclic and Knotted Proteins That Defines the Cyclic Cystine Knot Structural Motif." *Journal of Molecular Biology* 294 (5): 1327–36. https://doi.org/10.1006/jmbi.1999.3383.

Craik, D. J., Clark, R. J., and Daly, N. L. 2007. "Potential Therapeutic Applications of the Cyclotides and Related Cystine Knot Mini-Proteins." *Expert Opinion on Investigational Drugs* 16 (5): 595–604. https://doi.org/10.1517/13543784.16.5.595.

Craik, D. J., Lee, M. H., Rehm, F. B., Tombling, B., Doffek, B., and Peacock, H. 2018. "Ribosomally-Synthesised Cyclic Peptides from Plants as Drug Leads and Pharmaceutical Scaffolds." *Bioorganic and Medicinal Chemistry* 26 (10): 2727–37. https://doi.org/10.1016/j.bmc.2017.08.005.

Dall'Antonia, Y., Pavkov, T., Fuchs, H., Breiteneder, H., and Keller, W. 2005. "Crystallization Communications Crystallization and Preliminary Structure Determination of the Plant Food Allergen Pru Av 2 Crystallization Communications," 186–188. https://doi.org/10.1107/S1744309104033822.

Darwish, N. A., Khan, R. S., Ntui, V. O., Nakamura, I., and Mii, M. 2014. "Generation of selectable marker-free transgenic eggplant resistant to *Alternaria solani* using the R/RS site-specific recombination system". *Plant cell reports* 33 (3): 411-421.

Datta, K., Velazhahan, R., Oliva, N., Ona, I., Mew, T., Khush, G. S., ... and Datta, S. K. 1999. "Over-Expression of the Cloned Rice Thaumatin-like Protein (PR-5) Gene in Transgenic Rice Plants Enhances Environmental Friendly Resistance to *Rhizoctonia solani* Causing Sheath Blight Disease." *TAG Theoretical and Applied Genetics* 98 (6–7): 1138–45. https://doi.org/10.1007/s001220051178.

Dhatwalia, V. K., Sati, O. P., Tripathi, M. K., and Kumar, A. 2009. "Isolation, Characterization and Antimicrobial Activity at Diverse Dilution of Wheat Puroindoline Protein." *World Journal of Agricultural Sciences* 5 (3): 297–300.

Ding, S., Li, H., Su, C., Yu, J., and Jin, F. 2013. "Evolutionary Artificial Neural Networks: A Review." *Artificial Intelligence Review* 39 (3): 251–60. https://doi.org/10.1007/s10462-011-9270-6.

Dobslaff, K., Kreisig, T., Berthold, N., Hoffmann, R., and Zuchner, T. 2012. "Novel Peptide–Protein Assay for Identification of Antimicrobial Peptides by Fluorescence Quenching." *Analytical and Bioanalytical Chemistry* 403 (9): 2725–31. https://doi.org/10.1007/s00216-012-6050-3.

Domínguez, E., Heredia-Guerrero, J. A., and Heredia, A. 2015. "Plant Cutin Genesis: Unanswered Questions." *Trends in Plant Science* 20 (9): 551–58. https://doi.org/10.1016/j.tplants.2015.05.009.

Douliez, J. P., Michon, T., Elmorjani, K., and Marion, D. 2000. "Mini Review: Structure, Biological and Technological Functions of Lipid Transfer Proteins and Indolines, the Major Lipid Binding Proteins from Cereal Kernels." *Journal of Cereal Science* 32 (1): 1–20. https://doi.org/10.1006/jcrs.2000.0315.

Eddy, S. R. 1998. "Profile Hidden Markov Models." *Bioinformatics (Oxford, England)* 14 (9): 755–63.

Edqvist, J., Blomqvist, K., Nieuwland, J., and Salminen, T. A. 2018. "Plant Lipid Transfer Proteins: Are We Finally Closing in on the Roles of These Enigmatic Proteins?" *Journal of Lipid Research* 59 (8): 1374–82. https://doi.org/10.1194/jlr.R083139.

Edstam, M. M., Viitanen, L., Salminen, T. A., and Edqvist, J. 2011. "Evolutionary History of the Non-Specific Lipid Transfer Proteins." *Molecular Plant* 4 (6): 947–64.

Edstam, M. M., and Edqvist, J. 2014. "Involvement of GPI-anchored lipid transfer proteins in the development of seed coats and pollen in *Arabidopsis thaliana*". *Physiologia plantarum* 152 (1): 32-42.

Elsbach, P. 2003. "What Is the Real Role of Antimicrobial Polypeptides That Can Mediate Several Other Inflammatory Responses?" *Journal of Clinical Investigation* 111 (11): 1643–45. https://doi.org/10.1172/JCI18761.

Eltabakh, M. Y., Ouzzani, M., and Aref, W. G. 2006. "BDBMS--a Database Management System for Biological Data." *ArXiv Preprint Cs/0612127*.

Eng, J. K., Fischer, B., Grossmann, J., and MacCoss, M. J. 2008. "A Fast SEQUEST Cross Correlation Algorithm," 4598–4602.

Epple, P., Apel, K., and Bohlmann, H. 1997. "Overexpression of an endogenous thionin enhances resistance of *Arabidopsis* against *Fusarium oxysporum*". *The Plant Cell* 9 (4): 509-520.

Essig, A., Hofmann, D., Münch, D., Gayathri, S., Künzler, M., Kallio, P. T., ... and Aebi, M. 2014. "Copsin, a Novel Peptide-Based Fungal Antibiotic Interfering with the Peptidoglycan Synthesis." *Journal of Biological Chemistry* 289 (50): 34953–64. https://doi.org/10.1074/jbc.M114.599878.

Faccio, P., Vazquez-Rovere, C., Hopp, E., Gonzalez, G., Decima-Oneto, C., Favret, E., ... and Franzone, P. 2011. "Increased Tolerance to Wheat Powdery Mildew by Heterologous Constitutive Expression of the *Solanum chacoense* Snakin-1 Gene." *Czech Journal of Genetics and Plant Breeding* 47: 135–41.

Farrokhi, N., Whitelegge, J. P., and Brusslan, J. A. 2008. "Plant Peptides and Peptidomics." *Plant Biotechnology Journal* 6 (2): 105–34. https://doi.org/10.1111/j.1467-7652.2007.00315.x.

Fedders, H., and Leippe, M. 2008. "A Reverse Search for Antimicrobial Peptides in *Ciona intestinalis*: Identification of a Gene Family Expressed in Hemocytes and Evaluation of Activity." *Developmental and Comparative Immunology* 32 (3): 286–298. https://doi.org/10.1016/j.dci.2007.06.003.

Felício, M. R., Silva, O. N., Gonçalves, S., Santos, N. C., and Franco, O. L. 2017. "Peptides with Dual Antimicrobial and Anticancer Activities." *Frontiers in Chemistry* 5 (February). https://doi.org/10.3389/fchem.2017.00005.

Field, H. I., Fenyö, D., and Beavis, R. C. 2002. "RADARS, a Bioinformatics Solution That Automates Proteome Mass Spectral Analysis, Optimises Protein Identification, and Archives Data in a Relational Database." *Proteomics* 2 (1): 36–47. https://doi.org/10.1002/1615-9861(200201)2:1<36::AID-PROT36>3.0.CO;2-W.

Fierens, E., Rombouts, S., Gebruers, K., Goesaert, H., Brijs, K., Beaugrand, J., ... and Delcour, J. A. 2007. "TLXI , a Novel Type of Xylanase Inhibitor from Wheat (*Triticum aestivum*) Belonging to the Thaumatin Family" 591: 583–591. https://doi.org/10.1042/BJ20061291.

Fierens, E., Gebruers, K., Voet, A. R., De Maeyer, M., Courtin, C. M., and Delcour, J. A. 2009. "Biochemical and Structural Characterization of TLXI, the *Triticum aestivum* L. Thaumatin-like Xylanase Inhibitor" 24 (June): 646–654. https://doi.org/10.1080/14756360802321831.

Flaherty, R. A., Freed, S. D., and Lee, S. W. 2014. "The Wide World of Ribosomally Encoded Bacterial Peptides." Edited by Virginia Miller. *PLoS Pathogens* 10 (7): e1004221. https://doi.org/10.1371/journal.ppat.1004221.

Freddolino, P. L., Harrison, C. B., Liu, Y., and Schulten, K. 2010. "Challenges in Protein-Folding Simulations." *Nature Physics* 6 (10): 751–58. https://doi.org/10.1038/nphys1713.

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., ... and Mainz, D. T. 2006. "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes." *Journal of Medicinal Chemistry* 49 (21): 6177–96. https://doi.org/10.1021/jm051256o.

Fukuzawa, N., Tabayashi, N., Okinaka, Y., Furusawa, R., Furuta, K., Kagaya, U., and Matsumura, T. 2010. "Production of Biologically Active Atlantic Salmon Interferon in Transgenic Potato and Rice Plants." *Journal of Bioscience and Bioengineering* 110 (2): 201–7.

Furukawa, T., Sakaguchi, N., and Shimada, H. 2006. "Two OsGASR Genes, Rice GAST Homologue Genes That Are Abundant in Proliferating Tissues, Show Different Expression Patterns in Developing Panicles." *Genes and Genetic Systems* 81 (3): 171–80.

Galdino da Rocha Pitta, M., Galdino da Rocha Pitta, M., and Lins Galdino, S. 2010. "Development of Novel Therapeutic Drugs in Humans from Plant Antimicrobial Peptides." *Current Protein and Peptide Science* 11 (3): 236–47.

Games, P. D., Gonçalves daSilva, E. Q., de Oliveira Barbosa, M., Almeida-Souza, H. O., Fontes, P. P., Pereira, P. R. G., ... and Baracat-Pereira, M. C. 2016. "Computer Aided Identification of a Hevein-like Antimicrobial Peptide of Bell Pepper Leaves for Biotechnological Use." *BMC Genomics* 17 (S12). https://doi.org/10.1186/s12864-016-3332-8.

Gao, A. G., Hakimi, S. M., Mittanck, C. A., Wu, Y., Woerner, B. M., Stark, D. M., ... and Rommens, C. M. 2000. "Fungal Pathogen Protection in Potato by Expression of a Plant Defensin Peptide." *Nature Biotechnology* 18 (12): 1307–10. https://doi.org/10.1038/82436.

García-Olmedo, F., Molina, A., Alamillo, J. M., and Rodríguez-Palenzuéla, P. 1998. "Plant Defense Peptides." *Biopolymers* 47 (6): 479–91. https://doi.org/10.1002/(SICI)1097-0282(1998)47:6<479::AID-BIP6>3.0.CO;2-K.

García, A. N., Ayub, N. D., Fox, A. R., Gómez, M. C., Diéguez, M. J., Pagano, E. M., ... and Soto, G. 2014. "Alfalfa Snakin-1 Prevents Fungal Colonization and Probably Coevolved with Rhizobia." *BMC Plant Biology* 14 (1). https://doi.org/10.1186/s12870-014-0248-9.

Gaspar, Y. M., McKenna, J. A., McGinness, B. S., Hinch, J., Poon, S., Connelly, A. A., ... and Heath, R. L. 2014. "Field Resistance to *Fusarium oxysporum* and *Verticillium dahliae* in Transgenic Cotton Expressing the Plant Defensin NaD1." *Journal of Experimental Botany* 65 (6): 1541–50. https://doi.org/10.1093/jxb/eru021.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. 2005. "Protein Identification and Analysis Tools on the ExPASy Server." *The Proteomics Protocols Handbook*, 571–607. https://doi.org/10.1385/1-59259-890-0:571.

Gautier, M. F., Aleman, M. E., Guirao, A., Marion, D., and Joudrier, P. 1994. "*Triticum aestivum* Puroindolines, Two Basic Cystine-Rich Seed Proteins: CDNA Sequence Analysis and Developmental Gene Expression." *Plant Molecular Biology* 25 (1): 43–57.

Gazzana, G., and Borlak, J. 2007. "Improved Method for Proteome Mapping of the Liver by 2-DE MALDI-TOF MS." *Journal of Proteome Research* 6 (8): 3143–3151. https://doi.org/10.1021/pr070097p.

Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., ... and Bryant, S. H. 2004. “Open Mass Spectrometry Search Algorithm.” *Journal of Proteome Research* 3 (5): 958–964. https://doi.org/10.1021/pr0499491.

Germain, H., Chevalier, E., and Matton, D. P. 2006. “Plant Bioactive Peptides: An Expanding Class of Signaling Molecules.” *Canadian Journal of Botany* 84 (1): 1–19. https://doi.org/10.1139/b05-162.

Gerwick, W. H. 2013. “Plant Sources of Drugs and Chemicals.” In *Encyclopedia of Biodiversity*, 129–39. Elsevier. https://doi.org/10.1016/B978-0-12-384719-5.00111-8.

Ghag, S. B., Shekhawat, U. K. S., and Ganapathi, T. R. 2012. "Petunia floral defensins with unique prodomains as novel candidates for development of Fusarium wilt resistance in transgenic banana plants". *PLoS One* 7 (6): e39557.

Ghag, S. B., Shekhawat, U. K. S., and Ganapathi, T. R. 2014. "Transgenic banana plants expressing a Stellariamedia defensin gene (Sm-AMP-D1) demonstrate improved resistance to Fusariumoxysporum". *Plant Cell Tissue and Organ Culture* 119 (2): 247-255.

Ghoorah, A. W., Devignes, M. D., Smaïl-Tabbone, M., and Ritchie, D. W. 2013. “Protein Docking Using Case-Based Reasoning: Protein Docking Using Case-Based Reasoning.” *Proteins: Structure, Function, and Bioinformatics* 81 (12): 2150–58. https://doi.org/10.1002/prot.24433.

Ghosh, R., and Chakrabarti, C. 2008. “Crystal Structure Analysis of NP24-I: A Thaumatin-like Protein.” *Planta* 228 (5): 883–890. https://doi.org/10.1007/s00425-008-0790-5.

Gijzen, M., Kuflu, K., Qutob, D., and Chernys, J. T. 2001. “A Class I Chitinase from Soybean Seed Coat.” *Journal of Experimental Botany* 52 (365): 2283–89.

Gincel, E., Simorre, J. P., Caille, A., Marion, D., Ptak, M., and Vovelle, F. 1994. “Three-Dimensional Structure in Solution of a Wheat Lipid Transfer Protein from Multidimensional 1H-NMR Data. A New Folding for Lipid Carriers.” *European Journal of Biochemistry* 226 (2): 413–22.

Giroux, M. J., Sripo, T., Gerhardt, S., and Sherwood, J. 2003. “Puroindolines: Their Role in Grain Hardness and Plant Defence.” *Biotechnology and Genetic Engineering Reviews* 20 (1): 277–90. https://doi.org/10.1080/02648725.2003.10648047.

Gogoladze, G., Grigolava, M., Vishnepolsky, B., Chubinidze, M., Duroux, P., Lefranc, M. P., and Pirtskhalava, M. 2014. “DBAASP : Database of Antimicrobial Activity and Structure of Peptides.” *FEMS Microbiology Letters* 357 (1): 63–68. https://doi.org/10.1111/1574-6968.12489.

Göransson, U., Burman, R., Gunasekera, S., Strömstedt, A. A., and Rosengren, K. J. 2012. “Circular Proteins from Plants and Fungi.” *Journal of Biological Chemistry* 287 (32): 27001–6. https://doi.org/10.1074/jbc.R111.300129.

Goyal, R. K., and Mattoo, A. K. 2016. “Plant Antimicrobial Peptides.” In *Host Defense Peptides and Their Potential as Therapeutic Agents*, 111–36. Springer.

Gran, L. 1973. “On the Effect of a Polypeptide Isolated from ‘Kalata-Kalata’ (*Oldenlandia affinis* DC) on the Oestrogen Dominated Uterus.” *Acta Pharmacologica Et Toxicologica* 33 (5): 400–408.

Greenstein, S., Shadkchan, Y., Jadoun, J., Sharon, C., Markovich, S., and Osherov, N. 2006. “Analysis of the *Aspergillus nidulans* Thaumatin-like CetA Gene and Evidence for Transcriptional Repression of Pyr4 Expression in the CetA -Disrupted Strain” 43: 42–53. https://doi.org/10.1016/j.fgb.2005.10.001.

Grenier, J., Potvin, C., Trudel, J., and Asselin, A. 1999. “Some Thaumatin-like Proteins Hydrolyse Polymeric β-1,3-Glucans.” *Plant Journal* 19 (4): 473–480. https://doi.org/10.1046/j.1365-313X.1999.00551.x.

Grenier, J., Potvin, C., and Asselin, A. 2000. “Some Fungi Express Beta-1,3-Glucanases Similar to Thaumatin-like Proteins.” *Mycologia* 92 (5): 841. https://doi.org/10.2307/3761579.

Gustafson, K. R., Sowder, R. C., Henderson, L. E., Parsons, I. C., Kashman, Y., Cardellina, J. H., ... and Boyd, M. R. 1994. “Circulins A and B. Novel Human Immunodeficiency Virus (HIV)-Inhibitory Macrocyclic Peptides from the Tropical Tree *Chassalia parvifolia*.” *Journal of the American Chemical Society* 116 (20): 9337–38. https://doi.org/10.1021/ja00099a064.

Guzmán-Rodríguez, J. J., Ibarra-Laclette, E., Herrera-Estrella, L., Ochoa-Zarzosa, A., Suárez-Rodríguez, L. M., Rodríguez-Zapata, L. C., ... and López-Gómez, R. 2013. “Analysis of Expressed Sequence Tags (ESTs) from Avocado Seed (*Persea americana* var. drymifolia) Reveals Abundant Expression of the Gene Encoding the Antimicrobial Peptide Snakin.” *Plant Physiology and Biochemistry* 70 (September): 318–24. https://doi.org/10.1016/j.plaphy.2013.05.045.

Haas, J., Barbato, A., Behringer, D., Studer, G., Roth, S., Bertoni, M., ... and Schwede, T. 2018. “Continuous Automated Model Evaluation (CAMEO) Complementing the Critical Assessment of Structure Prediction in CASP12.” *Proteins: Structure, Function, and Bioinformatics* 86 (March): 387–98. https://doi.org/10.1002/prot.25431.

Hammami, R., Ben Hamida, J., Vergoten, G., and Fliss, I. 2009. “PhytAMP: A Database Dedicated to Antimicrobial Plant Peptides.” *Nucleic Acids Research* 37 (Database): D963–68. https://doi.org/10.1093/nar/gkn655.

Hammami, R., and Fliss, I. 2010. “Current Trends in Antimicrobial Agent Research: Chemo and Bioinformatics Approaches.” *Drug Discovery Today* 15 (13): 540–46. https://doi.org/10.1016/j.drudis.2010.05.002.

Hancock, R. E. 2001. “Cationic Peptides: Effectors in Innate Immunity and Novel Antimicrobials.” *The Lancet Infectious Diseases* 1 (3): 156–64.

Hashempour, H., Koehbach, J., Daly, N. L., Ghassempour, A., and Gruber, C. W. 2013. “Characterizing Circular Peptides in Mixtures: Sequence Fragment Assembly of Cyclotides from a Violet Plant by MALDI-TOF/TOF Mass Spectrometry.” *Amino Acids* 44 (2): 581–95. https://doi.org/10.1007/s00726-012-1376-x.

He, R., Wu, J., Zhang, Y., Agüero, C. B., Li, X., Liu, S., ... and Lu, J. 2017. “Overexpression of a Thaumatin-like Protein Gene from *Vitis amurensis* Improves Downy Mildew Resistance in *Vitis vinifera* Grapevine.” *Protoplasma* 254 (4): 1579–89. https://doi.org/10.1007/s00709-016-1047-y.

Henriques, S. T., Huang, Y. H., Chaousis, S., Sani, M. A., Poth, A. G., Separovic, F., and Craik, D. J. 2015. “The Prototypic Cyclotide Kalata B1 Has a Unique Mechanism of Entering Cells.” *Chemistry and Biology* 22 (8): 1087–97. https://doi.org/10.1016/j.chembiol.2015.07.012.

Herbel, V., Schäfer, H., and Wink, M. 2015. “Recombinant Production of Snakin-2 (an Antimicrobial Peptide from Tomato) in *E. coli* and Analysis of Its Bioactivity.” *Molecules* 20 (8): 14889–901. https://doi.org/10.3390/molecules200814889.

Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. 2008. “GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.” *Journal of Chemical Theory and Computation* 4 (3): 435–47. https://doi.org/10.1021/ct700301q.

Higgins, C. F., and Payne, J. W. 1982. “Plant Peptides.” In *Nucleic Acids and Proteins in Plants I*, 438–58. Springer.

Holaskova, E., Galuszka, P., Frebort, I., and Oz, M. T. 2015. “Antimicrobial Peptide Production and Plant-Based Expression Systems for Medical and Agricultural Biotechnology.” *Biotechnology Advances* 33 (6): 1005–23. https://doi.org/10.1016/j.biotechadv.2015.03.007.

Holly, M. K., Diaz, K., and Smith, J. G. 2017. “Defensins in Viral Infection and Pathogenesis.” *Annual Review of Virology* 4 (1): 369–91. https://doi.org/10.1146/annurev-virology-101416-041734.

Hon, W. C., Griffith, M., Mlynarz, A., Kwok, Y. C., and Yang, D. S. 1995. “Antifreeze Proteins in Winter Rye Are Similar to Pathogenesis-Related Proteins.” *Plant Physiology* 109 (3): 879–89. https://doi.org/10.1104/pp.109.3.879.

Hong, W., Zeng, J., and Xie, J. 2014. “Antibiotic Drugs Targeting Bacterial RNAs.” *Acta Pharmaceutica Sinica. B* 4 (4): 258–65. https://doi.org/10.1016/j.apsb.2014.06.012.

Hoskin, D. W., and Ramamoorthy, A. 2008. “Studies on Anticancer Activities of Antimicrobial Peptides.” *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1778 (2): 357–75. https://doi.org/10.1016/j.bbamem.2007.11.008.

Hughes, C., Ma, B., and Lajoie, G. A. 2010. “De Novo Sequencing Methods in Proteomics.” In *Proteome Bioinformatics*, edited by Simon J. Hubbard and Andrew R. Jones, 604:105–21. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-60761-444-9_8.

Iseli, B., Boller, T., and Neuhaus, J. M. 1993. “The N-Terminal Cysteine-Rich Domain of Tobacco Class I Chitinase Is Essential for Chitin Binding but Not for Catalytic or Antifungal Activity.” *Plant Physiology* 103 (1): 221–26. https://doi.org/10.1104/pp.103.1.221.

Iwai, T., Kaku, H., Honkura, R., Nakamura, S., Ochiai, H., Sasaki, T., and Ohashi, Y. 2002. “Enhanced Resistance to Seed-Transmitted Bacterial Diseases in Transgenic Rice Plants Overproducing an Oat Cell-Wall-Bound Thionin.” *Molecular Plant-Microbe Interactions* 15 (6): 515–21. https://doi.org/10.1094/MPMI.2002.15.6.515.

Iyer, S., and Acharya, K. R. 2011. “Tying the Knot: The Cystine Signature and Molecular-Recognition Processes of the Vascular Endothelial Growth Factor Family of Angiogenic Cytokines: Cystine-Knot Growth Factors and Angiogenesis.” *FEBS Journal* 278 (22): 4304–22. https://doi.org/10.1111/j.1742-4658.2011.08350.x.

Jami, S. K., Anuradha, T. S., Guruprasad, L., and Kirti, P. B. 2007. “Molecular , Biochemical and Structural Characterization of Osmotin-like Protein from Black Nightshade (*Solanum nigrum*)” 164: 238–252. https://doi.org/10.1016/j.jplph.2006.01.006.

Jennings, C., West, J., Waine, C., Craik, D., and Anderson, M. 2001. “Biosynthesis and Insecticidal Properties of Plant Cyclotides: The Cyclic Knotted Proteins from *Oldenlandia affinis*.” *Proceedings of the National Academy of Sciences* 98 (19): 10614–19. https://doi.org/10.1073/pnas.191366898.

Jennings, C. V., Rosengren, K. J., Daly, N. L., Plan, M., Stevens, J., Scanlon, M. J., ... and Craik, D. J. 2005. “Isolation, Solution Structure, and Insecticidal Activity of Kalata B2, a Circular Protein with a Twist: Do Möbius Strips Exist in Nature?” *Biochemistry* 44 (3): 851–60. https://doi.org/10.1021/bi047837h.

Jha, S., Tank, H. G., Prasad, B. D., and Chattoo, B. B. 2009. "Expression of Dm-AMP1 in rice confers resistance to *Magnaporthe oryzae* and *Rhizoctonia solani*". *Transgenic research* 18 (1): 59-69.

Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. 1997. “Development and Validation of a Genetic Algorithm for Flexible Docking.” *Journal of Molecular Biology* 267 (3): 727–48. https://doi.org/10.1006/jmbi.1996.0897.

Jorge, P., Lourenço, A., and Pereira, M. O. 2012. “New Trends in Peptide-Based Anti-Biofilm Strategies: A Review of Recent Achievements and Bioinformatic Approaches.” *Biofouling* 28 (10): 1033–61. https://doi.org/10.1080/08927014.2012.728210.

Jothi, A. 2012. “Principles, Challenges and Advances in *Ab initio* Protein Structure Prediction.” *Protein and Peptide Letters* 19 (11): 1194–1204.

Jung, Y. C., Lee, H. J., Yum, S. S., Soh, W. Y., Cho, D. Y., Auh, C. K., ... and Lee, S. C. 2005. “Drought-Inducible - but ABA-Independent - Thaumatin-like Protein from Carrot (*Daucus carota* L.).” *Plant Cell Reports* 24 (6): 366–373. https://doi.org/10.1007/s00299-005-0944-x.

Kaas, Q., Westermann, J. C., and Craik, D. J. 2010. “Conopeptide Characterization and Classifications: An Analysis Using ConoServer.” *Toxicon* 55 (8): 1491–1509. https://doi.org/10.1016/j.toxicon.2010.03.002.

Kader, J. C. 1996. “Lipid-Transfer Proteins in Plants.” *Annual Review of Plant Physiology and Plant Molecular Biology* 47 (June): 627–54. https://doi.org/10.1146/annurev.arplant.47.1.627.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. 2012. “Template-Based Protein Structure Modeling Using the RaptorX Web Server.” *Nature Protocols* 7 (8): 1511–22. https://doi.org/10.1038/nprot.2012.085.

Kanzaki, H., Nirasawa, S., Saitoh, H., Ito, M., Nishihara, M., Terauchi, R., and Nakamura, I. 2002. “Overexpression of the Wasabi Defensin Gene Confers Enhanced Resistance to Blast Fungus (*Magnaporthe grisea*) in Transgenic Rice.” *Theoretical and Applied Genetics* 105 (6–7): 809–14.

Karasuyama, M., and Takeuchi, I. 2010. “Multiple Incremental Decremental Learning of Support Vector Machines.” *IEEE Transactions on Neural Networks* 21 (7): 1048–59. https://doi.org/10.1109/TNN.2010.2048039.

Karplus, M., and McCammon, J. A. 2002. “Molecular Dynamics Simulations of Biomolecules.” *Nature Structural Biology* 9 (9): 646–52. https://doi.org/10.1038/nsb0902-646.

Kaufmann, K. W., Lemmon, G. H., de Luca, S. L., Sheehan, J. H., and Meiler, J. 2010. “Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You.” *Biochemistry* 49 (14): 2987–98. https://doi.org/10.1021/bi902153g.

Kaur, J., Thokala, M., Robert-Seilaniantz, A., Zhao, P., Peyret, H., Berg, H., ... and Shah, D. 2012. "Subcellular targeting of an evolutionarily conserved plant defensin M t D ef4. 2 determines the outcome of plant–pathogen interaction in transgenic Arabidopsis". *Molecular plant pathology* 13 (9): 1032-1046.

Kazan, K., Rusu, A., Marcus, J. P., Goulter, K. C., and Manners, J. M. 2002. “Enhanced Quantitative Resistance to *Leptosphaeria maculans* Conferred by Expression of a Novel Antimicrobial Peptide in Canola (*Brassica napus* L.).” *Molecular Breeding* 10 (1–2): 63–70.

Khan, R. S., Nishihara, M., Yamamura, S., Nakamura, I., and Mii, M. 2006. "Transgenic potatoes expressing wasabi defensin peptide confer partial resistance to gray mold (*Botrytis cinerea*)". *Plant biotechnology* 23 (2): 179-183.

Khan, R. S., Nakamura, I., and Mii, M. 2011. "Development of disease-resistant marker-free tomato by R/RS site-specific recombination". *Plant cell reports* 30 (6): 1041-1053.

Kitajima, S., and Sato, F. 1999. “Plant Pathogenesis-Related Proteins : Molecular Mechanisms of Gene Expression and Protein Function 1” 8: 1–8.

Koiwa, H., Kato, H., Nakatsu, T., Oda, J. I., Yamada, Y., and Sato, F. 1999. “Crystal Structure of Tobacco PR-5d Protein at 1.8 A Resolution Reveals a Conserved Acidic Cleft Structure in Antifungal Thaumatin-like Proteins.” *Journal of Molecular Biology* 286 (4): 1137–1145. https://doi.org/10.1006/jmbi.1998.2540.

Kolodziejek, I., and Van der Hoorn, R. A. 2010. “Mining the Active Proteome in Plant Science and Biotechnology.” *Current Opinion in Biotechnology* 21 (2): 225–33. https://doi.org/10.1016/j.copbio.2010.02.003.

Koo, J. C., Lee, S. Y., Chun, H. J., Cheong, Y. H., Choi, J. S., Kawabata, S. I., ... and Han, C. D. 1998. “Two Hevein Homologs Isolated from the Seed of *Pharbitis nil* L. Exhibit Potent Antifungal Activity.” *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1382 (1): 80–90. https://doi.org/10.1016/S0167-4838(97)00148-9.

Koo, J. C., Chun, H. J., Park, H. C., Kim, M. C., Koo, Y. D., Koo, S. C., ... and Lim, C. O. 2002. “Over-Expression of a Seed Specific Hevein-like Antimicrobial Peptide from *Pharbitis nil* Enhances Resistance to a Fungal Pathogen in Transgenic Tobacco Plants.” *Plant Molecular Biology* 50 (3): 441–52.

Kotilainen, M., Helariutta, Y., Mehto, M., Pöllänen, E., Albert, V. A., Elomaa, P., and Teeri, T. H. 1999. “GEG Participates in the Regulation of Cell and Organ Shape during Corolla and Carpel Development in *Gerbera hybrida*.” *The Plant Cell* 11 (6): 1093–1104.

Kovalskaya, N., Zhao, Y., and W Hammond, R. 2011. “Antibacterial and Antifungal Activity of a Snakin-Defensin Hybrid Protein Expressed in Tobacco and Potato Plants.” *The Open Plant Science Journal* 5 (1): 29–42. https://doi.org/10.2174/1874294701105010029.

Kraszewska, J., Beckett, M. C., James, T. C., and Bond, U. 2016. “Comparative Analysis of the Antimicrobial Activities of Plant Defensin-Like and Ultrashort Peptides against Food-Spoiling Bacteria.” Edited by J. Björkroth. *Applied and Environmental Microbiology* 82 (14): 4288–98. https://doi.org/10.1128/AEM.00558-16.

Kumar, K. R. R., and Kirti, P. B. 2011. “Differential Gene Expression in *Arachis diogoi* upon Interaction with Peanut Late Leaf Spot Pathogen, *Phaeoisariopsis personata* and Characterization of a Pathogen Induced Cyclophilin.” *Plant Molecular Biology* 75 (4–5): 497–513. https://doi.org/10.1007/s11103-011-9747-3.

Lacerda, A., Vasconcelos, É. A. R., Pelegrini, P. B., and Grossi-de-Sa, M. F. 2014. “Antifungal Defensins and Their Role in Plant Defense.” *Frontiers in Microbiology* 5. 116. https://doi.org/10.3389/fmicb.2014.00116.

Laverty, G., Gorman, S. P., and Gilmore, B. F. 2011. “The Potential of Antimicrobial Peptides as Biocides.” *International Journal of Molecular Sciences* 12 (10): 6566–96.

Lay, F. T. 2003. “Isolation and Properties of Floral Defensins from Ornamental Tobacco and Petunia.” *Plant Physiology* 131 (3): 1283–93. https://doi.org/10.1104/pp.102.016626.

Le Bihan, T., Blochet, J. É., Désormeaux, A., Marion, D., and Pézolet, M. 1996. “Determination of the Secondary Structure and Conformation of Puroindolines by Infrared and Raman Spectroscopy.” *Biochemistry* 35 (39): 12712–22. https://doi.org/10.1021/bi960869n.

Leah, R., Tommerup, H., Svendsen, I. B., and Mundy, J. 1991. “Biochemical and Molecular Characterization of Three Barley Seed Proteins with Antifungal Properties.” *The Journal of Biological Chemistry* 266 (3): 1564–73.

Lee, D. G., Shin, S. Y., Kim, D. H., Seo, M. Y., Kang, J. H., Lee, Y., ... and Hahm, K. S. 1999. “Antifungal Mechanism of a Cysteine-Rich Antimicrobial Peptide, Ib-AMP1, from *Impatiens balsamina* against *Candida albicans*.” *Biotechnology Letters* 21 (12): 1047–50.

Lee, H. T., Lee, C. C., Yang, J. R., Lai, J. Z., and Chang, K. Y. 2015. “A Large-Scale Structural Classification of Antimicrobial Peptides.” *BioMed Research International* 2015: 1–6. https://doi.org/10.1155/2015/475062.

Lee, H. H., Kim, J. S., Hoang, Q. T., Kim, J. I., and Kim, Y. S. 2018. “Root-Specific Expression of Defensin in Transgenic Tobacco Results in Enhanced Resistance against *Phytophthora parasitica* Var. Nicotianae.” *European Journal of Plant Pathology* 151 (3): 811–23. https://doi.org/10.1007/s10658-018-1419-6.

Lensink, M. F., and Wodak, S. J. 2013. “Docking, Scoring, and Affinity Prediction in CAPRI: Docking, Scoring, and Affinity Prediction in CAPRI.” *Proteins: Structure, Function, and Bioinformatics* 81 (12): 2082–95. https://doi.org/10.1002/prot.24428.

Lensink, M. F., Velankar, S., and Wodak, S. J. 2017. “Modeling Protein-Protein and Protein-Peptide Complexes: CAPRI 6th Edition: Modeling Protein-Protein and Protein-Peptide Complexes.” *Proteins: Structure, Function, and Bioinformatics* 85 (3): 359–77. https://doi.org/10.1002/prot.25215.

Leone, P., Menu-Bouaouiche, L., Peumans, W. J., Payan, F., Barre, A., Roussel, A., ... and Rougé, P. 2006. “Resolution of the Structure of the Allergenic and Antifungal Banana Fruit Thaumatin-like Protein at 1.7 A.” *Biochimie* 88 (1): 45–52. https://doi.org/10.1016/j.biochi.2005.07.001.

Li, G., and Asiegbu, F. O. 2004. “Use of Scots Pine Seedling Roots as an Experimental Model to Investigate Gene Expression during Interaction with the Conifer Pathogen *Heterobasidion annosum* (P-Type).” *Journal of Plant Research* 117 (2): 155–62. https://doi.org/10.1007/s10265-003-0140-4.

Li, Z., Zhou, M., Zhang, Z., Ren, L., Du, L., Zhang, B., ... and Xin, Z. 2011. "Expression of a radish defensin in transgenic wheat confers increased resistance to *Fusarium graminearum* and *Rhizoctonia cerealis*". *Functional and integrative genomics* 11(1), 63-70.

Lima, C. B., Benko-Iseppon A. M., Bezerra-Neto, J. P., Amorim, L. L., Neto, J. R., Crovella, S., and Pandolfi, V. 2016. “Plants Defense-Related Cyclic Peptides: Diversity, Structure and Applications.” *Current Protein and Peptide Science*, July.

Lin, Y., and Jeon, Y. 2006. “Random Forests and Adaptive Nearest Neighbors.” *Journal of the American Statistical Association* 101 (474): 578–90. https://doi.org/10.1198/016214505000001230.

Lindholm, P., Göransson, U., Johansson, S., Claeson, P., Gullbo, J., Larsson, R., ... and Backlund, A. 2002. “Cyclotides: A Novel Type of Cytotoxic Agents.” *Molecular Cancer Therapeutics* 1 (6): 365.

Lindorff-Larsen, K., and Winther, J. R. 2001. "Surprisingly high stability of barley lipid transfer protein, LTP1, towards denaturant, heat and proteases". *FEBS letters* 488 (3): 145-148.

Lipkin, A., Anisimova, V., Nikonorova, A., Babakov, A., Krause, E., Bienert, M., ... and Egorov, T. 2005. “An Antimicrobial Peptide Ar-AMP from Amaranth (*Amaranthus retroflexus* L.) Seeds.” *Phytochemistry* 66 (20): 2426–31. https://doi.org/10.1016/j.phytochem.2005.07.015.

Liu, J. J., Sturrock, R., and Ekramoddoullah, A. K. 2010. “The Superfamily of Thaumatin-like Proteins: Its Origin, Evolution, and Expression towards Biological Function.” *Plant Cell Reports* 29 (5): 419–36. https://doi.org/10.1007/s00299-010-0826-8.

Liu, J., Mehdi, S., Topping, J., Friml, J., and Lindsey, K. 2013. “Interaction of PLS and PIN and Hormonal Crosstalk in *Arabidopsis* Root Development.” *Frontiers in Plant Science* 4. https://doi.org/10.3389/fpls.2013.00075.

Liu, S., Fan, L., Sun, J., Lao, X., and Zheng, H. 2017. “Computational Resources and Tools for Antimicrobial Peptides: Computational Resources and Tools for Antimicrobial Peptides.” *Journal of Peptide Science* 23 (1): 4–12. https://doi.org/10.1002/psc.2947.

London, N., Raveh, B., and Schueler-Furman, O. 2013. “Peptide Docking and Structure-Based Characterization of Peptide Binding: From Knowledge to Know-How.” *Current Opinion in Structural Biology* 23 (6): 894–902. https://doi.org/10.1016/j.sbi.2013.07.006.

Lucau-Danila, A., Laborde, L., Legrand, S., Huot, L., Hot, D., Lemoine, Y., ... and Blervacq, A. S. 2010. “Identification of Novel Genes Potentially Involved in Somatic Embryogenesis in Chicory (*Cichorium intybus* L.).” *BMC Plant Biology* 10 (January): 122–37. https://doi.org/10.1186/1471-2229-10-122.

Mahdavi, F., Sariah, M., and Maziah, M. 2012. "Expression of rice thaumatin-like protein gene in transgenic banana plants enhances resistance to Fusarium wilt". *Applied biochemistry and biotechnology* 166 (4): 1008-1019.

Maldonado, A. M., Doerner, P., Dixon, R. A., Lamb, C. J., and Cameron, R. K. 2002. “A Putative Lipid Transfer Protein Involved in Systemic Resistance Signalling in *Arabidopsis*.” *Nature* 419 (6905): 399–403. https://doi.org/10.1038/nature00962.

Manners, J. M. 2009. “Primitive Defence: The MiAMP1 Antimicrobial Peptide Family.” *Plant Molecular Biology Reporter* 27 (3): 237–42. https://doi.org/10.1007/s11105-008-0083-y.

Mao, Z., Zheng, J., Wang, Y., Chen, G., Yang, Y., Feng, D., and Xie, B. 2011. “The New CaSn Gene Belonging to the Snakin Family Induces Resistance against Root-Knot Nematode Infection in Pepper.” *Phytoparasitica* 39 (2): 151–64. https://doi.org/10.1007/s12600-011-0149-5.

Marcus, J. P., Goulter, K. C., Green, J. L., Harrison, S. J., and Manners, J. M. 1997. “Purification, Characterisation and cDNA Cloning of an Antimicrobial Peptide from *Macadamia integrifolia*.” *European Journal of Biochemistry* 244 (3): 743–49.

Marion, D., Bakan, B., and Elmorjani, K. 2007. “Plant Lipid Binding Proteins: Properties and Applications.” *Biotechnology Advances* 25 (2): 195–97. https://doi.org/10.1016/j.biotechadv.2006.11.003.

Marmiroli, N., and Maestri, E. 2014. “Plant Peptides in Defense and Signaling.” *Peptides* 56 (June): 30–44. https://doi.org/10.1016/j.peptides.2014.03.013.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. 2000. “Comparative Protein Structure Modeling of Genes and Genomes.” *Annual Review of Biophysics and Biomolecular Structure* 29 (1): 291–325. https://doi.org/10.1146/annurev.biophys.29.1.291.

Mazliak, P., Douady, D., Demandre, C., and Kader, J. C. 1975. “Exchange processes between organelles involved in membrane lipid biosynthesis.” *Recent Advances in the Chemistry and Biochemistry of Plant Lipids. Ed. T. Gailliard and EJ Mercer*. Academic Press, New York.

McHugh, L., and Arthur, J. W. 2008. “Computational Methods for Protein Identification from Mass Spectrometry Data.” *PLoS Computational Biology* 4 (2). https://doi.org/10.1371/journal.pcbi.0040012.

McManus, A. M., Nielsen, K. J., Marcus, J. P., Harrison, S. J., Green, J. L., Manners, J. M., and Craik, D. J. 1999. “MiAMP1, a Novel Protein from *Macadamia integrifolia* Adopts a Greek Key β-Barrel Fold Unique amongst Plant Antimicrobial Proteins.” *Journal of Molecular Biology* 293 (3): 629–38. https://doi.org/10.1006/jmbi.1999.3163.

Mehboob-ur-Rahman, T. S., Mahmood-ur-Rahman, M. A. I., and Zafar, Y. 2016. “Bioinformatics: A Way Forward to Explore ‘Plant Omics’.” *Bioinformatics-Updated Features and Applications*, 203.

Melo, M. N., Ferre, R., and Castanho, M. A. 2009. “Antimicrobial Peptides: Linking Partition, Activity and High Membrane-Bound Concentrations.” *Nature Reviews Microbiology* 7 (3): 245.

Mendez, E., Moreno, A., Colilla, F., Pelaez, F., Limas, G. G., Mendez, R., ... and de Haro, C. 1990. “Primary Structure and Inhibition of Protein Synthesis in Eukaryotic Cell-Free System of a Novel Thionin, Gamma-Hordothionin, from Barley Endosperm.” *European Journal of Biochemistry* 194 (2): 533–39. https://doi.org/10.1111/j.1432-1033.1990.tb15649.x.

Menschaert, G., Vandekerckhove, T. T., Baggerman, G., Schoofs, L., Luyten, W., and Criekinge, W. V. 2010. “Peptidomics coming of age: a review of contributions from a bioinformatics angle.” *Journal of Proteome Research*, 9(5), 2051-2061. https://doi.org/10.1021/pr900929m.

Min, K., Ha, S. C., Hasegawa, P. M., Bressan, R. A., Yun, D. J., and Kim, K. K. 2004. “Crystal Structure of Osmotin, a Plant Antifungal Protein” 173 (June 2003): 170–173.

Misra, R. C., Kamthan, M., Kumar, S., and Ghosh, S. 2016. "A thaumatin-like protein of *Ocimum basilicum* confers tolerance to fungal pathogen and abiotic stress in transgenic *Arabidopsis*". *Scientific Reports* 6, 25340.

Mitsutake, A., and Takano, H. 2018. “Relaxation Mode Analysis for Molecular Dynamics Simulations of Proteins.” *Biophysical Reviews* 10 (2): 375–89. https://doi.org/10.1007/s12551-018-0406-7.

Mitulović, G., and Mechtler, K. 2006. “HPLC Techniques for Proteomics Analysis--a Short Overview of Latest Developments.” *Briefings in Functional Genomics and Proteomics* 5 (4): 249–60. https://doi.org/10.1093/bfgp/ell034.

Molesini, B., Treggiari, D., Dalbeni, A., Minuz, P., and Pandolfini, T. 2017. “Plant Cystine-Knot Peptides: Pharmacological Perspectives: Plant Cystine-Knot Proteins in Pharmacology.” *British Journal of Clinical Pharmacology* 83 (1): 63–70. https://doi.org/10.1111/bcp.12932.

Morris, C. F. 2002. “Puroindolines: The Molecular Genetic Basis of Wheat Grain Hardness.” *Plant Molecular Biology* 48 (5–6): 633–47.

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A. 2016. “Critical Assessment of Methods of Protein Structure Prediction: Progress and New Directions in Round XI: Progress in CASP XI.” *Proteins: Structure, Function, and Bioinformatics* 84 (September): 4–14. https://doi.org/10.1002/prot.25064.

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A. 2018. “Critical Assessment of Methods of Protein Structure Prediction (CASP)-Round XII.” *Proteins: Structure, Function, and Bioinformatics* 86 (March): 7–15. https://doi.org/10.1002/prot.25415.

Moyes, D. L., Wilson, D., Richardson, J. P., Mogavero, S., Tang, S. X., Wernecke, J., ... and Murciano, C. 2016. “Candidalysin Is a Fungal Peptide Toxin Critical for Mucosal Infection.” *Nature* 532 (7597): 64–68. https://doi.org/10.1038/nature17625.

Mulvenna, J. P., Wang, C., and Craik, D. J. 2006. “CyBase: A Database of Cyclic Protein Sequence and Structure.” *Nucleic Acids Research* 34 (90001): D192–94. https://doi.org/10.1093/nar/gkj005.

Munita, J. M., and Arias, C. A. 2016. “Mechanisms of Antibiotic Resistance.” *Microbiology Spectrum* 4 (2). https://doi.org/10.1128/microbiolspec.VMBF-0016-2015.

Muramoto, N., Tanaka, T., Shimamura, T., Mitsukawa, N., Hori, E., Koda, K., ... and Imaeda, T. 2012. “Transgenic Sweet Potato Expressing Thionin from Barley Gives Resistance to Black Rot Disease Caused by *Ceratocystis fimbriata* in Leaves and Storage Roots.” *Plant Cell Reports* 31 (6): 987–97. https://doi.org/10.1007/s00299-011-1217-5.

Murphy, E., Smith, S., and De Smet, I. 2012. “Small Signaling Peptides in *Arabidopsis* Development: How Cells Communicate Over a Short Distance.” *The Plant Cell* 24 (8): 3198–3217. https://doi.org/10.1105/tpc.112.099010.

Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. 1995. “SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures.” *Journal of Molecular Biology* 247 (4): 536–40. https://doi.org/10.1006/jmbi.1995.0159.

Nadal, A., Montero, M., Badosa, E., Messeguer, J., Montesinos, L., Montesinos, E., and Pla, M. 2012. “Constitutive Expression of Transgenes Encoding Derivatives of the Synthetic Antimicrobial Peptide BP100: Impact on Rice Host Plant Fitness.” *BMC Plant Biology* 12 (1): 159.

Nägele, E., Vollmer, M., Hörth, P., and Vad, C. 2004. “2D-LC/MS Techniques for the Identification of Proteins in Highly Complex Mixtures.” *Expert Review of Proteomics* 1 (1): 37–46. https://doi.org/10.1586/14789450.1.1.37.

Nahirñak, V., Almasia, N. I., Fernandez, P. V., Hopp, H. E., Estevez, J., Carrari, F., and Vazquez-Rovere, C. 2012. “Potato Snakin-1 Gene Silencing Affects Cell Division, Primary Metabolism, and Cell Wall Composition.” *Plant Physiology* 158 (1): 252–63. https://doi.org/10.1104/pp.111.186544.

Nawrot, R., Barylski, J., Nowicki, G., Broniarczyk, J., Buchwald, W., and Goździcka-Józefiak, A. 2014. “Plant Antimicrobial Peptides.” *Folia Microbiologica* 59 (3): 181–96. https://doi.org/10.1007/s12223-013-0280-4.

Nesvizhskii, A. I., Vitek, O., and Aebersold, R. 2007. “Analysis and Validation of Proteomic Data Generated by Tandem Mass Spectrometry.” *Nat. Methods* 4: 787–797.

Neu, H. C., and Gootz, T. D. 1996. “Antimicrobial Chemotherapy.” In *Medical Microbiology*, edited by Samuel Baron, 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston. http://www.ncbi.nlm.nih.gov/books/NBK7986/.

Nguyen, G. K. T., Zhang, S., Nguyen, N. T. K., Nguyen, P. Q. T., Chiu, M. S., Hardjojo, A., and Tam, J. P. 2011. “Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Albumin-1 Chain a and Cyclotide Domains in the Fabaceae Family.” *Journal of Biological Chemistry* 286 (27): 24275–87. https://doi.org/10.1074/jbc.M111.229922.

Niarchou, A., Alexandridou, A., Athanasiadis, E., and Spyrou, G. 2013. “C-PAmP: Large Scale Analysis and Database Construction Containing High Scoring Computationally Predicted Antimicrobial Peptides for All the Available Plant Species.” Edited by Jamunarani Vadivelu. *PLoS ONE* 8 (11): e79728. https://doi.org/10.1371/journal.pone.0079728.

Nikaido, H. 1994. “Prevention of Drug Access to Bacterial Targets: Permeability Barriers and Active Efflux.” *Science* 264 (5157): 382–88. https://doi.org/10.1126/science.8153625.

Northfield, S. E., Wang, C. K., Schroeder, C. I., Durek, T., Kan, M. W., Swedberg, J. E., and Craik, D. J. 2014. “Disulfide-Rich Macrocyclic Peptides as Templates in Drug Design.” *European Journal of Medicinal Chemistry* 77 (April): 248–57. https://doi.org/10.1016/j.ejmech.2014.03.011.

Ntui, V. O., Thirukkumaran, G., Azadi, P., Khan, R. S., Nakamura, I., and Mii, M. 2010. "Stable integration and expression of wasabi defensin gene in “Egusi” melon (*Colocynthis citrullus* L.) confers resistance to Fusarium wilt and Alternaria leaf spot". *Plant cell reports* 29 (9): 943-954.

Oddepally, R., and Guruprasad, L. 2015. “Isolation, Purification, and Characterization of a Stable Defensin-like Antifungal Peptide from *Trigonella foenum-graecum* (Fenugreek) Seeds.” *Biochemistry (Moscow)* 80 (3): 332–42. https://doi.org/10.1134/S0006297915030086.

Ogata, C. M., Gordon, P. F., de Vos, A. M., and Kim, S. H. 1992. “Crystal Structure of a Swe Tasting Protein Thaumatin I, at 165 A Resolution,” 893–908.

Ohno, A., Tate, S. I., Seeram, S. S., Hiraga, K., Swindells, M. B., Oda, K., and Kainosho, M. 1998. “NMR Structure of the Streptomyces Metalloproteinase Inhibitor, SMPI, Isolated from *Streptomyces nigrescens* TK-23: Another Example of an Ancestral Bγ-Crystallin Precursor Structure.” *Journal of Molecular Biology* 282 (2): 421–33. https://doi.org/10.1006/jmbi.1998.2022.

Ojola, P. O., Nyaboga, E. N., Njiru, P. N., and Orinda, G. 2018. “Overexpression of Rice Thaumatin-like Protein (Ostlp) Gene in Transgenic Cassava Results in Enhanced Tolerance to *Colletotrichum gloeosporioides* F. Sp. *Manihotis*.” *Journal of Genetic Engineering and Biotechnology* 16 (1): 125–31. https://doi.org/10.1016/j.jgeb.2017.12.002.

Oliveira-Lima, M., Maria Benko-Iseppon, A., Ribamar Costa Ferreira Neto, J., Rodriguez-Decuadro, S., Kido, E. A., Crovella, S., and Pandolfi, V. 2017. “Snakin: Structure, Roles and Applications of a Plant Antimicrobial Peptide.” *Current Protein and Peptide Science* 18 (4): 368–74. https://doi.org/10.2174/1389203717666160619183140.

Osherov, N., Mathew, J., Romans, A., and May, G. S. 2002. “Identification of Conidial-Enriched Transcripts in *Aspergillus nidulans* Using Suppression Subtractive Hybridization” 37: 197–204.

Pagadala, N. S., Syed, K., and Tuszynski, J. 2017. “Software for Molecular Docking: A Review.” *Biophysical Reviews* 9 (2): 91–102. https://doi.org/10.1007/s12551-016-0247-1.

Parisi, K., Shafee, T. M., Quimbar, P., van der Weerden, N. L., Bleackley, M. R., and Anderson, M. A. 2018. “The Evolution, Function and Mechanisms of Action for Plant Defensins.” *Seminars in Cell and Developmental Biology*, February. https://doi.org/10.1016/j.semcdb.2018.02.004.

Park, H. C., Kang, Y. H., Chun, H. J., Koo, J. C., Cheong, Y. H., Kim, C. Y., ... and Koo, Y. D. 2002. "Characterization of a stamen-specific cDNA encoding a novel plant defensin in Chinese cabbage". *Plant molecular biology* 50 (1): 57-68.

Park, S., Yoo, K. O., Marcussen, T., Backlund, A., Jacobsson, E., Rosengren, K. J., ... and Göransson, U. 2017. "Cyclotide Evolution: Insights from the Analyses of Their Precursor Sequences, Structures and Distribution in Violets (Viola)." *Frontiers in Plant Science* 8 (December). https://doi.org/10.3389/fpls.2017.02058.

Patel, S. U., Osborn, R., Rees, S., and Thornton, J. M. 1998. "Structural Studies of *Impatiens balsamina* Antimicrobial Protein (Ib-AMP1)." *Biochemistry* 37 (4): 983–90. https://doi.org/10.1021/bi971747d.

Patkar, R. N., and Chattoo, B. B. 2006. "Transgenic Indica Rice Expressing Ns-LTP-Like Protein Shows Enhanced Resistance to Both Fungal and Bacterial Pathogens." *Molecular Breeding* 17 (2): 159–71. https://doi.org/10.1007/s11032-005-4736-3.

Pearce, G., Bhattacharya, R., and Chen, Y. C. 2008. "Peptide Signals for Plant Defense Display a More Universal Role." *Plant Signaling and Behavior* 3 (12): 1091–92.

Pearson, W. R. 1990. "[5] Rapid and Sensitive Sequence Comparison with FASTP and FASTA." In *Methods in Enzymology*, 183:63–98. Elsevier. https://doi.org/10.1016/0076-6879(90)83007-V.

Pelegrini, P. B., and Franco, O. L. 2005. "Plant γ-Thionins: Novel Insights on the Mechanism of Action of a Multi-Functional Class of Defense Proteins." *The International Journal of Biochemistry and Cell Biology* 37 (11): 2239–53. https://doi.org/10.1016/j.biocel.2005.06.011.

Pérez-Bernal, M., Delgado, M., Cruz, A., Abreu, D., Valdivia, O., and Armas, R. 2017. "Marker-Free Transgenic Rice Lines with a Defensin Gene Are Potentially Active against Phytopathogenic Fungus *Sarocladium oryzae*." *Acta Phytopathologica et Entomologica Hungarica* 52 (2): 135–44. https://doi.org/10.1556/038.52.2017.021.

Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. 1999. "Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data." *Electrophoresis* 20 (18): 3551–3567. https://doi.org/10.1002/(sici)1522-2683(19991201)20:18<3551::aid-elps3551>3.0.co;2-2.

Pestana-Calsa, M. C., Ribeiro I. L., and Calsa Jr, T. 2010. "Bioinformatics-Coupled Molecular Approaches for Unravelling Potential Antimicrobial Peptides Coding Genes in Brazilian Native and Crop Plant Species." *Current Protein and Peptide Science* 11 (3): 199–209. https://doi.org/CPPS-38 [pii].

Peters, B. M., Shirtliff, M. E., and Jabra-Rizk, M. A. 2010. "Antimicrobial Peptides: Primeval Molecules or Future Drugs?" *PLoS Pathogens* 6 (10): e1001067.

Petrey, D., and Honig, B. 2014. "Structural Bioinformatics of the Interactome." *Annual Review of Biophysics* 43 (1): 193–210. https://doi.org/10.1146/annurev-biophys-051013-022726.

Peumans, W. J., De Ley, M., and Broekaert, W. F. 1984. "An Unusual Lectin from Stinging Nettle (*Urtica dioica*) Rhizomes." *FEBS Letters* 177 (1): 99–103. https://doi.org/10.1016/0014-5793(84)80989-8.

Peyraud, R., Dubiella, U., Barbacci, A., Genin, S., Raffaele, S., and Roby, D. 2016. "Advances on Plant–Pathogen Interactions from Molecular toward Systems Biology Perspectives." *The Plant Journal* 90 (4): 720–37. https://doi.org/10.1111/tpj.13429.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... and Schulten, K. 2005. "Scalable Molecular Dynamics with NAMD." *Journal of Computational Chemistry* 26 (16): 1781–1802. https://doi.org/10.1002/jcc.20289.

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., and Weng, Z. 2014. "ZDOCK Server: Interactive Docking Prediction of Protein-Protein Complexes and Symmetric Multimers." *Bioinformatics* 30 (12): 1771–73. https://doi.org/10.1093/bioinformatics/btu097.

Plan, M. R. R., Saska, I., Cagauan, A. G., and Craik, D. J. 2008. "Backbone Cyclised Peptides from Plants Show Molluscicidal Activity against the Rice Pest *Pomacea canaliculata* (Golden Apple Snail)." *Journal of Agricultural and Food Chemistry* 56 (13): 5237–41. https://doi.org/10.1021/jf800302f.

Plimpton, S. 1995. "Fast Parallel Algorithms for Short-Range Molecular Dynamics." *Journal of Computational Physics* 117 (1): 1–19. https://doi.org/10.1006/jcph.1995.1039.

Polyanovsky, V. O., Roytberg, M. A., and Tumanyan, V. G. 2011. "Comparative Analysis of the Quality of a Global Algorithm and a Local Algorithm for Alignment of Two Sequences." *Algorithms for Molecular Biology* 6 (1): 25. https://doi.org/10.1186/1748-7188-6-25.

Portieles, R., Ayra, C., Gonzalez, E., Gallo, A., Rodriguez, R., Chacón, O., ... and Enriquez, G. 2010. "NmDef02, a Novel Antimicrobial Gene Isolated from *Nicotiana megalosiphon* Confers High-Level Pathogen Resistance under Greenhouse and Field Conditions: NmDef02, a Novel Antimicrobial Protein." *Plant Biotechnology Journal* 8 (6): 678–90. https://doi.org/10.1111/j.1467-7652.2010.00501.x.

Porto, W. F., Pires, A. S., and Franco, O. L. 2017. "Computational Tools for Exploring Sequence Databases as a Resource for Antimicrobial Peptides." *Biotechnology Advances* 35 (3): 337–49. https://doi.org/10.1016/j.biotechadv.2017.02.001.

Poth, A. G., Colgrave, M. L., Philip, R., Kerenga, B., Daly, N. L., Anderson, M. A., and Craik, D. J. 2011. "Discovery of Cyclotides in the Fabaceae Plant Family Provides New Insights into the Cyclization, Evolution, and Distribution of Circular Proteins." *ACS Chemical Biology* 6 (4): 345–55. https://doi.org/10.1021/cb100388j.

Poth, A. G., Mylne, J. S., Grassl, J., Lyons, R. E., Millar, A. H., Colgrave, M. L., and Craik, D. J. 2012. "Cyclotides Associate with Leaf Vasculature and Are the Products of a Novel Precursor in *Petunia* (Solanaceae)." *Journal of Biological Chemistry* 287 (32): 27033–46. https://doi.org/10.1074/jbc.M112.370841.

Pränting, M., Lööv, C., Burman, R., Göransson, U. L. F., and Andersson, D. I. 2010. "The Cyclotide Cycloviolacin O2 from *Viola odorata* Has Potent Bactericidal Activity against Gram-Negative Bacteria." *Journal of Antimicrobial Chemotherapy* 65 (9): 1964–71. https://doi.org/10.1093/jac/dkq220.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... and Hess, B. 2013. "GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit." *Bioinformatics* 29 (7): 845–54. https://doi.org/10.1093/bioinformatics/btt055.

Raikhel, N. V., Lee, H. I., and Broekaert, W. F. 1993. "Structure and Function of Chitin-Binding Proteins." *Annual Review of Plant Physiology and Plant Molecular Biology* 44 (1): 591–615. https://doi.org/10.1146/annurev.pp.44.060193.003111.

Rajam, M. V., Chandola, N., Goud, P. S., Singh, D., Kashyap, V., Choudhary, M. L., and Sihachakr, D. 2007. "Thaumatin gene confers resistance to fungal pathogens as well as tolerance to abiotic stresses in transgenic tobacco plants". *Biologia Plantarum* 51 (1): 135-141.

Ramakrishnan, G., Chandra, N. R., and Srinivasan, N. 2014. "From Workstations to Workbenches: Towards Predicting Physicochemically Viable Protein–Protein Interactions across a Host and a Pathogen." *IUBMB Life* 66 (11): 759–74. https://doi.org/10.1002/iub.1331.

Rees, D. C., and Lipscomb, W. N. 1982. "Refined Crystal Structure of the Potato Inhibitor Complex of Carboxypeptidase A at 2.5 Å Resolution." *Journal of Molecular Biology* 160 (3): 475–98. https://doi.org/10.1016/0022-2836(82)90309-6.

Rentzsch, R., and Renard, B. Y. 2015. "Docking small peptides remains a great challenge: an assessment using AutoDock Vina". *Briefings in bioinformatics* 16 (6): 1045-1056.

Robustelli, P., Piana, S., and Shaw, D. E. 2018. "Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States." *Proceedings of the National Academy of Sciences* 115 (21): E4758–66. https://doi.org/10.1073/pnas.1800690115.

Rodríguez-Decuadro, S., Barraco-Vega, M., Dans, P. D., Pandolfi, V., Benko-Iseppon, A. M., and Cecchetto, G. 2018. "Antimicrobial and Structural Insights of a New Snakin-like Peptide Isolated from *Peltophorum dubium* (Fabaceae)." *Amino Acids*, June. https://doi.org/10.1007/s00726-018-2598-3.

Rogozhin, E. A., Slezina, M. P., Slavokhotova, A. A., Istomina, E. A., Korostyleva, T. V., Smirnov, A. N., ... and Odintsova, T. I. 2015. "A Novel Antifungal Peptide from Leaves of the Weed *Stellaria media* L." *Biochimie* 116 (September): 125–32. https://doi.org/10.1016/j.biochi.2015.07.014.

Rong, W., Qi, L., Wang, J., Du, L., Xu, H., Wang, A., and Zhang, Z. 2013. "Expression of a Potato Antimicrobial Peptide SN1 Increases Resistance to Take-All Pathogen *Gaeumannomyces graminis* Var. Tritici in Transgenic Wheat." *Functional and Integrative Genomics* 13 (3): 403–9. https://doi.org/10.1007/s10142-013-0332-5.

Roxrud, I., Lid, S. E., Fletcher, J. C., Schmidt, E. D., and Opsahl-Sorteberg, H. G. 2007. "GASA4, One of the 14-Member *Arabidopsis* GASA Family of Small Polypeptides, Regulates Flowering and Seed Development." *Plant and Cell Physiology* 48 (3): 471–83. https://doi.org/10.1093/pcp/pcm016.

Rubio-Infante, N., Govea-Alonso, D. O., Alpuche-Solís, Á. G., García-Hernández, A. L., Soria-Guerra, R. E., Paz-Maldonado, L. T., ... and Moreno-Fierros, L. 2012. "A Chloroplast-Derived C4V3 Polypeptide from the Human Immunodeficiency Virus (HIV) Is Orally Immunogenic in Mice." *Plant Molecular Biology* 78 (4–5): 337–49.

Rudd, J., Kanyuka, K., Hassani-Pak, K., Derbyshire, M., Devonshire, J., Saqi, M., ... and Bharti, A. 2015. "Transcriptome and Metabolite Pro Fi Ling of the Infection Cycle of *Zymoseptoria tritici* on Wheat Reveals a Biphasic Interaction with Plant Immunity Involving Differential Pathogen Chromosomal Contributions and a Variation on the Hemibiotrophic Lifestyle D" 167 (March): 1158–1185. https://doi.org/10.1104/pp.114.255927.

Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., ... and Morley, S. D. 2014. "RDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids." Edited by Andreas Prlic. *PLoS Computational Biology* 10 (4): e1003571. https://doi.org/10.1371/journal.pcbi.1003571.

Ruperti, B., Cattivelli, L., Pagni, S., and Ramina, A. 2002. "Ethylene-Responsive Genes Are Differentially Regulated during Abscission, Organ Senescence and Wounding in Peach (*Prunus persica*)." *Journal of Experimental Botany* 53 (368): 429–37. https://doi.org/10.1093/jexbot/53.368.429.

Ryan, C A. 1990. "Protease Inhibitors in Plants: Genes for Improving Defenses Against Insects and Pathogens." *Annual Review of Phytopathology* 28 (1): 425–49. https://doi.org/10.1146/annurev.py.28.090190.002233.

Sakamoto, Y., Watanabe, H., Nagai, M., Nakade, K., Takahashi, M., and Sato, T. 2006. "*Lentinula Edodes* Tlg1 Encodes a Thaumatin-Like Protein That Is Involved in Lentinan Degradation and Fruiting Body Senescence" 141 (June): 793–801. https://doi.org/10.1104/pp.106.076679.ture.

Salminen, T. A., Blomqvist, K., and Edqvist, J. 2016. "Lipid Transfer Proteins: Classification, Nomenclature, Structure, and Function." *Planta* 244 (5): 971–97. https://doi.org/10.1007/s00425-016-2585-4.

Sathish, K., Sriraman, R., Subramanian, B. M., Rao, N. H., Balaji, K., Narasu, M. L., and Srinivasan, V. A. 2011. "Plant Expressed EtMIC2 Is an Effective Immunogen in Conferring Protection against Chicken Coccidiosis." *Vaccine* 29 (49): 9201–8.

Schimoler-O'Rourke, R., Richardson, M., and Selitrennikoff, C. P. 2001. "Zeamatin Inhibits Trypsin and □α-Amylase Activities" 67 (5): 2365–2366. https://doi.org/10.1128/AEM.67.5.2365.

Schneider, D. J., and Collmer, A. 2010. "Studying Plant-Pathogen Interactions in the Genomics Era: Beyond Molecular Koch's Postulates to Systems Biology." *Annual Review of Phytopathology* 48: 457–79. https://doi.org/10.1146/annurev-phyto-073009-114411.

Schneider, T., Kruse, T., Wimmer, R., Wiedemann, I., Sass, V., Pag, U., ... and Neve, S. 2010. "Plectasin, a Fungal Defensin, Targets the Bacterial Cell Wall Precursor Lipid II." *Science* 328 (5982): 1168–72. https://doi.org/10.1126/science.1185723.

Seebah, S., Suresh, A., Zhuo, S., Choong, Y. H., Chua, H., Chuon, D., ... and Verma, C. 2007. "Defensins Knowledgebase: A Manually Curated Database and Information Source Focused on the Defensins Family of Antimicrobial Peptides." *Nucleic Acids Research* 35 (Database): D265–68. https://doi.org/10.1093/nar/gkl866.

Segura, A., Moreno, M., Madueño, F., Molina, A., and García-Olmedo, F. 1999. "Snakin-1, a Peptide from Potato That Is Active Against Plant Pathogens." *Molecular Plant-Microbe Interactions* 12 (1): 16–23. https://doi.org/10.1094/MPMI.1999.12.1.16.

Seo, H. H., Park, S., Park, S., Oh, B. J., Back, K., Han, O., ... and Kim, Y. S. 2014. "Overexpression of a defensin enhances resistance to a fruit-specific anthracnose fungus in pepper". *PloS* one 9 (5): e97936.

Shafee, T. M., Lay, F. T., Phan, T. K., Anderson, M. A., and Hulett, M. D. 2017. "Convergent Evolution of Defensin Sequence, Structure and Function." *Cellular and Molecular Life Sciences* 74 (4): 663–82. https://doi.org/10.1007/s00018-016-2344-5.

Sheoran, I. S., Ross, A. R., Olson, D. J., and Sawhney, V. K. 2009. "Compatibility of Plant Protein Extraction Methods with Mass Spectrometry for Proteome Analysis." *Plant Science* 176 (1): 99–104. https://doi.org/10.1016/j.plantsci.2008.09.015.

Shukurov, R. R., Voblikova, V. D., Nikonorova, A. K., Komakhin, R. A., Komakhina, V. V., Egorov, T. A., ... and Babakov, A. V. 2012. "Transformation of Tobacco and *Arabidopsis* Plants with *Stellaria media* Genes Encoding Novel Hevein-like Peptides Increases Their Resistance to Fungal Pathogens." *Transgenic Research* 21 (2): 313–25. https://doi.org/10.1007/s11248-011-9534-6.

Silva, O. N. O., Mulder, K. C., Barbosa, A. A., Otero-Gonzalez, A. J., Lopez-Abarrategui, C., Dias, S. C., ... and Franco, O. L. 2011. "Exploring the Pharmacological Potential of Promiscuous Host-Defense Peptides: From Natural Screenings to Biotechnological Applications." *Frontiers in Microbiology* 2: 232.

Simorre, J. P., Caille, A., Marion, D., Marion, D., and Ptak, M. 1991. "Two- and Three-Dimensional 1H NMR Studies of a Wheat Phospholipid Transfer Protein: Sequential Resonance Assignments and Secondary Structure." *Biochemistry* 30 (49): 11600–608.

Singh, V., Singh, K., Amdekar, S., Singh, D. D., Tripathi, P., Sharma, G. L., and Yadav, H. 2009. "Innate and Specific Gut-Associated Immunity and Microbial Interference." *FEMS Immunology and Medical Microbiology* 55 (1): 6–12. https://doi.org/10.1111/j.1574-695X.2008.00497.x.

Sjahril, R., Chin, D. P., Khan, R. S., Yamamura, S., Nakamura, I., Amemiya, Y., and Mii, M. 2006. "Transgenic Phalaenopsis plants with resistance to *Erwinia carotovora* produced by introducing wasabi defensin gene using Agrobacterium method". *Plant Biotechnology* 23 (2): 191-194.

Slavokhotova, A. A., Shelenkov, A. A., Andreev, Y. A., and Odintsova, T. I. 2017. "Hevein-like Antimicrobial Peptides of Plants." *Biochemistry (Moscow)* 82 (13): 1659–74. https://doi.org/10.1134/S0006297917130065.

Sousa, S. F., Fernandes, P. A., and Ramos, M. J. 2006. "Protein-Ligand Docking: Current Status and Future Challenges." *Proteins: Structure, Function, and Bioinformatics* 65 (1): 15–26. https://doi.org/10.1002/prot.21082.

Sousa, S. F., Ribeiro, A. J., Coimbra, J. T. S., Neves, R. P. P., Martins, S. A., Moorthy, N. S. H. N., ... and Ramos, M. J. 2013. "Protein-Ligand Docking in the New Millennium--a Retrospective of 10 Years in the Field." *Current Medicinal Chemistry* 20 (18): 2296–2314.

Spitzer, R., and Jain, A. N. 2012. "Surflex-Dock: Docking Benchmarks and Real-World Application." *Journal of Computer-Aided Molecular Design* 26 (6): 687–99. https://doi.org/10.1007/s10822-011-9533-y.

Stierle, A., Strobel, G., and Stierle, D. 1993. "Taxol and Taxane Production by *Taxomyces andreanae*, an Endophytic Fungus of Pacific Yew." *Science* 260 (5105): 214–16. https://doi.org/10.1126/science.8097061.

Stratmann, J. W. 2003. "Long Distance Run in the Wound Response – Jasmonic Acid Is Pulling Ahead." *Trends in Plant Science* 8 (6): 247–50. https://doi.org/10.1016/S1360-1385(03)00106-7.

Straus, S. K., and Hancock, R. E. 2006. "Mode of Action of the New Antibiotic for Gram-Positive Pathogens Daptomycin: Comparison with Cationic Antimicrobial Peptides and Lipopeptides." *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1758 (9): 1215–23.

Sun, J., Deng, Z., and Yan, A. 2014. "Bacterial Multidrug Efflux Pumps: Mechanisms, Physiology and Pharmacological Exploitations." *Biochemical and Biophysical Research Communications*, Integrative Glycobiology and Future Perspectives, 453 (2): 254–67. https://doi.org/10.1016/j.bbrc.2014.05.090.

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., ... and Ouellette, M. 2018. "Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis." *The Lancet Infectious Diseases* 18(3), 318-327.

Tachi, H., Fukuda-Yamada, K., Kojima, T., Shiraiwa, M., and Takahara, H. 2009. "Molecular Characterization of a Novel Soybean Gene Encoding a Neutral PR-5 Protein Induced by High-Salt Stress." *Plant Physiology and Biochemistry* 47 (1): 73–79. https://doi.org/10.1016/j.plaphy.2008.09.012.

Tailor, R. H., Acland, D. P., Attenborough, S., Cammue, B. P., Evans, I. J., Osborn, R. W., ... and Broekaert, W. F. 1997. "A Novel Family of Small Cysteine-Rich Antimicrobial Peptides from Seed of *Impatiens balsamina* Is Derived from a Single Precursor Protein." *The Journal of Biological Chemistry* 272 (39): 24480–87.

Takahashi, F., Suzuki, T., Osakabe, Y., Betsuyaku, S., Kondo, Y., Dohmae, N., ... and Shinozaki, K. 2018. "A Small Peptide Modulates Stomatal Control via Abscisic Acid in Long-Distance Signalling." *Nature* 556 (7700): 235–38. https://doi.org/10.1038/s41586-018-0009-2.

Tam, J. P., Wang, S., Wong, K. H., and Tan, W. L. 2015. "Antimicrobial Peptides from Plants." *Pharmaceuticals* 8 (4): 711–57. https://doi.org/10.3390/ph8040711.

Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., and Cammue, B. P. 2015. "The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions." *The Plant Cell* 27 (8): 2095–2118. https://doi.org/10.1105/tpc.15.00440.

Tennessen, J. A. 2008. "Positive Selection Drives a Correlation between Non-Synonymous/ Synonymous Divergence and Functional Divergence." *Bioinformatics* 24 (12): 1421–1425. https://doi.org/10.1093/bioinformatics/btn205.

Terras, F. R., Schoofs, H. M., De Bolle, M. F., Van Leuven, F., Rees, S. B., Vanderleyden, J., ... and Broekaert, W. F. 1992. "Analysis of Two Novel Classes of Plant Antifungal Proteins from Radish (*Raphanus sativus* L.) Seeds." *The Journal of Biological Chemistry* 267 (22): 15301–9.

Terras, F. R., Eggermont, K., Kovaleva, V., Raikhel, N. V., Osborn, R. W., Kester, A., ... and Vanderleyden, J. 1995. "Small cysteine-rich antifungal proteins from radish: their role in host defense". *The Plant Cell* 7 (5): 573-588.

Thao, H. T., Lan, N. N., Tuong, H. M., Thanh, N. V. T., Van Son, L., and Hoang, C. 2017. "Expression Analysis of Recombinant *Vigna radiata* Plant Defensin 1 Protein in Transgenic Tobacco Plants." *Journal of Applied Biology and Biotechnology*, 070–075. https://doi.org/10.7324/JABB.2017.50112.

Thell, K., Hellinger, R., Sahin, E., Michenthaler, P., Gold-Binder, M., Haider, T., ... and Schabbauer, G. 2016. "Oral Activity of a Nature-Derived Cyclic Peptide for the Treatment of Multiple Sclerosis." *Proceedings of the National Academy of Sciences* 113 (15): 3960–65. https://doi.org/10.1073/pnas.1519960113.

Thevissen, K., François, I. E., Sijtsma, L., Van Amerongen, A., Schaaper, W. M., Meloen, R., ... and Cammue, B. P. 2005. "Antifungal Activity of Synthetic Peptides Derived from *Impatiens balsamina* Antimicrobial Peptides Ib-AMP1 and Ib-AMP4." *Peptides* 26 (7): 1113–19. https://doi.org/10.1016/j.peptides.2005.01.008.

Thevissen, K., Kristensen, H. H., Thomma, B. P., Cammue, B. P., and Francois, I. E. 2007. "Therapeutic Potential of Antifungal Plant and Insect Defensins." *Drug Discovery Today* 12 (21–22): 966–71. https://doi.org/10.1016/j.drudis.2007.07.016.

Thomas, S., Karnik, S., Barai, R. S., Jayaraman, V. K., and Idicula-Thomas, S. 2010. "CAMP: A Useful Resource for Research on Antimicrobial Peptides." *Nucleic Acids Research* 38 (suppl_1): D774–80. https://doi.org/10.1093/nar/gkp1021.

Thompson, K. 1968. "Programming Techniques: Regular Expression Search Algorithm." *Communications of the ACM* 11 (6): 419–22. https://doi.org/10.1145/363347.363387.

Thompson, J. D., Schaeffer-Reiss, C., and Ueffing, M. 2008. *Functional Proteomics : Methods and Protocols*.

Torrent, M., Victoria Nogues, M., and Boix, E. 2012. "Discovering New *In Silico* Tools for Antimicrobial Peptide Prediction." *Current Drug Targets* 13 (9): 1148–57. https://doi.org/10.2174/138945012802002311.

Trudel, J., Grenier, J., Potvin, C., and Asselin, A. 1998. "Several Thaumatin-like Proteins Bind to Beta-1,3-Glucans." *Plant Physiology* 118 (4): 1431–38.

Tsai, W. C., Zhuang, Z. J., Lin, C. Y., and Chen, W. J. 2016. "Novel Antimicrobial Peptides with Promising Activity against Multidrug Resistant *Salmonella enterica* greater *Serovar choleraesuis* and Its Stress Response Mechanism." *Journal of Applied Microbiology* 121 (4): 952–965. https://doi.org/10.1111/jam.13203.

Tsuboi, S., Osafune, T., Tsugeki, R., Nishimura, M., and Yamada, M. 1992. "Nonspecific Lipid Transfer Protein in Castor Bean Cotyledon Cells: Subcellular Localization and a Possible Role in Lipid Metabolism." *Journal of Biochemistry* 111 (4): 500–508.

Umadevi, P., Soumya, M., George, J. K., and Anandaraj, M. 2018. "Proteomics Assisted Profiling of Antimicrobial Peptide Signatures from Black Pepper (*Piper nigrum* L.)." *Physiology and Molecular Biology of Plants* 24 (3): 379–87. https://doi.org/10.1007/s12298-018-0524-5.

Unal, E. B., Gursoy, A., and Erman, B. 2010. "VitAL: Viterbi Algorithm for de Novo Peptide Design." Edited by Joel S. Bader. *PLoS ONE* 5 (6): e10926. https://doi.org/10.1371/journal.pone.0010926.

van Parijs, J., Broekaert, W. F., Goldstein, I. J., and Peumans, W. J. 1991. "Hevein: An Antifungal Protein from Rubber-Tree (*Hevea brasiliensis*) Latex." *Planta* 183 (2): 258–64. https://doi.org/10.1007/BF00197797.

van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., ... and Bonvin, A. M. J. J. 2016. “The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes.” *Journal of Molecular Biology* 428 (4): 720–25. https://doi.org/10.1016/j.jmb.2015.09.014.

Vasavirama, K., and Kirti, P. B. 2012. "Increased resistance to late leaf spot disease in transgenic peanut using a combination of PR genes". *Functional and integrative genomics* 12 (4): 625-634.

Velazhahan, R., and Muthukrishnan, S. 2003. “Transgenic Tobacco Plants Constitutively Overexpressing a Rice Thaumatin-like Protein (PR-5) Show Enhanced Resistance to *Alternaria alternata*.” *Biologia Plantarum* 47 (3): 347–54. https://doi.org/10.1023/B:BIOP.0000023876.55053.5e.

Vriens, K., Cammue, B., and Thevissen, K. 2014. “Antifungal Plant Defensins: Mechanisms of Action and Production.” *Molecules* 19 (8): 12280–303. https://doi.org/10.3390/molecules190812280.

Waghu, F. H., Gopi, L., Barai, R. S., Ramteke, P., Nizami, B., and Idicula-Thomas, S. 2014. “CAMP: Collection of Sequences and Structures of Antimicrobial Peptides.” *Nucleic Acids Research* 42 (D1): D1154–58. https://doi.org/10.1093/nar/gkt1157.

Waghu, F. H., Barai, R. S., Gurung, P., and 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, X., Zafian, P., Choudhary, M., and Lawton, M. 1996. “The PR5K Receptor Protein Kinase from *Arabidopsis thaliana* Is Structurally Related to a Family of Plant Defense Proteins” 93 (March): 2598–2602.

Wang, Y., Nowak, G., Culley, D., Hadwiger, L. A., and Fristensky, B. 1999. "Constitutive expression of pea defense gene DRR206 confers resistance to blackleg (*Leptosphaeria maculans*) disease in transgenic canola (*Brassica napus*)". *Molecular plant-microbe interactions* 12 (5): 410-418.

Wang, Z. 2004. “APD: The Antimicrobial Peptide Database.” *Nucleic Acids Research* 32 (90001): 590D – 592. https://doi.org/10.1093/nar/gkh025.

Wang, P., Bang, J. K., Kim, H. J., Kim, J. K., Kim, Y., and Shin, S. Y. 2009. “Antimicrobial Specificity and Mechanism of Action of Disulfide-Removed Linear Analogs of the Plant-Derived Cys-Rich Antimicrobial Peptide Ib-AMP1.” *Peptides* 30 (12): 2144–49. https://doi.org/10.1016/j.peptides.2009.09.020.

Wang, G., ed. 2010. *Antimicrobial Peptides: Discovery, Design and Novel Therapeutic Strategies*. Wallingford: CABI. https://doi.org/10.1079/9781845936570.0000.

Wang, C. K., Wacklin, H. P., and Craik, D. J. 2012. “Cyclotides Insert into Lipid Bilayers to Form Membrane Pores and Destabilize the Membrane through Hydrophobic and Phosphoethanolamine-Specific Interactions.” *Journal of Biological Chemistry* 287 (52): 43884–98. https://doi.org/10.1074/jbc.M112.421198.

Wang, G. 2012. “Post-Translational Modifications of Natural Antimicrobial Peptides and Strategies for Peptide Engineering.” *Current Biotechnology E* 1 (1): 72–79. https://doi.org/10.2174/2211550111201010072.

Wang, G., Li, X., and Wang, Z. 2016. “APD3: The Antimicrobial Peptide Database as a Tool for Research and Education.” *Nucleic Acids Research* 44 (D1): D1087–93. https://doi.org/10.1093/nar/gkv1278.

Wasinger, V. C., and Corthals, G. L. 2002. “Proteomic Tools for Biomedicine.” *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* 771 (1–2): 33–48. https://doi.org/10.1016/S1570-0232(02)00125-3.

Webb, B., and Sali, A. 2016. “Comparative Protein Structure Modeling Using MODELLER: Comparative Protein Structure Modeling Using Modeller.” In *Current Protocols in Bioinformatics*, edited by Alex Bateman, William R. Pearson, Lincoln D. Stein, Gary D. Stormo, and John R. Yates, 5.6.1-5.6.37. Hoboken, NJ, USA: John Wiley and Sons, Inc. https://doi.org/10.1002/cpbi.3.

Weidmann, J., and Craik, D. J. 2016. “Discovery, Structure, Function, and Applications of Cyclotides: Circular Proteins from Plants.” *Journal of Experimental Botany* 67 (16): 4801–12. https://doi.org/10.1093/jxb/erw210.

Weiner, A. K., Sidoli, S., Diskin, S. J., and Garcia, B. 2017. “GiaPronto: A One-Click Graph Visualization Software for Proteomics Datasets.” *Molecular and Cellular Proteomics*, mcp-TIR117.000438. https://doi.org/10.1074/mcp.TIR117.000438.

Wiederstein, M., and Sippl, M. J. 2007. “ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins.” *Nucleic Acids Research* 35 (Web Server): W407–10. https://doi.org/10.1093/nar/gkm290.

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., and Williams, K. L. 1996. “Progress with Proteome Projects: Why All Proteins Expressed by a Genome Should Be Identified and How to Do It.” *Biotechnology and Genetic Engineering Reviews* 13 (1): 19–50. https://doi.org/10.1080/02648725.1996.10647923.

Wilson, A. J. 2009. “Inhibition of Protein–Protein Interactions Using Designed Molecules.” *Chemical Society Reviews* 38 (12): 3289. https://doi.org/10.1039/b807197g.

Wong, J. H., Xia, L., and Ng, T. B. 2007. “A Review of Defensins of Diverse Origins.” *Current Protein and Peptide Science* 8 (5): 446–59. https://doi.org/10.2174/138920307782411446.

Wong, K. H., Tan, W. L., Serra, A., Xiao, T., Sze, S. K., Yang, D., and Tam, J. P. 2016. “Ginkgotides: Proline-Rich Hevein-Like Peptides from Gymnosperm *Ginkgo biloba*.” *Frontiers in Plant Science* 7 (November). https://doi.org/10.3389/fpls.2016.01639.

Wu, S., Skolnick, J., and Zhang, Y. 2007. “*Ab initio* Modeling of Small Proteins by Iterative TASSER Simulations.” *BMC Biology* 5 (1): 17. https://doi.org/10.1186/1741-7007-5-17.

Wu, S., and Zhang, Y. 2008. “MUSTER: Improving Protein Sequence Profile-Profile Alignments by Using Multiple Sources of Structure Information.” *Proteins: Structure, Function, and Bioinformatics* 72 (2): 547–56. https://doi.org/10.1002/prot.21945.

Yang, Y., Faraggi, E., Zhao, H., and Zhou, Y. 2011. “Improving Protein Fold Recognition and Template-Based Modeling by Employing Probabilistic-Based Matching between Predicted One-Dimensional Structural Properties of Query and Corresponding Native Properties of Templates.” *Bioinformatics* 27 (15): 2076–82. https://doi.org/10.1093/bioinformatics/btr350.

Yeung, H., Squire, C. J., Yosaatmadja, Y., Panjikar, S., López, G., Molina, A., ... and Brimble, M. A. 2016. “Radiation Damage and Racemic Protein Crystallography Reveal the Unique Structure of the GASA/Snakin Protein Superfamily.” *Angewandte Chemie International Edition*, 1–5. https://doi.org/10.1002/anie.201602719.

Yoon, B. J. 2009. “Hidden Markov Models and Their Applications in Biological Sequence Analysis.” *Current Genomics* 10 (6): 402–15. https://doi.org/10.2174/138920209789177575.

Yu, J., Kan, Y., Rapp, M., Danner, E., Wei, W., Das, S., ... and Israelachvili, J. N. 2013. “Adaptive Hydrophobic and Hydrophilic Interactions of Mussel Foot Proteins with Organic Thin Films.” *Proceedings of the National Academy of Sciences* 110 (39): 15680–85. https://doi.org/10.1073/pnas.1315015110.

Zanetti, M. 2004. “Cathelicidins, Multifunctional Peptides of the Innate Immunity.” *Journal of Leukocyte Biology* 75 (1): 39–48. https://doi.org/10.1189/jlb.0403147.

Zeller, S. L., Kalinina, O., and Schmid, B. 2013. “Costs of Resistance to Fungal Pathogens in Genetically Modified Wheat.” *Journal of Plant Ecology* 6 (1): 92–100. https://doi.org/10.1093/jpe/rts013.

Zhang, N., Aebersold, R., and Schwikowski, B. 2002. “ProbID: A Probabilistic Algorithm to Identify Peptides through Sequence Database Searching Using Tandem Mass Spectral Data.” *Proteomics* 2 (10): 1406–1412. https://doi.org/10.1002/1615-9861(200210)2:10<1406::AID-PROT1406>3.0.CO;2-9.

Zhang, Y., Kolinski, A., and Skolnick, J. 2003. “TOUCHSTONE II: A New Approach to *Ab initio* Protein Structure Prediction.” *Biophysical Journal* 85 (2): 1145–64. https://doi.org/10.1016/S0006-3495(03)74551-2.

Zhang, Y. 2008. “I-TASSER Server for Protein 3D Structure Prediction.” *BMC Bioinformatics* 9 (1): 40. https://doi.org/10.1186/1471-2105-9-40.

Zhang, S., Yang, C., Peng, J., Sun, S., and Wang, X. 2009. “GASA5, a Regulator of Flowering Time and Stem Growth in *Arabidopsis thaliana*.” *Plant Molecular Biology* 69 (6): 745–59. https://doi.org/10.1007/s11103-009-9452-7.

Zhang, J., Martin, J. M., Balint-Kurti, P., Huang, L., and Giroux, M. J. 2011. “The Wheat Puroindoline Genes Confer Fungal Resistance in Transgenic Corn: The Puroindolines Confer Corn *SLB* Resistance.” *Journal of Phytopathology* 159 (3): 188–90. https://doi.org/10.1111/j.1439-0434.2010.01744.x.

Zhao, X., Wu, H., Lu, H., Li, G., and Huang, Q. 2013. “LAMP: A Database Linking Antimicrobial Peptides.” Edited by Bin Xue. *PLoS ONE* 8 (6): e66557. https://doi.org/10.1371/journal.pone.0066557.

Zhou, P., Yang, C., Ren, Y., Wang, C., and Tian, F. 2013. “What Are the Ideal Properties for Functional Food Peptides with Antihypertensive Effect? A Computational Peptidology Approach.” *Food Chemistry* 141 (3): 2967–73. https://doi.org/10.1016/j.foodchem.2013.05.140.

Zhou, P., and Huang, J. 2015. *Computational Peptidology*. Vol. 1268. Methods in Molecular Biology. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-2285-7.

Zhu, Y. J., Agbayani, R., and Moore, P. H. 2007. "Ectopic expression of *Dahlia merckii* defensin DmAMP1 improves papaya resistance to *Phytophthora palmivora* by reducing pathogen vigor". *Planta* 226 (1): 87-97.

Zimmermann, R., Sakai, H., and Hochholdinger, F. 2010. “The Gibberellic Acid Stimulated-Like Gene Family in Maize and Its Role in Lateral Root Development.” *PLANT PHYSIOLOGY* 152 (1): 356–65. https://doi.org/10.1104/pp.109.149054.

Anexo IV - Artigo submetido a revista DATABASES – B1

LAITOR4HPC: A text mining pipeline based on HPC for building interaction networks

Bruna Piereck Moura[1], Marx Oliveira Lima[1], Ana Maria Benko-Iseppon[1], Sarah Diehl[2], Reinhard Schneider[2], Ana Christina Brasileiro-Vidal[1] and Adriano Barbosa-Silva[2,3]

1- Universidade Federal de Pernambuco, Genetics Department, Recife, Pernambuco, Brazil.
2- University of Luxembourg, Luxembourg Centre for System Biology, Luxembourg.
3- Centre for Translational Bioinformatics, William Harvey Research, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M6BQ.

E-mail: ana.benko.iseppon@pq.cnpq.br (corresponding Author)

Background: The number of published full-text articles have increased dramatically over the last years, with the number of available articles trespassing 31 million available abstracts deposited on the MEDLINE database until June 2017. As primary information distribution source, scientific publication aggregates essential data for System Biology with the potential to improve the understanding about structure and dynamics of biological systems from cellular to organismal level. Thus, text mining tools configure as an important approach to build biological networks, updating databases and providing annotation for new pathways. PESCADOR is an online web server based on the text-mining tools LAITOR and NLProt, which retrieves protein-protein co-occurrences in a tabular-based format and adds a network schema to it.

Results: We present here a HPC-oriented version of the PESCADOR's native text-mining tool re-named to LAITOR4HPC. The priority was to take advantage of a parallel computing infrastructure available at an HPC facility in order to analyze the complete collection of MEDLINE abstracts, in a much shorter period of time when compared to the original LAITOR implementation. The presented use case illustrates how LAITOR4HPC retrieved more than two times the number of interactions available in Soycyc database, and how it allowed the construction of an exemplary pathway for *Arabidopsis thaliana*.

Conclusion: LAITOR4HPC can be used for efficient text-mining based construction of biological networks derived from big data sources such as MEDLINE. Time consumption and data input limitation will be dependent to the available resources at the HPC facility. Text-mining based analysis using the LAITOR4HPC has retrieved more protein-protein co-occurrences in Soybean than those deposited at specific databases for this crop, making evident that pathways enrichment can be achieved with this tool. In terms of identification of terms related to noncoding elements such as transposons, LAITOR4HPC tool is not promptly ready yet. This still demands the manual construction of specific dictionaries containing the names that should be tagged during the named-entity recognition step of the text-mining analysis.

Keywords: Soybean. Bioinformatics. System Biology.

Background

Scientific information used to be shared in the past via letters between peers. This evolved to printed journals and magazines, and, during the early days of computation, diskettes became a fashion way to exchange articles prior to the invention of the World Wide Web. Nowadays, in the digital era, the access to information has gained a paradigm and, in parallel, new challenges [1]. Likewise, keeping updated with the "state-of-the-art" on most, if not all, the subjects of study by using conventional search and retrieve methods have become obsolete, and a challenge since the twenty-first century information boom in the scientific publishing. According to NCBI resource coordinators (2017) [2], the number of full-text articles increased at a rate of 11.9% a year! Until September 2016, there were about 26 million abstracts indexed at the PubMed/MEDLINE database

Additionally, to understand Biology in all its complexity, it is necessary to understand the structure and dynamics of organisms from cellular to organismal levels. Thus, the focus shall change from one element (protein, gene, phenotype), to a multidimensional point of view. Systems Biology aims to access multi-OMICS data in a variety of experimental conditions and time series to exhaustively generate networks, that may offer an overview of mechanismal responses of organisms to different conditions [3].

Research information is mainly widespread on scientific journals, likewise important data useful to system biology studies [4]. The need for a more efficient way to explore the plethora of information buried in the various literature silos, motivated the application of Information Retrieval and Extraction techniques in Biology. The area of Text- or Literature-mining has emerged to fill this gap between published information and useful information contained in scientific journals. Text-mining is now capable of identify and extract biological entities co-occurrences in different levels such as cellular, tissue and organism-specific contexts [4-6].

Text mining tools follow three fundamental processes described by Krallinger & Valencia (2005) [7]: (i) information retrieval (finding relevant literature to be analyzed), (ii) biological entities identification (protein and gene names tagging) and (iii) biological interaction terms to relate / associate the tagged entities.

PESCADOR [5], a web server based on the text-mining tools LAITOR [6] and NLProt [8], uses a list of articles identifiers (PubMed IDs – PMIDS) as query to search and retrieve relevant abstracts, further, it tags bioentities or biointeractions terms mentioned in the text collection (*corpus*) and finally identifies biological concepts and

their co-occurrences along with bioentities. These co-occurrences are classified according to its reliability in four types: (type 1) – bioentities names co-occur in the same sentence with biointeraction term(s) in between them; (type 2) – bioentities names in the same sentence with biointeraction term in any position; (type 3) – bioentities names in the same sentence, permissive identification of biointeraction terms; and (type 4) – all biological entities of the abstract are retained, not being mandatory their co-occurrence in the same sentence. Co-occurrences of biological concepts are taken in consideration and reported for co-occurrences of types 1-3. Other online tools like iHOP [9] and STRING [10-12] have different approaches. The former is based on keyword search and the latter looks for co-occurrences based on a protein query with two text mining steps added after the update, respectively.

Thus, text mining tools have been a valuable approach to support Systems Biology, not only for updating databases but also for providing annotation of new pathways *ab initio*, by using automated processing of texts [13,14]. In this study, we aimed to improve LAITOR text-mining tool by adding customized programming functions suitable for HPC (High-performance computing) environments, making it possible to analyze a higher amount of text in a shorter period of time.

HPC parallelization

Abstracts preparation

In this new version, the abstracts must be provided as NCBI-PubMed XML format, the files can be downloaded from the PubMed server by doing a search using key-words or accessed on the MEDLINE FTP servers. A python 2.7 script was written to parse the XML tree structure to recover the PMID, title and abstract of each record. The script provides an output which is used as NLProt input [8]. We used the head node to run the parser, which is composed by: Bull B500, 2 * Intel Xeon L5640 @ 2,26 GHz, 12 cores and 2880 Gb of RAM.

Parallelization

To parallelize the analysis the GNU Parallel software [16] (https://www.gnu.org/software/parallel/) was used, with the flag "-j N", where **N** represents the number of cores to be used, where each core is running the i_{th} input file at

a time. To do that, a file containing the list of paths for all the input files was generated and shared across the cores to be used.

Implementation

The approach described here uses all the papers from our selected *corpus* (i.e. MEDLINE) retrieved as described on the above session. The XML files for the corpus is then parsed, and the parsed output is used as input for the NLProt software to highlight all bioentities (i.e. genes, proteins, species names, tissues and cell types). We used NLProt 1.0.2 described at [15] made available by Rostlab.

The final step is to run LAITOR4HPC. This tool was originally developed using PHP (http://www.php.net/) and had its database designed using MySQL as database management system (https://www.mysql.com/) [6]. The LAITOR4HPC implementation is intent to be used as a stand-alone application, differently to the version integrated to PESCADOR, mainly due to the fact that jobs originated from web servers are normally executed in a dedicated machine (or virtual machine) rather than in an HPC environment. Nevertheless, some of the new implemented features can also run in a single core, such as the in-memory database query and the name tagging recovery.

In order to distribute the LAITOR program as a parallel process, it was necessary to make sure that processes running on different nodes could query the bioentities and biointeraction dictionaries seamlessly. However, by using MySQL it would be necessary to install it in every node, which is possible but against the user practices in most HPC systems, including ours. Therefore, we chose to switch the original disk-stored LAITOR databases (MySQL) by an in-memory database approach. For that purpose, we used SQLite (version 3.0): a self-contained, high-reliability, embedded, full-featured, public-domain, SQL database engine (https://www.sqlite.org). Consequently, we needed to adapt the queries from the former system to the latter (Figure 1).

```
###OLD MySQL QUERIES
(A)    #Connect to database
       $conn=mysql_connect($server,$user,$pass);
       $sele=mysql_select_db($db);

(B)    #Preparing and executing query
       $query=mysql_query("select name_txt from ".$table_genes.
                          "where tax_id=\"".$tax_id."\"");
       if(mysql_num_rows($query)>0){
(C)    #Fetching the results
              $result=mysql_fetch_array($query);
              return($result['name_txt']);
       }
       else{
              return(FALSE);
       }


###NEW SQLite QUERIES
(D)    #Creating the in-memory SQL database
       $pdo = new PDO('sqlite::memory:');
       $pdo->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_EXCEPTION);
       $pdo->exec('ATTACH "./laitor_nocase.db" as laitor_db');

(E)    #Preparing the SQL query
       $sql = $pdo->prepare("select * from ".$table_genes.
                          "where tax_id=\"".$tax_id."\"");
       $sql->setFetchMode(PDO::FETCH_ASSOC);

(F)    #Executing the SQL query
       $sql->execute();

(G)    #Fetching the results
       $result = $sql->fetchAll();
```

Figure 1: LAITOR4HPC database management system updates. The principles are the same, however MySQL connects to a server where the database is stored in the disk (A) whereas SQLite loads the database file in the RAM memory of the node executing the query (D). The remaining processes are similar when using both technologies: (B, E and F) preparing and executing the query; and (C and G) retrieving the results.

HPC execution

The NLProt step (bioentity tagging) analysis was launched as four distinct jobs, with 15 cores each (60 core in total). LAITOR was run as a single job, using a total of with 20 cores.

Study case

For this purpose, the taxonomy identifier (*taxid*) filter option of LAITOR was used to check all soybean (*Glycine max* – Taxonomy ID: 3847) interactions. The analyses were performed using the 1,034 XML files downloaded from MEDLINE and submitted to the pipeline. The second test was performed by using keywords, in order to build a small pathway of the regulation of plant defensin (PDF) in *Arabidopsis thaliana*. Two keywords were chosen with the purpose to build the respective pathways, "Plant AND Defensin" (PDF) and "Plant AND Transposon" (PTP).

Results and Discussion

All analyses were performed on the Gaia Cluster maintained by the University of Luxembourg High Performance Computing Department. System configuration and cluster organization can be accessed online [19]. The Python parser script has been run on the interactive node against the 1,034 XML files during about 5 min. Whereas, the NLProt and LAITOR runs took approximately 6 days, in total. A rate of approximately 0,017 s per abstract for processing. This is a speed-up of approximately 117 times faster in comparison to the original implementation per abstract, which only the NLProt tagging took nearly 2 s to complete [5]. Running time should vary depending on node configuration and cores available on the HPC, but it is obviously faster than using a single core approach.

Figure 2 represents the general pipeline obtained for the preparation of the MEDLINE abstracts as input for the LAITOR4HPC text-mining process. After download of the full MEDLINE collection we obtained a dataset of 1,034 XML files each containing approximately 30,000 PMIDS (Figure 2A). These files were transferred to the HPC environment via SCP (secure copy protocol) over a SSH (secure shell) protocol (Figure 2B). The Python parser converted these records into NLProt readable MEDLINE input files (Figure 2C), this step took approximately 5 min in the interactive (head) node. Next, the NLProt job was launched (Figure 2D) where 60 computing nodes where used to run i NLProt processes (where: $\{i \in Z \mid \{0 < i < 1{,}305\}$) in order to tag the bioentities names present on those 1,304 input files (Figure 2E). Upon conclusion, those 1,304 NLProt output files were made available on the head node (Figure 2F) ready for the LAITOR4HPC step.

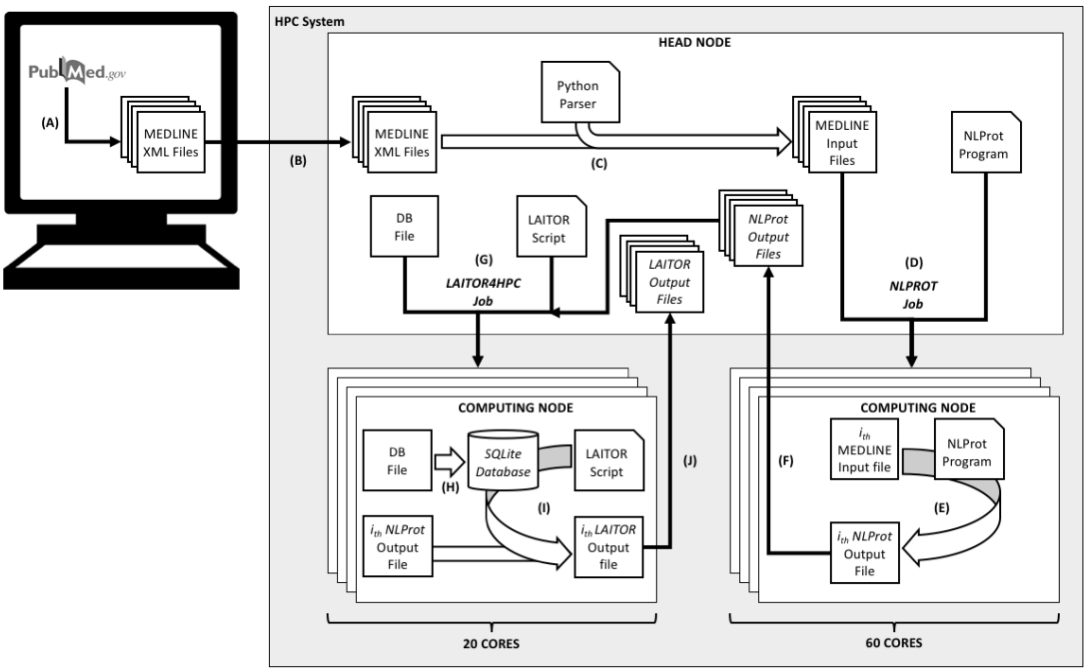


Figure 2: **Complete text-mining pipeline using NLProt and LAITOR4HPC. A)** MEDLINE files are downloaded from NCBI FTP as XML files; **B)** a Python parser is executed to convert the XML files into input files for NLProt which are then **C)** transferred into the interactive (head) node of the HPC system. **D)** A job is then started and *i* different processes are launched in parallel on 60 computing nodes (where: $\{i \in Z \mid \{0 < i < 1{,}305\}\}$). **E)** In each node, the corresponding *i*-th MEDLINE input file is tagged by NLProt which generates **(F)** an *i*-th NLProt output file, which is then placed back to the head node together with the other outputs. **G)** These files are used together with the DB file as input for the LAITOR4HPC job; **H)** which loads an in-memory database prior to the **(I)** tagging of the bioentities and biointeraction present in the corpus. **J)** After completion, the results are placed back to the head node and made available for downstream applications.

The LAITOR4HPC job execution used the DB file and the NLProt output files as inputs (Figure 2G). The jobs were launched from the head node, to be executed in 20 cores. Each *i*-th process is directed to a corresponding computing node together with the DB file, the LAITOR script and the NLProt output. Every computing node loads the DB file as an SQLite in-memory database on that node during execution (Figure 2H). Then the LAITOR4HPC script receives the *i*-th process and analyses it against the loaded in-memory database, which contains the bioentities and biointeraction dictionaries (Figure 2I). Once the results are obtained, they are made available back to the head node (Figure 2J). At the end of the job, all the LAITOR output files are retrieved back to the head node and can be copied by SCP or similar method to user-client computer; from there, users can further explore the text-mining outputs to create co-occurrence networks, for example.

By switching from MySQL to SQLite we avoid HPC limitations during the database querying in the HPC architecture as mentioned previously. Using SQLite in-memory, a new database is created purely in the memory of the computing nodes, this database ceases to exist as soon as the database connection is closed. As the database is self-contained in a text file, this file needs to be distributed across the computing nodes together with the input file to be analyzed.

The proposed parallel approach allowed us to query the whole *corpus* and to extract all its bioentities co-occurrences. In this context: the more articles available, the better the result. LAITOR4HPC running time is drastically decreased by the parallelization approach, as in comparison to the original version used by the PESCADOR website, where only a maximum of 1,000 papers could be read on-the-fly. Parallel SQL limitations caused by competitive accesses on the HPC environment were avoided by loading the database in the RAM memory of each computing node.

After soybean analysis, the pipeline tagged 7,894 co-occurrences, more than two times the amount of interactions available for this species in Soycyc (3,539 entries), a specific database that contains pathway data for soybean. This shows the efficiency of the pipeline. LAITOR tags four types of co-occurrences, according to the scope by which they were found in the text, being type 1 more and type 4 less likely to correspond to effective bioentities interactions. Consequently, to build reliable pathways, manual curation is advised [6].

By the keywords (Plant AND Defensin) search on PubMed, 770 papers were retrieved and analyzed. However, only five had interactions related to the regulation of PDF. This specific regulation pathway contains 14 elements among different proteins and hormones, showing that even the genes related to plant responses to biotic stress may use factors that are not directly related to this stress, regarding the known cross-talk response. This mechanism has been recognized, especially for plant antimicrobial peptides like PDF [19-21] where genes related to other functions are induced, like EIN3, related to salt tolerance [22], which can also induce defense against biotic factors (Figure 3).

The query “Plant AND Transposon” (PTP) didn’t retrieve abstracts where NLProt could identify bioentities names with precision. Consequently, it lacked to establish bioentities co-occurrences for this corpus. To succeed with this task, it will be necessary to build specific dictionaries and, likely, to use machine learning methods so that the text mining tools succeed to work with noncoding, repetitive sequences terms. It is an essential

demand to prepare those tools to pursue such a challenge, due to the influence of TEs in gene regulation, especially in eukaryotic organisms [23-28].

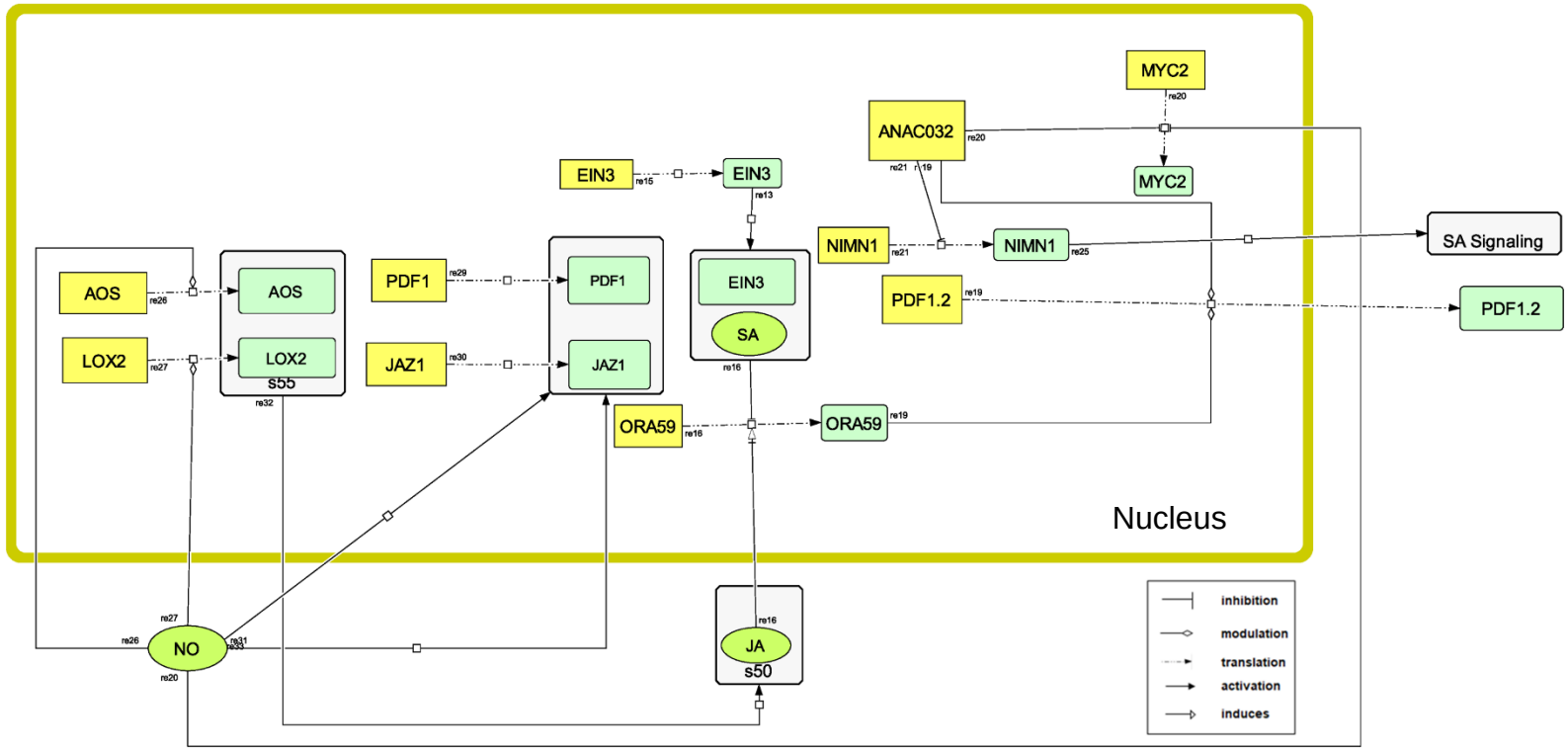


Figure 3: Regulation pathway of the PDF gene recovered by LAITOR4HPC. The pathway evidenced the elements that may induce (like the hormone jasmonic acid - JA), or repress (like the nitric oxide - NO), the expression of the gene in the nucleus. The yellow rectangles are the genes and the light green are the proteins.

Conclusion

The improvement of LAITOR, with the achievement of LAITOR4HPC decreased significantly the computing time, due to the implementation of parallelization. Such an improvement resulted not only in much faster run time, but also maintained the consistence of LAITOR4HPC considering the expected results. Time will vary according to available hardware resources, especially considering memory capacity and the number of available cores. Since this improved online tool include only data from abstracts, it is important to consider manual data curation to confirm predicted protein-protein interactions from term co-occurrences.

Text mining-based analysis using LAITOR4HPC has identified almost two times more putative protein-protein interactions in soybean than the ones deposited in the specific database Soycyc, making it a suitable tool for the enrichment of previously defined pathways, as well as for the establishment new ones. Therefore, it is expected that LAITOR4HPC will be useful not only to search for plant interactions, as in the present work, but also considering many other research questions, including human research and interactions involving multiple taxonomic categories.

Availability and requirements

Project name: BioComp - InterSys

Project home page: Will be available on GitHub

Operating system(s): Linux and MacOSX

Programming language: PHP and Python

Other requirements: SQLite / PHP interpreter

License: Not applicable

Restriction: Not applicable

References

[1] Andrade AM, Bork P. Automated extraction of information in molecular biology. FEBS Letters. 2000; doi: 10.1016/S0014-5793(00)01661-6

[2] NCBI Resource Coordinators. Database resources of the National Center for Biotechnology information. Nucleic Acids Res. 2017; doi: 10.1093/nar/gkw1071

[3] Kitano H. Systems biology: a brief overview. Science. 2002, doi: 10.1126/science.1069492

[4] Rebholz-Schuhmann D, Oellrich A, Hoehndorf R. Text-mining solutions for biomedical research: enabling integrative biology. Nat Rev Genet. 2012; doi: 10.1038/nrg3337

[5] Barbosa-Silva A, Fontaine J-F, Donnard ER, Stussi F, Ortega JM, Andrade-Navarro MA. PESCADOR, a web-based tool to assist text-mining of biointeractions extracted from PubMed queries. BMC Bioinformatics. 2011; doi: 10.1186/1471-2105-12-435

[6] Barbosa-Silva A, Soldatos TG, Magalhães ILF, Pavlopoulos GA, Fontaine J-F, Andrade-Navarro MA, Schneider R, Ortega M. LAITOR – Literature Assistant for Identification of Terms co-occurrences and Relationships. BMC Bioinformatics. 2010; doi: 10.1186/1471-2105-11-70

[7] Krallinger M, Valencia A. Text-mining and information-retrieval services for molecular biology. Genome Biol. 2005; doi: 10.1186/gb-2005-6-7-224

[8] Mika S, Rost B. NLProt: extracting protein names and sequences from papers. Nucleic Acids Res. 2004; doi: 10.1093/nar/gkh427

[9] Fernández JM, Hoffmann R, Valencia A. iHOP web services. Nucleic Acids Res. 2007; doi: 10.1093/nar/gkm298

[10] Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighborhood of a gene. Nucleic Acids Res. 2000; doi: 10.1093/nar/28.18.3442

[11] Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, Mering CV. STRING 8 – a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009; doi: 10.1093/nar/gkn760

[12] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, Mering CV. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017; doi: 10.1093/nar/gkw937

[13] Fluc, J, Hofmann-Apitius M. Text mining for systems biology. Drug Discov Today. 2014; doi: 10.1016/j.drudis.2013.09.012

[14] (Fluck and Hofmann-Apitius 2014; Trindade et al. 2015)Trindade D, Orsine LA, Barbosa-Silva A, Donnard E R, Ortega JM. A guide for building biological pathways along with two case studies: Hair and breast development. Methods. 2015; 74:16-35; doi: 10.1016/j.ymeth.2014.10.006

[15] This wiki is Open – NLProt. https://rostlab.org/owiki/index.php/NLProt. Accessed 20 May 2018

[16] Tange O. GNU Parallel - The Command-Line Power Tool. USENIX. 2011; doi: 10.5281/zenodo

[17] Schläpfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, et al. Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. Plant Physiol. 2017; 173:2041–59. doi:10.1104/pp.16.01942.

[18] HPC @ Uni.lu – High Performance Computing in Luxembourg. https://hpc.uni.lu. Acessed 20 May 2018

[19] Do HM, Lee SC, Jung HW, Sohn KH, Hwang BK. Differential expression and in situ localization of a pepper defensin (CADEF1) gene in response to pathogen infection, abiotic elicitors and environmental stresses in Capsicum annuum. Plant Sci. 2004; doi: 10.1016/j.plantsci.2004.01.008

[20] Fraire-Velázquez, S, Rodríguez-Guerra R. Abiotic and Biotic Stress Response Crosstalk in Plants. 1st ed. Mexico: Physiological, Biochemical and Genetic Perspectives; 2011

[21] Vos IA, Moritz L, Pieterse CMJ, Van Wees SCM (2015) Impact of hormonal crosstalk on plant resistance and fitness under multi-attacker conditions. Front Plant Sci 6:1–13. doi: 10.3389/fpls.2015.00639

[22] Ryu H, Cho YG. Plant hormones in salt stress tolerance. J Plant Biol. 2015; doi: 10.1007/s12374-015-0103-z

[23] Mette MF, van-der-Winden J, Matzke M, Matzke AJ. Short RNAs can identify new candidate transposable element families in Arabidopsis. Plant Physiol. 2002; doi: 10.1104/pp.007047

[24] Gonzáles J, Karasov TL, Messer PW and Petrov DA. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. PLoS Genet. 2010; doi: 10.1371/journal.pgen.1000905

[25] Li Y, Li C and Jin Y. Domestication of transposable elements into MicroRNA genes in plants. PLoS One. 2011; doi: 10.1371/journal.pone.0019212

[26] Lisch D and Bennetzen JL. Transposable element origins of epigenetic gene regulation. Curr Opin Plant Biol. 2011; doi: 10.1016/j.pbi.2011.01.003

[27] Wang X, Weigel D, Smith LM. Transposon variants and their effects on gene expression in Arabidopsis. PLoS Genet. 2013; doi: 10.1371/journal.pgen.1003255

[28] Hanada K, Vallejo V, Nobuta K, Slotkin RK, Lisch D, Meyers BC, Shiu S-H, Jiang N. The functional role of Pack-MULEs in rice inferred from purifying selection and expression profile. Plant Cell. 2009; doi: 10.1105/tpc.108.063206

Anexo V

Patente submetida a Diretoria de Inovação e Empreendedorismo (DINE)

Novo peptídeo antimicrobiano modificado a partir da espécie *Cajanus cajan* com potencial atividade antimicrobiana

Autores: Ana Maria Benko-Iseppon; Lívia Maria Batista Vilela; Carlos André dos Santos Silva; Sergio Crovella; **Marx Oliveira de Lima**

UFPE POSITIVA DIRETORIA DE INOVAÇÃO

Universidade Federal de Pernambuco
Diretoria de Inovação

COMPROVANTE DE RECEBIMENTO DE DOCUMENTOS

Declaro para os devidos fins que esta diretoria recebeu no dia 17/05/2018, Comunicado de Invenção, Termo de Sigilo e Anexo de Inventores da solicitação para depósito do Pedido de Patente de título: **"Novo peptídeo antimicrobiano modificado a partir da espécie *Cajanus cajan* com potencial atividade antimicrobiana".**

Para:
(X) estudo por esta diretoria;
(X) busca prévia;
(X) análise e emissão de parecer de patenteabilidade;
(X) depósito dos documentos para Pedido de Patente de Invenção.

A presente solicitação pertence a um grupo tendo como inventora vinculada à UFPE Ana Maria Benko Iseppon, Portador do RG: 9.111.082 SSP/SP e CPF: 9.111.083, nascida em São Paulo - SP, estado civil: casada. A referida é professora da UFPE, do Departamento de Genética, do Centro de Biociências.

Os outros membros do Grupo, coinventores da tecnologia são: Lívia Maria Batista Vilela, Carlos André dos Santos Silva, Sergio Crovella, Marx Oliveira de Lima.

Recife, 14 de junho de 2018.

Atenciosamente,

Ana Carolina Borba Fernandes Padrão
Assistente em Administração
SIAPE 2156155
UFPE

Anexo VI

Patente submetida a Diretoria de Inovação e Empreendedorismo (DINE)

Novo peptídeo sintético modificado a partir de uma defensina de *Manihot esculenta*

Autores: Ana Maria Benko-Iseppon; Lívia Maria Batista Vilela; Carlos André dos Santos Silva; Sergio Crovella; **Marx Oliveira de Lima**

UFPE

POSITIVA
DIRETORIA DE INOVAÇÃO

Universidade Federal de Pernambuco
Diretoria de Inovação

COMPROVANTE DE RECEBIMENTO DE DOCUMENTOS

Declaro para os devidos fins que esta diretoria recebeu no dia 17/05/2018, Comunicado de Invenção, Termo de Sigilo e Anexo de Inventores da solicitação para depósito do Pedido de Patente de título: **"Novo peptídeo sintético modificado a partir de uma defensina de *Manihot esculenta*".**

Para:

(X) estudo por esta diretoria;

(X) busca prévia;

(X) análise e emissão de parecer de patenteabilidade;

(X) depósito dos documentos para Pedido de Patente de Invenção.

A presente solicitação pertence a um grupo tendo como inventora vinculada à UFPE Ana Maria Benko Iseppon, Portador do RG: 9.111.082 SSP/SP e CPF: 9.111.083, nascida em São Paulo - SP, estado civil: casada. A referida é professora da UFPE, do Departamento de Genética, do Centro de Biociências.

Os outros membros do Grupo, coinventores da tecnologia são: Carlos André dos Santos Silva, Lívia Maria Batista Vilela, Sergio Crovella, Marx Oliveira de Lima.

Recife, 14 de junho de 2018.

Atenciosamente,

Ana Carolina Borba Fernandes Padrão
Assistente em Administração
SIAPE 2156155
UFPE

Anexo VII – Resultados parciais

Trabalho em fase de conclusão

Prospecção, expressão e análise genômica: Uma busca por esnaquinas no transcriptoma do feijão caupi (*Vigna unguiculata*).

Marx Oliveira-Lima; Valesca Pandolfi; Ana Maria Benko-Iseppon*

- Foram utilizadas 30 sequencias *seed* para a prospecção no transcriptoma, onde foram recuperados 24 homólogos não redundantes que foram anotados para: presença e composição do domínio, peso molecular, pI, localização subcelular, provável atividade antimicrobiana e presença do peptídeo sinal.
- Os transcritos recuperados foram avaliados para o seu perfil de expressão baseado no log do *fold change* nas situações: infectado com Poty vírus 60 minutos e 16 horas, infectado com vírus do mosaico severo 60 minutos e 16 horas, Défict hídrico 25 minutos e 150 minutos. Onde foram gerados *heat maps* que serviram a seleção de 10 representantes para validação via RT-qPCR a qual encontra-se em andamento.
- Na análise genômica avaliamos a estrutura dos 24 trancritos mapeados, onde observamos a conservação das estruturas gênicas com dois, três ou quatro éxons. Esta análise encontra-se em andamento para a comparação com outras leguminosas modelo e/ou de interesse econômico.
- Por fim pretendemos observar a estrutura dos peptídeos recuperados via modelagem molecular, bem como avaliar sua estabilidade.