Universidad de la República

Facultad de Ciencias - PEDECIBA Biología

Prospección de defensinas y esnaquinas de plantas nativas para el desarrollo de nuevos agentes antimicrobianos

Mag. Susana Rodríguez Decuadro

Tesis presentada para la obtención del título de

Doctor en Ciencias Biológicas

Orientadores:

Dra. Ana Maria Benko Iseppon

Dra. Gianna Cecchetto

Tesis presentada para la obtención del título de Doctor en Ciencias Biológicas

Prospección de defensinas y esnaquinas de plantas nativas para el desarrollo de nuevos agentes antimicrobianos

Mag. Susana Rodríguez Decuadro

Orientadores:

Dra. Gianna Cecchetto, Microbiología-IQB, Facultad de Ciencias, Universidad de la República

Dra. Ana María Benko Iseppon, Departamento de Genética, Universidad Federal de Pernambuco

Tribunal:

Dra. Mónica Marín, Facultad de Ciencias, Universidad de la República (Presidente)

Dra. María Julia Pianzzola, Facultad de Química, Universidad de la República (Vocal)

Dra. Sabina Vidal, Facultad de Ciencias, Universidad de la República (Vocal)

«A los hombres fuertes les pasa lo que a las cometas; se elevan cuando es mayor el viento que se opone a su ascenso.»

José Ingenieros

Agradecimientos

Quisiera agradecer a todas las personas que estuvieron cerca y aportaron de una forma u otra a esta tesis de doctorado:

a mis tutoras de tesis, Gianna y Ana, por aceptar dirigir este trabajo de tesis, aún sin conocerme. Por transmitirme sus conocimientos y por el estímulo para continuar con esta temática tan apasionante.

a Mari que me recibió y me mostró el funcionamiento del laboratorio 12 de Micro, ayudándome en todo

a Gabi, que me dio una mano en los últimos experimentos

a todos mis compañeros del Laboratorio 12, especialmente otra vez a Mary y Gabi y también a MIL, Vale, César, Sonia, María Inés y María Julia por las horas compartidas, siempre dispuestos a ayudarme y a hacerme sentir como una más

a tod@s MICRO por recibirme y ayudarme cuando lo precisé

a Pablo Dans por su colaboración en la predicción de la estructura 3D de los péptidos y en la redacción del artículo científico ya publicado

a Pablo Smircich por su colaboración en el análisis de los transcriptomas

al grupo de Bioquímica, especialmente a Laura Franco-Fraguas por su orientación con los extractos proteicos y a Valentín Castilla al que recurría a cada momento para consultarlo por la expresión y purificación de los péptidos

a las personas que integran el Departamento de Biología Vegetal de Facultad de Agronomía, mis compañeros de trabajo, por apoyarme en la realización de este doctorado. Especiamente a Silvia Ross por ayudarme con la germinación de semillas y a Magdalena Vaio por el apoyo emocional de siempre

a mi familia, muy especialmente a mis padres, por apoyarme a lo largo de toda mi vida y alentarme para que siguiera adelante. Principalmente después que nació Renata mi familia ha sido una tribu que me ha ayudado incondicionalmente

a Daniel por estar siempre y por hacerme reír, aún en los momentos difíciles…

a Renata, por cambiarme la vida, por acompañarme desde la panza en los útimos experimentos de la tesis y ya en la etapa de redacción hacerme los días más divertidos y agotadores también. Por sorprenderme cada día….

Resumen

Las plantas producen un repertorio diverso de péptidos antimicrobianos (AMPs), estructural y funcionalmente distintos, que forman parte de la inmunidad innata actuando contra una amplia gama de patógenos. Presentan actividad contra fitopatógenos pero también contra patógenos del ser humano. Por ello, han sido considerados compuestos prometedores para el desarrollo de productos con importantes aplicaciones biotecnológicas. Además de su tamaño pequeño, los AMPs presentan algunas características comunes como carga neta positiva a pH fisiológico y un número par de residuos cisteína, pero difieren en tamaño, composición aminoacídica y estructura molecular. Los AMPs vegetales han sido clasificados en varias familias, de acuerdo a su homología de secuencia, sus motivos cisteínas y su estructura 3D, entre las que se incluyen defensinas, esnaquinas, tioninas, proteínas de transferencia de lípidos, ciclótidos y proteínas *hevein-like*. El objetivo de este trabajo fue aislar y caracterizar estructural y funcionalmente AMPs de tipo defensina y esnaquina, presentes en especies nativas de Sudamérica con usos tradicionales en la medicina popular: *Peltophorum dubium* (ibirapitá), *Maytenus ilicifolia* (congorosa) and *Erythrina crista-galli* (ceibo). Varios genes de estas familias fueron aislados usando dos estrategias: amplificación por PCR usando *primers* degenerados o específicos y secuenciación masiva de transcriptomas (RNA-seq) usando ensamblado *de novo*. Un gen de tipo esnaquina (*PdSN1*) fue aislado de *P. dubium* con *primers* degenerados diseñados en base a secuencias disponibles de otras especies y un gen de tipo defensina (*EcgDf1*) fue aislado de *E. crista-galli*, con *primers* específicos diseñados para *Vigna unguiculata.* Además, varios genes Snakin/GASA y de defensinas fueron identificados en los transcriptomas *de novo* de *P. dubium* y *M. ilicifolia*. Algunos de estos genes fueron verificados en ADN genómico y caracterizados estructuralmente. Todos los péptidos encontrados presentaron patrones comunes de cada familia de AMP, con una distribución de cisteínas típica de cada familia, con la excepción de algunas defensinas putativas (tanto en *M. ilicifolia* como en *P. dubium*) que tienen un motivo αβ modificado o el motivo γ-core está ausente. Algunas defensinas putativas encontradas en los transcriptomas estudiados difieren en varios residuos, incluso aquellos altamente conservados cuando se toman en cuenta las defensinas depositadas en PDB. Para la caracterización funcional de PdSN1 y EcgDf1, éstos fueron expresados de forma recombinante. Ambos péptidos maduros fueron producidos en *E. coli* mayormente de forma soluble y luego de un paso de purificación por cromatografía de afinidad se obtuvieron péptidos biológicamente activos. La actividad antimicrobiana fue evaluada *in vitro* contra varios patógenos humanos y de plantas, incluyendo hongos filamentosos, levaduras y bacterias. Ambos péptidos presentaron actividad antimicrobiana, generalmente con mejor performance para hongos, especialmente en el caso de la defensina. Se destaca que EcgDf1 mostró actividad contra *Penicillium expansum,* un patógeno que causa importantes enfermedades post-cosecha; mientras que PdSN1 presentó fuerte actividad contra la bacteria *Streptomyces scabies*, importante patógeno de papa. También es interesante recalcar que ambos péptidos tuvieron actividad contra *Candida albicans* y *Aspergillus niger,* patógenos humanos que producen infecciones en pacientes inmunodeprimidos. Varios de los péptidos encontrados en este trabajo pueden ser considerados como candidatos para el desarrollo de productos útiles para el tratamiento o el control efectivo de enfermedades. Investigaciones enfocadas a las relaciones estructura-función de los péptidos de *P. dubium*, *E. crista-galli* y *M. ilicifolia*, permitirán evaluar su potencial futura aplicabilidad.

Summary

Plants produce a diverse repertoire of structurally and functionally diversified antimicrobial peptides (AMPs) which are an inherent part of innate plant immunity by acting against a wide range of pathogens. They have activities against phytopathogens, as well as against human pathogens. Accordingly, plant AMPs are considered promising compounds with important biotechnological applications. In addition to their small size, they have some features in common like net positive charge at physiological pH and an even number of cysteine residues, but they differ in size, amino acid composition, and molecular structure. Based on amino acid sequence homology, cysteine motifs, and three-dimensional structures, plant AMPs are grouped into several families, including defensins, snakins, thionins, lipid transfer proteins, cyclotides, and hevein-like proteins. The aim of this work was to isolate and characterize AMPs structurally and functionally, with emphasis on those of the defensin and snakin families, present in three South American native species traditionally used in folk medicine: *Peltophorum dubium* (ibirapitá), *Maytenus ilicifolia* (congorosa) and *Erythrina crista-galli* (ceibo). Several genes of these families were isolated using two strategies: PCR amplification with degenerated or specific primers and transcriptome massive sequencing (RNAseq) using a *de novo* assembly. A snakin gene (*PdSN1*) was isolated from *P. dubium* with degenerated primers designed based on available sequences of other species. Also, a defensin gene (*EcgDf1*) was isolated from *E. crista-galli*, with primers designed for *V. unguiculata*. In addition, several defensins and Snakin/GASA genes were identified from *de novo P. dubium* and *M. ilicifolia* transcriptomes. Some of the identified genes were verified in genomic DNA and structurally characterized. All peptides found, presented common structural patterns of each family, wiht a typical cysteine distribution, with the exception of some putative defensins (both in *M. ilicifolia* and *P. dubium*) that exhibited a modified αβ motif or the γ-core motif is absent. Some putative defensins found in the transcriptomes of *P. dubium* and *M. ilicifolia* differ in several residues, even those expected to be highly conserved, considering defensins deposited in PDB. For the functional characterization of PdSN1 and EcgDf1, both peptides were produced recombinantly. Both mature peptides were expressed in *E. coli* mostly in a soluble form, which after a affinity chromatographic step, resulted in the purification of a biologically active peptides. The antimicrobial activity was evaluated *in vitro* against several human and plant pathogenic microorganisms, including filamentous fungi, yeast and bacteria. Both peptides showed antimicrobial activity, generally with a better performance for fungi, especially in the case of the defensin. It is emphasized that EcgDf1 exhibited activity against *Penicillium expansum*, a pathogen that causes important post-harvest diseases while PdSN1 showed a potent activity against the bacterium *Streptomyces scabies*, an important pathogen of potato. It is also interesting to note that both peptides showed activity against *Candida albicans* and *Aspergillus niger*, both human pathogens that produce infections in immunosuppressed patients. Several of the peptides found in this work can be considered good candidates for the development of useful products for the treatment or effective control of diseases. Focused investigations regarding structure-function relationships of *P. dubium*, *E. crista-galli* and *M. ilicifolia* peptides, will allow to evaluate their potential future applicability.

Tabla de contenido

Introducción general

Importancia de la búsqueda de nuevos productos bioactivos para el control de enfermedades

Algunos hongos, virus y bacterias son agentes de enfermedades, en ocasiones devastadoras en humanos plantas y animales. En producción agrícola, el control de patógenos es un factor determinante para asegurar la cosecha, mejorar rendimientos y aumentar la vida útil poscosecha. Estos patógenos pueden deteriorar la calidad del producto (valor nutricional, características organolépticas), llegando en algunos casos a ser responsables de efectos nocivos para los consumidores debido a la producción de toxinas y alergenos (Díaz-Dellavalle et al. 2011). Uno de los mayores desafíos a nivel mundial ha sido y será en los próximos años, lograr un aumento en la producción agrícola para alimentar a una población en constante crecimiento, contribuir al desarrollo general de los que dependen de la agricultura en países en desarrollo y adoptar métodos de producción más eficientes y sostenibles (FAO, 2009). El manejo de plagas y enfermedades ha jugado un papel importante en la duplicación de la producción de alimentos en los últimos 40 años. Sin embargo, cada año se estima que un 10-16% de la cosecha mundial se pierde debido a las enfermedades de los cultivos, sin contar las pérdidas post-cosecha (Chakraborty y Newton 2011). El aumento constante de resistencia a las drogas disponibles para combatir enfermedades, genera la necesidad de descubrir nuevos compuestos con actividad antimicrobiana. Además, el uso de agroquímicos, muchos de ellos perjudiciales para el ambiente y la salud de productores y consumidores, está siendo cuestionado desde diversos ámbitos sociales. La demanda de productos "más verdes" para el manejo de plagas ha llevado al incremento del desarrollo de nuevos bio-plaguicidas (Cantrell et al. 2012).

Péptidos antimicrobianos: potenciales agentes terapéuticos innovadores

La búsqueda de nuevos productos ha aumentado el interés por los péptidos antimicrobianos (AMPs, *antimicrobial peptides*), componentes del sistema inmune de todos los organismos vivos. Los AMPs se consideran entre los compuestos más prometedores para su uso en medicina y como agentes terapéuticos para prevenir y tratar enfermedades en cultivos de importancia comercial (Kaur et al. 2011, Larrañaga et al. 2011, Shah y Read 2013). A nivel clínico, un alto número de patentes documenta que, en muchos casos, los AMPs muestran una acción incrementada contra bacterias resistentes a antibióticos. Diferentes aplicaciones terapéuticas de estos compuestos, desde administración tópica hasta tratamientos sistémicos de las infecciones han sido evaluadas por varias compañías biotecnológicas (Zasloff 2002, Dalla Rizza

et al. 2008). Además, AMPs sintéticos están siendo explorados como nuevos fármacos (Shah y Read 2013). Dado que este tipo de péptidos está ampliamente distribuido en el reino vegetal, podrían ser útiles para mejorar los rendimientos agrícolas, aumentando la resistencia de los cultivos al ataque de patógenos, ya sea mediante mejoramiento genético, desarrollo de cultivos transgénicos, o de productos que sustituyan los actuales agroquímicos. Podrían utilizarse además para incrementar la tolerancia a distintos tipos de estrés abiótico, ya que se ha mostrado que algunos están involucrados en las respuestas de tolerancia a estrés por salinidad, calor, seca, etc. (Carvalho y Gomes 2011; Nahirñak et al. 2012b). Como los AMPs vegetales no sólo presentan actividad contra patógenos de plantas, sino también contra patógenos de humanos, están siendo estudiados para el desarrollo de nuevos fármacos con acción antibiótica y anti-inflamatoria (Thevissen et al. 2007).

Presentación de los péptidos antimicrobianos vegetales

Las plantas están continuamente siendo atacadas por patógenos. Sin embargo, pocos ataques llegan a producir una infección sistémica y desarrollo de enfermedad ya que las plantas se defienden de los mismos mediante mecanismos tanto constitutivos como inducibles. Una vez que las plantas perciben al patógeno, se desencadena una cascada de señales que llevan a la formación de barreras físicas (como el reforzamiento de la pared celular), la producción de metabolitos secundarios y la síntesis de proteínas relacionadas a la patogénesis (proteínas PR). Dentro de los diferentes grupos de proteínas PR, algunas presentan un peso molecular menor a los 10 KDa, entre las que se encuentran los comúnmente denominados péptidos antimicrobianos (Sels et al. 2008; Benko-Iseppon et al. 2010).

Los AMPs son parte esencial de la respuesta inmune innata no sólo de vegetales, sino también de vertebrados e invertebrados. Son moléculas evolutivamente ancestrales que tienen un rol fundamental en la formación de barreras de defensa, pre y post infección, contra una amplia variedad de patógenos incluyendo bacterias, hongos, virus y protozoarios. Promueven una respuesta relativamente rápida del organismo con un costo energético comparativamente menor al uso del sistema inmune adaptativo de los vertebrados superiores (Zasloff 2002). Dicha respuesta refleja la estructura genética y proteica de estas moléculas ya que muchos de ellos comparten varias características: tamaño pequeño (menor a 10 KDa), carga general positiva, estereogeometría anfipática, tolerancia a solventes ácidos y orgánicos, estabilidad térmica y amplia actividad biológica (Yeaman y Yount 2003; Brogden 2005). La naturaleza catiónica de los AMPs en condiciones fisiológicas, asociados con una tendencia a adoptar estructuras anfipáticas podría facilitar su interacción e inserción dentro de las paredes celulares negativas y las membranas fosfolipídicas de los microorganismos (Oren y Shai 1998) por lo que se ha

postulado que su actividad biológica podría explicarse por dos mecanismos de acción principales: a) una interacción no específica con las membranas microbianas llevando a la permeabilización por formación de poros o b) interacciones específicas con moléculas blanco, intracelulares o asociados a membranas, resultando en la inhibición de procesos metabólicos vitales como la síntesis de ácidos nucleicos, de proteínas o de la pared celular (Figura 1) (Tossi y Sandri 2002; Brogden 2005; Peters et al. 2010).

En plantas, estos péptidos contienen frecuentemente un número par de residuos de cisteína que estabilizan la estructura a través de la formación de puentes disulfuro (Lay y Anderson 2005). A pesar de su tamaño pequeño, que impone limitaciones al repertorio de estructuras que pueden adoptar, estos AMPs muestran una alta variabilidad de secuencia en sitios no conservados y tasas de sustitución de aminoácidos consistentes con la hipótesis de presiones evolutivas o coevolutivas durante la interacción huésped/patógeno (Hughes 1999, Blaser 2002). Los principales grupos de éste tipo de péptidos en plantas son: defensinas, tioninas, esnaquinas, proteínas de transferencia de lípidos, ciclótidos y heveínas (Benko-Iseppon et al. 2010; Nawrot et al. 2014).

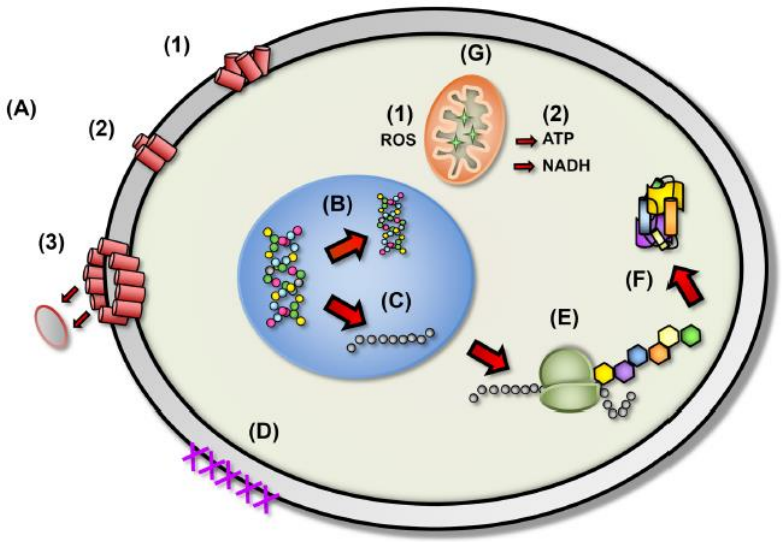


Figura 1. Diferentes mecanismos de acción propuestos para los AMPs en células microbianas. **(**A) Disrupción de la integridad de la membrana celular: (1) inserción aleatoria en la membrana; (2) interacción entre sí a través de sus secuencias hidrofóbicas (3) eliminación de secciones de membrana y formación de poros. (B) Inhibición de la síntesis de ADN. (C) Bloqueo de la síntesis de ARN. (D) Inhibición de las enzimas necesarias para la unión de proteínas estructurales de la pared celular. (E) Inhibición de la función ribosomal y síntesis de proteínas. (F) Bloqueo de proteínas chaperonas necesarias para un correcto plegamiento de las proteínas (G) Blanco en la mitocondria: (1) inhibición de la respiración celular e inducción de la formación de ROS y (2) ruptura de la membrana mitocondrial y eflujo de ATP y NADH. Extraído de Peters et al. 2010.

Poniendo el foco en dos tipos de AMPs de plantas: defensinas y esnaquinas

DEFENSINAS

Entre los AMPs vegetales descritos hasta hoy, sólo las defensinas parecen estar conservados entre organismos tan distintos como plantas, invertebrados y vertebrados, constituyendo una superfamilia de proteínas (Thomma et al. 2002, Charlet et al. 1996, Oard y Karki, 2006). En el reino vegetal se presentan aparentemente distribuidos en forma ubicua. Las primeras defensinas vegetales aisladas fueron obtenidas en el inicio de los años 90 a partir de granos de trigo y cebada (Mendez et al. 1990) pero actualmente se han reportado genes de defensinas en prácticamente todas las familias de plantas. Forman parte de una familia multigénica, con más de 300 genes tipo-defensina descritos para el genoma de *Arabidopsis thaliana* (Silverstein et al. 2005). La estructura de estos genes está integrada por dos exones y un intrón (Carvalho y Gomes 2009). La estructura primaria generalmente consiste en un péptido señal en el extremo amino terminal y un péptido maduro básico (45 a 54 aminoácidos), rico en cisteínas. Se han identificado, sin embargo, defensinas con estructuras alternativas, por ejemplo, las defensinas florales NaD1 (*Nicotiana alata*) y PhD1 y PhD2 (*Petunia hybrida*) que presentan un prodominio C-terminal; o AtPDF3.1 y AtPDF3.2, de *A. thaliana* que contienen un dominio rico en cisteínas entre el péptido señal y el péptido maduro (De Coninck et al. 2013). Análisis filogenéticos a partir de la estructura primaria (secuencia proteica) de varias defensinas ha mostrado que defensinas con actividad similar tienden a aparecer en un mismo grupo. Sin embargo, una limitante a este tipo de análisis es que muchas defensinas han sido evaluadas únicamente para uno o dos patógenos, por lo que la actividad reportada podría no reflejar su función inhibitoria principal. Aunque es necesario caracterizar muchas más defensinas vegetales para desarrollar pautas que permitan predecir la actividad de secuencias nuevas, esta aproximación podría resultar útil para identificar relaciones estructura-función (Van der Weerden y Anderson, 2013).

La estructura terciaria está definida por la presencia de una hélice alfa y tres hojas beta antiparalelas, estabilizadas por cuatro puentes disulfuro (Figura 2a) (Lay y Anderson 2005). Esta estructura compacta les conferiría resistencia a temperaturas extremas y a degradación mediada por pH o proteasas (Hegedus y Marx 2012). Se ha determinado además mediante cristalografía de rayos X que algunas defensinas forman dímeros (Figura 2b), lo que aumenta su actividad antifúngica (Song et al. 2011; Lay et al. 2012). En el caso de NaD1, la mutagénesis dirigida de la lisina 4 redujo su capacidad de dimerizar, reduciendo además su capacidad de inhibir el crecimiento de *F. oxysporum* f. sp. vasinfectum (Lay et al. 2012). Las defensinas animales, vegetales y fúngicas han sido clasificadas recientemente en dos superfamilias, las defensinas *cis* y *trans*, cada una de las cuales tiene un origen evolutivo independiente. Las

defensinas vegetales pertenecen a la subfamilia *cis*, donde dos disulfuros paralelos unen la cadena β final a la α-hélice, mientras que los miembros de la superfamilia trans tienen dos disulfuros análogos que apuntan en direcciones opuestas desde la cadena β final y, por lo tanto, se unen a diferentes elementos de la estructura secundaria (Figura 3) (Shafee et al. 2017).

A pesar de que han sido mayoritariamente aisladas de semillas, estudios de localización de proteínas y expresión génica muestran que pueden encontrarse en todos los tejidos vegetales, incluyendo frutas, flores, polen, hojas, cotiledones, raíces, tubérculos (Carvalho y Gomes 2009). Algunas defensinas se expresan exclusivamente en partes muy específicas de un tejido mientras que otras se expresan constitutivamente en una amplia variedad de tejidos. La expresión constitutiva de estos genes en los tejidos mencionados constituye una defensa primaria de la planta contra el ataque de microorganismos (Carrasco et al. 1981). Además, la producción de estas moléculas puede ser inducida también por otros factores, como daño mecánico, frío (Koike et al. 2002), salinidad o sequía (Do et al. 2004). El péptido señal ubicado en el extremo amino-terminal, la ausencia de alguna señal de retención conocida, y estudios de inmunolocalización, indican que son secretadas al exterior celular(De Coninck et al. 2013). Sin embargo, se ha encontrado en tabaco, una defensina específica de tejidos florales con localización vacuolar (Lay et al. 2003).

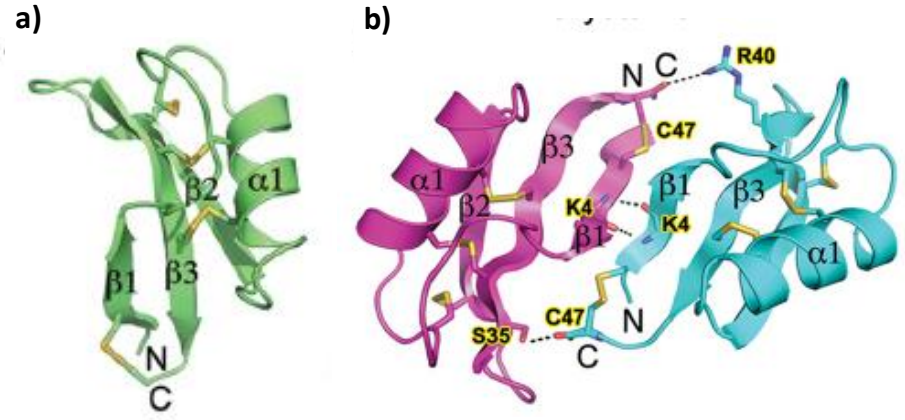


Figura 2. Estructura cristalina de NaD1, defensina de *Nicotiana alata*. a) representación esquemática de la forma cristalina monomérica. Los enlaces disulfuro se muestran como barras. b) representación esquemática de la forma cristalina dimérica B. Los residuos involucrados en la interfaz del dímero se muestran como barras, y los enlaces de hidrógeno y los puentes salinos se muestran en negro como líneas de puntos. Adaptado de Lay et al. 2012.

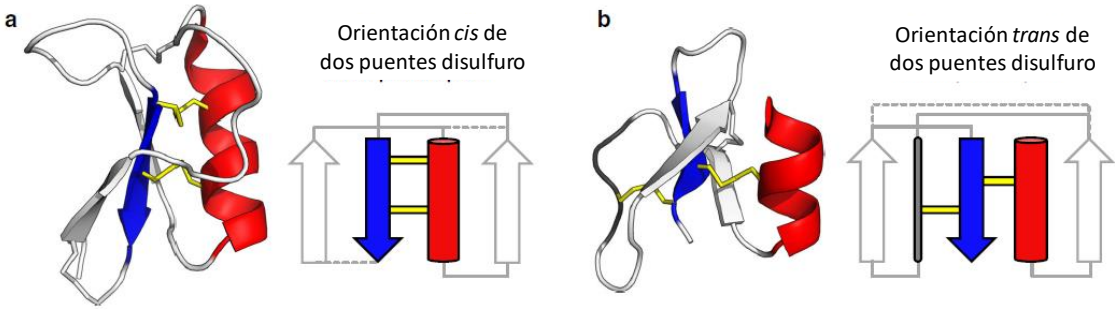


Figura 3. Arquitectura de *cis*- y *trans* defensinas. a) La defensina vegetal NaD1 (PDB: 1MR4) es una *cis*-defensina típica en la que los dos puentes disulfuro más conservados (amarillo) de la hoja β final (azul) apuntan en la misma dirección y se unen a la α-hélice (rojo). b) La β-defensina humana HBD-1 (PDB: 1IJV) es una *trans*-defensina típica en la que los dos disulfuros de la hoja β final apuntan en direcciones opuestas, por lo tanto, se unen a diferentes elementos de la estructura secundaria. Los enlaces S-S no conservados se representan como líneas discontinuas en los diagramas de estructura secundaria. Adaptado de Shafee et al. 2017.

La mayoría de las defensinas vegetales conocidas presenta actividad *in vitro* contra una gama de microorganismos, especialmente hongos y levaduras (Carvalho y Gomes 2009; Carvalho y Gomes 2011; Parisi et al. 2018) (Tabla 1). Es interesante destacar que algunas defensinas vegetales, como HsAFP1 y RsAFP2 presentan actividad contra patógenos humanos como *Candida* spp. y *Aspergillus* spp. (Thevissen et al. 2007), demostrándose además que HsAFP1 producida de forma recombinante inhibe la formación de biofilms de *C. albicans* (Vriens et al. 2015). El carácter catiónico de las defensinas vegetales hace pensar que las mismas, al igual que otros AMPs podrían interactuar con componentes de la membrana plasmática de microorganismos, cargada negativamente. El modo de acción más ampliamente aceptado apunta a la interacción de estos péptidos con componentes específicos de membranas y paredes celulares de hongos, excluyendo la hipótesis de una interacción inespecífica entre las defensinas y los fosfolípidos de membrana ocasionando la formación de poros, que llevaría a la desintegración de la membrana, lisis y muerte celular (Spelbrink et al. 2004, De Coninck et al. 2013). Luego de la interacción con su molécula blanco, las defensinas o bien permanecen en la superficie celular e inducen la muerte celular a través de la inducción de una cascada de señales como se ha detectado para RsAFP2 (Thevissen et al. 2012), o son internalizadas e interactúan con dianas intracelulares, como es el caso de NaD1 (Van der Weerden et al. 2008) y MtDef4 (Sagaram et al. 2013). Los mecanismos propuestos para la internalización incluyen la translocación de la membrana (es decir, la permeabilización transitoria de la membrana y la formación de poros asistida por lípidos), la internalización mediada por receptor y la

translocación o permeabilización no específica de membranas. La producción de especies reactivas del oxígeno (ROS) parecerían jugar un rol en la muerte celular mediada por las defensinas, sugiriendo que la permeabilización de las membranas podría ser un efecto secundario en lugar de la clave de la destrucción microbiana, ya que se produce en concentraciones muy encima de la concentración requerida para la inhibición del crecimiento (Vriens et al. 2014).

Se ha observado que la permeabilización de la membrana no es el único mecanismo de internalización de defensinas. Es el caso de MtDef4, que parece utilizar mecanismos de inhibición del crecimiento diferentes en dos hongos Ascomicetes: *Fusarium graminearum* y *Neurospora crassa*. En el caso de *F. graminearum*, la permealización de membrana es necesaria para su actividad, pero no para *N. crassa*, donde la internalización de MtDef4 es dependiente de energía e implica endocitosis (El-Mounadi et al. 2016). En relación a los blancos intracelulares, se observó para Psd1 de *Pisum sativum* que esta defensina es capaz de interaccionar con una ciclina F, localizada en el núcleo, causando interferencias en el ciclo celular del hongo *N. crassa*. La interferencia con el ciclo celular fue comprobada además usando células de rata como sistema modelo (Lobo et al. 2007). Este descubrimiento, asociado al hecho de que las defensinas vegetales no parecen presentar efectos tóxicos sobre las células de mamíferos (Thevissen et al. 2004) apunta a una posible aplicabilidad terapéutica, no sólo en el tratamiento de dolencias fúngicas humanas, sino también en el combate de algunos tipos de cáncer que presentan como causa primaria la expresión descontrolada de ciclinas (Thomma et al. 2001).

En la última década, la sobreexpresión de defensinas mediante transgénesis ha resultado en plantas con resistencia incrementada contra varias enfermedades fúngicas (Gao et al. 2000; Portieles et al. 2010). Esto, junto a la observación de que los genes de defensinas son inducidos por un amplio rango de estreses bióticos, apoya la teoría de que estos péptidos tendrían un rol *in vivo* en la respuesta de defensa de la planta contra el ataque de microorganismos. Se ha postulado que las defensinas vegetales están relacionadas además con la regulación del crecimiento, el desarrollo y la fertilización (De Coninck et al. 2013).

Tabla 1. Lista de defensinas vegetales y patógenos susceptibles (adaptado de Parisi et al. 2018)

Nombre	Especie	Tejido	Actividad antimicrobiana
NaD1	*Nicotiana alata*	flores	*Aspergillus nidulans, Botrytis cinerea, Candida albicans, Colletotrichum graminicola, Cryptococcus gattii, Cryptococcus neoformans, Fusarium graminearum, F. oxysporum, Puccinia coronate, P. sorghi, Saccharomyces cerevisiae, Thielaviopsis basicola, Verticillium dahliae*
RsAFP2	*Raphanus sativa*	semillas	*Alternaria longipes, A. solani, A.brassicola, Ascochyta pisi, Aspergillus flavus, B. cinerea, C. albicans, C.krusei. C. glabrata, Cercospora beticola, Cladosporium sphaerosperm, Colletotrichum lindemuthianum, Fusarium culmorum F. graminearum, F. solani, F. oxysporum, Leptosphaeria maculans, Mycosphaerella fijiensis, Rhizoctonia solani, Sclerotinia sclerotiorum Septoria nodorum, Septoria tritici, Trichoderma hamatum, Trichophyton mentagrophytes, Trichoderma viride, Verticilium alboatrum, V. dahlia, Venturia inaequalis*
Psd1	*Pisum sativum*	semillas	*Aspergillus niger, Avicularia. versicolor, Fusarium solani, F. moniliformae, F. oxysporum, Neurospora crassa, S. cerevisiae, T. mentagrophytes*
HsAFP1	*Heuchera sanguinea*	semillas	*A. flavus, B. cinerea, C. albicans, C. glabrata, Candida krusei, Cladsporium sphaerospermum, F. culmorum, F. solani, Leptosphaeria maculans, N. crassa, Penicllium digitatum, Septoria tritici, Verticillium albo-atrum, T. viride*
PvD1	*Phaseolus vulgaris*	semillas	*C. albicans, Candida tropicalis, F. oxysporum, F. solani, Fusarium laterithium, Kluyveromyces marxiannus, S. cerevisiae*
MsDef1	*Medicago sativa*	semillas	*F. graminearum, N. crassa V. dahlia*
MtDef4	*Mendicago trucatula*	semillas	*F. graminearum, N. crassa, Puccinia tritici*
DmAMP1	*Dahlia merkii*	semillas	*Alternaria brassicicola, A. flavus, B. cinerea, C. albicans, C. glabrata, C. krusei, C. sphaerospermum, F. solani, F. culmorum, L. maculans, N. crassa, P. digitatum, S. cerevisiae, S. tritici, Trichoderma viride, V. albo-atrum*
NsD1, NsD2	*Nigella sativa*	semillas	*A. niger, B. cinerea, Bipolaris sorokiniana, F. culmorum, F. graminearum, F. oxysporum*
PhD1, PhD2	*Petunia hybrida*	flores	*B. cinerea, F. oxysporum*
Zm-ESR6	*Zea mays*	semillas	*B. cinerea, F. oxysporum, Plectosphaerella cucumerina*
PTH1	*Solanum tuberosum*	tubérculos	*Clavibacter michiganesis, F. solani, Pseudomonas solanacearum*
TvD1	*Tephrosia villosa*	semillas	*Alternaria helianthi, B. cinerea, Curvularia sp, Fusarium moniliforme, F. oxysporum, Pheaoisariopsis personata*
Vv-AMP1	*Vigna unguiculata*	Semillas	*Alternaria longipes, B. cinerea, F. graminearum, F. oxysporum, F. solani, V. dahliae*

ESNAQUINAS

Las esnaquinas constituyen un grupo de AMPs que se ha encontrado únicamente en plantas y presentan un amplio espectro de actividades antimicrobianas *in vitro* (Oliveira-Lima et al. 2017). Esta familia ha sido definida a partir de dos péptidos aislados de tubérculos de papa: Esnaquina-1 (StSN1) (Segura et al. 1999) y Esnaquina-2 (StSN-2) (Berrocal-Lobo et al. 2002). Estos péptidos pertenecen a un grupo de proteínas codificadas por genes snakin/GASA y se caracterizan por tener tres dominios definidos: un péptido señal putativo de 18-29 residuos, una región variable (alta divergencia en tamaño y composición de secuencia) y una región C-terminal o dominio GASA, de aproximadamente 60 aminoácidos y 12 residuos cisteína en posiciones conservadas (Nahirñak et al. 2012b). En relación a la estructura de los genes que codifican StSN1 y StSN2, el primero está formado por dos exones y un intrón mientras que el segundo posee tres exones y dos intrones. Cuando se analizaron estos genes, en varios genotipos de *Solanum tuberosum*, se observó que ambos se encuentran altamente conservados en la especie, sugiriendo que podrían estar relacionados con alguna función biológica importante (Meiyalaghan et al. 2014). El gen de la Esnaquina-2 es parálogo del de la StSN1, con una baja similitud de secuencia (38%). Genes ortólogos han sido detectados en otras plantas, incluyendo tomate (Shi et al. 2002) y *Arabidopsis thaliana* (Herzog et al. 1995).

La identificación de los genes Snakin/GASA en especies distantes sugiere que estas proteínas cumplen funciones esenciales en las plantas. Varias evidencias indicarían que la mayoría de estos genes son regulados por hormonas y participan en las vías de señalización hormonal participando en los procesos de desarrollo y la tolerancia al estrés a través de su participación en la homeostasis redox y hormonal (Figura 4) (Nahirñak et al. 2012b). La asignación de función para algunos miembros de este grupo de proteínas ha sido basada en análisis de perfiles de expresión, caracterización fenotípica de mutantes y plantas transgénicas. Algunas de ellas estarían involucradas en procesos de arresto o promoción de la elongación y división celular y en procesos de desarrollo como formación de raíces, tiempo de floración y maduración de frutos. Se ha mostrado además que algunas están involucradas en respuestas a estrés abiótico (tolerancia a calor, sal y estrés oxidativo) y de defensa frente a patógenos (Nahirñak et al. 2012b). Análisis de expresión de los genes de StSN1 y StSN2, en diferentes tejidos ha mostrado resultados no del todo coincidentes, usando diferentes técnicas: northern blot (Segura et al. 1999; Berrocal-Lobo et al. 2002), RNAseq (Meiyalaghan et al. 2014) y qPCR (Nahirñak et al. 2016). Las diferencias encontradas, podrían deberse a diferentes cultivares utilizados en los análisis, niveles de ploidía de los mismos y condiciones de tratamiento/crecimiento; además de las diferencias en las sensibilidades de las técnicas usadas. Con respecto a la inducción de estos

genes frente al tratamiento con hormonas y al estrés biótico, en los tres trabajos mencionados anteriormente, los resultados tampoco fueron coincidentes. Meiyalaghan et al. (2014) no detectaron inducción significativa en ninguno de los tratamientos, ni para StSN1 ni para StSN2. Segura et al. (1999) tampoco habían encontrado que StSN1 respondiera a estrés biótico o al tratamiento con hormonas, mientras que Berrocal Lobo et al. (2002) detectaron activación de los mismos durante la infección por el hongo *Botrytis cinerea* o la inducción de heridas mientras que su expresión disminuye durante la infección con bacterias patógenas de papa (Segura et al. 1999; Berrocal-Lobo et al. 2002).

Por otro lado, StSN1 y StSN2 de papa, únicos péptidos de la familia snakin/GASA aislados hasta ahora, tienen un amplio espectro de actividad antimicrobiana *in vitro*, y aumentan la resistencia a patógenos de relevancia comercial cuando son sobreexpresados en papa (Almasia et al. 2008) y en tomate (Balaji et al. 2012), sugiriendo que poseen además actividad antifúngica y antibacteriana *in vivo*. La sobreexpresión constitutiva de SN1 no parece alterar el fenotipo de la planta, pero su silenciamiento afecta la división celular, el metabolismo de la hoja y la composición de la pared celular (Nahirñak et al. 2012a). Además, el silenciamiento de genes homólogos a esnaquina-2 en *Nicotiana benthamiana* aumenta la susceptibilidad a *Clavibacter michiganensis* (Balaji y Smart 2011). Además, genes de tipo Snakin/GASA han sido expresados de forma heteróloga y se ha evaluado su actividad *in vitro*. Así, una esnaquina de *Capsicum annum* induce resistencia a la infección por nemátodos (Mao et al. 2011) y StSN1 expresada en *Escherichia coli* mostró una fuerte actividad contra la bacteria *Clavibacter michiganensis* subsp. *sepedonicus* y los hongos *Colletotrichum coccoides* y *B. cinerea* (Kovalskaya y Hammond 2009). La producción recombinante de una SN2 de tomate, en *E. coli*, produjo un péptido con actividad antibacteriana y antifúngica, explicada por una permeabilización de las membranas de los microorganismos (Herbel et al. 2015). StSN1 también fue producida en *Pichia pastoris*, estudiando el mecanismo de acción contra la bacteria *E. coli*. Estos ensayos demostraron que este péptido puede actuar afectando la integridad de las membranas externa y citoplasmática de *E. coli* (Kuddus et al. 2016).

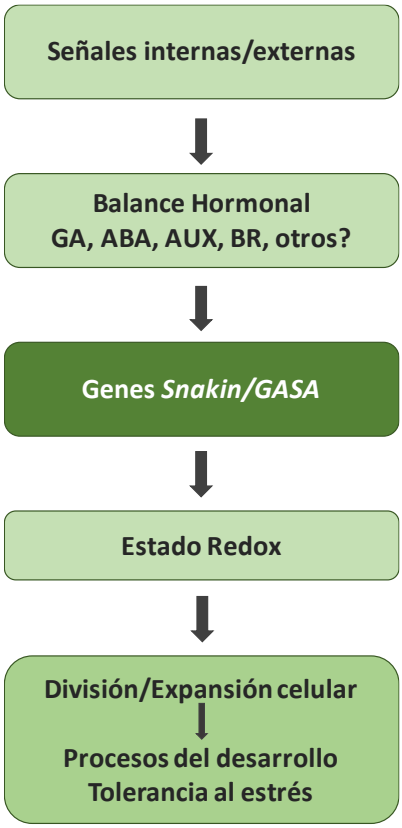


Figura 4. Modelo simplificado del papel de las proteínas Snakin/GASA. La expresión de los genes Snakin/GASA se modula en respuesta a diferentes hormonas y/o señales externas/internas. Las proteínas codificadas regulan la homeostasis redox a través de su dominio conservado rico en cisteína y, en consecuencia, participan en los procesos de desarrollo y la tolerancia al estrés mediante el ajuste del equilibrio de la promoción y la inhibición del crecimiento celular. Adaptado de Nahirñak et al. 2012b.

Estudios sobre la estructura terciaria de las esnaquinas y las relaciones estructura-función son aún muy escasos. El hecho de que el número y posiciones de las cisteínas se conserven en la evolución de estos péptidos sugeriría que estos residuos juegan un rol central y podrían ser importantes en determinar sus funciones bioquímicas y estructurales. Porto y Franco (2013) predijeron la estructura tridimensional StSN1 a través de un modelo *ab initio* y comparativo en combinación con un predictor de puentes disulfuro. La determinación de rayos X de StSN1 (Yeung et al. 2016) mostró una estructura con un motivo *helix-turn-helix* estabilizado por seis enlaces disulfuro entre cisteínas específicas (Figura 5). El modelo StSN1 de Porto y Franco sólo comparte dos enlaces disulfuro con la estructura de rayos X, pero también presentó un motivo *helix-turn-helix*. Esto podría sugerir que una conexión precisa entre pares específicos de cisteína puede no ser necesaria para que el péptido se pliegue al estado nativo y funcional. Existen evidencias que señalan que la forma reducida de los péptidos StSN1 y StSN2 (donde todas las cisteínas están libres; no hay enlaces disulfuro) eran igualmente activos contra los patógenos como las formas nativas (Harris et al. 2014).

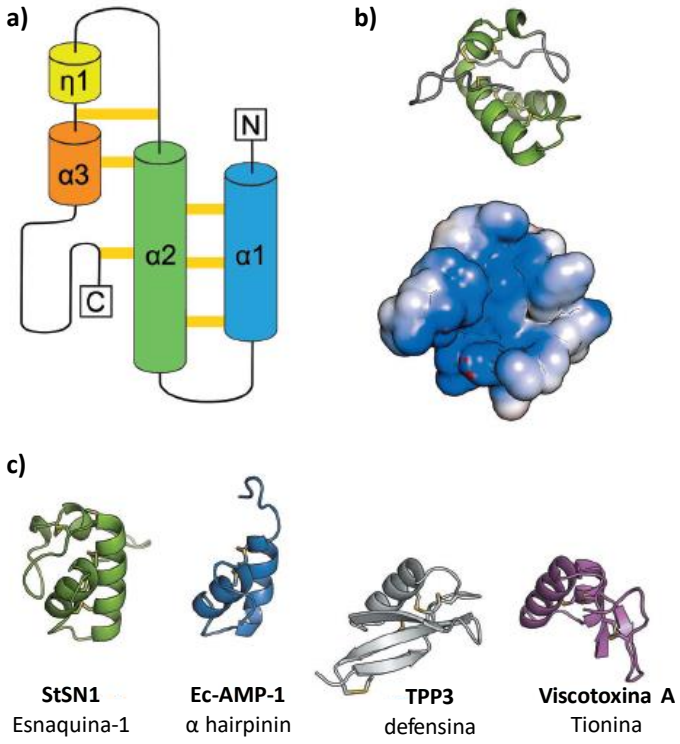


Figura 5. Estructura 3D de esnaquina-1 de papa (StSN1). a) Diagrama de topología mostrando la conectividad de la red troncal de StSN1. Los puentes disulfuro se indican en amarillo. b) Representación en cinta (arriba) y superficie electrostática en la misma orientación (parte inferior), mostrando una hendidura altamente cargada positivamente (azul). c) Se compara la estructura de StSN1 con las estructuras de otros péptidos antimicrobianos ricos en cisteínas. StSN1 (verde), EcAMP1 (horquilla α-helicoidal; azul), TPP3 (defensina vegetal; gris), y viscotoxina A3 (tionina; violeta). Adaptado de Yeung et al. 2016.

Aislamiento y producción de AMPs vegetales

El alto potencial de los AMPs como agentes terapéuticos ha incentivado a los investigadores a encontrar, caracterizar y sintetizar péptidos antimicrobianos eficaces. En la última década, las herramientas bioinformáticas han permitido identificar genes que codifican AMPs en varias especies vegetales. Debido a la mayor disponibilidad de datos moleculares, las plantas cultivadas y las especies modelo han sido el principal blanco. Las especies nativas que se destacan por su importancia en la medicina popular y constituyen la mayor biodiversidad vegetal, continúan siendo poco exploradas debido principalmente a que sus ómicas no están disponibles (Pestana-Calsa et al. 2010). Actualmente, las tecnologías de secuenciado masivo (NGS-*Next Generation Sequencing*), permiten a costos razonables, la caracterización completa y el análisis global de genomas y transcriptos en una célula o tejido, aun sin ninguna información genómica previa. Entre la diversidad de aplicaciones que tienen estas nuevas tecnologías se

incluye el descubrimiento de genes y transcriptos (Lister et al. 2009). Es así que estas tecnologías abren nuevas puertas para encontrar genes que codifiquen péptidos novedosos en los transcriptomas de especies nativas. Koehbach et al. (2013) han combinado datos transcriptómicos y espectrometría de masas para identificar y caracterizar nuevos ciclótidos de la familia Rubiaceae. SuperSAGE, una variante mejorada de SAGE (*serial analysis of gene expression*), en combinación con NGS ha sido utilizado para identificar genes relacionados a defensa de tipo inhibidores de proteasas (Silva et al. 2013). Recientemente se identificaron transcriptos que codifican AMPs en trancriptomas de plántulas de la especie silvestre de la familia Poaceae *Leymus arenarius* (Slavokhotova et al. 2015) y de la maleza, *Stellaria media* (L.) Vill (Caryophyllaceae) (Slavokhotova et al. 2017) mediante RNA-seq y ensamblado *de novo*.

Los métodos usados para recuperación y producción de AMPs han sido: aislamiento directo de fuentes naturales, síntesis química y expresión recombinante. Los dos primeros son muy útiles para producción en pequeña escala, pero presentan limitantes relacionadas con el costo y el bajo rendimiento, por lo que el método usado más ampliamente para producción a mediana y gran escala es la expresión heteróloga. En las últimas décadas, varios sistemas (organismo productor/vector) se han desarrollado para alcanzar una producción costo-efectiva y a gran escala de varias proteínas, en diversos hospederos como bacterias, levaduras, hongos y plantas. Entre los sistemas microbianos, *Escherichia coli, Pichia pastoris* y *Saccharomyses cerevisiae*, son los más populares para expresar proteínas recombinantes, debido a sus altas tasas de crecimiento, requerimientos de bajo costo, vasto conocimiento sobre su genética y fisiología. Se han generado varios sistemas cepa/vector con funciones alteradas que favorecen la producción como ser cambios en ARNt, introducción de modificaciones postraduccionales, etc (Thevissen et al. 2007; Padovan et al. 2010a; Silva et al. 2011). Existen varios antecedentes de producción de AMPs en sistemas bacterianos y en levaduras (Vijayan et al. 2008; Kovalskaya y Hammond 2009; Kant et al. 2009; de Breer y Vivier 2011; Kovaleva et al. 2011; Herbel et al. 2015; Kuddus et al. 2016, Guillén-Chable et al. 2017). A escala de laboratorio, la producción en *E. coli*, es la primera opción en cuanto a facilidad de manejo y costo, para la caracterización de sus propiedades, a partir de las cuales se identifican los posibles usos. Finalmente, una vez identificado y caracterizado un nuevo péptido con propiedades biológicas interesantes, su desarrollo para usos en agricultura y usos terapéuticos en medicina presenta algunos desafíos relacionados a su posible toxicidad intrínseca, su estabilidad y costos relacionados a su producción a gran escala.

Objetivos

Objetivo General

El constante desarrollo de resistencia por parte de algunos microorganismos patógenos y el alto potencial de los AMP vegetales como agentes terapéuticos en agricultura y en medicina, ha estimulado su búsqueda, caracterización y síntesis/producción. Como aporte a esta línea de investigación nos hemos planteado el siguiente objetivo:

> Identificar y caracterizar péptidos antimicrobianos de tipo defensinas y esnaquinas presentes en especies vegetales de la flora nativa de Uruguay, evaluando su actividad antimicrobiana *in vitro*, con vistas al desarrollo de nuevos agentes de control de enfermedades.

Objetivos Específicos

1. Aislar genes de defensinas y esnaquinas de tres especies de la flora nativa: *Peltophorum dubium* (ibirapitá)*, Erythrina crista-galli* (ceibo) y *Maytenus ilicifolia* (congorosa)*.*
2. Caracterizar de forma estructural los genes aislados y sus respectivos péptidos.
3. Producir un péptido de tipo esnaquina y otro de tipo defensina mediante expresión heteróloga.
4. Evaluar la actividad antimicrobiana *in vitro* de la defensina y esnaquina producidas de forma recombinante.

En relación a los objetivos planteados nos planteamos las siguientes preguntas e hipótesis:

- ¿Qué características estructurales tendrán los péptidos codificados en los genomas de las plantas nativas seleccionadas? En este sentido, consideramos que mostrarán patrones estructurales comunes a otros péptidos antimicrobianos reportados en la literatura, pero con una mayor diversidad de secuencia.

- ¿Los genes vegetales aislados, pueden ser expresados de forma heteróloga en un sistema procariota como *E. coli,* obteniendo el péptido de forma soluble y sin perder sus propiedades funcionales? ¿Contra qué microorganismos presentarán actividad? De acuerdo con los antecedentes, creemos que es posible producirlos en esta bacteria de forma mayormente soluble y con actividad, usando un sistema cepa/vector que permita modificaciones postraduccionales como la formación de puentes disulfuro y que evite al menos en forma parcial la formación de agregados del péptido en los cuerpos de inclusión y la posible toxicidad de los péptidos. Los péptidos producidos de forma recombinante presentarán actividad contra un amplio rango de patógenos, pero serán más efectivos para algunos de ellos, observando diferencias de actividad, dependiendo del péptido y el microorganismo analizados.

- Alguno de los péptidos identificados, ¿tendrá características originales a nivel de su secuencia y/o de su perfil de actividad? ¿Puede considerarse como un buen candidato para el desarrollo de un nuevo bioactivo? Nuestra hipótesis es que al menos algunos de los péptidos identificados serán estructural y/o funcionalmente originales y podrán ser tenidos en cuenta para la selección de futuros candidatos para el desarrollo de productos útiles para el diagnóstico o control eficaz de enfermedades. Para ello deberán ser estudiados con mayor profundidad para comprender sus mecanismos de acción, y su posible interacción con células vegetales o de mamíferos.

Estrategia de Investigación

En este trabajo se buscaron genes que codifiquen AMPs novedosos, particularmente defensinas y esnaquinas, en el genoma de especies poco conocidas como son las plantas nativas. Se buscaron plantas usadas en la medicina popular, eligiendo *P. dubium, E. crista-galli* y *M. ilicifolia* conocidas por propiedades: anti-infectivas, anti-ulcerogénicas, regenerativas, antitumorales. Para la selección de las especies a analizar, se tuvo en cuenta además la abundancia de secuencias de AMPs de la familia a la que pertenecen en las bases de datos. *P. dubium* y *E. crista-galli* son leguminosas, familia para la que se encuentra abundante información de secuencias, existiendo genomas completamente secuenciados (por ej: *Medicago truncatula* y *Glycine max*). *M. ilicifolia* pertenece a la familia Celastraceae, para la cual no existe información en las bases de datos. Para asegurar la factibilidad de alcanzar los objetivos planteados, dentro de las especies no exploradas, se eligieron dos plantas de una familia para la cual ya hay antecedentes de éxito y otra más alejada filogenéticamente a partir de la cual se espera encontrar secuencias más novedosas (menor similitud).

Para cumplir con el objetivo 1 se siguieron dos estrategias. La primera se basó en el rastreo de genes de defensinas y esnaquinas putativos en los genomas de las especies seleccionadas, por amplificación con cebadores degenerados. La segunda estrategia se basó en el ensamblado de transcriptomas (secuenciado masivo-RNA-seq) para la identificación de transcriptos que codifiquen proteínas con motivos cisteínas típicos de AMPs.

La primera estrategia, implica rastrear el genoma de las plantas usando *primers* dirigidos a zonas conservadas de los AMPs, a partir de secuencias conocidas de otras especies. Las regiones conservadas se identifican a partir de alineamientos múltiples utilizando secuencias anotadas en las bases de datos. Alternativamente, pueden emplearse cebadores de bibliografía utilizados para la amplificación de estos genes en otras plantas. Usando diferentes combinaciones de *primers*, ciclos anidados y diferentes condiciones de reacción de amplificación: variación de temperaturas de *annealing* y concentraciones de $MgCl_2$, tiempos de elongación, etc, se busca encontrar aquellas combinaciones que permitan obtener las secuencias de interés. Si bien esta metodología se limita a encontrar secuencias con cierta similitud en zonas más o menos conservadas, debido a la variabilidad de las secuencias primarias de estos AMPs, se espera encontrar genes con variaciones en la secuencia que queda limitada ente las zonas conservadas. Con esta estrategia se aisló un gen de esnaquina del genoma de *P. dubium* (utilizando *primers*

degenerados) y una defensina de *E. crista-galli* (usando cebadores específicos para otra especie leguminosa). Ambos genes, *PdSN1* y *EcgDf1* fueron clonados y su secuencia fue confirmada. Esta estrategia fue la inicialmente utilizada y no fue todo lo satisfactoria esperable, especialmente para *M. ilicifolia*, para la que no se obtuvieron secuencias del tipo AMP.

Con la finalidad de aumentar el número de secuencias de defensinas y esnaquinas putativas, se incluyó en este trabajo la segunda estrategia (RNA-seq). Más allá de poder encontrar más secuencias, permite además tener un panorama más amplio del potencial de estas plantas como productores de AMPs. Esta metodología tiene la ventaja de que se levantan las restricciones de la metodología anterior dadas por los *primers* y/o condiciones. Por otro lado, permite ver lo que se esté expresando en la muestra analizada por lo que los resultados dependerán del transcriptoma analizado es decir de qué genes se expresen en determinado tejido, estado del desarrollo, estímulo hormonal, situaciones de estrés biótico o abiótico. Una forma de seleccionar los transcriptomas a secuenciar, es analizar la actividad de extractos peptídicos de diferentes tejidos/condiciones como guía para ver si alguno de ellos presenta actividad que sobresalga. La otra alternativa es secuenciar el transcriptoma de semillas germinadas ya que varios antecedentes reportan a las semillas como productores de AMPs; o hacer pool de varios transcriptomas para tener un panorama más amplio. El objetivo es obtener un listado de secuencias que codifiquen AMPs putativos; no se busca cuantificar la expresión de estos genes de acuerdo a la condición particular, por lo que no es necesario comparar los resultados del tejido/condición de expresión contra un blanco. Por razones de costos para esta estrategia además de congorosa se eligió sólo una de las especies leguminosas (ibirapitá).

Con respecto al ensamblado de secuencias provenientes de RNA-seq, para el mismo se puede usar un genoma de referencia. Por no contar con el genoma de la propia especie, en ibirapitá se podrían usar genomas de otras leguminosas (por ejemplo, *Glycine max*, *Medicago truncatula*). En congorosa tendríamos que seleccionar un genoma por fuera de su familia, donde el más cercano podría ser *Populus trichocarpa* del orden Malpighiales. Actualmente se pueden analizar transcriptos aun sin ninguna información genómica previa usando ensamblado *de novo*. Una vez ensamblados, la identificación de secuencias de interés puede hacerse por homología y también, en el caso de AMPs, utilizado expresiones regulares basadas en los motivos cisteína que definen cada tipo de AMP. Para el ensamblado de transcriptomas y la búsqueda de motivos cisteína se contó con la colaboración de Pablo Smircich y Santiago Radío. Finalmente, para asegurar que las secuencias realmente existen en el genoma de la planta, descartando artefactos de ensamblado, las mismas deben validarse mediante PCR a partir de ARN o ADN genómico de la planta en cuestión. Esta estrategia permitió efectivamente ampliar el número

de secuencias encontradas en los genomas de las especies *P. dubium* y *M. ilicifolia* a partir de transcriptomas ensablados *de novo.* Varios genes de tipo esnaquina y defensina fueron validados mediante PCR a partir de ADN genómico, lo que permite, además de ratificar la fiabilidad del ensamblado, confirmar la secuencia codificante del transcripto y cotejar la presencia y la ubicación de los intrones.

Para la caracterización estructural de los péptidos deducidos de los genes aislados (objetivo 2), se usaron herramientas bioinformáticas para la predicción de características como peso molecular y punto isoeléctrico, presencia de péptido señal, localización subcelular y distribución de residuos cisteína. Además, se realizaron análisis de similitud con péptidos reportados en las bases de datos (BLAST) y alineamientos múltiples para evaluar la conservación de ciertos aminoácidos (por ejemplo, presencia de motivos cisteína) y para la construcción de árboles basados en matrices de distancias para el análisis del grado de similitud de las secuencias. Previamente, para los genes *PdSN1* y *EcgDf*, que fueron clonados desde ATG hasta stop a partir de ADN genómico, se debe confirmar la secuencia en la región de los cebadores y clonar la secuencia completa del marco de lectura abierto, sin la interrupción de intrones. Para ello, el ADN copia de ambos se obtuvo utilizando la técnica RACE (*Rapid Amplification of cDNA Ends*). Para estos péptidos se predijo además la estructura 3D, utilizando herramientas de modelado y dinámica molecular, gracias a la colaboración del Dr. Pablo Dans. Varios péptidos deducidos del transcriptoma de *P. dubium* fueron validados en ADN genómico, en estos casos también se pudo conocer la estructura génica (exones-intrones).

Los objetivos 3 y 4 se realizaron para los genes *PdSN1* y *EcgDf1*. Para la expresión heteróloga de los mismos se consideraron dos posibles opciones: un sistema procariota (*E. coli*) y un sistema eucariota (*Pichia pastoris*). *E. coli* es el sistema más popular utilizado para expresar proteínas recombinantes debido a su rápida tasa de crecimiento, bajo costo y gran disponibilidad de vectores de expresión. Además, se ha aplicado a innumerables estudios de expresión de AMP de diferentes organismos. La limitación de su uso es que pueda realizar modificaciones postraduccionales necesarias para la estabilidad y la actividad biológica de ciertos péptidos. Para superar esta limitación, se pueden usar cepas especiales que permitan esas modificaciones. Alternativamente, la levadura *Pichia pastoris* representa uno de los hospederos eucariotas más comunes utilizado para la expresión de proteínas heterólogas. En un principio decidimos probar con diferentes cepas de *E. coli*, y considerar el sistema *P. pastoris* GS115/pPICZα, si los resultados con *E. coli* no eran los esperados. Los péptidos maduros PdSN1 y EcgDf fueron producidos entonces en el sistema procariota, utilizando tres diferentes cepas de *E. coli* y el vector pET102/D que incluye la fusión N-terminal del péptido con la proteína

tiorredoxina (disminuye la formación de agregados del péptido en los cuerpos de inclusión) y un sitio de reconocimiento de la proteasa enteroquinasa para la escisión del péptido recombinante de la tiorredoxina. Además, tiene una etiqueta C-terminal 6xHis que permite la purificación mediante cromatografía de afinidad. Ambos péptidos fueron obtenidos de forma mayormente soluble.

La actividad antimicrobiana de los péptidos recombinantes fue evaluada *in vitro* contra varias especies de bacterias y hongos, tanto patógenos de humanos como de algunos cultivos de importancia comercial. La inhibición del crecimiento puede ser detectada en ensayos de difusión en placa o mediante diluciones seriadas y medidas de micro-espectrofotometría. El segundo nos permite construir una curva de inhibición, usando menores cantidades de péptidos, por lo que fue el método finalmente usado para la evaluación de la actividad antimicrobiana. Estos ensayos permitieron detectar actividad en ambos péptidos contra varios patógenos.

Capítulo 1

Identificación y caracterización estructural y funcional de una esnaquina de ibirapitá

Este capítulo incluye la búsqueda de genes de tipo esnaquina en los genomas de ibirapitá, congorosa y ceibo, el clonado de un gen de *P. dubium* con similitud con esnaquinas de tipo I (*PdSN1*) por amplificación con *primers* degenerados y un artículo publicado en la revista Amino Acids, que describe el aislamiento de *PdSN1*, su expresión relativa en dos tejidos y su expresión heteróloga en *E. coli*. En el mismo, además, se predijo la estructura 3D del péptido nativo y recombinante y se evaluó la actividad antimicrobiana del péptido recombinante purificado, contra un grupo de microorganismos patógenos. En el artículo, los análisis de expresión génica fueron llevados a cabo con la colaboración de la Mag. Mariana Barraco-Vega, que realizó los ensayos de RT-qPCR. La predicción de la estructura 3D de PdSN1 nativo y recombinante fue realizada por el Dr. Pablo Dans que utilizó técnicas de modelado *ab initio* y por homología, sumado a simulaciones de dinámica molecular para dichas predicciones.

Búsqueda de genes de tipo esnaquina en los genomas de ibirapitá, congorosa y ceibo

Para el aislamiento de genes de tipo esnaquina en las especies vegetales seleccionadas, se buscaron secuencias de esnaquinas en las bases de datos (GenBank, Uniprot, Phytozome) y se alinearon para identificar regiones conservadas (Figura 1). Para el alineamiento se usaron todas las secuencias identificadas como esnaquinas hasta ese momento. Se diseñaron *primers* degenerados (Tabla 1), para las dos subfamilias (subfamilia I a la que pertenece StSN1 y subfamilia II, StSN2) donde el primer F en principio se diseñó desde ATG inicial y el R en los últimos 6 aminoácidos. Además, se incluyeron *primers* en zonas conservadas internas, dentro del péptido maduro (Figura 2). Se utilizaron varios pares de *primers* y diferentes condiciones (gradientes de temperatura, concentraciones de $MgCl_2$) a partir de ADN genómico. Se usaron varias estrategias de amplificación, como enriquecimiento dentro de bandas en secuencia de interés usando *primers* anidados. Además, se usaron *primers* específicos diseñados para la secuencia de *Solanum commersonii* (Tabla 1).

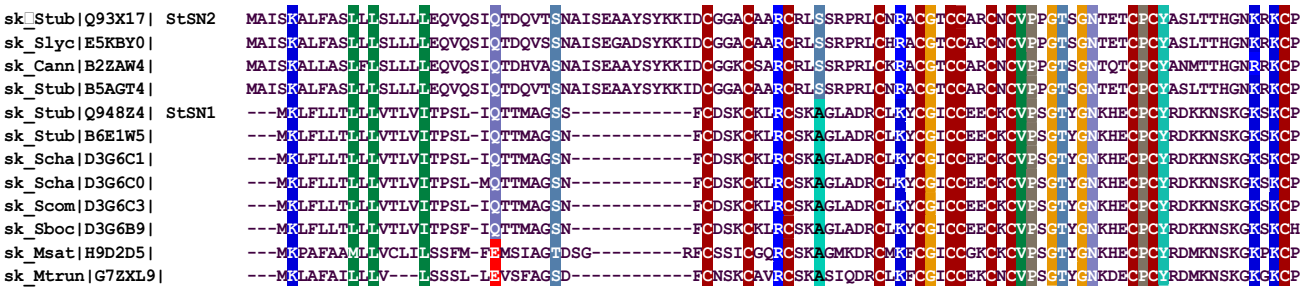


Figura 1. Alineamiento de esnaquinas de tipo 1 y 2 obtenidas de las bases de datos. Stub: *Solanum tuberosum*; Slyc: *Solanum lycopersicum*; Cann: *Capsicum annum*; Scha; *Solanum chacoense*; Scom: *Solanum commersonii*; Msat: *Medicago sativa*; Mtrun: *Medicago truncatula*

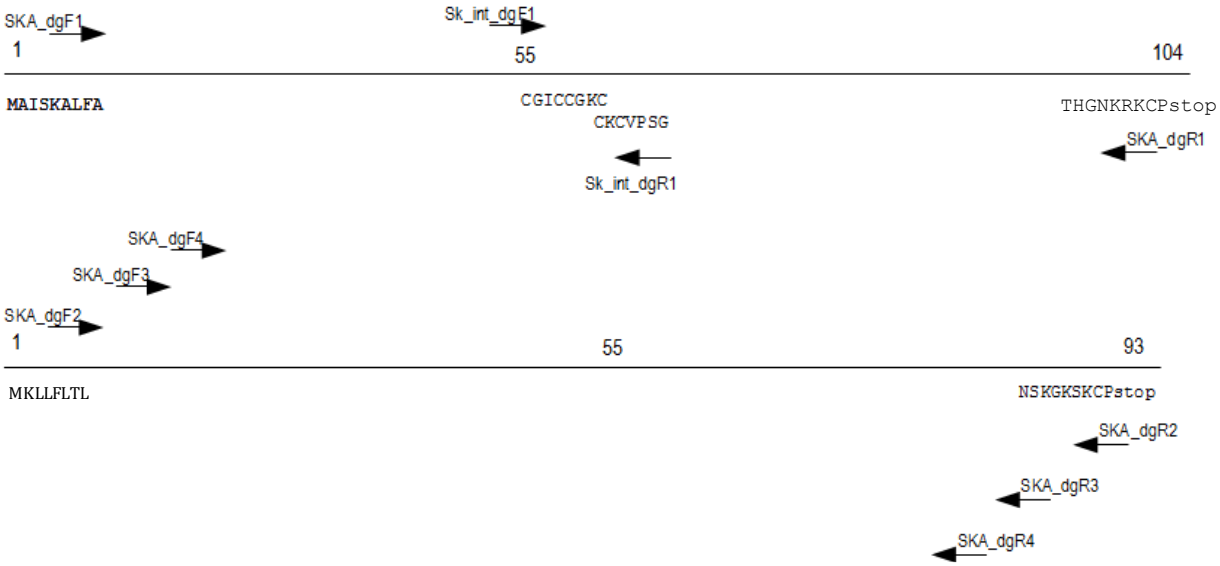


Figura 2. Esquemas de ubicación de los distintos *primers* degenerados utilizados para amplificación de secuencias tipo esnaquinas

Los amplicones encontrados fueron clonados y varios clones (de cada banda de amplificación) fueron secuenciados. Entre los clones analizados, se detectaron tres, provenientes de la amplificación de ADN de *Peltophorum dubium* con *primers* degenerados anidados (Figura 3), con alta similitud con genes de esnaquinas de tipo I. Los 3 clones correspondieron a la misma secuencia genómica (*PdSN1*), uno de ellos con la secuencia desde ATG a stop. Para *M. ilicifolia* y *E. crista-galli* no se encontraron secuencias con similitud con esnaquinas.

Tabla 1. *Primers* degenerados y específicos usados para la búsqueda de genes de tipo esnaquina en los genomas de *P. dubium*, *M. ilicifolia* y *E. crista-galli*

Nombre	Secuencia
SK_int_dgF1	TGTGGRATWTGTTGTGRARAATG
SK_int_dgR1	CCWGAWGGCACACABTTRCA
SKA_dgF1	ATGAAGCTAGTTCTAKYARCTCTGC
SKA_dgR1	TCAAGGGCATTTWGRYTTKC
SKA_dgF2	ATGAAGCCAGCATTTKCARCTMT
SKA_dgR2	TTAAGGGCATTTWGSCTTKCC
SKA_dgF3	TTTGCAGCTMTGCTWYTKGT
SKA_dgR3	TTGCCCTTRGARTTCTTCWTG
SKA_dgF4	TTGTCTCTCTTGTYMTTASCYC
SKA_dgR4	GAGTTCTTCWTGTCYCTRTARC
SKsp_Scom_F1	ATGAAGTTATTTCTATTAACTCTGCTT
SKsp_Scom_R1	TCAAGGGCATTTAGACTTGC

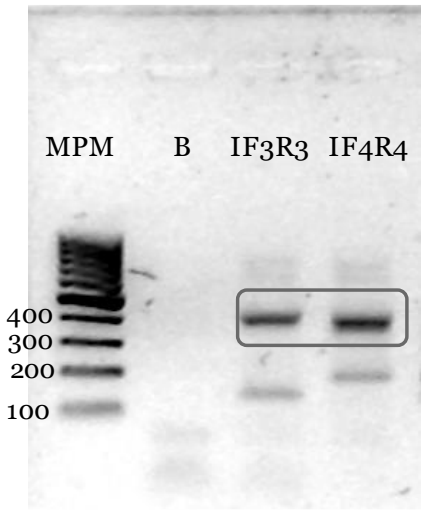


Figura 3. Amplificación por PCR usando *primers* degenerados diseñados para esnaquinas a partir de ADN genómico de ibirapitá. Electroforesis en gel de agarosa 2 %. MPM: marcador de peso molecular 100 pb (GeneRuler 100 bp DNA Ladder; Thermo Scientific). B: blanco sin ADN

Amino Acids
https://doi.org/10.1007/s00726-018-2598-3

ORIGINAL ARTICLE

Antimicrobial and structural insights of a new snakin-like peptide isolated from *Peltophorum dubium* (Fabaceae)

Susana Rodríguez-Decuadro[1] · Mariana Barraco-Vega[2] · Pablo D. Dans[3,4] · Valesca Pandolfi[5] · Ana Maria Benko-Iseppon[5] · Gianna Cecchetto[2,6]

Received: 21 March 2018 / Accepted: 31 May 2018

Abstract

Snakins are antimicrobial peptides (AMPs) found, so far, exclusively in plants, and known to be important in the defense against a wide range of pathogens. Like other plant AMPs, they contain several positively charged amino acids, and an even number of cysteine residues forming disulfide bridges which are considered important for their usual function. Despite its importance, studies on snakin tertiary structure and mode of action are still scarce. In this study, a new snakin-like gene was isolated from the native plant *Peltophorum dubium*, and its expression was verified in seedlings and adult leaves. The deduced peptide (PdSN1) shows 84% sequence identity with potato snakin-1 mature peptide, with the 12 cysteines characteristic from this peptide family at the GASA domain. The mature PdSN1 coding sequence was successfully expressed in *Escherichia coli*. The purified recombinant peptide inhibits the growth of important plant and human pathogens, like the economically relevant potato pathogen *Streptomyces scabies* and the opportunistic fungi *Candida albicans* and *Aspergillus niger*. Finally, homology and ab initio modeling techniques coupled to extensive molecular dynamics simulations were used to gain insight on the 3D structure of PdSN1, which exhibited a helix–turn–helix motif conserved in both native and recombinant peptides. We found this motif to be strongly coded in the sequence of PdSN1, as it is stable under different patterns of disulfide bonds connectivity, and even when the 12 cysteines are considered in their reduced form, explaining the previous experimental evidences.

Keywords Pathogenesis-related peptide · Heterologous expression · *Escherichia coli* · Ab initio and homology modeling · Molecular dynamics simulations

Introduction

Plants and animals own a diverse group of small proteins with antimicrobial activity, so-called antimicrobial peptides (AMPs). These peptides are evolutionarily ancient, being important in the defense against a wide range of

Handling Editor: M. S. Palma.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00726-018-2598-3) contains supplementary material, which is available to authorized users.

✉ Gianna Cecchetto
gianna.cecchetto@gmail.com

[1] Departamento de Biología Vegetal, Facultad de Agronomía, Universidad de la República, Garzón 780, 12900 Montevideo, Uruguay

[2] Departamento de Biociencias, Facultad de Química, Universidad de la República, General Flores 2124, 11800 Montevideo, Uruguay

[3] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona, Spain

[4] Joint BSC-IRB Research Program in Computational Biology, Baldiri Reixac 10-12, 08028 Barcelona, Spain

[5] Universidade Federal de Pernambuco, Centro de Biociências, Av. Prof. Moraes Rego, 1235, Recife, PE CEP 50.670-420, Brazil

[6] Instituto de Química Biológica, Facultad de Ciencias, Facultad de Química, Universidad de la República, General Flores 2124, 11800 Montevideo, Uruguay

Published online: 08 June 2018

pathogens, including bacteria, fungi, viruses, and protozoa (Zasloff 2002). Known AMPs differ in size, composition, and molecular structure, although they share some features like a small size (< 10 kDa), an amphipathic structure, and a net positive charge at physiological pH (Yeaman and Yount 2003; Brogden 2005). This high structural diversity allows AMPs to act against a wide range of microbial agents in diverse physiological environments (Tossi and Sandri 2002). In plants, these peptides often contain a high number of cysteine residues that stabilize protein structure through disulfide bond formation (Lay and Anderson 2005). Moreover, disulfide bonds may act catalytically and can be reversibly reduced and oxidized, playing a role in redox regulation (Nahirñak et al. 2012b). The main groups of plant AMPs are defensins, thionins, snakins, lipid transfer proteins, and cyclotides (Benko-Iseppon et al. 2010).

Snakins constitute a family of AMPs described from potato Snakin-1 (StSN1, Segura et al. 1999) and Snakin-2 (StSN2, Berrocal-Lobo et al. 2002), the first peptides isolated of this family. These peptides belong to a group of proteins encoded by Snakin/GASA genes, that are characterized by having a GASA (Gibberellic Acid Stimulated in Arabidopsis) domain of approximately 60 amino acids, with 12 highly conserved cysteine residues that may be involved in the formation of up to six disulfide bonds (Nahirñak et al. 2012b). Mass spectrometry was first applied to decipher the disulfide connectivity, with the hypothesis that they may be required for the generation or stabilization of a particular 3D motif responsible for the function. Accordingly, a partial connectivity pattern with three disulfide bonds was identified for StSN2 (Harris et al. 2014). Nearly, at the same time, Porto and Franco (2013) predicted the StSN1 three-dimensional structure through an ab initio and comparative modeling in combination with a predictor of disulfide bridges. Although the arrangement of cysteines along the sequence is conserved in the snakin family, the StSN1 model only shared one disulfide bond with the mass spectrometry data obtained for StSN2. The disagreement between experiments and modeling became evident with the X-ray determination of StSN1 (Yeung et al. 2016) that fully supported the mass spectrometry data, reporting a structure with a helix–turn–helix (HTH) motif stabilized by six disulfide bonds between specific cysteine residues. Surprisingly, the StSN1 model of Porto and Franco that only shared two disulfide bonds with the X-ray structure, also presented an HTH motif, suggesting that a precise connection between specific cysteine pairs may not be necessary for the peptide to fold to the native and functional state. Furthermore, some evidence exist pointing out that the reduced form of the StSN1 and StSN2 peptides (where all cysteines are free, and there are no disulfide bonds) were equally active against pathogens as the native forms (Harris et al. 2014).

The role assignment for snakins has been based on expression profiling analysis (Segura et al. 1999; Berrocal-Lobo et al. 2002; Meiyalaghan et al. 2014; Nahirñak et al. 2016; Herbel et al. 2017) and transgenic expression of these genes (Almasia et al. 2008; Balaji and Smart 2012; Rong et al. 2013; Mohan et al. 2014). Overexpression of snakin-1 or 2 in transgenic potato and tomato plants increased resistance to pathogens (Almasia et al. 2008; Balaji and Smart 2012; Mohan et al. 2014). In vitro, StSN1 and StSN2 displayed a broad spectrum of antimicrobial activity (Segura et al. 1999; Berrocal-Lobo et al. 2002). On the other hand, silencing of snakins-2 homologous genes in *Nicotiana benthamiana* increases susceptibility to *Clavibacter michiganensis* ssp. *michiganensis* (Balaji et al. 2011). In addition, potato snakin-1 gene silencing seems to influence cell wall composition, cell division, and leaf primary metabolism in potato plants, suggesting that StSN1 may also be involved in several cellular processes (Nahirñak et al. 2012a).

The biological properties of snakins make them attractive biotechnological targets, especially for the development of novel disease control agents (Oliveira-Lima et al. 2017). Heterologous expression is the most widely used method for medium and large-scale production. Several systems (producing organism/vector) have been developed to achieve a cost-effective and large-scale production of several proteins, in hosts such as bacteria, yeasts, fungi, and plants (Thevissen et al. 2007; Padovan et al. 2010; Silva et al. 2011). Specifically, recombinant snakins have been successfully produced in *E. coli*, *Pichia pastoris,* and more recently in baculovirus/insect cells (Almasia et al. 2017), displaying the expected antibacterial and antifungal activity (Kovalskaya and Hammond 2009; Mao et al. 2011; Herbel et al. 2015; Kuddus et al. 2016).

In the last decade, the bioinformatics tools have identified AMP-coding genes in several plant species. Due to the increasing availability of their omics, cultivated plants and model species have been the main target. Native species that stand out for their importance in folk medicine and exhibit considerable plant biodiversity remain underexplored, mainly because their molecular data are generally unavailable (Pestana-Calsa et al. 2010). In this work, we report the first snakin-like gene *PdSN1* isolated from *Peltophorum dubium*, a non-cultivated South American tree with medicinal properties. Our expression analysis showed that *PdSN1* is strongly expressed during seedling development and could be a constitutive element of the defense mechanisms of the storage organs. The snakin peptide was successfully produced through recombinant expression in *E. coli*, purified and characterized for its activity against a set of microorganisms. In addition, we report the first extensive analysis on the importance of the disulfide bridges on the stability of key structural marks present in the snakin family, like the HTH motif found in both the native and recombinant PdSN1

peptides. Supported by our structural analyses, we found that disulfide bridges might not be essential for the activity of the snakin family. In addition, we hypothesize that these peptides possibly could, in addition to the previously proposed lipid-membrane targeting mechanism of action, alter the microbial gene expression by binding to DNA/RNA.

Materials and methods

Biological material

Peltophorum dubium seeds and leaves were obtained from Facultad de Agronomía Garden (Montevideo, Uruguay). Seeds were immersed in concentrate grade H_2SO_4 during 15 min for scarification. Surface sterilized seeds were germinated on Whatman paper soaked with distilled water in Petri dishes at 28 °C. Leaves and 5-day-old seedlings (ca. 6-cm long) were frozen in liquid nitrogen and stored at − 70 °C until further usage.

Bacteria and fungi strains used for bioactivity assays were obtained from the Laboratory of Microbiología Facultad de Química collection. Fungal cultures (*Candida albicans* CCMG13, *Aspergillus niger* CCMG17, *Botrytis cinerea* CCMG14 g, *Alternaria alternata* CBS916.96, and *Penicillium expansum* CCMG14s) and *Streptomyces scabies* DSM41658 were grown on potato dextrose agar (PDA) at 28 °C. *Staphylococcus aureus* ATCC6538P and *Escherichia coli* CCMG50 were grown on Tryptone soya agar (TSA) at 37 °C. *C. michiganensis* ssp. *michiganensis* MAI1008 and *Xanthomonas versicatoria* MAI2020 were grown on Nutrient Agar at 28 °C. *E. coli* TOP10 (Invitrogen, Carlsbad, USA), Shuffle (New England Biolabs, Ipswich, USA) and Rosetta-gami(DE3)pLysS (Novagen, Madison, USA) were used as cloning and expression host, respectively, being grown on Luria–Bertani Agar.

Isolation of DNA and RNA from *P. dubium*

Genomic DNA was extracted from leaves using the standard cetyl-trimethylammonium bromide (CTAB) method (Doyle 1991). Total RNA was extracted using Qiagen RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and treated with RNase-free DNase (Invitrogen, Carlsbad, USA) to eliminate any residual DNA. Genomic and plasmid DNA were visualized under UV illumination after electrophoresis on agarose gels in 1× TBE buffer, stained with GoodView (Ecoli s.r.o., Bratislava, Slovak Republic).

PdSN1 gene cloning

PdSN1 gene was PCR amplified from ATG to stop codon, from *P. dubium* genomic DNA using degenerate primers SN_dgF 5′-ATGAAGCCAGCATTTKCARCTMT-3′ and SN_dgR 5′-TTAAGGGCATTTWGSCTTKCC-3′. Primers design was based on the alignment from plant snakin encoding sequences that were identified in the UniProt database (http://www.uniprot.org/) and the GenBank database of the National Center for Biotechnological Information (NCBI) (https://www.ncbi.nlm.nih.gov/genbank/). The PCR reaction was performed in a 20-µl reaction containing: 1× buffer with 2-mM $MgCl_2$, 0.2-mM dNTPs, 2 µM of each primer, 50-ng template DNA, and 0.5 U Taq DNA polymerase (Invitrogen, Carlsbad, USA). The PCR program was as follows: 94 °C for 3 min, followed by 35 cycles of 94 °C for 30 s, 54 °C for 40 s and 72 °C for 40 s. PCR products were purified according to Richero et al. (2013) and cloned into a pGEM-T easy vector (Promega Corporation, Madison, USA). *PdSN1* gene sequence was confirmed by sequencing three clones at Macrogen Inc. (Seoul, Korea) and analyzed using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

PdSN1-coding sequence (termed *PdSN1*c) was verified by the synthesis of 5′ and 3′ cDNA regions from total RNA extracted from *P. dubium* seedlings. In the 5′-RACE experiment (Rapid Amplification cDNA Ends), the cDNA was obtained using the 5′ RACE kit (Invitrogen) and the specific oligonucleotides PdSNc_R1 5′-AGTTCTTGAGGTCTCTGTAGCA-3′ and PdSNc_R2 5′-CTCTGTAGCAAGGGCACTCGT-3′ (Fig. 1). The 3′ region was obtained by 3′-RACE according to Frohman et al. (1988) using the specific oligonucleotides PdSNc_F1 5′-GTGCCTCTTTCGTTGAGGTC-3′ and PdSNc_F2 5′-GCCTTGTACTCAGTGCCTCTT-3′. PCR products were cloned in the pGEM-T easy vector (Promega Corporation, Madison, USA) and sequenced using the Sanger method. The resulting plasmid (pG-*PdSN1*c) contained the complete coding sequence of the mature peptide. The cDNA sequence was deposited to GenBank with accession number MG229644.

Bioinformatics analysis of the PdSN1 sequence

The deduced amino acid sequences of PdSN1 were created in BioEdit (Hall 1999) and analyzed with the Expasy-ProtParam tool (http://web.expasy.org/protparam/) to obtain the different peptide parameters. The peptide structure was evaluated for the presence of a signal peptide sequence with SignalP (http://www.cbs.dtu.dk/services/SignalP). The deduced amino acid sequences encoding for the mature *P. dubium* peptide were aligned against a set of mature plant proteins with GASA domain. All sequences used were taken from the UniProt database (http://www.uniprot.org/) (Table S1). Alignment with the newly isolated snakin was performed in ClustalW from BioEdit program. An unrooted tree was generated using the Neighbor-Joining method of the MEGA package version 5 (Tamura et al. 2011), with 5000 bootstrap replicates.

Fig. 1 Nucleotide genomic sequence of *PdSN1* and amino acidic sequence of PdSN1. The lower case letters indicate the 5′ UTR, intron, and 3′ UTR regions. Amino acid sequence of PdSN1 mature peptide indicated by gray-shaded is preceded by a signal peptide sequence predicted using the SignalP (http://www.cbs.dtu.dk/services/SignalP). Oligonucleotide sequences are indicated by horizontal arrows

Gene expression analysis

Quantitative reverse transcription-PCR (RT-qPCR) was employed to determine the relative expression levels of *PdSN1* gene in *P. dubium*'s adult leaves and seedlings (inoculated with *A. niger* and mock-not inoculated control). Five-day-old seedlings were inoculated at five points with 20 µl per point of a spore suspension of *A. niger* (1×10^6 conidia ml^{-1}). Seedling samples were collected 48 h after inoculation and frozen in liquid nitrogen. Mock inoculations were done with Tween 10^{-4}. RT synthesis was performed using MMLV reverse transcriptase (Invitrogen, Carlsbad, USA) and random primers (QIAGEN, Hilden, Germany). Specific primers (PdSN1_qF1 ACTCAGTGC CTCTTTCGTTG and PdSN1_R1 CATTCCTGCTTTCGA GCATCT) were used to amplify the cDNA fragments. PCR reactions were carried out using SensiMixPlus SYBR PCR Master Mix (2×) (Quantace, London, UK) in a Corbett Rotor Gene™ 6000 as follows: 15-min pre-denaturing at 95 °C, followed by 30 cycles of 94 °C for 15 s, 59 °C for 1 min. Three biological and two technical replicates were performed for each sample. The specificity of each reaction was confirmed by inspection of the melting curve profiles.

To choose the reference genes (RGs), the expression of actin (*PdACT*), elongation factor 1-α (*PdEF1*), F-box/kelch-repeat protein (*PdF-box*), phosphoenolpyruvate carboxylase-related kinase 1 (*PdPEPKR*), protein phosphatase 2A (*PdPP2A*), and tubulin A (*PdTUA*) were evaluated for all conditions. The expression stability of RGs was assessed using geNorm algorithm (Vandesompele et al. 2002). Previously, the candidate genes were partially cloned by 5′- and 3′-RACE using reported primers from *Caragana korshinskii* (Zhu et al. 2013; Table S2). qPCR specific primers of RGs of *P. dubium* were then located in the first cloned region of each one (Table S3).

Partial cDNA sequences of RGs were deposited in GenBank with accession numbers MG397035, MG397036, MG397037, MG397038, MG397039, and MG397040.

Construction of the expression plasmid

The pET102/D-TOPO vector (Invitrogen, Carlsbad, USA) was used for expression of *PdSN1* gene. This vector allows the expression of PdSN1 mature peptide fused to a thioredoxin protein (N-ter) and a His-tag (C-ter) to facilitate solubility and downstream purification using affinity chromatography. The mature coding sequence of *PdSN1* was PCR amplified from pG-*PdSN1*c using the following primers: PdSN1_pEF 5′-CACCGGTTCTGAGTTCTGT GACTCCAAGTGCGCG-3′ (TOPO cloning nucleotides CACC are underlined) and PdSN1_pER 5′-CCTTCCCTC GAT GGGGCATTTGGGCTTGCC-3′. Reverse primer has 12 nucleotides which code for the factor Xa recognition site. Blunt-end PCR product was obtained using a Phusion High-Fidelity DNA Polymerase (Thermo-Scientific, Waltham, USA). The PCR conditions were 95 °C for 1 min followed by 30 cycles (95 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min) and finally 72 °C for a 5-min extension. The TOPO cloning reaction and the transformation of *E. coli* were performed according to the manufacturer's instructions. The resulting plasmid pE-PdSN1 was confirmed by DNA sequencing. In parallel, a pET102/D without PdSN1 sequence (pE-Trx*) was obtained.

Expression and purification of recombinant PdSN1

The *PdSN1* carrying plasmid pE-PdSN1 was introduced into *E. coli* Rosetta-gami (DE3)pLysS and Shuffle strains. A transformation with pE-Trx* was performed as a control. Cells were grown at 37 °C and 150 rpm in the presence of 100-µg ml^{-1} ampicillin (Amresco, Solon, USA) until the culture density reached an OD600 of 0.5–0.7. Induction was carried on with 1-mM IPTG (Sigma-Aldrich, Saint Louis, USA) varying time (2, 4, and 24 h) and temperature (20, 28, and 30 °C). After incubation, the bacterial cells were harvested by centrifugation at 4000×*g* for 15 min at 4 °C, washed twice in 50-mM potassium phosphate buffer, pH 7.8, and stored at − 70 °C until protein extraction.

The cell pellet was resuspended in 2 ml of ice-cold lysis buffer (50-mM potassium phosphate buffer, pH 7.5, 500-mM NaCl, 10% glycerol, 10-mM imidazole). The cells were lysed by sonication 4× pulse at 4 °C (30% amplitude; 6 cycles of 6-s pulse-on, 9-s pulse-off; 15 s) (Cole-Parmer, Vernon Hills, USA). The lysate was centrifuged 15 min at 4 °C and 10,000 rpm, and the supernatant was filtered through a 0.22-µm filter (Millipore). The recombinant protein was purified using a HisTrap-FF (GE Healthcare, Little Chalfont, UK) Ni-affinity chromatography, according to the supplier's recommendations. Unbound proteins were eluted by adding 20- and 30-mM imidazole to the buffer, while 300-mM imidazole was used to elute the bound proteins. To remove the thioredoxin (Trx) fusion fragment, the immobilized Trx-PdSN1protein was subjected to EkMax Enterokinase (Invitrogen) digestion for 15 h at room temperature. Eluted fractions were analyzed by 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using Tris–Tricine buffer (Schägger and von Jagow 1987). Fractions containing recombinant peptide were pooled, dialyzed against distilled water, and lyophilized. Subsequently, the peptide was resuspended in water and the protein concentration was measured by Bradford assay (Bradford 1976) using bovine serum albumin as a standard.

Identification of recombinant PdSN1 peptide

The peptide sequence was confirmed by peptide mass fingerprinting and subsequent fragmentation of selected peptides (MS/MS). Disulfide bonds were reduced by incubation with DTT (100 mM) for 1 h at 56 °C and Cys were alkylated with iodoacetamide (300 mM). Tryptic digestion was performed overnight at 37 °C using sequencing grade trypsin (Promega). Peptide mixtures were mixed with matrix solution (-cyano-4-hydroxycinnamic acid in 60% acetonitrile, 0.1% TFA) directly onto the sample plate. Spectra were acquired in positive reflector mode and were externally calibrated using a standard mixture of peptides (Applied Biosystems). Peptide sequences were assigned using an in-house Mascot v.2.3 version (Matrix Science) for searching a local database that includes the sequence of PdSN1. The following parameters were used with variable modifications: oxidation (M), carbamidomethyl (C), ammonia loss (N-term C); mass values: monoisotopic; peptide mass tolerance: ± 0.05 Da; and fragment mass tolerance: ± 0.45 Da. Selected *m*/*z* values were further fragmented, and the assignment to PdSN1 amino acid sequence was validated by manual inspection of MS/MS spectra.

Antimicrobial activity assay

Antimicrobial activity was estimated microspectrophotometrically (Broekaert et al. 1990), determining the IC50 value (peptide concentration at which 50% inhibitions was reached). Fungal spore suspension (1×10^4 spores ml^{-1}) or bacterial suspension (1×10^5 bacteria ml^{-1}) was cultivated in 100-µl reaction, containing Potato Dextrose Broth (for fungi and *S. scabies*), Nutrient Broth or Tryptone Soya Broth (for bacteria), and serial dilutions of recombinant PdSN1. The maximum peptide concentration used in all assays was 1.8 µM. Control wells contained no peptide. The optical density of fungal or bacterial suspension was measured in a 96-well microtiter plate under 595 nm, after 48–72 h of incubation (or 24 h for *E. coli* and *S. aureus*) in the dark. Antimicrobial activity of PdSN1 was expressed as a percentage of growth inhibition. Serial dilutions of Gentamicin were used as positive control for bacteria and propiconazole for fungi. The maximum concentrations used were 30 and 100 µM, respectively. All experiments were performed in three replicates.

3D structure prediction and molecular dynamics simulations

The prediction of the 3D structure of the PdSN1 peptide was achieved using three different approaches: (1) ab initio or de novo 3D modeling; (2) homology modeling; and (3) multi-microsecond long, unbiased molecular dynamics. In all cases, the histidine residues were considered as protonated in ε.

Ab initio predictions

For the ab initio 3D determination, we followed the work of Porto and Franco (2013), who used a combination of tools, including the QUARK ab initio molecular modeling server (Xu and Zhang 2012), the DIANNA disulfide bond predictor (Ferrè and Clote 2006), the MODELLER 9.14 modeling tool (Webb and Sali 2014), and short unbiased molecular dynamics simulations. The mature form of the native *PdSN1* sequence was submitted to the QUARK server, which delivered an initial guess of the 3D structure. Then,

the disulfide bond pattern was determined using five different predictors: DIANNA, DISULFIND (Ceroni et al. 2006), CYSCON (Yang et al. 2015), DIPRO (Cheng et al. 2006), and DISLOCATE (Savojardo et al. 2011). These predictors gave different connectivity patterns between the six pairs of cysteines; hence, an individual model was built according to each prediction (see Table S4). Each 3D structure was then subjected to microsecond long unbiased MD simulation with options as described below.

Homology modeling

Using the recently determined X-ray structure of the potato snakin-1 peptide (PDB id: 5E5Q) as a template, we build a 3D model by homology using MODELLER (Webb and Sali 2014). This peptide shares 84% similarity in sequence with the mature form of the native *PdSN1* (63 residues). The connectivity pattern of disulfide bonds was taken from 5E5Q, and the final structure was subject to unbiased MD simulations (labeled $PdSN1_{5E5Q}$). A model of the full recombinant peptide (104 residues) was also built, where the GASA domain was obtained from 5E5Q, and the remaining amino acids were modeled using the QUARK server ($PdSN1_{5E5Q\text{-}ALL}$).

Long unbiased MD simulations

To assess the importance of the disulfide bonds on the stability and conformation of the 3D and secondary structure, we took the model generated by homology modeling and run a multi-microsecond long MD simulation considering all the cysteines in their reduced form (no disulfide bonds, $PdSN1_{5E5Q\text{-}RED}$).

MD simulation protocol

All models were minimized in vacuo, neutralized with 8 Cl^-, solvated (with explicit water and 0.15 M of K^+Cl^-), and minimized in solution with positional restraints on the peptide using our well-established multi-step protocol (Pérez et al. 2007; Dans et al. 2016). To produce the final models, the minimized structures were thermalized to 298 °C at NVT, and then simulated during at least 1 µs using molecular dynamics simulations at NPT ($P = 1$ atm). The first 100 ns of simulations were considered as an equilibration step and were discarded for further analysis. For representing the snakin peptide, we used the state-of-the-art ff14SB force field (Maier et al. 2015), surrounded by a truncated octahedral box of ~ 5000 TIP3P water molecules (Jorgensen et al. 1983), and Dang parameters for ions (Smith and Dang 1994). Ions were initially placed randomly, at a minimum distance of 5 Å from the solute and 3.5 Å from one another. All systems were simulated using the Berendsen algorithm (Berendsen et al. 1984) to control the temperature and the pressure, with a coupling constant of 5 ps. Center of mass motion was removed every 10 ps to limit the build-up of the translational kinetic energy of the solute. SHAKE (Ryckaert et al. 1977) was used to keep all bonds involving hydrogen at their equilibrium values, allowing the use a 2-fs step for the integration of Newton equations of motion. Long-range electrostatic interactions were accounted for using the particle mesh Ewald method (Darden et al. 1993) with standard defaults and a real-space cutoff of 9 Å. All simulations were carried out using the PMEMD CUDA code module (Salomon-Ferrer et al. 2013) of AMBER 16 (Case et al. 2017) and analyzed with CPPTRAJ (Roe and Cheatham III 2013). Structures and MD trajectories were visually analyzed using VMD 1.9 (Humphrey et al. 1996).

Results

Isolation and sequence analysis of *PdSN1*

PdSN1 genomic sequence isolated by PCR presents 424 nt, from de ATG start codon to stop codon, with one intron of 157 nt (Fig. 1). The coding *PdSN1* sequence, isolated from seedlings cDNA, allowed for the confirmation of the sequence and localization of the intron. Analysis of the deduced amino acid sequence (88 residues) showed a signal peptide sequence of 25 amino acids, followed by a mature peptide with 63 residues. The mature peptide had a calculated mass of 6978.1 Da and a basic isoelectric point of 8.99.

PdSN1 BLAST search showed high similarity to Snakin/GASA proteins, especially with snakin type 1 (i.e., 76% sequence identity to StSN1 from potato and 74% to MsSN1 from *Medicago sativa*). It has a GASA domain (84% sequence identity to StSN1), characteristic from this protein family, with 12 cysteines in conserved positions ($XCX_3CX_3CX_8CX_3CX_2CCX_2CXCX_{11}CXCX_{12}CX$). Alignment analysis of PdSN1 mature sequence and plant mature proteins with GASA domain revealed that the newly isolated gene shared a higher homology with both StSN1 and MsSN1 snakins-1, with a peamaclein (PMLN) from *Prunus persica* and with Gibberellin regulated proteins 7, 8, and 10 from *Arabidopsis thaliana* (GASA7; GASA8 and GASAA). All these peptides belong to Snakin/GASA subfamily I, according to the classification of Berrocal-Lobo et al. (2002) (Fig. 2). This analysis was restricted to sequences with reported experimental evidence such as antimicrobial activity, including a snakin-2 from tomato (|E5KBY0|E5KBY0_SOLLC; Herbel et al. 2015), a snakin-1 from alfalfa (|H9D2D5|H9D2D5_MEDSA; García et al. 2014), a snakin-1 from pepper (|B2ZAW4|B2ZAW4_CAPAN; Mao et al. 2011), a snakin-3 from potato

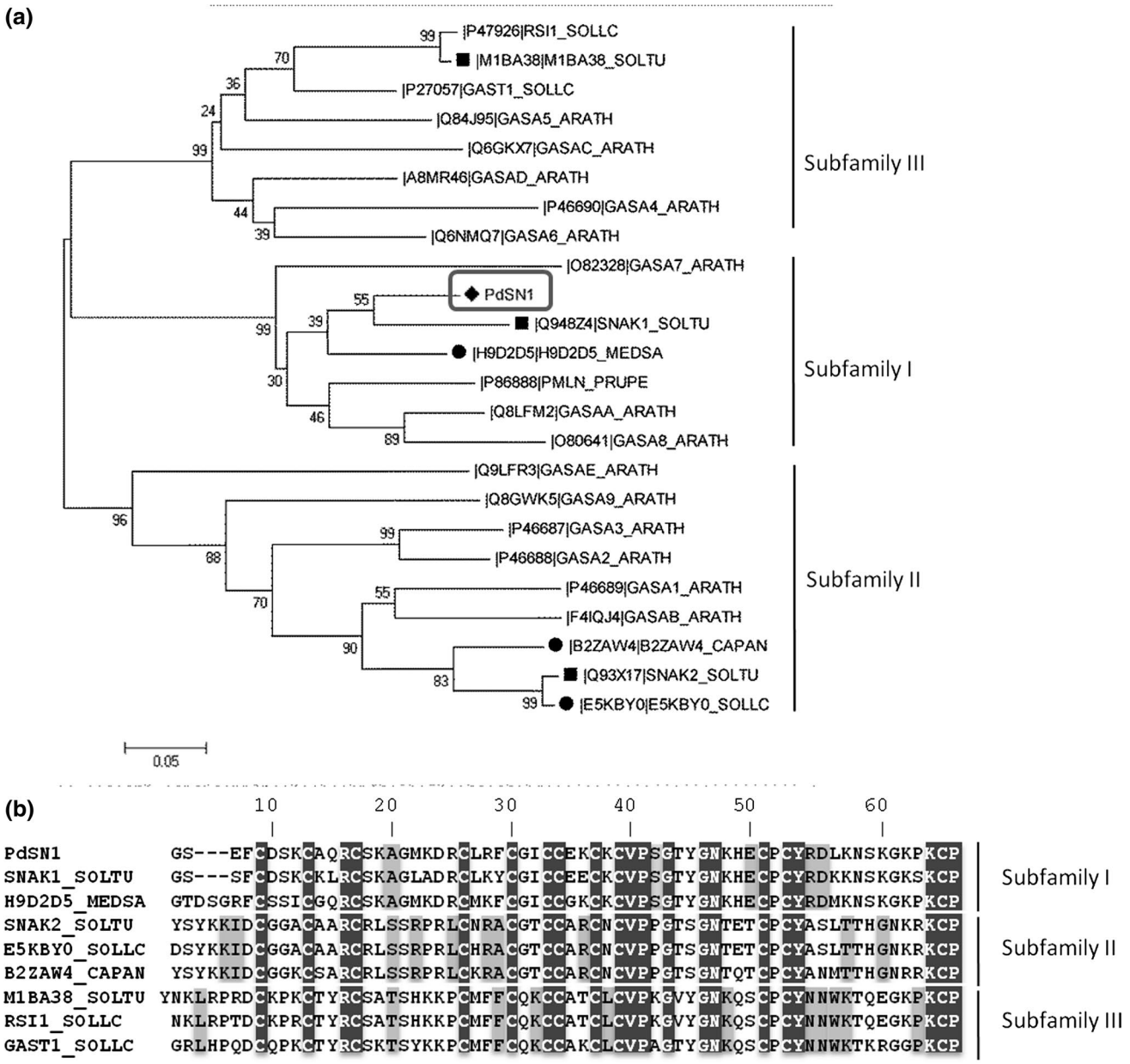


Fig. 2 Sequence alignments and distance matrix of PdSN1 peptide and members of the plant Snakin/GASA protein family. The deduced amino acid sequence of the newly isolated gene (*PdSN1*) was aligned with a set of mature plant proteins with GASA domain. Sequences used were manually annotated and reviewed from the UniProt database (http://www.uniprot.org/). Four snakins with reported experimental evidence were also included (|E5KBY0|E5KBY0_SOLLC, |H9D2D5|H9D2D5_MEDSA, |B2ZAW4|B2ZAW4_CAPAN, |M1BA38|M1BA38_SOLTU). Snakin/GASA proteins grouped following their classification into three subfamilies (I–III) as proposed by Berrocal-Lobo et al. (2002). **a** Neighbor-Joining unrooted tree. Values in the nodes regard bootstrap values (5000 replicates). Each sequence was named according to its UniProt Entry identifier, followed by a Mnemonic identifier of a UniProtKB entry. ARATH: *A. thaliana*, SOLTU: *S. tuberosum*, SOLLC: *Solanum lycopersicum*, PRUPE: *P. persica*. MEDSA: *M. sativa*, CAPAN: *Capsicum annum*. Black diamond indicates PdSN1; black squares indicate snakin-1, snakin-2, and snakin-3 (M1BA38) from *S. tuberosum* and black circles indicate snakin-1 (H9D2D5) from *M. sativa* and snakin-2 from *S. lycopersicum* (E5KBY0) and *C. annum* (B2ZAW4). **b** Alignment of mature sequences of representative snakin/GASA proteins, following their classification in subfamilies I, II, and III (indicated at the right side of the alignment). Amino acids conserved across all the family members are black shaded. Most conserved residues which are relevant for subfamily classification (according to Berrocal-Lobo et al. 2002) are shaded in gray

(|M1BA38|M1BA38_SOLTU; Nahirñak et al. 2016), and sequences obtained from the curated UniProt database (http://www.uniprot.org/).

PdSN1 gene expression in plant

PdSN1 mRNA steady states from leaves and seedlings (inoculated with *A. niger* and uninoculated) were measured by RT-qPCR using three reference genes. The geNorm tool identified *PdACT*, *PdEF1,* and *PdPP2A* as the most stable genes (expression stability value [*M*]: 0.705, 0.903, and 1.069, respectively), among six candidates analyzed in all conditions (methodology described in "Gene expression analysis"). The geometric mean of these three genes was used for normalization of *PdSN1* expression, as recommended by Vandesompele et al. (2002). The expression of *PdSN1* was 40 fold higher in seedling than in adult leaves (Fig. 3). The presence of *A. niger* did not produce significant variations of seedlings' mRNA steady-state levels 48 h after inoculation.

Expression and purification of recombinant PdSN1 in *E. coli*

To reach the highest yield of the soluble Trx-PdSN1 fusion protein, its expression was tested by varying temperatures (30, 28, and 21 °C) and times after induction (2, 4, and 24 h), using *E. coli* strains Rosetta-gami pLysS and Shuffle. Both *E. coli* strains were transformed with pE-PdSN1 and pE-Trx* (control without PdSN1). A schematic representation of the peptides resulting from these vectors, Trx-PdSN1 and Trx*, respectively, is shown in Fig. 4a. The SDS-PAGE analysis of total lysates from induced cells revealed the presence of recombinant proteins Trx-PdSN1 and Trx* with an expected molecular weight of ~ 25 and ~ 17 kDa, respectively (Fig. 4b, lanes 1–4). No significant peptide bands corresponding to the recombinant protein were observed in non-induced cultures (Fig. 4b, lane 2). Some amount of the recombinant protein was present in inclusion bodies, in all conditions. We found that the highest yield (0.9 mg l^{-1} bacterial culture) was obtained with Rosetta-gami(DE3)pLysS at 28 °C and 24 h of induction, while with Shuffle strain, we found a lower percentage of recombinant protein in the soluble fraction (data not shown). The recovery of the 25-kDa band at the end of the purification by Ni-affinity chromatography confirmed that this band contains the recombinant peptide (Fig. 4b, lane 6).

For practical reasons a cleavage on-column, Trx was conducted before peptide elution. A previous pilot test, carried out with several concentrations of enterokinase, showed that three µg of the fusion protein was digested entirely with one unit of the enzyme overnight at room temperature (data not shown). The developed protocol resulted in an efficient separation of the recombinant PdSN1 peptide from Trx (Fig. 4b, lanes 7–8).

The band of the PdSN1 peptide seemed to be larger (ca. 14 kDa) than the expected (11 kDa), possibly due to an abnormal migration which is sometimes observed with cationic peptides on the SDS-PAGE (Shi et al. 2012; Herbel et al. 2015). The results obtained by the mass fingerprinting of the peptide confirmed that its sequence corresponded to PdSN1 (100% coverage, Table S5; Fig. S1).

Antimicrobial activity of the recombinant peptide PdSN1

The antimicrobial activity of the recombinant peptide was tested in vitro against several bacterial and fungal species, as detailed in "Materials and methods". To the maximum concentration (1.8 µM) of PdSN1 tested, significant inhibition of growth (99.7%) was observed for *S. scabies,* while for *S. aureus*, *C. michiganensis* ssp. *michiganensis*, *C. albicans*, *B. cinerea*, *A. niger*, and *A. alternata*, the inhibition was intermediate (60.0, 56.3, 65.5, 53.6, 56.7, and 58.0%, respectively) (Fig. 5). The percentage of inhibition decays at lower concentrations of the peptide for all organisms. *S. aureus* was excluded from Fig. 5, because it only displayed inhibition at the maximum concentration tested (1.80 µM, 56.3% inhibition). No activity was observed for *E. coli*, *X. versicatoria,* and the fungus *P. expansum* at this concentration. IC50 values of recombinant PdSN1 for *C. michiganensis* ssp. *michiganensis, S. scabies, C. albicans, B. cinerea, A. alternata,* and *A. niger* were estimated to be 1.7, 0.3, 1.2, 0.4, 0.4, and 1.4 µM, respectively. IC_{50} value of propiconazole, which was used against *A. niger* as a positive control, was 3.0 µM. IC_{50} value of Gentamicin, used against *C. michiganensis* ssp. m*ichiganensis* was 0.5 µM.

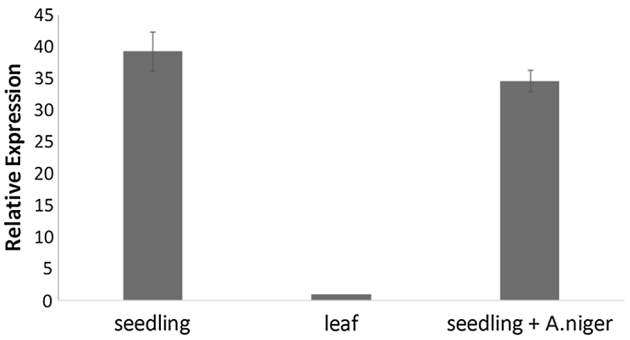


Fig. 3 Expression of *PdSN1* gene in *P. dubium* by RT-qPCR. The geometric mean of *PdACT, PdEF1,* and *PdPP2A* genes was used for normalization of *PdSN1* expression from leaves, mock seedlings, and seedling inoculated with *A. niger*. The final data were obtained by rescaled normalized expression: $(Q_{sample}/NF_{sample})/Min(Q_{sample}/NF_{sample})$. Each bar graph represents the mean relative fold change ± SD of three independent biological replicates

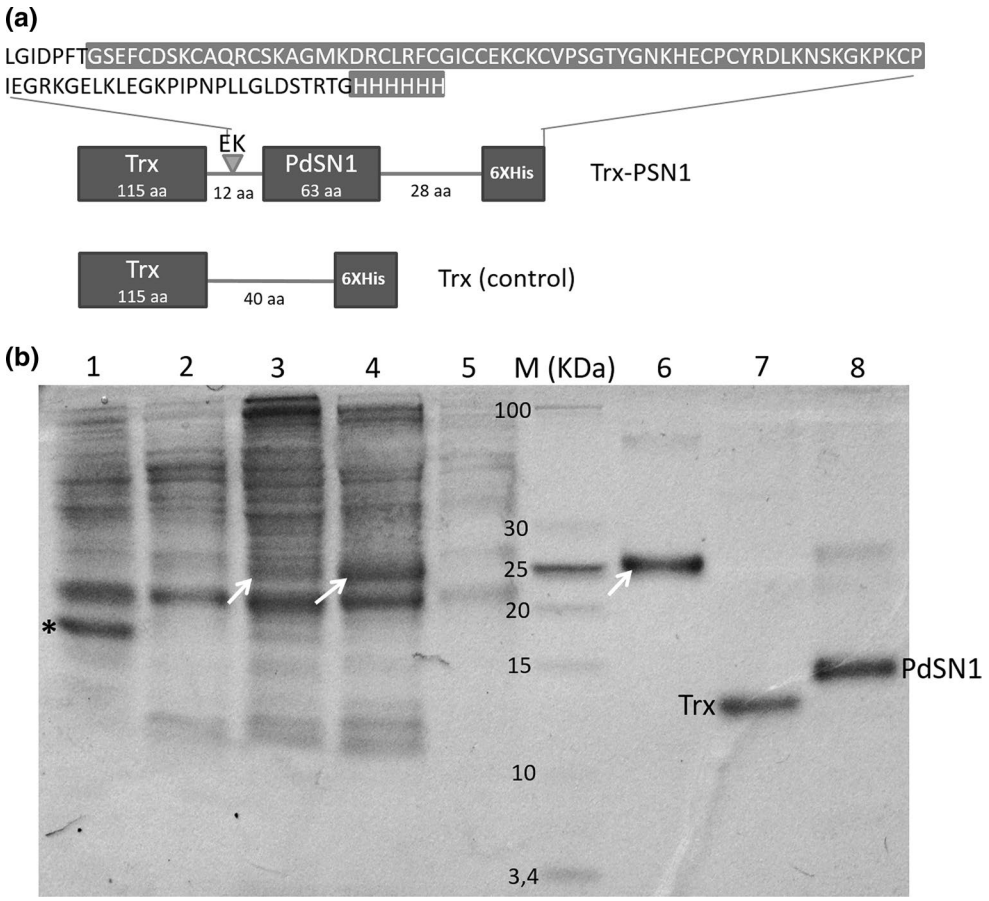


Fig. 4 Affinity purification of the recombinant PdSN1 peptide. **a** Schematic representation of recombinant Trx-PdSN1 and Trx* (control) proteins. **b** Tris–Tricine SDS-PAGE (15%) analysis of the purification steps of the Trx–PdSN1fusion protein and PdSN1 peptide. Protein fractions from *E. coli* Rosetta-gami (DE3) cells transformed with pE-Trx expression plasmid (used as a control) (1), pE-PdSN1 expression plasmid incubated in the absence (2) or the presence of IPTG (3: crude extract; 4: soluble fraction); Unbound proteins removed by the first washing step in which the fusion protein is bound to the column (5); final elution step, where the purified Trx-PdSN1is released from the Ni-affinity column (6); Trx protein after cleavage with enteroquinase (7); elution of PdSN1 peptide (8). *M* PageRuler Low Range Unstained Protein Ladder (ThermoFisher). Asterisk indicates expression of Trx* protein; white arrows show Trx-PdSN1fusion protein

3D structure determination

Using the recently determined X-ray structure of the potato snakin-1 peptide, the homology modeling of the 3D structure of the native PdSN1 was solved. Connection pattern of the disulfide bonds was also taken from the experimental structure, leading after extensive unbiased molecular dynamics simulations to a final model exhibiting an HTH motif (two α-helixes: H1 and H2, linked by a small turn), and two other small helixes (H3 and H4) in the region of the C-terminal flexible loops (Fig. 6a). The extra amino acids present in the recombinant PdSN1 peptide did not alter the HTH domain (present in the GASA domain), nor the presence of the third and fourth helixes, although H4 appeared in a different region nearest to the C-terminal end (Fig. 6b). 3D predictions coming from pure ab initio methods (a combination of QUARK and MODELLER) and different disulfide bond predictors (see Table S4) also conserved the HTH motif (Fig. 6c), although the connectivity pattern of disulfide bonds was substantially different across the models (Table S4). Furthermore, the model built without disulfide bonds, where all cysteines were represented in their reduced form, also showed a stable HTH motif in the microsecond timescale (Fig. 6c). The time evolution of the root mean squared displacements of the HTH backbone in respect to the X-ray conformation showed that all structures converged to a stable HTH motif (after 650 ns) which was less than 2.75 Å apart from the crystal (Fig. 6d). The superposition of the final structures obtained from MD clearly highlights the similarities within the HTH motifs of all models, independent of the presence of disulfide bridges or the specific disulfide bond connectivity.

While the HTH was present in all cases given to all models a very similar 3D conformation on the N-terminal side,

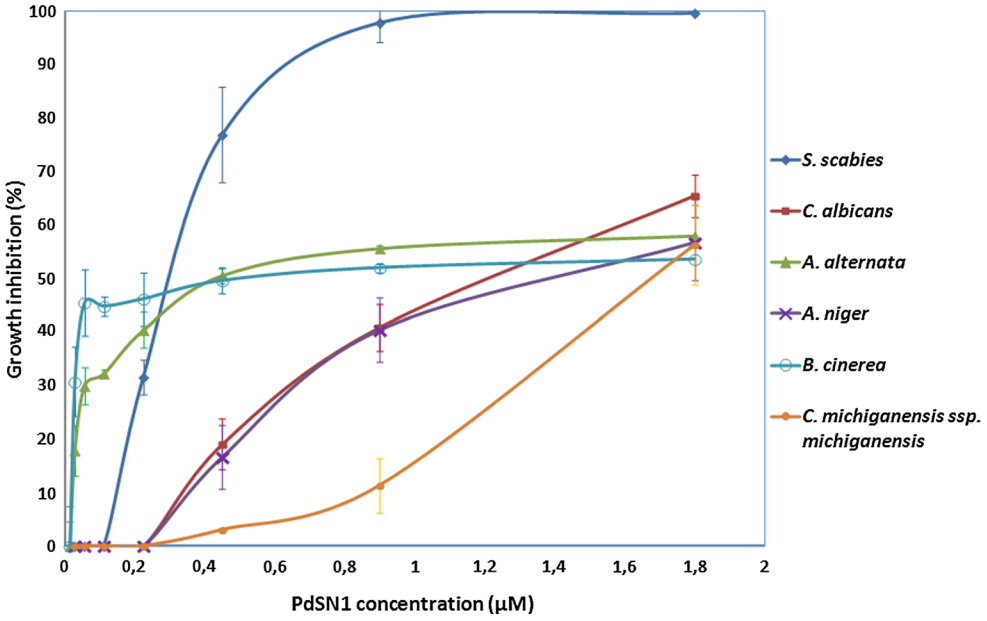


Fig. 5 In vitro antimicrobial activity of PdSN1. Dose-dependent growth inhibition curves of the following bacterial and fungal pathogens: *C. michiganensis* ssp. *michiganensis* (black circles), *S. scabies* (diamonds), *A. niger* (crosses), *B. cinerea* (open circles), *A. alternata* (triangles), and *C. albicans* (squares) were determined by measuring the cultures at OD595, in the presence of different amounts of recombinant PdSN1. The data are mean ± SD ($n = 3$)

the C-terminal region was quite variable (see the red arrows in Fig. 6a–c). The model built by homology modeling was similar to the X-ray structure, showing two small helixes embedded in between flexible loops (Fig. 6a), while the extra amino acids present in the recombinant PdSN1 sequence increased the separation between H3 and H4 with more disordered loops (Fig. 6b). The ab initio models have different C-terminal regions, one of which showed no H3 and H4 helixes (see $PdSN1_{Q\text{-}CYS}$ in Fig. 6c). Interestingly, the model with reduced cysteines exhibited a helix in the C-terminal region which seems to be a condensation of the 3_{10} helix (H3) and α-helix (H4) present in the X-ray structure of the potato snakin-1 peptide. Finally, the computed molecular electrostatic potential of the HTH motifs and extra helixes present in the recombinant PdSN1 peptide also showed a positively charged cleft (see the blue arrow in Fig. 6e), as in the X-ray structure, and a tail of histidines (only present in our recombinant peptide) that could be protonated, enhancing the electrophilic tendency of the peptide (Fig. 6e).

Discussion

In this study, a new snakin gene from *P. dubium* was isolated, aiming to explore the potential of this South American native plant as a producer of novel AMPs with antimicrobial activity. For this purpose, a PCR-based isolation strategy was used to amplify putative snakins, using degenerate primers designed from known snakin sequences of other plants. With the isolated sequence, the potential antimicrobial activity of the new snakin was tested in vitro, using a heterologous expression system for its production. In addition, an introductory analysis was performed to study the expression of this new snakin gene in different plant tissues. Finally, using comparative and ab initio modeling together with molecular dynamics simulations, we explored the structure–activity relationship of this peptide.

Alignment of the snakin genes from the NCBI and UniProt databases revealed a low level of similarity in the sequence that encodes the signal peptide. For this reason, we choose to use degenerate primers, which allowed for the obtention of a sequence, named *PdSN1* (snakin gene from *P. dubium*). This new snakin exhibits an ORF that codes a peptide of 88 residues including a GASA domain with 12 cysteines in conserved positions, sharing 76% identity with potato StSN1. After comparing PdSN1 against UniProt curated sequences bearing a GASA domain, the most significant similarity was obtained with StSN1 from potato, MsSN1 from *M. sativa*, GASA7, GASA8, and GASA10 (GASAA) from Arabidopsis and PMLN from *P. persica*. Of these proteins, only StSN1 (Segura et al. 1999; Kovalskaya and Hammond 2009; Kuddus et al. 2016) and MsSN1 (García et al. 2014) were reported to have antimicrobial activity. GASA10, expressed in *E. coli*, showed toxicity to the host cell, while no significant effect on several microorganisms growth was observed when expressed in *P. pastoris* (Trapalis et al. 2017). From its side, Peamaclein (PMLN), found in protein extracts from peach peel and pulp, was characterized as a new allergen (Tuppo et al. 2013). An NJ analysis done with MsSN1 and additional type 1 snakins reported that MsSN1 could be ortholog to snakin-1 from *Solanum tuberosum* and, ortholog to the GASA7 protein from *A. thaliana* (García et al. 2014), suggesting that they can share with PdSN1 a similar function in plants.

The expression of *PdSN1* gene was analyzed in seedlings and leaves with expression in both, but steady-state

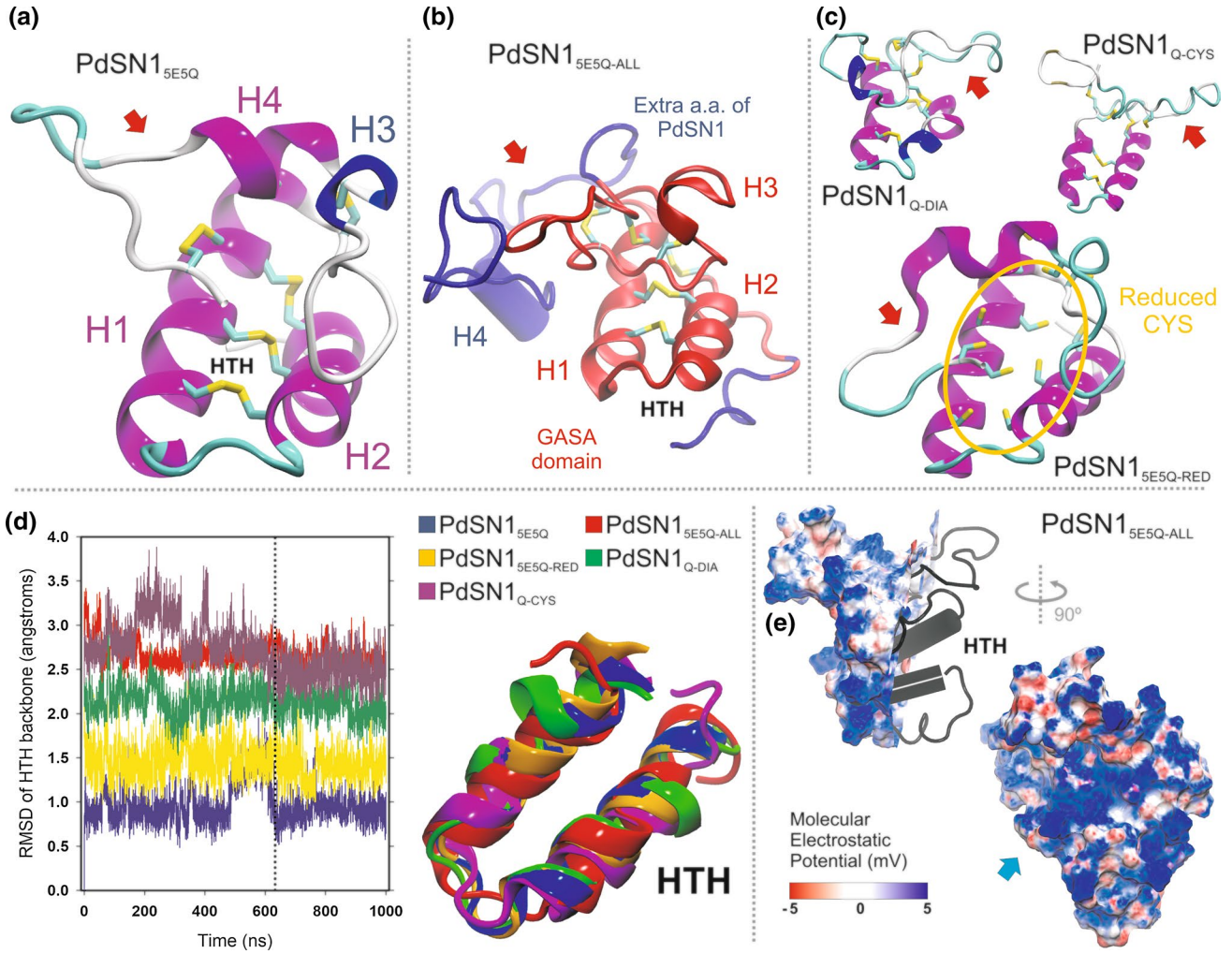


Fig. 6 Homology modeling and 3D structure prediction of the native and recombinant PdSN1 peptide. **a** Structural prediction of the native PdSN1 with his GASA domain, obtained using as template the X-ray structure with PDB id 5E5Q, showing two long helixes (H1 and H2) that define the helix–turn–helix motif (HTH), and two short helixes (H3 and H4) located in the flexible loops of the C-terminal region. **b** Model of the full PdSN1 peptide. The GASA domain is depicted in red while remaining amino acids are shown in blue. **c** 3D structure predictions combining QUARK, DIANNA and/or CYSCON, servers (top); and, bottom, the last structure of the long MD simulation (1.5 µs) of the PdSN1 GASA domain, where all cysteines were considered in their reduced form with no disulfide bonds (see "Materials and methods"). **d** Root mean square deviations in Å of the backbone atoms of the HTH motif respect to the X-ray structure (5E5Q) along time. Structural superposition of the HTH motifs obtained from the last frame of the corresponding MD simulations. **e** Molecular electrostatic potential (MEP) of the full recombinant PdSN1 peptide (front and side view). Note that a clipping plane was used in the top structure to cut the MEP allowing to visually locating the HTH motif (color figure online)

levels of *PdSN1* mRNA were higher in seedlings than in adult leaves. The results obtained for leaves are in agreement with recently reported data from *StSN1* (Meiyalaghan et al. 2014; Nahirñak et al. 2016). However, in the past, Segura et al. (1999) and Berrocal-Lobo et al. (2002) did not detect *StSN1* mRNA in potato leaves. The high levels of mRNA detected in seedlings suggest that expression of AMPs genes are strong during seedling development, justified by their susceptibility to the attacked by a large number of pathogens present in the soil. Strikingly, there was no significant increase in expression 48 h after inoculation with *A. niger*, a pathogen commonly found in soil. Plant AMPs are found constitutively in storage organs like seeds and generative tissues (reproductive organs, fruits, and flowers; revised by Benko-Iseppon et al. 2010), so our data suggest that PdSN1 could be a component of the constitutive defense barriers organs and tissues during their early development, besides fulfilling a physiological role in seedlings (Nahirñak et al. 2012b). A more exhaustive analysis should be done, including a follow-up in time after inoculation with the fungus, something beyond the scope of the present contribution.

The mature PdSN1 peptide was successfully produced in *E. coli*. Owing to the codon bias and the inability of *E. coli* to form disulfide bridges within its cytoplasm (Lobstein

et al. 2012), we decided to work with Rosetta-gami and Shuffle strains that allow for the expression of eukaryotic proteins that contain codons rarely used in *E. coli*, while enhancing disulfide bond formation. Kovalskaya and Hammond (2009) reported that the use of a recombinant expression system like *E. coli* for the production of potato defensin PTH1 and snakin-1 yielded inactive protein aggregates which needed to be denatured and refolded for gaining activity. To minimize protein aggregates into inclusion bodies, a polypeptide fusion partner such as the hydrophilic tag thioredoxin (Trx) (Terpe 2003) was used. Such tags could also serve to avoid the possible inherent sensitivity of *E. coli* towards the expression of peptides with antibacterial activity. We used pET102/D which included a N-terminal Trx protein, a C-terminal 6xHis tag, and a protease recognition site for cleavage of the fusion Trx from the recombinant peptide PdSN1. This last feature is important, since some fused proteins may not possess antimicrobial activity. For example, a scots pine defensin expressed in *E. coli*, fused to glutathione S-transferase (GST), was biologically inactive until GST was removed (Kovaleva et al. 2011). Using Rosetta-gami/pET102/D, we obtained the snakin peptide mostly in a soluble form, which after a single chromatographic step, resulted in the purification of a biologically active peptide.

The purified recombinant PdSN1 showed variable in vitro antibacterial and antifungal activities, depending on the microorganism analyzed. The Gram-negative bacteria *X. versicatoria* and *E. coli*, besides the fungus *P. expansum,* were not sensitive to PdSN1 at the maximum concentration tested (1.8 µM). Higher concentrations should be used to affirm that PdSN1 do not affect their growth. In fact, the recombinant StSN1 showed activity against the Gram-negative bacteria *Salmonella enterica* and *E. coli* at concentrations of 5 and 10 µM (Kuddus et al. 2016). The PdSN1 IC_{50} values are comparable to others, either native or recombinant snakin peptides. In our experiments, IC_{50} for the fungus *B. cinerea* was 0.4 µM, while for native StSN1 and StSN2, the effective concentration for 50% inhibition was 0.8 and 2.0 µM, respectively (Berrocal-Lobo et al. 2002). When StSN1 was produced in *E. coli,* IC_{50} for *B. cinera* was 9.0 µM (Kovalskaya and Hammond 2009); this peptide was obtained exclusively in the form of inclusion bodies and, therefore, needed to be solubilized and refolded for activity. Kuddus et al. (2016) reported a recombinant StSN1 expressed in *P. pastoris* with minimum fungicidal concentration for *Candida parapsilosis* and *P. pastoris* yeasts, of 5.0 and 10.0 mM, respectively. Our PdSN1 had an IC_{50} value for the yeast *C. albicans* in the concentration of 1.2 µM. A snakin-2 from tomato showed activity against all tested microorganisms (Gram-negative bacteria, Gram-positive bacteria, and fungi) with IC_{50} values between 0.1 µM for the Gram-positive bacteria *Micrococcus luteus* and 1.6 µM for the mold *Fusarium solani* (Herbel et al. 2015). Interestingly, our recombinant PdSN1 showed potent activity (99.7% inhibition with 1.8 µM of the peptide) against the Gram-positive bacteria *S. scabies*, a plant pathogen that causes the economically relevant potato disease named 'common scab'. PdSN1 also displayed intermediate activity against other critical plant pathogens like *C. michiganensis* ssp. *michiganensis, A. alternata,* and *B. cinerea* while inhibiting the growth of opportunistic pathogens such as *C. albicans* and *A. niger*.

The 3D structure obtained from the PdSN1 sequence using the X-ray structure of the potato snakin-1 as a template showed a clear HTH motif in the GASA domain that was preserved in the microsecond long MD simulation of the peptide that was performed in near physiological conditions. As previously observed (Yeung et al. 2016), the HTH motif is also present in the plant thionins and in the α-helical hairpin protein classes of cysteine-rich antimicrobial peptides. These other classes exhibited not only the same two helixes of the HTH motif in the same relative orientation but also shared some of the disulfide bond connections (Vila-Perelló et al. 2005). It has been suggested, based on the antimicrobial activity of truncated thionin containing only the HTH motif (Vila-Perelló et al. 2005), that these two helixes linked by a short turn seem to be essential for the activity. Our results show that the HTH motif is present independently of the specific connectivity pattern of the disulfide bonds. Moreover, the complete absence of those disulfide bonds, after modeling the PdSN1 peptide with all the cysteines in their reduced form, also lead to an HTH motif stable in the microsecond timescale (this simulation was extended to 1.5 µs with the same result).

Altogether, these results suggest that the HTH motif is strongly determined by the sequence, and do not primarily rely on the disulfide bonds, explaining the observed antimicrobial activity of unfolded synthetic potato snakin-1 and -2, where the cysteines were in their reduced form (Harris et al. 2014). Furthermore, recombinant peptides produced in BL21(DE3) *E. coli*, were still active (Mao et al. 2011; Herbel et al. 2015). In this strain, the folding could not be guided by the formation of disulfide bridges, and reinforcing the hypothesis that disulfide bonds may be not essential for antimicrobial activity. The HTH is an evident evolutionary conserved motif of plant defense peptides, and the disulfide bonds could have appeared later in the evolution, collaborating to further stabilize the structural motif and providing and entropic advantage when considering the free energy of binding of snaking-1 to its biological targets.

As in the case of the potato snakin-1, PdSN1 also displayed a large positive electrostatic surface, with a pronounced electrophile cleft. It has been proposed that this may be particularity important in targeting these peptides to their site of action (Yeung et al. 2016). Since many antimicrobial peptides exert their function by disrupting (or interacting with) the negatively charged microbial membrane surfaces,

a lipid-membrane-targeting mechanism is the currently most accepted hypothesis about their mode of action. Nevertheless, the other natural target for positively charged peptides is DNA, which is the most negatively charged polymer found in nature (Cuervo et al. 2014). Moreover, the HTH motif is a well-established motif found in many proteins that regulate gene expression, like transcription factors, or the famous λ-repressor (Wintjens and Rooman 1996), where the motif is known to bind to the major groove of DNA.

The positive molecular electrostatic potential (MEP), added to the apparent requirement of the HTH motif for antimicrobial activity, leads us to hypothesize about a possible additional mechanism of action based on the deregulation of the microbial gene expression, which could be acting together with the lipid-membrane-targeting mechanism hypothesis. Indeed, some evidence already exists suggesting that snakin peptides could disrupt the microbial membrane (Herbel et al. 2015; Kuddus et al. 2016), besides some other types of AMPs that are known to interact with intracellular targets like DNA, RNA, or proteins (Brogden 2005; Lay and Anderson 2005). Considering defensins, for example, uncertainties have arisen on which are the main antimicrobial strategies: membrane permeabilization and leakage of cytoplasmic contents, or intracellular interaction with DNA/RNA and proteins from the synthesis machinery (Lay and Anderson, 2005). All these hypotheses about the mode of action clearly deserve future studies for confirmation, which are beyond the present contribution.

Conclusions

Given that snakins with antimicrobial activity are strong candidates for the development of novel biotechnological products meant for the control of diseases, we focused our work on the isolation of a new snakin gene from an unexplored genome such as a native, not cultivated tree with ethnobotanical tradition. Using a PCR-based strategy with degenerate primers, we successfully isolated a new snakin-like gene, named *PdSN1*. This gene was expressed in *E. coli*, allowing for the obtention of small amounts of a purified recombinant peptide with promising antimicrobial activity against relevant plant pathogens and opportunistic human fungi. Our 3D structural analysis brought some insights about possible relationships between the HTH motif present in the GASA domain and PdSN1 activity. Additional studies are ongoing in our laboratories to further characterize the antimicrobial activity of PdSN1 and to verify its possible membrane and/or DNA interactions.

Acknowledgements The authors thank CSIC (Comisión Sectorial de Investigación Científica, Universidad de la República, Uruguay), PEDECIBA (Programa de Desarrollo de las Ciencias Básicas, Uruguay), ANII (Agencia Nacional de Investigación e Innovación, Uruguay), CNPq (National Council for Scientific and Technological Development, Brazil), and CAPES (Coordination for the Improvement of Higher Education Personnel, Brazil) for financial support and fellowships. We also thank Dr. Rosario Durán and Madelón Portela for MALDI-TOFF analysis. SRD was supported by a doctoral fellowship from CSIC. IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO (Ministerio de Economía, Industria y Competitividad. Government of Spain). PDD and GC are PEDECIBA and SNI (Sistema Nacional de Investigadores, Uruguay) researchers.

Compliance with ethical standards

Conflict of interest The authors declare that no competing interests exist.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Almasia NI, Bazzini AA, Hopp HE, Vazquez-Rovere C (2008) Overexpression of snakin-1 gene enhances resistance to *Rhizoctonia solani* and *Erwinia carotovora* in transgenic potato plants. Mol Plant Pathol 9:329–338. https://doi.org/10.1111/J.1364-3703.2008.00469.X

Almasia NI, Molinari MP, Maroniche GA, Nahirñak V, Barrios Barón MP, Taboga OA, Vazquez-Rovere C (2017) Successful production of the potato antimicrobial peptide Snakin-1 in baculovirus-infected insect cells and development of specific antibodies. BMC Biotechnol 17:1–11. https://doi.org/10.1186/s12896-017-0401-2

Balaji V, Smart CD (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 Res 21:23–37. https://doi.org/10.1007/s11248-011-9506-x

Balaji V, Sessa G, Smart CD (2011) Silencing of host basal defense response-related gene expression increases susceptibility of *Nicotiana benthamiana* to *Clavibacter michiganensis* subsp. *michiganensis*. Phytopathology 101:349–357. https://doi.org/10.1094/PHYTO-05-10-0132

Benko-Iseppon AM, Galdino SL, Calsa T, Kido EA, Tossi A, Belarmino LC, Crovella S (2010) Overview on plant antimicrobial peptides. Curr Protein Pept Sci 11:181–188. https://doi.org/10.2174/138920310791112075

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3688. https://doi.org/10.1063/1.448118

Berrocal-Lobo M, Segura A, Moreno M, López G, García-Olmedo F, Molina A (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. https://doi.org/10.1104/pp.010685.1

Bradford MM (1976) A rapid and sensitive method for the quantitation microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254. https://doi.org/10.1016/0003-2697(76)90527-3

Broekaert WF, Terras F, Cammue BP, Vandedeyden J (1990) An automated quantitative assay for fungal growth inhibition. FEMS Microbiol Lett 69:55–60. https://doi.org/10.1016/0378-1097(90)90412-J

Brogden KA (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3:238–250. https://doi.org/10.1038/nrmicro1098

Case DA, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, Izadi S, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein D, Merz KM, Monard G, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Simmerling CL, Botello-Smith WM, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Xiao L, York DM, Kollman PA (2017) AMBER 2017. University of California, San Francisco

Ceroni A, Passerini A, Vullo A, Frasconi P (2006) DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. Nucleic Acids Res 34:W177–W181. https://doi.org/10.1093/nar/gkl266

Cheng J, Saigo H, Baldi P (2006) Large-scale prediction of disulphide bridges using kernel methods, two-dimensional recursive neural networks, and weighted graph matching. Proteins: structure, function. Bioinformatics 62:617–629. https://doi.org/10.1002/prot.20787

Cuervo A, Dans PD, Carrascosa JL, Orozco M, Gomila G, Fumagalli L (2014) Direct measurement of the dielectric polarization properties of DNA. Proc Natl Acad Sci 111:E3624–E3630. https://doi.org/10.1073/pnas.1405702111

Dans PD, Danilāne L, Ivani I, Dršata T, Lankaš F, Walther J, Illa Pujagut R, Battistini F, Gelpí JL, Lavery R, Orozco M (2016) Long-timescale dynamics of the Drew–Dickerson dodecamer. Nucleic Acids Res 44:4052–4066. https://doi.org/10.1093/nar/gkw264

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10095. https://doi.org/10.1063/1.464397

Doyle J (1991) DNA protocols for plants-CTAB total DNA isolation. In: Hewitt GM, Johnston A (eds) Molecular techniques in taxonomy. Springer, Berlin, pp 283–293

Ferrè F, Clote P (2006) DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. Nucleic Acids Res 34:W182–W185. https://doi.org/10.1093/nar/gkl189

Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci 85:8998–9002

García AN, Ayub ND, Fox AR, Gómez MC, Diéguez MJ, Pagano EM, Berini CA, Muschietti JP, Soto G (2014) Alfalfa snakin-1 prevents fungal colonization and probably coevolved with rhizobia. BMC Plant Biol 14:248. https://doi.org/10.1186/s12870-014-0248-9

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Harris PWR, Yang SH, Molina A, Lopez G, Middleditch M, Brimble MA (2014) Plant antimicrobial peptides snakin-1 and snakin-2: chemical synthesis and insights into the disulfide connectivity. Chemistry 20:5102–5110. https://doi.org/10.1002/chem.201303207

Herbel V, Schäfer H, Wink M (2015) Recombinant production of snakin-2 (an antimicrobial peptide from tomato) in *E. coli* and analysis of its bioactivity. Molecules 20:14889–14901. https://doi.org/10.3390/molecules200814889

Herbel V, Sieber-frank J, Wink M (2017) The antimicrobial peptide snakin-2 is upregulated in the defense response of tomatoes (*Solanum lycopersicum*) as part of the jasmonate-dependent signaling pathway. J Plant Physiol 208:1–6. https://doi.org/10.1016/j.jplph.2016.10.006

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935. https://doi.org/10.1063/1.445869

Kovaleva V, Krynytskyy H, Gout I, Gout R (2011) Recombinant expression, affinity purification and functional characterization of Scots pine defensin 1. Appl Microbiol Biotechnol 89:1093–1101. https://doi.org/10.1007/s00253-010-2935-2

Kovalskaya N, Hammond RW (2009) Expression and functional characterization of the plant antimicrobial snakin-1 and defensin recombinant proteins. Protein Expr Purif 63:12–17. https://doi.org/10.1016/j.pep.2008.08.013

Kuddus MR, Rumi F, Tsutsumi M, Takahashi R, Yamano M, Kamiya M, Kikukawa T, Demura M, Aizawa T (2016) Expression, purification and characterization of the recombinant cysteine-rich antimicrobial peptide snakin-1 in *Pichia pastoris*. Protein Expr Purif 122:15–22. https://doi.org/10.1016/j.pep.2016.02.002

Lay FT, Anderson MA (2005) Defensins—components of the innate immune system in plants. Curr Protein Pept Sci 6:85–101. https://doi.org/10.2174/1389203053027575

Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M (2012) SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact 11:753. https://doi.org/10.1186/1475-2859-11-56

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99sb. J Chem Theor Comput 11:3696–3713. https://doi.org/10.1021/acs.jctc.5b00255

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

Meiyalaghan S, Thomson SJ, Fiers MW, Barrell PJ, Latimer JM, Mohan S, Jones EE, Conner AJ, Jacobs JM (2014) 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 Genom 15:2. https://doi.org/10.1186/1471-2164-15-2

Mohan S, Meiyalaghan S, Latimer JM, Gatehouse ML, Monaghan KS, Vanga BR, Pitman AR, Jones EE, Conner AJ, Jacobs JME (2014) GSL2 over-expression confers resistance to *Pectobacterium atrosepticum* in potato. Theor Appl Genet 127:677–689. https://doi.org/10.1007/s00122-013-2250-2

Nahirñak V, Almasia NI, Fernandez PV, Hopp HE, Estevez JM, Carrari F, Vazquez-Rovere C (2012a) Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. Plant Physiol 158:252–263. https://doi.org/10.1104/pp.111.186544

Nahirñak V, Almasia NI, Hopp HE, Vazquez-Rovere C (2012b) Involvement in hormone crosstalk and redox homeostasis Snakin/GASA proteins. Plant Signal Behav 7:1004–1008. https://doi.org/10.4161/psb.20813

Nahirñak V, Rivarola M, Gonzalez de Urreta M, Paniego N, Hopp HE, Almasia NI, Vazquez-Rovere C (2016) Genome-wide analysis of the snakin/gasa gene family in *Solanum tuberosum* cv. Am J Potato Res, Kennebec. https://doi.org/10.1007/s12230-016-9494-8

Oliveira-Lima M, Benko-Iseppon AM, Ferreira Costa RJ, Rodríguez-Decuadro S, Kido EA, Crovella S, Pandolfi V (2017) Snakin: structure, roles and applications of a plant antimicrobial peptide. Curr Protein Pept Sci 18:1–7. https://doi.org/10.2174/1389203717666161606191

Padovan L, Crovella S, Tossi A, Segat L (2010) Techniques for plant defensin production. Curr Protein Pept Sci 11:231–235. https://doi.org/10.2174/138920310791112101

Pérez A, Luque FJ, Orozco M (2007) Dynamics of B-DNA on the microsecond time scale. J Am Chem Soc 129:14739–14745. https://doi.org/10.1021/ja0753546

Pestana-Calsa MC, Ribeiro ILAC Jr (2010) Bioinformatics-coupled molecular approaches for unravelling potential antimicrobial peptides coding genes in Brazilian native and crop plant species. Curr Protein Pept Sci 11:199–209. https://doi.org/10.2174/138920310791112138

Porto WF, Franco OL (2013) Theoretical structural insights into the snakin/GASA family. Peptides 44:163–167. https://doi.org/10.1016/j.peptides.2013.03.014

Richero M, Barraco-Vega M, Cerdeiras MP, Cecchetto G (2013) Development of SCAR molecular markers for early and late differentiation of *Eucalyptus globulus* ssp globulus from *E. globulus* ssp *maidenii*. Trees 27:249–257. https://doi.org/10.1007/s00468-012-0792-6

Roe DR, Cheatham TE III (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 9:3084–3095. https://doi.org/10.1021/ct400341p

Rong W, Qi L, Wang J, Du L, Xu H, Wang A, Zhang Z (2013) Expression of a potato antimicrobial peptide SN1 increases resistance to take-all pathogen *Gaeumannomyces graminis* var. *tritici* in transgenic wheat. Funct Integr Genom 13:403–409. https://doi.org/10.1007/s10142-013-0332-5

Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. J Comput Phys 23:327–341. https://doi.org/10.1016/0021-9991(77)90098-5

Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond molecular dynamics simulations with amber on gpus. 2. Explicit solvent particle mesh Ewald. J Chem Theory Comput 9:3878–3888. https://doi.org/10.1021/ct400314y

Savojardo C, Fariselli P, Alhamdoosh M, Martelli PL, Pierleoni A, Casadio R (2011) Improving the prediction of disulfide bonds in Eukaryotes with machine learning methods and protein subcellular localization. Bioinformatics 27:2224–2230. https://doi.org/10.1093/bioinformatics/btr387

Schägger G, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 379:368–379. https://doi.org/10.1016/0003-2697(87)90587-2

Segura A, Moreno M, Madueño F, Molina A, García-Olmedo F (1999) Snakin-1, a peptide from potato that is active against plant pathogens. Mol Plant Microbe Interact 12:16–23. https://doi.org/10.1094/MPMI.1999.12.1.16

Shi Y, Mowery RA, Ashley J, Hentz M, Ramirez AJ, Bilgicer B, Slunt-brown H, Borchelt DR, Shaw BF (2012) Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. Protein Sci 21:1197–1209. https://doi.org/10.1002/pro.2107

Silva ON, Mulder KCL, Barbosa AEAD, Otero-Gonzalez AJ, Lopez-Abarrategui C, Rezende TMB, Dias SC, Franco OL (2011) Exploring the pharmacological potential of promiscuous host-defense peptides: from natural screenings to biotechnological applications. Front Microbiol 2:1–14. https://doi.org/10.3389/fmicb.2011.00232

Smith DE, Dang LX (1994) Computer simulations of NaCl association in polarizable water. J Chem Phys 100:3757–3762. https://doi.org/10.1063/1.466363

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods research resource. Mol Biol Evol 28:2731–2739. https://doi.org/10.1093/molbev/msr121

Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 60:523–533. https://doi.org/10.1007/s00253-002-1158-6

Thevissen K, Kristensen H, Thomma BPHJ, Cammue BPA, Francois IEJA (2007) Therapeutic potential of antifungal plant and insect defensins. Drug Discov Today 12:966–971. https://doi.org/10.1016/j.drudis.2007.07.016

Tossi A, Sandri L (2002) Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. Curr Pharm Des 8:743–761. https://doi.org/10.2174/1381612023395475

Trapalis M, Li SF, Parish RW (2017) The Arabidopsis GASA10 gene encodes a cell wall protein strongly expressed in developing anthers and seeds. Plant Sci 260:71–79. https://doi.org/10.1016/j.plantsci.2017.04.003

Tuppo L, Alessandri C, Pomponi D, Picone D, Tamburrini M, Ferrara R, Petriccione M, Mangone I, Palazzo P, Liso M, Giangrieco I, Crescenzo R, Bernardi M, Zennaro D, Helmer-Cittrerich M, Mari A, Ciardiello M (2013) Peamaclein—a new peach allergenic protein: similarities, differences and misleading features compared to Pru p 3. Clin Exp Allergy 43:128–140. https://doi.org/10.1111/cea.12028

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:Research 0034. https://doi.org/10.1186/gb-2002-3-7-research0034

Vila-Perelló M, Sanchez-Vallet A, Garcıa-Olmedo F, Molina A, Andreu D (2005) Structural dissection of a highly knotted peptide reveals minimal motif with antimicrobial activity. J Biol Chem 280:1661–1668. https://doi.org/10.1074/jbc.M410577200

Webb B, Sali A (2014) Comparative protein structure modeling using modeller. Curr Protoc Bioinform. https://doi.org/10.1002/0471250953.bi0506s15

Wintjens R, Rooman M (1996) Structural classification of hth dna-binding domains and protein–dna interaction modes. J Mol Biol 262:294–313. https://doi.org/10.1006/jmbi.1996.0514

Xu D, Zhang Y (2012) *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80:1715–1735. https://doi.org/10.1002/prot.24065

Yang J, He BJ, Jang R, Zhang Y, Shen HB (2015) Accurate disulfide-bonding network predictions improve ab initio structure prediction of cysteine-rich proteins. Bioinformatics 31:3773–3781. https://doi.org/10.1093/bioinformatics/btv459

Yeaman MR, Yount NY (2003) Mechanisms of antimicrobial peptide action and resistance. Pharmacol Rev 55:27–55. https://doi.org/10.1124/pr.55.1.2

Yeung H, Squire CJ, Yosaatmadja Y, Panjikar S, López G, Molina A, Baker EN, Harris PWR, Brimble MA (2016) Radiation damage and racemic protein crystallography reveal the unique structure of the gasa/snakin protein superfamily. Angew Chem Int Ed 55:7930–7933. https://doi.org/10.1002/ange.201602719

Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature 415:389–395. https://doi.org/10.1038/415389a

Zhu J, Zhang L, Li W, Han S, Yang W, Qi L (2013) Reference gene selection for quantitative real-time pcr normalization in *Caragana intermedia* under different abiotic stress conditions. PLoS One 8:e53196. https://doi.org/10.1371/journal.pone.0053196

Capítulo 2

Identificación y caracterización estructural y funcional de una defensina de ceibo

Este capítulo incluye la búsqueda de genes de tipo defensina en los genomas de ibirapitá, congorosa y ceibo, el clonado de un gen de *E. crista-galli* con similitud defensinas (*EcgDf1*) por amplificación con *primers* específicos y el borrador de un artículo científico que describe el aislamiento de *EcgDf1* y su expresión en *E. coli*. Además, se predijo la estructura 3D del péptido y se evaluó la actividad antimicrobiana del péptido recombinante purificado, contra un grupo de microorganismos patógenos. La producción recombinante de EcgDf1 fue realizada por la Lic. Alejandra Borba en el marco de su tesina de grado, con mi co-dirección. La predicción de la estructura 3D de EcgDf1 fue realizada por el Dr. Pablo Dans que utilizó técnicas de modelado por homología y simulaciones de dinámica molecular para dichas predicciones.

Búsqueda de genes de tipo defensina en los genomas de ibirapitá, congorosa y ceibo

Para el aislamiento de genes de tipo defensina en las especies vegetales seleccionadas, se buscaron secuencias de defensinas en las bases de datos (GenBank, Uniprot, Phytozome) y se alinearon para identificar regiones conservadas. Para diseñar primers, debido al alto número de secuencias de defensina, sólo se utilizaron secuencias dentro de la subclase Rosidae, a la que pertenecen las tres especies elegidas (Figura 1). Se diseñaron *primers* degenerados para las regiones del principio y fin del péptido precursor (varios pares, por ejemplo, dentro de la familia Fabidae, o más amplios) y también para regiones conservadas dentro del péptido maduro (Tabla 1). Se utilizaron varios pares de *primers* combinados (por ejemplo, primer F interno y primer R externo y viceversa) (Figura 2) y diferentes condiciones y estrategias de amplificación a partir de ADN genómico. Además, se usaron *primers* específicos diseñados para una defensina de *Medicago sativa* (alfalfa) y una de *Vigna unguiculata* (poroto tape) diseñados por Padovan et al. (2010b) (Tabla 1).

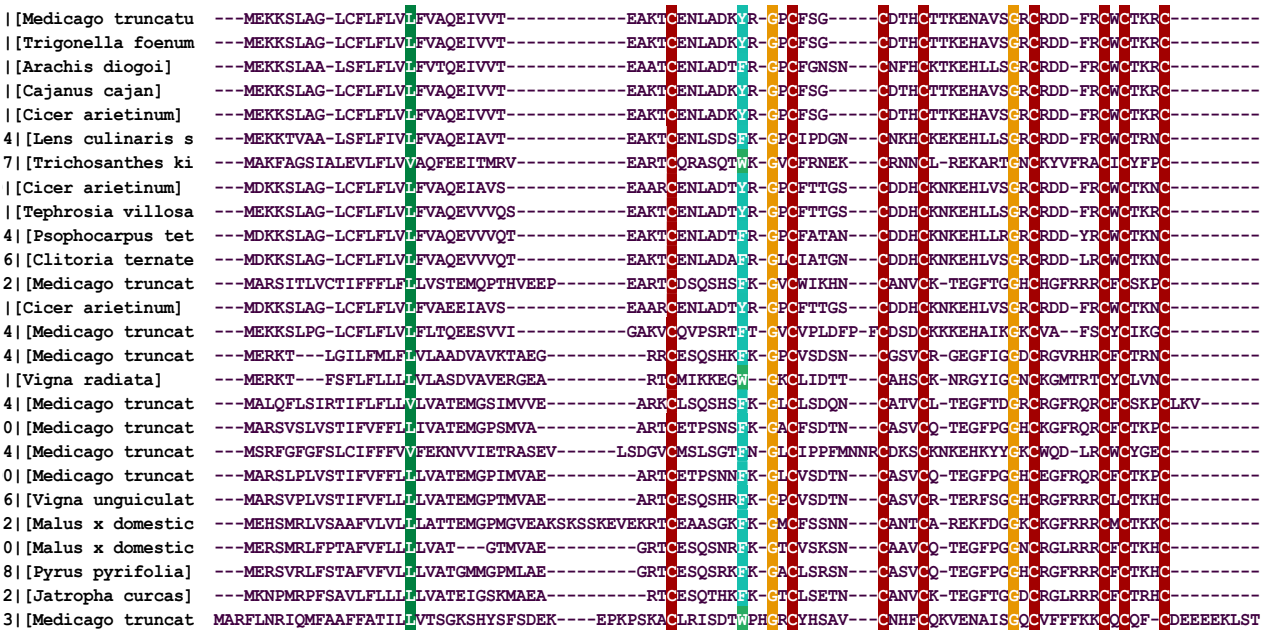


Figura 1. Alineamiento representativo de defensinas pertenecientes a especies vegetales de la subclase Rosidae, obtenidas de las bases de datos. Se incluyeron varias defensinas de *Medicago truncatula*, cuyo genoma está secuenciado

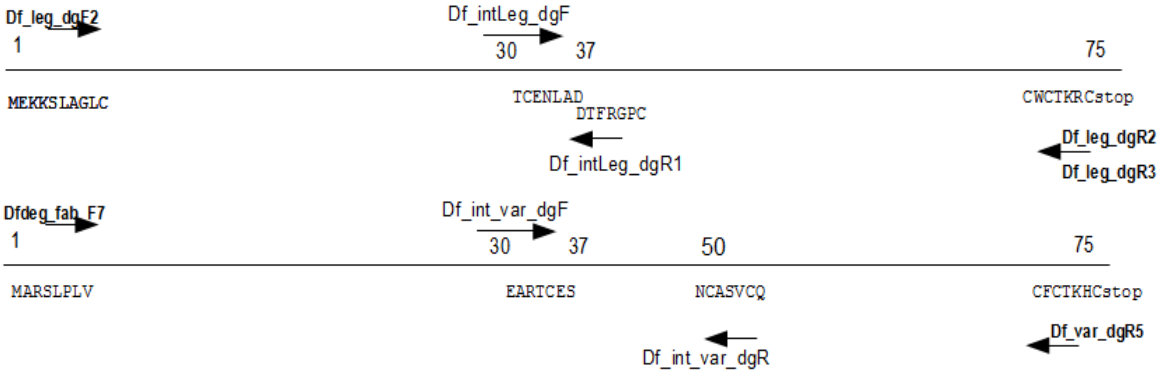


Figura 2. Esquemas de ubicación de los distintos *primers* degenerados diseñados para defensinas

Tabla 1. *Primers* degenerados y específicos usados para la búsqueda de genes de tipo defensina en los genomas de *P. dubium* y *M. ilicifolia y E. crista-galli*

Nombre	Secuencia
DfA_dgF1	ATGGCTAAGTTTGYYWCCAT
DfA_dgR1	TTAACATGGGAMRTARCAGATACA
Df_leg_dgF2	ATGGAGAAGAAAWCABTAGCYG
Df_leg_dgR2	TTAACATYTTTTRGTACACCARC
Df_leg_dgR3	TTAACATYTTTTAGTGCACCARC
Df_jat_dgF4	ATGAAGAATCCTATGCGTYYRTT
Df_var_dgR5	TAACARTGYYTAGTGCANAAGCA
Df_ros_dgF6	ATGGAGCGNTCCRTRCGTTTR
Df_fab_dgF7	ATGGCTCGYTCWSTKCCTTTG
Df_intLeg_dgF1	ACTTGYGAGAAYYTGGCWGA
Df_intLeg_dgR1	CATGGTCCCCTRWAYKTATC
Df_intvar_dgF1	GAGGCWAGGACHTGYGAGTCT
Df_intvar_dgR2	TYGGCAAACAKWWGCACAGTT
Df_leg_dgF2	ATGGAGAAGAAAWCABTAGCYG
Df_leg_dgR2	TTAACATYTTTTRGTACACCARC
Df_leg_dgR3	TTAACATYTTTTAGTGCACCARC
Df_jat_dgF4	ATGAAGAATCCTATGCGTYYRTT
Df_var_dgR5	TAACARTGYYTAGTGCANAAGCA
Df_ros_dgF6	ATGGAGCGNTCCRTRCGTTTR
Df_fab_dgF7	ATGGCTCGYTCWSTKCCTTTG
Df_intLeg_dgF1	ACTTGYGAGAAYYTGGCWGA
Df_intLeg_dgR1	CATGGTCCCCTRWAYKTATC
Df_intvar_dgF1	GAGGCWAGGACHTGYGAGTCT
Df_intvar_dgR2	TYGGCAAACAKWWGCACAGTT
Dfsp_Msat_F1	ATGGACAAGAAATCACTAGCAG
Dfsp_Msat_R1	TTAACATCTTTTAGTACACCAACAACG
Df_PadF	TCCATGGCTCGCTCTGTGTCTT
Df_PadR	TGAAGTTTTAACAGTGTTTGGTGCACAAG

En las condiciones en las que se obtuvo amplificación generalmente se observaron perfiles de múltiples bandas (Figura 3). Se eligieron bandas entre 200 y 900 pb considerando aproximadamente 250 pb de péptido y la posible presencia de intrón. Cuando los *primers* se ubicaban en el péptido maduro se cortaron bandas en rango de 100-800 pb. Cada producto de amplificación en esos rangos fue purificada por separado y clonada; varios clones (de cada banda de amplificación) fueron secuenciados. En congorosa casi no se detectó amplificación con las distintas combinaciones de *primers.* Entre todos los clones analizados, sólo se detectó similitud con genes de defensinas cuando se usaron los *primers* específicos diseñados para *V. unguiculata* (Padovan et al. 2010b) proveniente de ceibo. La banda recortada presentó un tamaño mayor a lo esperado (Figura 4) ya que el gen identificado de defensina de *V. unguiculata* presenta en total unos 387 pb, incluyendo un intrón de aproximadamente 150 pb (Padovan et

al. 2010b). Para *M. ilicifolia* y *P dubium* no se encontraron genes con similitud con defensinas mediante esta metodología.

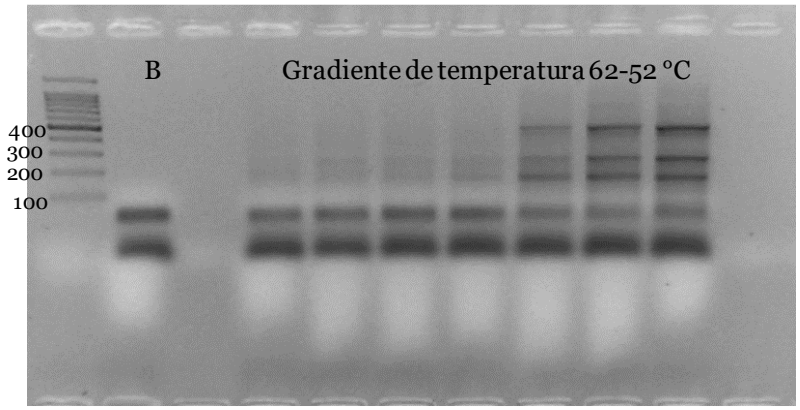


Figura 3. Amplificación por PCR usando *primers* degenerados diseñados para defensinas a partir de ADN genómico de ibirapitá. Electroforesis en gel de agarosa 2 %. MPM: marcador de peso molecular 100 pb (GeneRuler 100 bp DNA Ladder; Thermo Scientific). B: blanco. Se usó gradiente de temperaturas

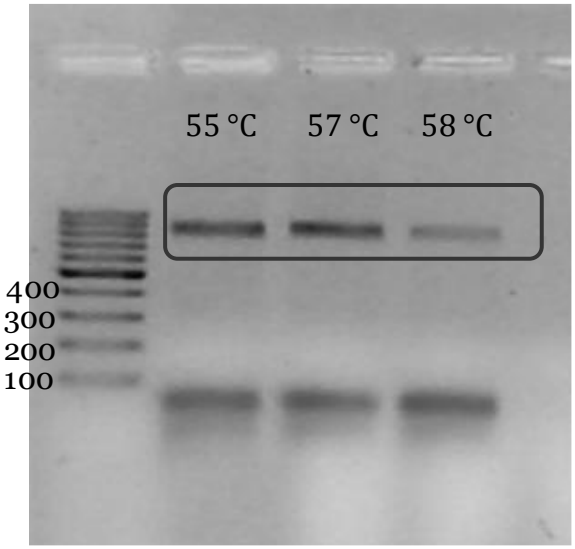


Figura 4. Amplificación por PCR a partir de ADN genómico de ceibo usando *primers* específicos de defensina de *V. unguiculata*. Electroforesis en gel de agarosa 2 %. MPM: marcador de peso molecular 100 pb (GeneRuler 100 bp DNA Ladder; Thermo Scientific). Se muestran tres temperaturas de *annealing*

Gene isolation and recombinant production of a new legume defensin with a broad spectrum of antimicrobial activity

Abstract

Plant defensins comprise a large family of antimicrobial peptides with a wide range of biological functions. They are cysteine-rich molecules, highly sequence-diverse but with a conserved and stable structure. In this work, a defensin gene (*EcgDf1*) was isolated from *Erythrina crista-galli*, a legume tree native from South America. *EcgDf1* gene structure is similar to other defensins, with two exons and one intron whereas the second exon entirely encodes the mature peptide The predicted EcgDf1 peptide present features common to all of plant defensins with 3D structural information, presenting eight cysteines, with a γ-core motif $GXCX_{3-9}C$ and six cysteines distributed like the typical defensin αβ motif. The mature EcgDf1 coding sequence was heterologously expressed in *Escherichia coli* strains and purified by affinity chromatography. Possible dimer and oligomers of EcgDf1 were visible in SDS gel electrophoresis. Moreover, its 3D structure, determined by homology modeling and molecular dynamics simulation, was found to be compatible with the formation of homo-dimers. The purified recombinant peptide inhibits the growth of some critical plant and human pathogens, such us the opportunistic fungi *Candida albicans* and *Aspergillus niger* and the plant pathogens *Clavibacter michiganensis* ssp. *michiganensis, Penicillium expansum*, *Botrytis cinerea,* and *Alternaria alternata.* Since EcgDf could be a promising candidate for the development of antimicrobial products for use in agriculture and medicine, further studies are necessary to understand their mode of action and their potential toxicity.

Keywords (4-6) *Erythrina crista-galli*, antimicrobial peptides, heterologous expression, homology modeling, dimerization.

Introduction

The production of antimicrobial peptides (AMPs) is an important mechanism of eukaryotic innate immunity. Among these peptides, defensins seem to be the only class that is conserved between plants, invertebrates, and vertebrates (Thomma et al. 2002). They belong to the PR-12 (Pathogenesis-Related) class (Sels et al. 2008) and, like other AMPs, they are small (~5 kDa), basic, cysteine-rich and exhibit an amphipathic structure (Yeaman and Yount 2003; Brogden 2005). Their three-dimensional structure presents one α-helix and three antiparallel β-sheets that are stabilized by four intramolecular disulfide bonds formed by eight highly conserved cysteine residues (Lay and Anderson 2005; Carvalho and Gomes 2009). This compacted structure gives them resistance to extreme temperatures and degradation mediated by proteases (Hegedüs and Marx 2013). Despite the similarity and conservation of their three-dimensional structures, plant defensins have surprisingly, significant amino acid sequence variability (Lay and Anderson 2005). The primary structure of these peptides generally consists of a signal peptide at the amino-terminal that targets the peptide to the extracellular space, and a basic, cysteine-rich mature peptide (45 to 54 amino acids). However, defensins with alternative structures have been identified (De Coninck et al. 2013).

Initially isolated from seeds, studies of protein localization and gene expression show that plant defensins can be found in all plant organs, including fruits, flowers, pollen, leaves, cotyledons, roots, tubers (Carvalho and Gomes 2009). Some defensins are expressed exclusively in very specific parts of one organ while others are constitutively expressed in a wide range of tissues (De Coninck et al. 2013). The constitutive expression of defensin genes constitutes a primary defense of the plant against pathogen attack (Broekaert et al. 1995), but the production of these molecules can also be induced by a wide range of biotic stresses (Lay and Anderson 2005). In addition to their function as defense molecules, defensins perform other roles related to plant physiology including growth regulation, development, fertilization (Hegedüs and Marx 2013) and resistance to different types of abiotic stresses such as cold (Koike et al. 2002), salinity and drought (Do et al. 2004).

Most plant defensins exhibit *in vitro* activity against a range of microorganisms, especially filamentous fungi, and yeasts (Carvalho and Gomes 2009). Furthermore, the overexpression of defensins in transgenic plants increased resistance against various fungal diseases (Gao et al. 2000; Portieles et al. 2010). Such previous evidences, as well as their induction under biotic stress support the theory that these peptides play a role in plant defense response against pathogen attack. These biological properties make defensins as promising molecules for the

development of therapeutic agents, not only for agriculture but also in medicine since these peptides display antimicrobial activity also against human fungal pathogens like *Candida* spp. and *Aspergillus* spp. (Thevissen et al. 2007). It has also been reported that plant defensins can inhibit certain types of cancer (Wang et al. 2009; Hegedüs and Marx 2013).

The high potential of defensins as antimicrobials has encouraged research initiatives for the development of novel products with biotechnological applications, resulting in a significant number of new defensin sequences available in databases. Strategies of gene isolation joined with recombinant production have been increasingly used for the characterization of defensin peptides, replacing the laborious and low-yield classic strategy of purification of endogenous defensins from plants (Osborn et al. 1995; Games et al. 2008; Slavokhotova et al. 2011). There are several examples for defensin production in bacterial and yeast systems (Cabral et al. 2003; Vijayan et al. 2008; Kant et al. 2009; Kovalskaya and Hammond 2009; Portieles et al. 2010; de Beer and Vivier 2011; Kovaleva et al. 2011; Mello et al. 2014; Shenkarev et al. 2014; Ferreira Lacerda et al. 2016; Guillén-Chable et al. 2017). In all of them, recombinant defensins with antimicrobial activity were obtained.

Erythrina crista-galli is a legume tree belonging to the Papilionoideae subfamily. Most of Erythrina L. are trees or shrub and are distributed throughout the tropics and subtropics. Particularly, *E. crista-galli* forms extensive populations along the estuary of the Rio de La Plata in extra-tropical South America (Neill 1988). It has been used in traditional medicine as antimicrobial for the treatment of infections, wound healing, and throat infections and it has been observed that hydromethanolic extracts of *E. crista-galli* exhibited antiviral activity (Simoes et al. 1999). In this work, we describe the successful isolation of a plant defensin gene from this South American legume and its heterologous expression in *Escherichia coli*. The purified recombinant peptide was biologically active against a set of microorganisms, including plant pathogens and opportunistic human fungi like *Candida albicans* and *Aspergillus niger*. The structure of the new defensin, determined by homology modeling and molecular dynamics simulation, was found to be compatible with the formation of homo-dimers.

Materials and Methods

Biological Material

E. crista-galli seeds and leaves were obtained from Facultad de Agronomía Garden (Montevideo, Uruguay). Surface sterilized seeds were germinated on Whatman paper soaked with distilled

water in Petri dishes at 28 ºC. Leaves and four days old seedlings were frozen in liquid nitrogen and stored at -70 ºC until further usage.

Bacteria and fungi strains used for bioactivity assays were obtained from Laboratory of Microbiología Facultad de Química collection. Fungal cultures (*Candida albicans* CCMG13, *Aspergillus niger* CCMG17, *Botrytis cinerea* CCMG14g, *Alternaria alternata* CBS916.96, *Penicillium expansum* CCMG14s) were grown on potato dextrose agar (PDA) at 28 ºC. *Clavibacter michiganensis* ssp. *michiganensis* MAI1008 and *Xanthomonas versicatoria* MAI2020 were grown on Nutrient Agar at 28 ºC.

E. coli TOP10 (Invitrogen, Carlsbad, USA), BL21(DE3) (New England Biolabs, Ipswich, USA) and Rosetta-gami(DE3)pLysS (Novagen, Madison, USA) were used as cloning or expression hosts respectively and were grown on Luria–Bertani.

Isolation of *E. crista-galli* DNA and RNA

Genomic DNA was extracted from fresh leaves using the standard cetyl-trimethylammonium bromide (CTAB) method (Doyle 1991). Total RNA was extracted using Qiagen RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and treated with RNase-free DNase (Invitrogen, Carlsbad, USA) to eliminate any residual DNA. Genomic and plasmid DNA were visualized under UV illumination after electrophoresis on agarose gels in 1X TBE buffer, stained with GoodView (Ecoli s.r.o., Bratislava, Slovak Republic).

EcgDf1 gene isolation and cloning

EcgDf1 gene was amplified, from *E. crista-galli* genomic DNA by using primers designed by Padovan et al. (2010) for another legume, *Vigna unguiculata*: Forward 5'-TCCATGGCTCGCTCTGTGTCTT-3' and Reverse 5'-TGAAGTTTTAACAGTGTTTGGTGCACAAG-3'. The PCR reaction was performed in a 20 µl reaction containing: 1X buffer with 2 mM MgCl2, 0.2 mM dNTPs, 2 µM of each primer, 50 ng template DNA and 0,5 U Taq DNA polymerase (Invitrogen, Carlsbad, USA). The PCR program was as follows: 94 °C for 3 min; followed by 30 cycles of 94 °C for 30 sec, 57 °C for 40 sec and 72 °C for 40 sec. PCR products were purified according to Richero et al. (2013) and cloned into a pGEM-T easy vector (Promega Corporation, Madison, USA). *EcgDf1* gene sequence was confirmed by sequencing four clones at Macrogen Inc. (Seoul, Korea) and analyzed using BLAST (Altschul et al. 1997).

EcgDf1 coding sequence corresponding to the mature peptide was verified by the synthesis of 3' cDNA region from total RNA extracted from *E. crista-galli* seedlings. The 3' region was obtained by 3'-RACE according to Frohman *et al*. (1988) using the same forward oligonucleotide mentioned above. PCR products were cloned in the pGEM-T easy vector (Promega Corporation,

Madison, USA) and sequenced. The resulting plasmid (pG-EcgDf1c) contains the complete coding sequence of the mature peptide.

Sequence analysis.

The deduced amino acid sequence of EcgDf1 was analyzed with the Expasy-ProtParam tool(Gasteiger et al. 2005) to obtain the different peptide parameters. The presence of a signal peptide sequence was evaluated with SignalP 7.0.5 (Nielsen 2017). The deduced amino acid sequence encoding for the mature *E. crista-galli* peptide was aligned against a set of mature plant defensins taken from the manually annotated and reviewed UniProt database(Apweiler et al. 2004). Alignment was performed in ClustalW from BioEdit program (Hall 1999). An unrooted tree was generated using the Neighbor-Joining method of the MEGA package version 10.0.4 (Tamura et al. 2011), with 5000 bootstrap replicates.

Recombinant production of EcgDf1 in *E. coli*

Escherichia coli Rosetta-gami and BL21(DE3) strains and pET102/D-TOPO (Invitrogen, Carlsbad, USA) were used as a host/vector system for expression of *EcgDf1* gene. The expression plasmid was constructed using a restriction-free cloning strategy (Van Den Ent et al. 2006). The mature coding sequence of *EcgDf1* was PCR amplified from pG-EcgDf1c using the following primers: EcgDf1_pEF 5'-TCCGGTGATGACGATGACAAGAGAACATGCGAGTCTCAAAGC -3' and EcgDf1_pER 5'-GGATCAAACTCAATGGTGATGGTGATGATGCCTTCCCTCGATACAGTGTTTGGTGCAGAAGCAT-3'. Reverse primer was 12 nucleotides long which code for the factor Xa recognition site. Blunt-end PCR products were cloned in pET102/D using a Phusion High-Fidelity DNA Polymerase (Thermo-Scientific, Waltham, USA). pET102/D without *EcgDf1* sequence (pE_Trx) was constructed previously (Rodríguez-Decuadro et al. 2018) as a control. The transformation of *E. coli* TOP10 was performed according to the manufacturer's instructions. The resulting plasmid pE_EcgDf1 was verified by DNA sequencing. pE_EcgDf1 and pE_Trx (control) were introduced into *E. coli* Rosetta-gami (DE3)pLysS and BL21(DE3) strains. Cells were grown and harvested as described in Rodríguez-Decuadro et al. (2018) with varying post-induction time (4 and 24 h) and temperature (20, 28 °C), being stored at −70 °C until protein extraction.

Purification of recombinant EcgDf1

The cell pellet was resuspended in 2 ml of ice-cold lysis buffer (50 mM potassium phosphate buffer, pH 7.5, 500 mM NaCl, 10% glycerol, 10 mM imidazole). The cells were lysed by sonication 4x pulse at 4 °C (30% amplitude; 6 cycles of: 6 s pulse-on, 9 s pulse-off; 15 sec) (Cole-Parmer, Vernon Hills, USA). The lysate was centrifuged 15 min at 4 ºC and 10,000 rpm,

and the supernatant was filtered through a 0.22-µm filter (Millipore). Aliquots of the crude extract or culture supernatant were analyzed by 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using Tris-Glycine buffer (Laemmli 1970).

Supernatant from the expression culture was applied to a Chelating Sepharose Fast Flow column (GE Healthcare Little Chalfont, United Kingdom) loaded with Cu2+ and equilibrated with lysis buffer. The soluble fraction was left on the column for one hour in batch at 4 °C. Unbound proteins were eluted by adding 20 and 30 mM imidazole in the lysis buffer, while 300 mM imidazole was used to elute the bound proteins. In order to remove the thioredoxin fusion fragment, the Trx-EcgDf1 protein was subjected to EkMax Enterokinase (Invitrogen) digestion for 15 and 120 h at room temperature. Eluted fractions were analyzed by 15 % sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using Tris-Tricine buffer (Schägger and von Jagow 1987). Fractions containing recombinant peptide were pooled and dialyzed against distilled water. The protein concentration was estimated by Bradford assay (Bradford 1976), using bovine serum albumin as a standard. The recombinant peptide was lyophilized and storage at -20 °C for further usage.

Identification of recombinant EcgDf1 peptide.

Fractions containing lyophilized recombinant peptide were resuspended in H_20. The molecular mass of recombinant peptide was determined by MALDI-TOF mass spectrometry on a 4800 MALDI TOF/TOF Analyzer (Abi Sciex). Mass spectrometry analysis was performed in a positive linear mode using a solution of -cyano-4-hydroxycinnamic acid in 60% acetonitrile, 0.1% TFA as a matrix. For external calibration, a standard of cytochrome C was used (Abi Sciex).

The peptide sequence was confirmed by peptide mass fingerprinting and subsequent fragmentation of selected peptides (MS/MS). Disulfide bonds were reduced by incubation with DTT (100mM) for 1h at 56 ºC and Cys were alkylated with iodoacetamide (300 mM) Tryptic digestion was performed overnight at 37 ºC using sequencing grade trypsin (Promega). Peptide mixtures were mixed with matrix solution -cyano-4-hydroxycinnamic acid in 60 % acetonitrile, 0.1 % TFA) directly onto the sample plate. Spectra were acquired in positive reflector mode and were externally calibrated using a standard mixture of peptides (Applied Biosystems). Peptide sequences were assigned using an in-house Mascot v.2.3 version (Matrix Sience) for searching a local database that includes the sequence of EcgDf1. The following parameters: were used with variable modifications: Oxidation (M),Carbamidomethyl (C),Ammonia-loss (N-term C); mass values: Monoisotopic; peptide mass tolerance: ± 0.05 Da; Fragment Mass Tolerance: ± 0.45 Da. Selected m/z values were further fragmented, and the assignment to EcgDf1 amino acid sequence was validated by manual inspection of MS/MS spectra.

3D structure prediction

A protein-protein BLAST analysis was achieved against the Protein Data Bank (PDB), finding 81% of identity of the EcgDf1 sequence with the defensin of *Medicago truncatula*, which was determined by solution NMR (PDB id: 2LR3). The program Modeller 9.20 was used to build 10 minimized structures using the first structures deposited as template (Webb and Sali 2014). The two most distinct structures obtained with Modeller were simulated at physiological conditions by means of short, unbiased Molecular Dynamics simulations. To do so, the two models were minimized *in vacuo*, neutralized with 16 Cl^- (note that H++ server predicted all histidines, including those in the extra tail, to be protonated at pH 6.5), solvated (with explicit water and 0.15 M of K^+Cl^-), and minimized in solution with positional restraints on the peptide using our well-established multi-step protocol (Pérez et al. 2007; Dans et al. 2016). To produce the final models, the minimized structures were thermalized to 298 ºC at NVT, and then simulated during at least 100 ns using Molecular Dynamics simulations at NPT (P = 1 atm). For representing the peptide, we used the state-of-the-art ff14SB force field (Maier et al. 2015), surrounded by a truncated octahedral box of ~5,000 TIP3P water molecules (Jorgensen et al. 1983), and Dang parameters for ions (Smith and Dang 1994). Ions were initially placed randomly, at a minimum distance of 5 Å from the solute and 3.5 Å from one another. All systems were simulated using the Berendsen algorithm (Berendsen et al. 1984) to control the temperature and the pressure, with a coupling constant of 5 ps. Center of mass motion was removed every 10 ps to limit the build-up of the translational kinetic energy of the solute. SHAKE (Ryckaert et al. 1977) was used to keep all bonds involving hydrogen at their equilibrium values, allowing the use of a 2 fs step for the integration of Newton equations of motion. Long-range electrostatic interactions were accounted for by using the Particle Mesh Ewald method (Darden et al. 1993) with standard defaults and a real-space cutoff of 9 Å. All simulations were carried out using the PMEMD CUDA code module (Salomon-Ferrer et al. 2013) of AMBER 16 (Case et al. 2017) and were analyzed with CPPTRAJ (Roe and Cheatham III 2013). Structures and MD trajectories were visually analyzed using VMD 1.9 (Humphrey et al. 1996).

Antimicrobial activity assay

Antimicrobial activity was estimated microspectrophotometrically (Broekaert et al. 1990), determining the IC50 value (peptide concentration at which 50 % inhibitions was reached). Fungal spore suspension (1×10^4 spores ml^{-1}) or bacterial suspension (1×10^5 bacteria ml^{-1}) were cultivated in 100 µl reaction, containing Potato Dextrose Broth (for fungus), Nutrient Broth or Tryptone Soya Broth (for bacteria) and serial dilutions of recombinant EcgDf1. The maximum peptide concentration used in all assays was 1.5 µM. Control wells contained no peptide. The

microbial growth was measured in a 96-well microtitre plate under 595 nm, after 48-72 h (appropriate for each microorganism) of incubation in the dark. Antimicrobial activity of EcgDf1 was expressed as a percentage of growth inhibition. All experiments were performed in three replicates.

After 15 days of incubation in the dark, the growth of *Candida albicans* was further observed in comparison with the growth control under a stereomicroscope (Nikon SMZ 1270). Photographs were at 3.15x, 20x, and 80x magnifications.

Results and Discussion

Isolation and sequence analysis of *EcgDf1*

EcgDf1 genomic sequence, 886 nt from de ATG start codon to stop codon (Fig. 1), was isolated by PCR based methodology using primers designed for the legume *Vigna unguiculata* (Padovan et al. 2010). The coding *EcgDf1* sequence, isolated from seedlings cDNA, allowed the confirmation of the sequence in the ORF (Open Reading Frame) and localization of one intron of 649 nt. The intron size is quite larger than those found in *A. thaliana* defensins, where more than 300 defensin-like genes have introns that range between 75 and 275 bp in size (Silverstein et al. 2005). *EcgDf1* gene structure is similar to other reported defensins, with two exons and one intron where the mature peptide is entirely encoded in the second exon. Analysis of the deduced amino acid sequence (78 residues) showed a signal peptide sequence of 31 amino acids, followed by a mature peptide of 47 residues, containing eight cysteine residues. This peptide has several characteristics in common with defensins with three-dimensional structural information, with six cysteines distributed like the αβ motif ($CX_nCX_3CX_nCX_nCXC$ where C=Cys, X=any amino acid, and n indicates a non-conserved number of amino acids; Silverstein et al. 2005) and with a γ-core motif $GXCX_{3-9}C$ (Yount and Yeaman 2004). The mature peptide, had a theoretical mass of 5427.19 Da and basic isoelectric point (pI) of 9.37, with an αβ motif $CX_5CXXXCX_7C_6CXC$, and a γ-core motif $GXCX_6C$ (see Fig. 2).

Fig. 1 Nucleotide and amino acidic sequences of EcgDf1. The lower case letters indicate the intron and 3' UTR regions. Amino acid sequence of EcgDf1 mature peptide indicated by gray-shaded is preceded by a signal peptide sequence predicted using the SignalP tool

EcgDf1 Blast search showed high similarity to defensin-like peptides, with 78 % sequence identity to defensin from *Vigna unguiculata* reported by Padovan et al. (2010). The Neighbor-Joining tree built from the alignment analysis of EcgDf1 mature sequence and a set of plant mature defensins manually annotated and reviewed from the UniProt database showed two large groups: Group 1 and Group 2 (Fig. 2). However, the bootstrap values at the nodes show that only a few subgroups present values higher than 80%, indicating significant similarity. EcgDf1 showed greater similarity with members of group 1. Sequences from group 2 have eight residues between the cysteine 4 and the glycine of the γ-core motif, unlike the group 1 that presents seven residues, except for two defensins from *Zea mays* and *Sorghum bicolor*, respectively. When a molecular phylogeny tree was constructed for over 80 defensin sequences from different plant species, division into two main groups was not obvious (Thomma et al. 2002). These authors discussed that the resolving power of molecular defensin phylogeny is poor because sequences are small, share only a few conserved residues, and have accumulated multiple substitutions at the same site. In solanaceous species, defensins can be divided into two classes according to the structure of the precursor proteins. The first (Class I) is the largest group, composed for endoplasmic reticulum (ER) signal sequence and a mature defensin domain entering the secretory pathway. The second class (Class II) include members produced as larger precursors with an additional C-terminal prodomain (Lay and Anderson 2005). When the mature peptides were aligned, the Neighbor-Joining tree (Fig. 2) includes class I and class II defensins, both within group 1 but class II is located in a branch separated from the rest of defensins from group 1.

Group 1 comprises representatives of 10 of the 12 plant families included in this analysis and Group 2 has only six families represented. Species within families such as Fabaceae and Brassicaceae have been much more studied and therefore more likely to detect sequence diversity. This may also explain why these plant families are found in both groups. For example, more than 300 defensins have been found in the *A. thaliana* genome (only three were included in our analysis as representative of different subgroups) divided into 46 subgroups (Silverstein et al. 2005). On the other hand, defensins of *Aesculus hippocastanum* (Sapindaceae), *Clitoria ternatea* (Fabaceae), *Dahlia merckii* (Asteraceae) and *Heuchera sanguinea* (Saxifragaceae) belonging to four different families (Osborn et al. 1995), are all in Group 2. In these defensins, sequences similarity seems to be related to the protein extraction method that could restrict the recovered repertoire.

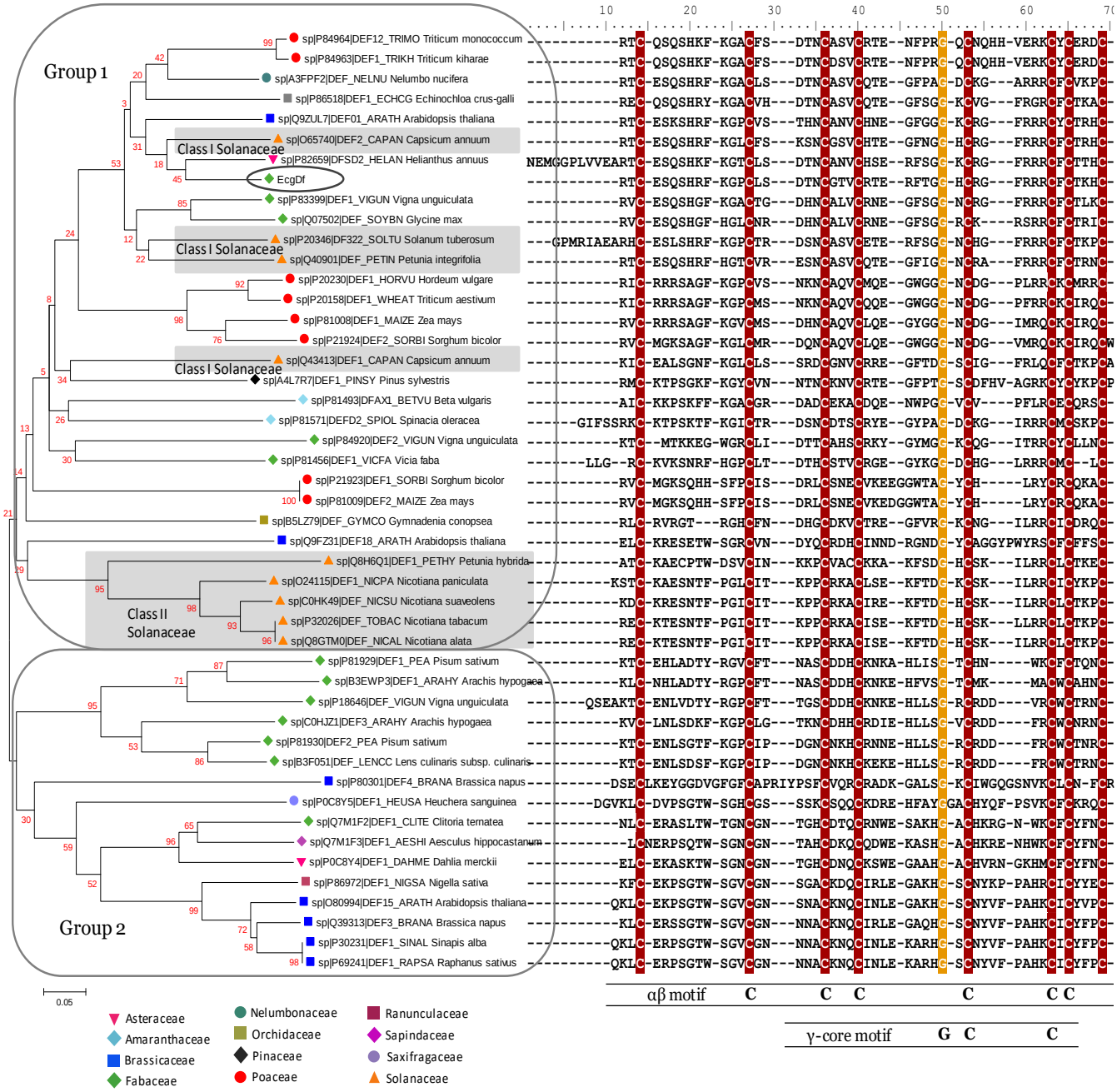


Fig. 2 Sequence alignments and distance matrix of EcgDf1 peptide and members of the plant defensin protein family. The deduced amino acid sequence of the newly isolated gene (*EcgDf1*) was aligned with a set of mature plant defensins. Sequences used were taken from the manually annotated and reviewed UniProt database. Each sequence was named according to its UniProt Entry identifier, followed by a Mnemonic identifier of a UniProtKB entry. An unrooted tree was generated using the Neighbor-Joining method. Values in the nodes regard bootstrap values (5000 replicates). In gray, class I and class II Solanaceae defensins

Expression and purification of recombinant EcgDf1 in *E. coli*

The mature EcgDf1 peptide was heterologously produced in *E. coli*. We choose recombinant DNA expression (using a bacterial host) to obtain our peptide as an easy-to-handle and fast alternative to complex extraction procedure from their natural source or their costly chemical synthesis. Owing to the codon bias and the inability of *E. coli* to form disulfide bridges within its cytoplasm (Lobstein et al. 2012), we decided to contrast Rosetta-gami pLysS and Bl21(DE3) strains. Unlike BL21(DE3), Rosetta-gami strain allows for the expression of eukaryotic proteins that contain codons rarely used in *E. coli*, while enhancing disulfide bond formation. To increase the solubility and to avoid toxicity for the host cells, a non-toxic polypeptide fusion partner such as the hydrophilic tag thioredoxin (Trx) (Terpe 2003) was used. We used pET102/D which included an N-terminal Trx protein a C-terminal 6xHis tag, and a protease recognition site for cleavage of the fusion Trx from the recombinant peptide EcgDf1. In order to reach the highest yield of the soluble Trx-EcgDf1 fusion protein, its expression was tested by two temperatures (28 and 20 °C) and incubation times post induction (4 and 24 h), using both *E. coli* Rosetta-gami and BL21(DE3) strains. *E. coli* strains were transformed with pE-EcgDf1 and pE-Trx (control without EcgDf1). The SDS-PAGE analysis of total lysates from induced cells revealed the presence of recombinant proteins Trx (control) and Trx-EcgDf1 with the expected molecular weight, ~17 kDa (Fig. 3a,3b-lanes1-2) and ~19.5 kDa respectively (Fig. 3a,3b-lanes 5-6). No significant peptide bands, corresponding to the recombinant protein, were observed in non-induced cultures (Fig. 3a-b lanes 3-4) except for Rosetta-gami at 24 hpi. The higher yield was obtained with Rosetta-gami(DE3)pLysS at 28 ºC and after 24 h of induction. With BL21(DE3) strain, we found a slightly lower percentage of recombinant protein in the soluble fraction (better observed at 4 hpi; see Fig. 3a, 3b, lines 5,6). The recovery of the 19.5 kDa band at the end of the purification by Ni-affinity chromatography confirmed that this band contains the recombinant peptide (Fig 3c). Some background has shown that both the use of *E. coli* for the production of plant defensins yielded inactive protein aggregates which needed to be denatured and refolded for gaining activity in the systems BL21(DE3)/pET26b (Kovalskaya and Hammond 2009) and Origami (DE3)/pQE30 (Guillen-Chable et al. 2017). Others works obtained a soluble defensin, that was biologically active using *E. coli* BL21(DE3)/ pET42a (Kovaleva et al. 2011) and Rosetta Gami 2 (DE3)pLysS)/ pET-32 EK/LIC systems. In these cases, the peptide was fused to glutathione S-transferase or thioredoxin, respectively.

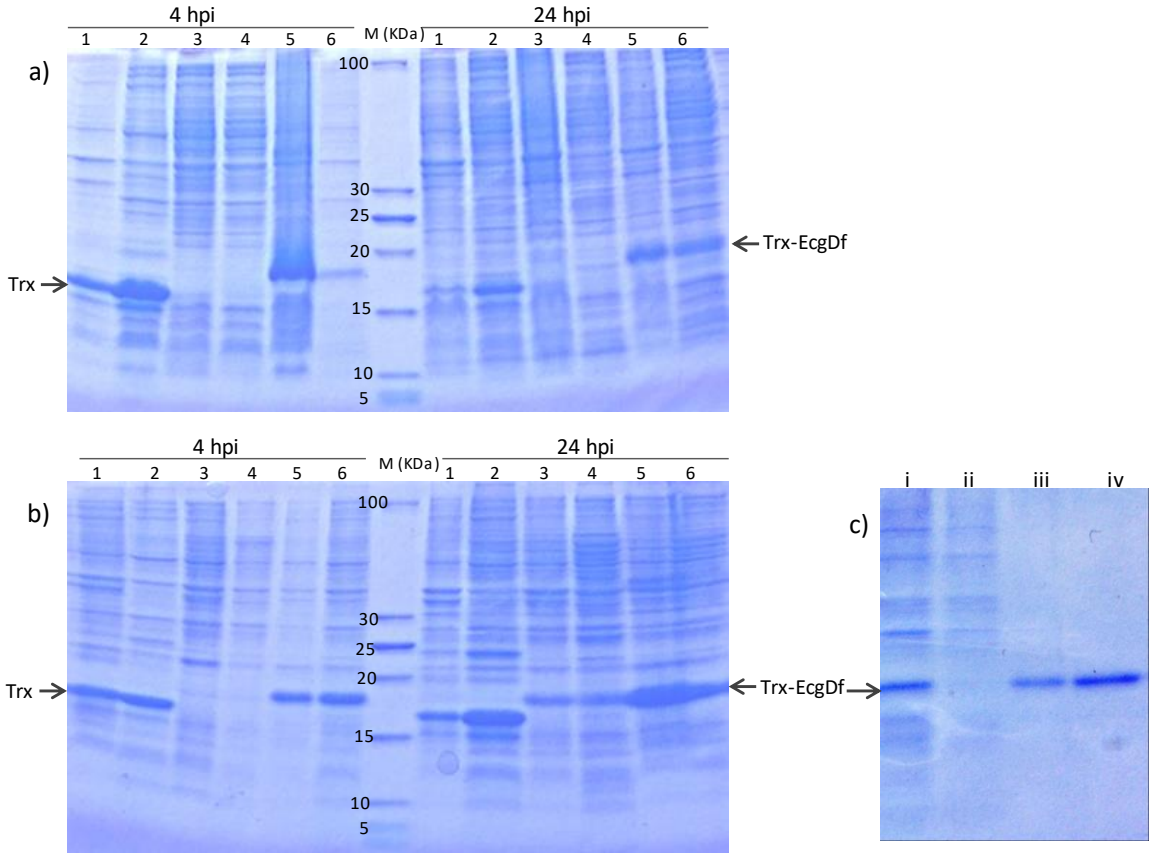
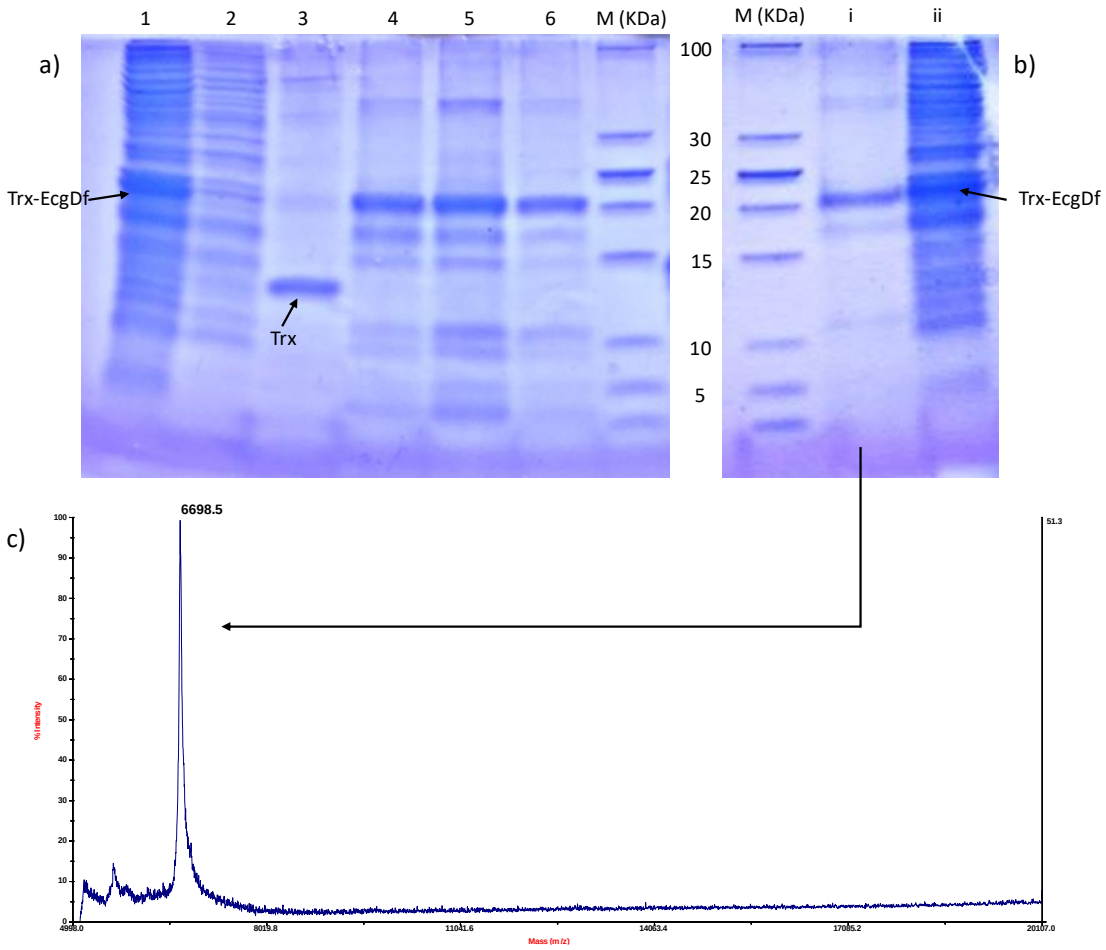


Fig. 3 Heterologous expression of Trx-EcgDf1 fusion protein. Tris-Glycine SDS-PAGE (15%) analysis of Protein fractions from *E. coli* **a)** BL21(DE3) and **b)** (Rosetta-gami (DE3) cells transformed with pE-Trx expression plasmid (control) (1,2), pE-EcgDf1 expression plasmid incubated in the absence (3,4) or presence of IPTG (5,6). Lines 1, 3, and 5 are crude extracts; lines 2, 4, and 6 are soluble fractions. 4 hpi: 4 h post-induction; 24 hpi: 24 hours post induction. Temperature post induction: 28 ºC. **c)** Affinity purification of the recombinant Trx-EcgDf1-6His protein. Soluble fraction (i); unbound proteins removed by the first washing step in which the fusion protein is bound to the column (ii); elution steps where the purified Trx-EcgDf1 is released from the Ni-affinity column (iii and iv); M: PageRuler Low Range Unstained Protein Ladder (ThermoFisher). Arrows indicate expression of Trx protein (control) and Trx-EcgDf1 fusion protein

For practical reasons, on-column Trx cleavage was conducted before peptide elution. A previous pilot test, carried out with several enterokinase concentrations, showed that 2,5 µg of the fusion protein was entirely digested with one unit of the enzyme, overnight at room temperature (data not shown). The thioredoxin fusion fragment was removed with 30 mM imidazole (Fig. 4 line 3). Several bands were observed when the proteins bound to the column were eluted with 300 mM imidazole (Fig. 4a line 4-6). To rule out partial digestion, a second in batch digestion was conducted for 120 h at room temperature (Fig. 4b). Several bands remained at these conditions, including the one of ~20 kDa, which seem to correspond to Trx-EcgDf1. MALDI-TOF mass

analysis of EcgDf1 peptide fractions after digestion for 120 h at room temperature is presented in Fig. 5. The mass spectrum had one major peak at an m/z value of 6698.5. The result was consistent with the molecular mass calculated from the amino acid sequence of the recombinant EcgDf1 forming six disulfide bonds (6697,55 Da). The results obtained by the peptide mass fingerprinting are shown in Fig. 4d. These results confirm that the recombinant peptide exhibits the expected amino acid composition. No other significant peaks were observed, indicating that the purification was effective. Furthermore, no peak was observed at 19500 Da, indicating that the digestion was total or almost total. One possibility for the presence of multiple bands is the formation of SDS-resistant oligomers. it has been detected that some proteins are resistant to sodium dodecyl sulfate (SDS)-induced denaturation. These stable proteins are characterized not only by SDS resistance but also by stability against proteases and an increased biological half-life, being most of them oligomeric β-sheet proteins (Gentile et al. 2002; Manning and Colón 2004). Defensins could be an example of this kind of proteins since they have a compact structure that confers resistance to extremes temperatures and degradation mediated by proteases (Hegedüs and Marx 2013). In addition, X-ray crystallography indicates that some defensins form dimers, enhancing their antifungal activity (Song et al. 2011; Lay et al. 2012). In fact, defensin oligomers have been reported on Tris-Tricine SDS-PAGE gels (de Beer and Vivier 2011; Guillen-Chable et al. 2017).

On the other hand, no band seems to match to the expected of~6000 Da possibly due to an abnormal migration. Migration on SDS-PAGE that does not correlate with formula molecular weights ("gel shifting") is sometimes observed with cationic peptides (Shi et al. 2012; Herbel et al. 2015). Moreover, increased SDS-PAGE migration rates were traceable to a compact S-S bridged conformation in proteins with disulfide bonds (Rath et al. 2009). Previous work with a scorpion neurotoxin with four disulfide bridges has demonstrated that close folded isoforms (different in cysteine pairing of disulfides) are present in recombinant peptides purified from inclusion bodies (Estrada et al. 2007). Then we think that multiple bands observed in SDS-PAGE could be due, on the one hand, to the formation of EcgDf1 oligomers and, on the other, to different conformations given by the number of disulfide bridges formed.



d) **RTCESQSHR**FK **GPCLSDTNCG TVCR**TER**FTG GHCR**GFRRRC FCTK**HCIEGR HHHHHH**

Fig. 4 Affinity purification and MALDI-TOF mass spectra of the recombinant EcgDf1 peptide. **a)** Tris-Tricine SDS-PAGE (15 %) analysis of the purification steps of the EcgDf1 peptide after 15 h EK digestion. Protein soluble fraction from *E. coli* Rosetta-gami (DE3) cells transformed with pE-EcgDf1 expression plasmid incubated in the presence of IPTG (1); Unbound proteins removed by the first washing step in which EcgDf1 is bound to the column (2); Trx protein after cleavage with enterokinase (3); elution of EcgDf1 peptide (4-6) after 15 h digestion. **b)** Tris-Tricine SDS-PAGE (15 %) analysis of the purification steps of the EcgDf1 peptide after 120 h EK digestion. Elution of EcgDf1 peptide after 120 h digestion (i); soluble protein fraction from *E. coli* Rosetta-gami (DE3) cells transformed with pE-EcgDf1 expression plasmid incubated in the presence of IPTG (ii) M: PageRuler Low Range Unstained Protein Ladder (ThermoFisher). **c)** MALDI-TOF mass spectra of EcgDf1 after Trx cleavage with enterokinase (120 h digestion). **d)** Peptide sequence confirmed by mass fingerprinting were the identified matched peptides are shown in red

3D structure determination

Using the recently determined nuclear magnetic resonance (NMR) structure of the *Medicago truncatula* defensin, we build by means of homology modeling the 3D structure of the native EcgDf1 peptide. The high sequence identity of EcgDf1 with *M. truncatula* defensin determined

that our native peptide belongs to the cys-defensin superfamily (Shafee et al. 2016). This implies a given secondary structure orientation, cysteine motifs, and disulfide bond connectivities. The structure is composed of three antiparallel β-sheets and one α-helix (α1) which is connected by two disulfide bonds to the β2-sheet (Figure 5a). This conformation, also shared by other defensins (Sagaram et al. 2013), is thought to be stabilized by four disulfide bonds due to the lack of a stabilizing hydrophobic core. Our peptide contains a histidine tail at the C-terminal, composed of 10 extra amino acids (six histidine residues). The MD simulations of the two most distinct conformations obtained through the homology modeling gave rise to two very different orientations of the extra tail (Figure 5a). In one case the tail interacts with α1, while in the other, the histidine tail interacts with β1. In both cases, the cysteine-stabilized α/β (CSα/β) motif stays completely stable along the simulation with an averaged root mean square displacements of the Cα atoms below 1.2 Å from the starting structure. The flexibility of the extra tail is important to explain the observed formation of EcgDf1 homodimers, since to date, in the experimentally determined dimers, both monomers interact either in a head-to-tail manner through antiparallel β1-β1 interactions (as in NaD1; Lay et al 2012), or through α1-β1 (as in SPE10; Song et al 2011). Both types of dimer formation are compatible with the structure determined herein for EcgDf1 defensin, and further docking experiments will be necessary to confirm the actual monomer orientation of our peptide. Finally, as previously described for other defensins (e.g., Sagaram et al. 2013), EcgDf1 displays an overall electrophilic surface produced by the excess of positively charged amino acids (Figure 5b). This cationic nature of defensins, known to be relevant for their biological activity, could be enhanced in EcgDf1 if the histidine residues in the extra tail get protonated in both ε and δ atoms as actually predicted by the titration algorithm (see Methods).

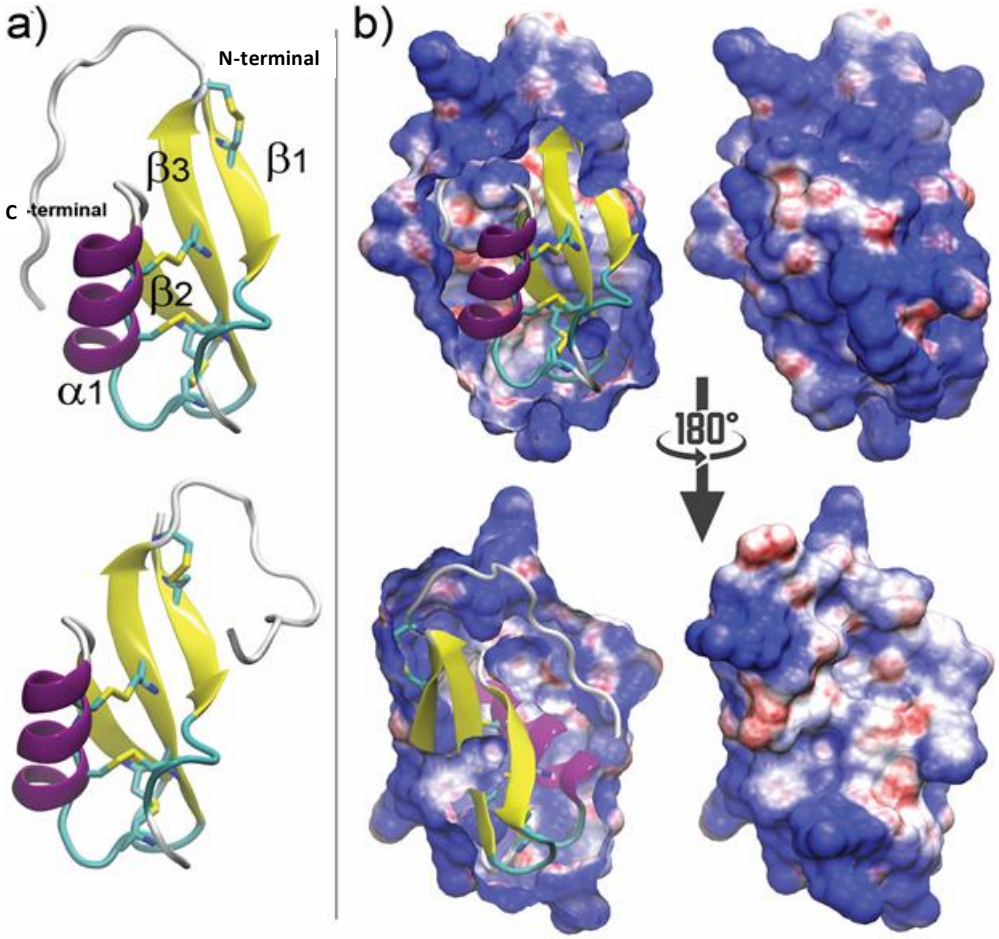


Fig. 5 Structure and electrostatic potential of EcgDf1. **a)** Structural prediction of the native EcgDf1 peptide exhibiting one α-helix and three β-sheets fixed by four disulfide bonds obtained using as template the solution NMR structure with PDB id 2L3R. The final structure (after 100 ns of production run) of the two simulated models are shown. **b)** Molecular electrostatic potential (MEP) of the full recombinant EcgDf1 peptide (front and side view). Note that a clipping plane was used to cut the MEP allowing to visually locating the αβ motif. Note that the histidine residues of the extra tail were considered neutral (not protonated) for the MEP calculation

Antimicrobial activity of the recombinant peptide EcgDf1

The antimicrobial activity of the recombinant EcgDf1 was tested *in vitro* against the bacterial and fungal species detailed in Materials and Methods. The purified recombinant EcgDf1 has *in vitro* antimicrobial, especially antifungal activity. These results are consistent with those found for other reported plant defensins, for which several *in vitro* biological activities have been recognized, but the best attributed activity is their ability to inhibit the growth of a broad range of filamentous fungi and yeasts (Carvalho and Gomes 2009; Kant et al. 2009; Kovaleva et al. 2011; Ferreira Lacerda et al. 2016), being also known for their antimicrobial activity at low micromolar concentrations (Shafee et al. 2017). To the maximum concentration (1,5 μM; ~10 μg/ml) of EcgDf1 tested, we found significant inhibition of growth for the fungi *A. niger, C.*

albicans, B. cinerea, P. expansum, and *A. altenata*, reaching between 96 and 99% (Fig. 6). IC50 values of recombinant EcgDf1 was estimated and found as 0.14, 0.16, 0.38, 0.62 and 0.64 µM, respectively. Other plant defensins have inhibition values similar to those found for EcgDf1. For example, for *B. cinerea*, the IC50 for a potato defensin (StPTH1) was 1 µM (Berrocal-Lobo et al. 2002). Like floral defensins from *Petunia hybrida* (PhD1, PhD2) and *Nicotiana alata* (NaD1), EcgDf inhibited almost completely growth of *B. cinerea* at ~10 µg/ml concentration (Lay et al. 2003). However, for the yeast *C. albicans*, EcgDf1 shows significant inhibition (99%) at the maximum concentration, while the growth inhibition of this microorganism for another defensin PvD1 from *Phaseolus vulgaris*, was 90% with a concentration 10 times higher than EcgDf1 (100 µg/ml) (Mello et al. 2014). According to Carvalho and Gomes (2011), the concentrations required for defensin inhibitory activity depend on the tested fungus, confirming the variations observed when we tested EcgDf with different microorganisms. In addition, concentrations necessary for inhibition will depend on the plant defensin analyzed. These authors propose that this phenomenon may be explained to the fact that a plant expresses a set of AMPs cooperatively (of the same or different families) with overlapping spectra of antimicrobial activities but with specificity against particular pathogen groups. No activity was observed for the gram negative *X. versicatoria* at the maximum concentration tested. This result was expected, since only a few plant defensins have been shown to inhibit bacterial growth, even less in gram negative bacteria (Carvalho and Gomes, 2009; Guillen-Chable et al. 2017). Nevertheless, for the gram positive *C. michiganensis ssp. michiganensis,* 92 % activity was reached at the maximum concentration (IC50=0.96 µM). This activity decayed rapidly and in the third dilution it was no longer detected.

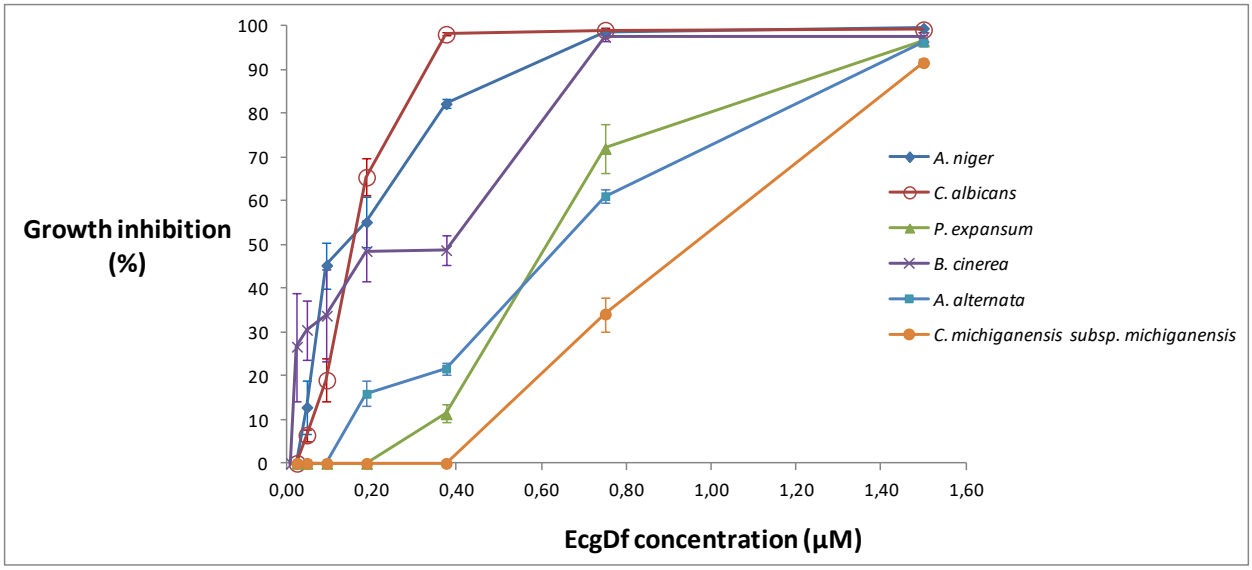


Fig. 6 *In vitro* antimicrobial activity of EcgDf1. Dose-dependent growth inhibition curves of the following bacterial and fungal pathogens: *A. niger* (diamonds), *C. albicans* (open circles), *P. expansum* (triangles), *B. cinerea* (crosses), *A. alternata* (squares), *C. michiganensis* ssp. *michiganensis* (black circles) was determined by measuring the cultures at OD595, in the presence of different amounts of recombinant PdSN1. The data are mean ± SD (n = 3)

It is interesting to note that EcgDf1 presented activity against two human opportunistic pathogens: *A. niger* and *C. albicans* (Fig. 7). For the yeast *C. albicans*, we further observe their growth in presence of EcgDf1, after 15 days of incubation, by a estereomicroscope. After that time, no growth was observed with the maximum concentration of EcgDf1 (1.5 µM) while with the first two serial dilutions, the growth was less than in the growth control (Fig. 7b). Thus, EcgDf1 seems to have a fungicidal effect on *C. albicans*, since 15 days after inoculation, no growth of the yeast was observed. The interaction between *C. albicans* and another defensin, NaD1, involves yeast killing induced by oxidative damage (Hayes et al. 2013). Future studies, as confocal microscopy, are necessary to characterize the mode of action of EcgDf1 further, taking into account that not all defensins have the same mode of action (Parisi et al. 2018) and ever more, for the same defensin, antifungal mechanisms can differ depending on the pathogen (El-Mounadi et al. 2016).

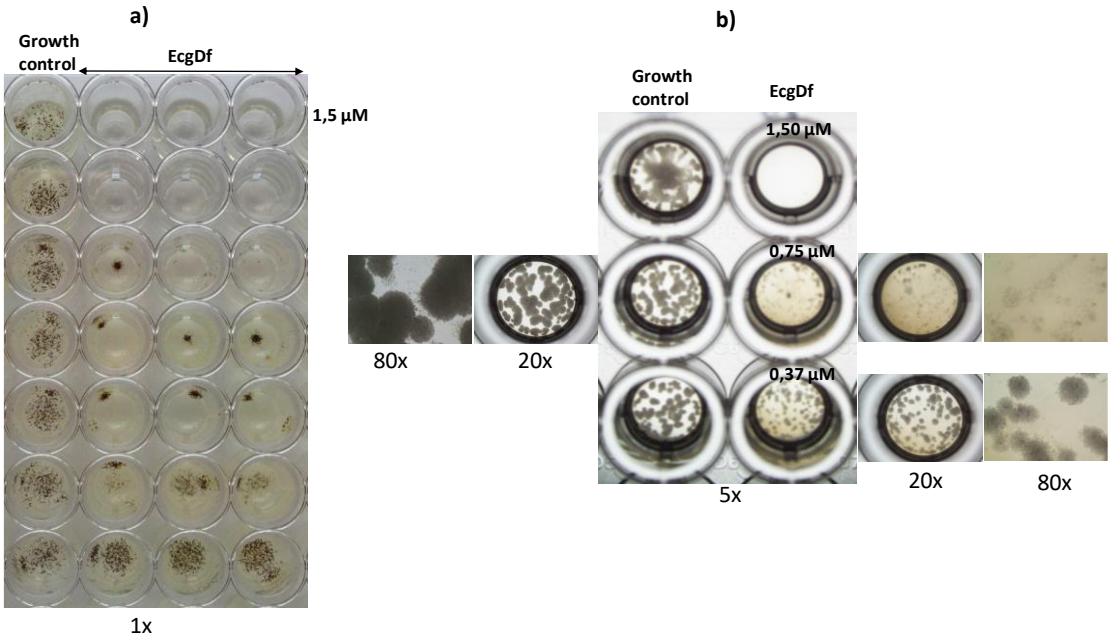


Fig. 7 EcgDf1 inhibits *A. niger* and *C. albicans* growth *in vitro*. Purified EcgDf1 was evaluated in a 96-well microtiter plate for *in vitro* antimicrobial activity against **a)** *A. niger* after 72 h of incubation **b)** *C. albicans* after 15 days of incubation followed by a stereomicroscope image. Serial dilutions of recombinant EcgDf1 were used, with a maximum of 1.5 µM

Conclusions

In the current study, a new defensin gene *EcgDf1* was isolated from a native Fabaceae plant, and the antimicrobial activity of the predicted peptide was characterized. EcgDf1 has several characteristics common to most of the reported defensins such as gene structure, size, charge cysteine motif and a 3D structure with one α-helix and three β-sheets. Using a bacterial expression system, we obtained a purified recombinant EcgDf1, and we presented evidence that this recombinant peptide could form stable oligomers under denaturant and reducing Tricine SDS- PAGE. The purified peptide was active against a set of microorganisms, including plant pathogens like *C. michiganensis* ssp. *michiganensis, P. expansum*, *B. cinerea, A. alternata* and also some opportunistic human pathogens as *C. albicans* and *A. niger*. We highlight the activity found for *P. expansum* since we did not find other reports of defensins with activity against such a broad host range of pathogens that cause the most economically damaging post-harvest diseases. EcgDf1 could then be an attractive molecule, for the development of antimicrobial agents for use in plant protection both in the field and in post-harvest, as an alternative to the use of agrochemicals. Also, since EcgDf1 displays antimicrobial activity against human fungal pathogens, in the future, if it is proven that it does not show toxicity, it could be a substitute or a coadjutant for classic antibiotics.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, Donovan CO, Redaschi N, Yeh LL (2004) UniProt : the Universal Protein knowledgebase. Nucleic Acids Res 32:D115–D119. doi: 10.1093/nar/gkh131

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684-3688. doi: 10.1063/1.448118

Berrocal-Lobo M, Segura A, Moreno M, López G, García-Olmedo F, Molina A (2002) Snakin-2 , an Antimicrobial Peptide from Potato Whose Gene Is Locally Induced by Wounding and Responds to Pathogen Infection 1. Plant Physiol 128:951–961. doi: 10.1104/pp.010685.1

Bradford MM (1976) A Rapid and Sensitive Method for the Quantitation Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. Anal Biochem 72:248–254

Broekaert WF, Terras F., Cammue BP., Vandedeyden J (1990) An automated quantitative assay for fungal growth inhibition. FEMS Microbiol Lett 69:55–60

Broekaert WF, Terras FR, Cammue BP, Osborn RW (1995) Plant defensins: novel antimicrobial peptides as components of the host defense system. Plant Physiol 108:1353–1358. doi: 10.1104/pp.108.4.1353

Brogden KA (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3:238–250. doi: 10.1038/nrmicro1098

Cabral KM., Almeida MS, Valente AP, Almeida FC., Kurtenbach E (2003) Production of the active antifungal *Pisum sativum* defensin 1 (Psd1) in *Pichia pastoris*: overcoming the inefficiency of the STE13 protease. Protein Expr Purif 31:115–122. doi: 10.1016/S1046-5928(03)00136-0

Carvalho ADO, Gomes VM (2011) Plant Defensins and Defensin-Like Peptides - Biological Activities and Biotechnological Applications Plant Defensins and Defensin-Like Peptides - Biological Activities and Biotechnological Applications. Curr Pharm Des 17:4270–4293. doi: 10.2174/138161211798999447

Carvalho DO, Gomes VM (2009) Peptides Plant defensins-Prospects for the biological functions and biotechnological properties. Peptides 30:1007–1020. doi: 10.1016/j.peptides.2009.01.018

Case DA, Cerutti DS, Cheatham III TE, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, Izadi S, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko

T, Luo R, Mermelstein D, Merz KM, Monard G, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Simmerling CL, Botello-Smith WM, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Xiao L, York DM, Kollman PA (2017) AMBER 2017, University of California, San Francisco.

Dans PD, Danilāne L, Ivani I, Dršata T, Lankaš F, Walther J, Illa Pujagut R, Battistini F, Gelpí JL, Lavery R, Orozco M (2016) Long-timescale dynamics of the Drew-Dickerson dodecamer. Nucl Acids Res 44:4052-4066. doi: 10.1093/nar/gkw264

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. J Chem Phys 98:10089-10095. doi: 10.1063/1.464397

de Beer A, Vivier MA (2011) Four plant defensins from an indigenous South African Brassicaceae species display divergent activities against two test pathogens despite high sequence similarity in the encoding genes. BMC Res Notes 4:459. doi: 10.1186/1756-0500-4-459

De Coninck B, Cammue BPA, Thevissen K (2013) Modes of antifungal action and in planta functions of plant defensins and defensin-like peptides. Fungal Biol Rev 26:109–120. doi: 10.1016/j.fbr.2012.10.002

Do HM, Lee SC, Jung HW, Sohn KH, Hwang BK (2004) 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 166:1297–1305. doi: 10.1016/j.plantsci.2004.01.008

Doyle J (1991) DNA protocols for plants-CTAB total DNA isolation. In: Hewitt GM, Johnston A (eds) Molecular Techniques in Taxonomy, Springer, Berlin, pp. 283–293

El-Mounadi K, Islam KT, Hernández-Ortiz P, Read ND, Shah DM (2016) Antifungal mechanisms of a plant defensin MtDef4 are not conserved between the ascomycete fungi *Neurospora crassa* and *Fusarium graminearum*. Mol Microbiol 100:542-559. doi: 10.1111/mmi.13333

Estrada G, Garcia B, Schiavon E, Ortiz E, Cestèle S, Wanke E, Possani L, Corzo G (2007) Four disulfide-bridged scorpion beta neurotoxin CssII : heterologous expression and proper folding *in vitro.* Biochim. Biophys. Acta 1770, 1161–1168.

Ferreira Lacerda A, Perseghini Del Sarto R, Santos Silva M, Rosas de Vasconcelos E, Ramos Coelho R, Olinto dos Santos V, Vieira Godoy C, Santos Seixas C, Mattar da Silva M, Grossi-de-Sa M (2016) The recombinant pea defensin Drr230a is active against impacting soybean and cotton pathogenic fungi from the genera Fusarium, Colletotrichum and Phakopsora. 3 Biotech 6:59. doi: 10.1007/s13205-015-0320-7

Games PD, dos Santos IS, Mello ÉO, Diz MSS, Carvalho AO, de Souza-Filho GA, Da Cunha M, Vasconcelos IM, Ferreira B dos S, Gomes VM (2008) Isolation, characterization and cloning of a cDNA encoding a new antifungal defensin from *Phaseolus vulgaris* L. seeds. Peptides

29:2090–2100. doi: 10.1016/j.peptides.2008.08.008

Gao AG, Hakimi SM, Mittanck CA, Wu Y, Woerner M, Stark DM, Rommens CM (2000) Fungal pathogen protection in potato by expression of a plant defensin peptide. Nature Biotechnol 18:1307-1310.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, New York, pp 571–607

Gentile F, Amodeo P, Febbraio F, Picaro F, Motta A, Formisano S, Nucci R (2002) SDS-resistant active and thermostable dimers are obtained from the dissociation of homotetrameric β-glycosidase from hyperthermophilic *Sulfolobus solfataricus* in SDS: Stabilizing role of the A-C intermonomeric interface. J Biol Chem 277:44050–44060. doi: 10.1074/jbc.M206761200

Guillen-Chable F, Arenas-Sosa I, Islas-flores I, Corzo G, Martinez-liu C, Estrada G (2017) Antibacterial activity and phospholipid recognition of the recombinant defensin J1-1 from Capsicum genus. Protein Expr Purif 136:45–51. doi: 10.1016/j.pep.2017.06.007

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98

Hayes BM, Bleackley MR, Wiltshire JL, Anderson MA, Traven A, van der Weerden NL (2013) Identification and mechanism of action of the plant defensin NaD1 as a new member of the antifungal drug arsenal against *Candida albicans*. Antimicrob. Agents Chemother 57:3667–3675

Hegedüs N, Marx F (2013) Antifungal proteins: More than antimicrobials? Fungal Biol Rev 26:132–145 . doi: 10.1016/j.fbr.2012.07.002

Herbel V, Schäfer H, Wink M (2015) Recombinant production of snakin-2 (an antimicrobial peptide from tomato) in *E. Coli* and analysis of its bioactivity. Molecules 20:14889–14901. doi: 10.3390/molecules200814889

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38. doi: 10.1016/0263-7855(96)00018-5

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–35. doi: 10.1063/1.445869

Kant P, Liu WZ, Pauls KP (2009) PDC1, a corn defensin peptide expressed in *Escherichia coli* and *Pichia pastoris* inhibits growth of *Fusarium graminearum.* Peptides 30:1593–1599. doi: 10.1016/j.peptides.2009.05.024

Koike M, Okamoto T, Tsuda S, Imai R (2002) A novel plant defensin-like gene of winter wheat is specifically induced during cold acclimation. Biochem Biophys Res Commun 298:46–53

Kovaleva V, Krynytskyy H, Gout I, Gout R (2011) Recombinant expression, affinity purification and functional characterization of Scots pine defensin 1. Appl Microbiol Biotechnol 89:1093–1101. doi: 10.1007/s00253-010-2935-2

Kovalskaya N, Hammond RW (2009) Expression and functional characterization of the plant antimicrobial snakin-1 and defensin recombinant proteins. Protein Expr Purif 63:12–7. doi: 10.1016/j.pep.2008.08.013

Laemmli UK (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature 227:680–685

Lay FT, Anderson MA (2005) Defensins-components of the innate immune system in plants. Curr Protein Pept Sci 6:85–101. doi: 10.2174/1389203053027575

Lay FT, Brugliera F, Anderson MA (2003) Isolation and Properties of Floral Defensins from Ornamental Tobacco and Petunia 1. 131:1283–1293. doi: 10.1104/pp.102.016626.)

Lay FT, Mills GD, Poon IKH, Cowieson NP, Kirby N, Baxter AA, Van Der Weerden NL, Dogovski C, Perugini MA, Anderson MA, Kvansakul M, Hulett MD (2012) Dimerization of plant defensin NaD1 enhances its antifungal activity. J Biol Chem 287:19961–19972. doi: 10.1074/jbc.M111.331009

Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M (2012) SHuffle , a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact 11:753. doi: 10.1186/1475-2859-11-56

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99sb. J Chem Theor Comput 11:3696-3713. doi: 10.1021/acs.jctc.5b00255

Manning M, Colón W (2004) Structural basis of protein kinetic stability: Resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward β-sheet structure. Biochemistry 43:11248–11254. doi: 10.1021/bi0491898

Mello ÉDO, Santos IS, Carvalho ADO, de Souza LS, de Souza-filho GA, do Nascimento V V, Machado OL, Zottich U, Gomes VM (2014) Functional expression and activity of the recombinant antifungal defensin PvD1r from *Phaseolus vulgaris* L. (common bean) seeds. BMC Biochem 15:1–13. doi: 10.1186/1471-2091-15-7

Neill DA (1988) Experimental Studies on Species Relationships in Erythrina (Leguminosae: Papilionoideae). Ann Missouri Bot Gard 75:886–969

Nielsen H (2017) Predicting secretory proteins with SignalP. In: D. K (ed) Methods in Molecular Biology. New York, pp 59–73

Osborn RW, De Samblanx GW, Thevissen K, Goderis I, Torrekens S, Van Leuven F, Attenborough S, Rees SB, Broekaert WF (1995) Isolation and characterisation of plant

defensins from seeds of Asteraceae, Fabaceae, Hippocastanaceae and Saxifragaceae. FEBS Lett 368:257–262. doi: 10.1016/0014-5793(95)00666-W

Padovan L, Segat L, Tossi A, Calsa T, Ederson AK, Brandao L, Guimarães RL, Pandolfi V, Pestana-Calsa MC, Belarmino LC, Benko-Iseppon AM, Crovella S (2010) Characterization of a new defensin from cowpea (*Vigna unguiculata* (L.) Walp.). Protein Pept Lett 17:297–304. doi: 10.2174/092986610790780350

Parisi K, Shafee TMA, Quimbar P, Weerden NL Van Der, Bleackley MR, Anderson MA (2018) The evolution, function and mechanisms of action for plant defensins. Semin Cell Dev Biol. doi: 10.1016/j.semcdb.2018.02.004

Pérez A, Luque FJ, Orozco M (2007) Dynamics of B-DNA on the microsecond time scale. J Am Chem Soc 129:14739-14745.doi: 10.1021/ja0753546

Portieles R, Rodriguez M, Castillo J, Pujol M, Enriquez G, Borroto C, Trujillo L, Thomma BPHJ, Borras-Hidalgo O (2010) NmDef02 , a novel antimicrobial gene isolated from *Nicotiana megalosiphon* confers high-level pathogen resistance under greenhouse and field conditions. Plant Biotechnol J 678–690. doi: 10.1111/j.1467-7652.2010.00501.x

Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci 106:1760–1765. doi: 10.1073/pnas.0813167106

Roe DR, Cheatham III TE (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 9:3084–3095. doi: 10.1021/ct400341p

Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341. doi: 10.1016/0021-9991(77)90098-5

Sagaram US, El-Mounadi K, Buchko GW, Berg HR, Kaur J, Pandurangi RS, Smith TJ, Shah DM (2013) Structural and functional studies of a phosphatidic acid-binding antifungal plant defensin MtDef4: Identification of an RGFRRR motif governing fungal cell entry. PLoS One 8:1–22. doi: 10.1371/journal.pone.0082485

Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh Ewald. J Chem Theory Comput 9:3878–3888. doi:10.1021/ct400314y

Schägger G, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 379:368–379. doi: 10.1016/0003-2697(87)90587-2

Sels J, Mathys J, De Coninck BMA, Cammue BPA, De Bolle MFC (2008) Plant pathogenesis-

related (PR) proteins: A focus on PR peptides. Plant Physiol Biochem 46:941–950. doi: 10.1016/j.plaphy.2008.06.011

Shafee TMA, Lay FT, Phan TK, Anderson MA, Hulett MD (2017) Convergent evolution of defensin sequence, structure and function. Cell Mol Life Sci 74:663–682. doi: 10.1007/s00018-016-2344-5

Shenkarev ZO, Gizatullina AK, Finkina EI, Alekseeva EA, Balandin S V., Mineev Konstantin S. Arseniev AS, Ovchinnikova T V (2014) Heterologous Expression and Solution Structure of Defensin from *Lentil lens*. Biochem Biophys Res Commun 451:252–257. doi: 10.1016/j.bbrc.2014.07.104

Shi Y, Mowery RA, Ashley J, Hentz M, Ramirez AJ, Bilgicer B, Slunt-brown H, Borchelt DR, Shaw BF (2012) Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. Protein Sci 21:1197–1209. doi: 10.1002/pro.2107

Silverstein K a T, Graham M a, Paape TD, VandenBosch K a (2005) Genome organization of more than 300 defensin-like genes in Arabidopsis. Plant Physiol 138:600–610. doi: 10.1104/pp.105.060079

Simoes CMO, Faikenberg M, Mentz LA, Schenkef EP, Amoros M, Girre L (1999) Antiviral activity of South Brazilian medicinal plant extracts. Phytomedicine 6:205–214. doi: 10.1016/S0944-7113(99)80010-5

Slavokhotova AA, Odintsova TI, Rogozhin EA, Musolyamov AK, Andreev YA, Grishin E V., Egorov TA (2011) Isolation, molecular cloning and antimicrobial activity of novel defensins from common chickweed (*Stellaria media* L.) seeds. Biochimie 93:450–456. doi: 10.1016/j.biochi.2010.10.019

Smith DE, Dang LX (1994) Computer simulations of NaCl association in polarizable water. J Chem Phys 100:3757-3762. doi: 10.1063/1.466363

Song X, Zhang M, Zhou Z, Gong W (2011) Ultra-high resolution crystal structure of a dimeric defensin SPE10. FEBS Lett 585:300–306. doi: 10.1016/j.febslet.2010.12.039

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5 : Molecular Evolutionary Genetics Analysis Using Maximum Likelihood , Evolutionary Distance , and Maximum Parsimony Methods Research resource. Mol Biol Evol 28:2731–2739. doi: 10.1093/molbev/msr121

Terpe K (2003) Overview of tag protein fusions : from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 60:523–533. doi: 10.1007/s00253-002-1158-6

Thevissen K, Kristensen H, Thomma BPHJ, Cammue BPA, Francois IEJA (2007) Therapeutic

potential of antifungal plant and insect defensins. Drug Discov Today 12:966–971. doi: 10.1016/j.drudis.2007.07.016

Thomma BPHJ, Cammue BPA, Thevissen K (2002) Plant defensins. Planta 216:193–202 . doi: 10.1007/s00425-002-0902-6

Vijayan S, Guruprasad L, Kirti PB (2008) Prokaryotic expression of a constitutively expressed *Tephrosia villosa* defensin and its potent antifungal activity. Appl Microbiol Biotechnol 80:1023–1032. doi: 10.1007/s00253-008-1648-2

Wang S, Rao P, Ye X (2009) Isolation and biochemical characterization of a novel leguminous defense peptide with antifungal and antiproliferative potency. Appl Microbiol Biotechnol 82:79–86. doi: 10.1007/s00253-008-1729-2

Webb B, Sali A (2014) Comparative protein structure modeling using modeller. Curr Protoc Bioinformatics, John Wiley & Sons, Inc., 5.6.1-5.6.32. doi:10.1002/0471250953.bi0506s15

Yeaman MR, Yount NY (2003) Mechanisms of antimicrobial peptide action and resistance. Pharmacol Rev 55:27–55. doi: 10.1124/pr.55.1.2

Yount NY, Yeaman MR (2004) Multidimensional signatures in antimicrobial peptides. Proc Natl Acad Sci 101:7363–7368. doi: 10.1073/pnas.0401567101

Capítulo 3

Búsqueda de péptidos antimicrobianos en transcriptomas de ibirapitá y congorosa

Este capítulo incluye el borrador de un artículo científico que describe el ensamblado *de novo* del transcriptoma de brotes de ibirapitá. Además, incluye algunos resultados preliminares del ensamblado *de novo* del transcriptoma de hojas/brotes/flores de congorosa. Previamente se describe brevemente la metodología usada para la selección de transcriptomas, la secuenciación del ARN y el tipo de ensamblado usado. Mayores detalles sobre el ensamblado *de novo* se encuentran en el borrador del artículo. El ensamblado de los transcriptomas y el *script* de búsqueda de motivos cisteína fue realizado por el Dr. Pablo Smircich y el Lic. Santiago Radío.

Selección de transcriptomas

Para la selección de los posibles transcriptomas a analizar, se evaluó la actividad antimicrobiana de extractos proteicos crudos de diferentes tejidos (hojas, flores y semillas germinadas) de *P. dubium* y de *M. ilicifolia,* mediante ensayos de difusión en placa.

Los extractos se obtuvieron usando buffer fosfato (10 mM Na2HPO4-15 mM NaH2PO4 pH 7; 100 mM KCl 2mM EDTA 1.5% PVPP polivinilpolipirrolidona) y posterior precipitación con sulfato de amonio (saturación 70 %). La concentración final de proteínas en los extractos, medida mediante absorbancia a 280 nm luego de diálisis, liofilización y resuspensión en mínimo volumen de agua, fue de entre 8-20 mg/ml. El volumen de extracto sin diluir utilizado fue de 20 y 30 µl.

Estos extractos fueron enfrentados a diferentes microorganismos tanto bacterias como levaduras y hongos filamentosos, usando en cada caso el medio y condiciones adecuados. Para *Escherichia coli, Staphylococcus aureus,* los ensayos se hicieron con inoculación por hisopado de una suspensión de $3x10^8$ células/mL en TSA 24 horas a 37 °C, mientras que para *Xanthomona versicatoria* se usó Agar Nutriente a 28 °C durante 48 horas. Las condiciones de crecimiento para *Candida albicans*; *Aspergillus niger, Botrytis cinerea* y *Penicillium expansum* sembrados a partir de una suspensión de $1x10^7$ esporas/mL, fueron Medio Agar Malta a 28 °C durante 5 días.

Los extractos obtenidos de hoja y de semillas germinadas de *P. dubium* y *M. ilicifolia* mostraron tener actividad contra los tres hongos filamentosos analizados. No se observó actividad antibacteriana ni contra la levadura *C. albicans*. En la Figura 1 se muestra un ejemplo de los resultados obtenidos con *P expansum* y *S. aureus*.

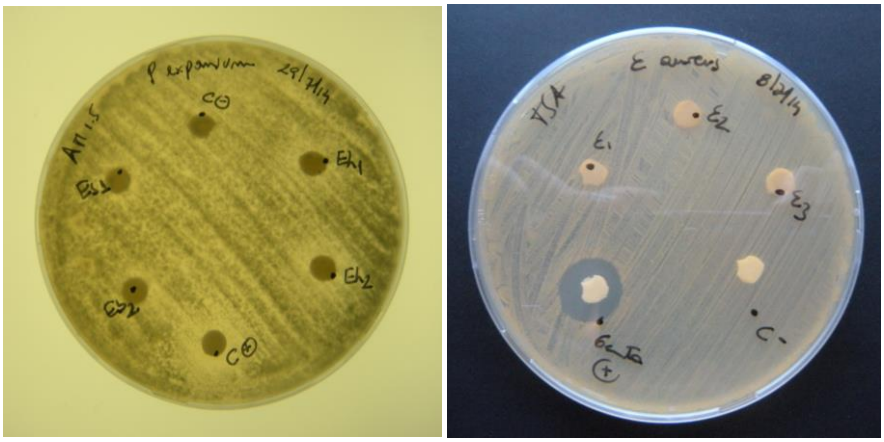


Fig. 1. Ensayos de difusión en placa. Izquierda: Extractos peptídicos de semilla germinada (ES1-congorosa, ES2-ibirapitá) y de hoja (Eh1-congorosa, Eh2-ibirapitá) enfrentados a *Penicillium expansum* C+: control positivo propiconazol, C-: control negativo agua. Derecha: extractos peptídicos de hoja (E1-congorosa, E2-ibirapitá, E3-ceibo) enfrentados a *Staphylococcus aureus*. C+: control positivo, gentamicina, C-: control negativo, agua.

La actividad observada en todos los extractos fue similar, ninguno mostró una actividad sobresaliente que amerite su análisis en particular. Estos resultados. junto a los antecedentes de expresión de genes de AMPs en diversos órganos-tejidos, nos llevó a plantearnos la posibilidad de analizar transcriptomas de hojas, flores y brotes de ambas plantas, priorizando las muestras de semillas germinadas (brotes) dado que diversos trabajos han demostrado la alta expresión de AMPs en este estadío.

Las extracciones de ARN se realizaron por dos métodos: extracción con Trizol (Thermo Fisher Scientific) y con el kit RNeasy Plant Mini Kit (Qiagen). En ambos casos la calidad no fue buena en hojas y flores de *P. dubium* mientras que en *M. ilicifolia* la cantidad fue el factor limitante en brotes y flores (debido a la poca cantidad de muestra inicial de tejido). Finalmente se secuenciaron los transcriptos de semillas germinadas (brotes) de ibirapitá, y de un *pool* de hojas/brotes/flores de congorosa. La preparación de las bibliotecas de ADNc y la secuenciación fueron llevadas a cabo por Macrogen Inc. (Seúl, Corea).

Ensamblado de transcriptomas

A partir de los datos de RNA-seq obtenidos usando la plataforma Illumina HiSeq2000, se realizó el ensamblado de *reads*. En un principio se probó el ensamblado con genomas de referencia. Por tratarse de especies nativas, sus genomas no están secuenciados por lo que se usaron en el caso de *P. dubium* dos genomas: el de *Medicago truncatula* y el de *Glycine max*, dos especies de la misma familia que ibirapitá. En el caso de congorosa, no existen genomas secuenciados de especies de la misma familia, entonces se probó el genoma de *Populus trichocarpa*,

perteneciente al oden Malpighiales, el más cercano al orden Celastrales (Zhang y Simmons 2006) al que pertenece la familia de *M. ilicifolia* (Celastraceae). Los resultados no fueron satisfactorios por lo que decidimos cambiar a un ensamblado *de novo*, para lo cual contamos con la colaboración del Dr. Pablo Smircich. Se utilizó una lista de motivos que de acuerdo con la disposición de los residuos cisteína, definen distintos tipos de AMP (Slavokhotova et al. 2015), para crear un *script* de búsqueda de motivos cisteína, con la colaboración de Santiago Radío.

De novo seedlings transcriptome assembly for prediction of antimicrobial peptides in *Peltophorum dubium*, a South American legume

Abstract

Antimicrobial peptides (AMPs) comprise an essential plant defense mechanism against invading pathogens. Due to their biological properties, these molecules have been considered useful for drug development, as novel agents in disease therapeutics, applicable to both agriculture and medicine. New technologies of massive sequencing open opportunities to discover novel AMPs genes in wild plant species. This work aimed to identify cysteine-rich AMPs from *Peltophorum dubium*, a legume tree from South America. We performed whole-transcriptome sequencing of *P. dubium* seedlings followed by *de novo* transcriptome assembly, uncovering 68 transcripts classified into four families: hevein-like, lipid transfer proteins, defensins, and Snakin/GASA peptides. The classification was based on their cysteine motifs and sequence similarity. No transcripts with similarity to cyclotide, alpha-hairpin or thionin genes were identified in seedling *de novo* transcriptome. We have validated the presence of 18 genes encoding six putative defensins and 12 Snakin/GASA peptides. PCR confirmation from genomic DNA also allowed the characterization of exons-introns structure of the analyzed genes. The present work also allowed to demonstrate that AMP prediction from a wild species is possible from RNA sequencing using a *de novo* assembly strategy, being a starting point for new studies focused on AMP gene evolution and expression. Moreover, given that AMPs are strong candidates for the development of novel biotechnological products, functional studies regarding their potential antimicrobial activity and mode of action are mandatory.

Keywords: Fabaceae, AMPs, RNA-seq, cysteine motifs.

Introduction

Plants are continuously being attacked by pathogens. However, few attacks can produce a systemic infection and disease development since the plants defend themselves by both constitutive and inducible mechanisms. After perception of a pathogen, a signaling cascade is triggered, inducing the strengthening of the cell wall, the production of secondary metabolites and the synthesis of pathogenesis-related (PR) proteins (Benko-Iseppon et al. 2010). Among the different PR protein classes, the so-called PR peptides stand out, due to their size (lower than 10 kDa; Sels et al. 2008) and some features in common like net positive charge at physiological pH and an even number of cysteine residues (Lay and Anderson 2005). In turn, AMPs differ in size, amino acid composition, and molecular structure. Based on homology of amino acid sequences, cysteine motifs, and the three-dimensional structures, a number of distinct groups have been defined, including defensins, thionins, lipid transfer proteins, snakins, cyclotides, and hevein-like proteins (Odintsova and Egorov 2012; Nawrot et al. 2014).

Playing essential roles in plant defense, AMPs have high potential as therapeutic agents. Therefore, they could be included in products used for disease resistance, control, and management in agricultural production, reducing the use of agrochemicals (Montesinos 2007). Further, previous reports show that plant defensins, for example, show activity not only against phytopathogens but also against human pathogens being considered among the most promising molecules for the development of new drugs with antibiotic action (Thevissen et al. 2007). The isolation, characterization, and synthesis of a wide range of effective AMPs would then be essential for the continued development of important products used in medicine and plant protection.

Gene isolation strategies are being increasingly used for the characterization of several AMPs, especially owing to the high number of AMP nucleotide sequences available coming from "omics" data. Analysis of the sequenced plant genomes revealed that AMPs have been under-predicted in model plants like *Arabidopsis thaliana* and *Oryza sativa*, accounting for 2–3% of the gene repertoire of each model species (Silverstein et al. 2007). The sequenced *Medicago truncatula* and *A. thaliana* genomes, for example, possess >300 defensin-like genes, present as multigene families (Graham et al. 2004; Silverstein et al. 2005). In spite of a large number of results available concerning plant AMPs, there is little information on such peptides derived from native species, that constitute the most significant plant biodiversity (Pestana-Calsa et al. 2010). With the advent of the new technologies of massive sequencing (NGS-Next Generation Sequencing), the complete characterization and the global analysis of genomes and transcriptomes are now possible at reasonable costs, even without any previous genomic

information (Martin and Wang 2011). Therefore, genome-scale sequencing technologies open new opportunities to discover novel AMPs from wild plant species. Koehbach et al. (2013) have combined transcriptome mining and mass spectrometry to identify and characterize new cyclotides from Rubiaceae family. A whole-transcriptome sequencing and *de novo* assembly was performed to identify transcripts encoding AMPs from seedlings of the wild growing Poaceae *Leymus arenarius* (Slavokhotova et al. 2015) and the weed *Stellaria media* (L.) Vill (Caryophyllaceae) (Slavokhotova et al. 2017).

Peltophorum dubium (Fabaceae) is a native tree from South America with pharmaceutical use in folk medicine, with reported applications against respiratory, gastrointestinal and skin diseases (Bolson et al. 2015). However, it remains poorly explored for drug development. Interestingly, some works have isolated trypsin inhibitors from *P. dubium* seeds, with activity against insects and rat lymphoma cells (Macedo et al. 2003; Troncoso et al. 2003). Moreover, it is poorly studied genetically, and there is no transcriptomic or genomic data for this legume. Nevertheless, transcriptome sequencing for this non-model plant is now accessible using a *de novo* assembly; allowing AMP prospection from AMP-like transcripts. In this work we performed RNA-seq and *de novo* assembly analysis of *P. dubium* seedlings for AMP prediction, focusing on defensin and Snakin/GASA genes.

Material and methods

Biological Material

Seeds and leaves of *P. dubium* were obtained from Facultad de Agronomía Garden (Montevideo, Uruguay). Seeds were immersed in H_2SO_4 (concentrated grade) during 15 min for scarification. Surface sterilized seeds were germinated on Whatman paper soaked with distilled water in Petri dishes at 28 ºC. Five days old seedlings (ca. 6 cm long) were frozen in liquid nitrogen and stored at -70 ºC until further usage.

RNA and DNA isolation

Total RNA was extracted from five days old seedlings using Qiagen RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Genomic DNA was extracted from leaves using the standard cetyl-trimethylammonium bromide (CTAB) method (Doyle, 1991).

Library preparation and Illumina sequencing

The quality of total RNA was checked using an Agilent 2100 Bioanalyzer (Agilent, USA). The RNA Integrity Number (RIN) of the extracted RNA was 7.8. Library preparation and sequencing were carried out by Macrogen Inc. (Seoul, Korea). 10 µg of total RNA was used for cDNA library construction, using the TruSeq stranded mRNA LT Sample Prep. Kit according to manufacturer's instructions. The library was sequenced using Illumina HiSeq 2000 sequencer in a paired-end 100 bp run.

Transcriptome assembly

Raw reads were filtered with Trimmomatic (Bolger et al. 2014), and the resulting good quality read pairs were assembled using Trinity (Grabherr et al. 2011). The quality of the assembled transcriptome was checked with Transrate (Smith-Unna et al. 2016). For each transcript, all open reading frames longer than 90 nucleotides were obtained using getorf (Rice et al. 2000). Annotation was assigned by performing BLASTp against SwissProt and TrEMBL databases (maximum expected value of $10e^{-5}$).

AMP prediction in *P. dubium* seedlings transcriptome

A list of motifs that define distinct types of AMPs was obtained from Slavokhotova et al. (2015) (Annexe) and used to search the *in-silico* translated ORFs by using regular expressions. Positive peptides were grouped using cd-hit, and for clusters of identical peptides, only one was chosen for further analysis. The presence of signal peptide in a sequence was analyzed using SignalP program (Nielsen 2017). Subcellular localization was predicted with DeepLoc (Almagro Armenteros et al. 2017). Multiple alignments with *P. dubium* deduced peptides (Table S2) and representative AMP family members from other plants (Table S3) were performed in ClustalW from BioEdit program (Hall 1999). The deduced amino acid sequences encoding for the mature *P. dubium* Snakin/GASA peptides were aligned against a set of mature plant proteins with GASA domain. All sequences used were taken from the UniProt database (Apweiler et al. 2004) (Table S3). An unrooted tree was generated using the Neighbor-Joining method of the MEGA package version 5 (Tamura et al. 2011), with 5000 bootstrap replicates.

PCR validation of defensin and Snakin/GASA genes

Six defensin and 12 Snakin/GASA genes were PCR amplified from *P. dubium* genomic DNA by using primers designed from transcript sequences, using Primer3. The primers were designed in 5' and 3' UTRs unless the transcript was not complete (Table S4). The PCR reactions were performed in a 20 µl reaction containing: 1X buffer with 2 mM $MgCl_2$, 0.2 mM dNTPs, 2 µM of each primer, 50 ng template DNA and 0.5 U Taq DNA polymerase (Invitrogen, Carlsbad, USA).

The PCR program was as follows: 94°C for 3 min; followed by 35 cycles of 94°C for 30 sec, 52-58°C for 40 sec and 72°C for 40 sec. The amplified fragments were visualized by agarose gel electrophoresis and purified according to Richero et al. (2013) and cloned into a pGEM-T easy vector (Promega Corporation, Madison, USA). AMP gene sequences were confirmed by sequencing at Macrogen Inc. (Seoul, Korea).

Results and Discussion

Sequencing and *de novo* assembly of *P. dubium* transcriptome.

Illumina sequencing of cDNA library obtained from *P. dubium* seedlings mRNA generated 52 million 100 bp-reads (Table 1). Good quality reads were assembled with Trinity into approx. 127K transcripts. This assembly was evaluated and filtered with the Transrate tool that estimates individual contig quality. Its use reduced the final number of non-redundant transcripts to approx. 108K with an N50 statistic of 1347 (see Table S1).

Table 1. Raw data statistics of Illumina sequencing

Total read bases (pb)	Total reads	GC(%)	AT(%)	Q20(%)	Q30(%)
5,268,156,364	52,159,964	49.12	50.88	94.84	89.17

AMP identification

According to their cysteine residues (cysteine motifs) arrangement, AMPs can be divided into several families including defensins, snakins, lipid-transfer proteins, thionins, alpha-hairpinins, hevein-like, and cyclotides. We predicted potential AMPs in *P. dubium* seedlings transcriptome using previously described "cysteine motifs" obtained from Slavokhotova et al. (2015). As a result, 2644 transcripts with one or more cysteine motif were identified. From these, only 441 were annotated against SwissProt and TrEMBL databases (maximum expected value of 10 e^{-5}). In 63 transcripts, the annotation corresponded to known AMPs, in 101 cases to uncharacterized proteins and in 277 were annotated as other protein families. Of the 441 transcripts with annotation, only 152 deduced peptides presented less than 200 amino acids. We did not discard those cases in which ORFs were not complete (from methionine to stop), but that contained the complete cysteine motif.

THIONINS

Thionins comprise small AMPs of the PR-13 protein family (Sels et al. 2008). Like other plant AMPs, thionins have broad in vitro antifungal and antibacterial activity (Stec 2006). No thionins were identified in *P. dubium* transcriptome. Thionin genes have been identified in 15 different plant species (reviewed by Stec 2006) of the families Santalaceae, Brassicaceae, Poaceae, Ranunculaceae, and Liliaceae. To date, no reports are available describing thionins in Fabaceae species, indicating that this protein group is not present in legumes, as is the case of *P. dubium*. Alternatively, these genes may have a discrete expression or, still, can be absent in seedlings, or the *de novo* assembly did not allow them to be detected.

ALPHA-HAIRPININS

Recently a new trypsin inhibitor with helical hairpin structure was reported as belonging to a new plant AMP family (Oparin et al. 2012). Although variable in amino acid sequences, this α-hairpinin family comprises peptides with the same cysteine motif ($CX_3CX_{1-15}CX_3C$) that share a helix-loop-helix fold stabilized by two disulfide bridges C1–C4 and C2–C3 (Slavokhotova et al. 2014). We found more than 100 transcripts with alpha hairpinin motif but without similarity with the α-hairpinins found in *Fagopyrum esculentum, Echinochloa crus-galli, Zea mays* and *Stellaria media* (reviewed by Slavokhotova et al. 2017). These four species belong to the Poaceae family and it is possible that alpha hairpinins are grass-specific. Several transcripts with α-hairpinin motif had similarity to snakin/GASA genes and were therefore classified within this family in this work. Two of the transcripts found (with the cysteine motif $CX_3CX_2CX_3C$) had similarity with Kunitz trypsin inhibitor.

CYCLOTIDES

Cyclotides are short cyclic peptides (~ 30 amino acids) that have a head-to-tail cyclized backbone and three conserved disulfide-bonds in a knotted arrangement (Craik et al. 1999). They were found in several species of the Rubiaceae, Cucurbitaceae, Violaceae, Solanaceae, Poaceae and Fabaceae families (Koehbach et al. 2013). Poth et al. (2011) reported the isolation of novel cyclotides from *Clitoria ternatea* seeds, being the first report of cyclotides in the Fabaceae. In legumes, according to the findings in *C. terneata*, the biosynthetic origin of a mature cyclotide is an albumin precursor protein, unlike all previously reported cyclotides (Poth et al. 2011). Even though we expected to find cyclotides in *P. dubium*, no transcripts with cyclotides cysteine motifs were found in our seedling transcriptome. Besides, we could not identify transcripts with similarity to their protein precursor: albumin-1. This protein is present in other legumes such as *G. max*, *Pisum sativum*, and *Vigna radiata*. All of them, as well as *C. terneata* belong to Papiloniodeae subfamily, while *P. dubium* belongs to the subfamily

Caesalpinioideae. Albumin-1 could be specific to subfamily Papilionoideae. We only found a transcript with 61 % similarity with cliotide T18 (a cyclotide from *C. terneata*) but with only five cysteines (Cys) (Fig. 1a). This transcript was not complete at 5' and it was scarcely assembled. Alternatively, it could be a pseudogene whose expression may be due to insufficient time for the complete degeneration of regulatory signals (Zou et al. 2009).

We also found one transcript with similarity to an albumin 2S, a seed storage protein. The sunflower trypsin inhibitor peptide SFTI-1, a 14 amino acid cyclic peptide with a single disulfide bond (GRCTKSIPPICFPD), is embedded within a 2S albumin (Mylne et al. 2011). *P. dubium* putative 2S albumin lacks an SFTI-1 like domain (Fig. 1b). 2S albumin large chain seems conserved in *P. dubium*, *H. annuus*, and *G. max*, but the small chain seems to be more variable. *P. dubium* putative albumin exhibits three cysteines between the signal peptide and the Large Chain while the 2S sunflower albumin presents four Cys (two within SFTI-1 and two within the Small chain). In turn, *G. max* 2S albumin has only two Cys (within Small Chain). Thus *G. max* and *P. dubium*, both seem not to contain a cyclic peptide between the signal peptide and the short chain of the 2S albumin. Finally, it cannot be discarded that *P. dubium* seedling transcriptome could have cyclic peptides without similarity with those previously reported.

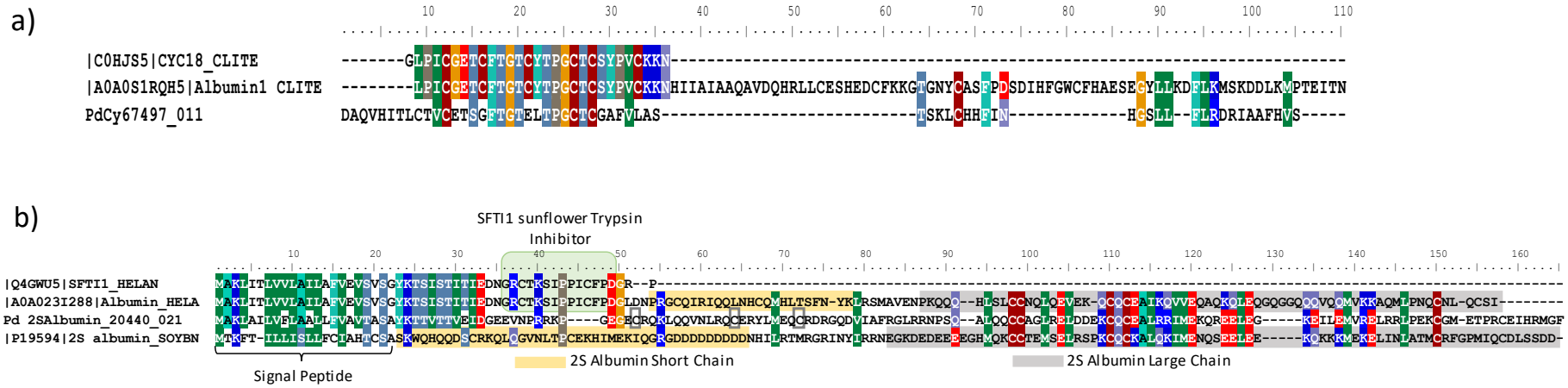


Fig. 1 Sequence comparison of putative *P. dubium* cyclotide precursors with corresponding cyclotide members from other well-studied and model plants. a) *P. dubium* predicted peptide (PdCy67497_011) with similarity with mature cyclotide precursor (albumin1) from *C. terneata.* b) *P. dubium* predicted peptide with similarity with albumin 2S from *H. annuus* and *G. max*. In Pd2S albumin (Pd2S albumin_20440_021) three cysteines between signal peptide and Large Chain are highlighted with gray rectangles. Sequence names to known proteins correspond to its UniProt Entry identifier. CYC18: cliotide 18 from *C. terneata*; CLITE: *C. terneata*; HELAN: *H. annuus*; SOYBN: *G. max*. Threshold (%) for Identity/Similarity shading was 75 %

HEVEIN-LIKE

Hevein-like peptides received this name from hevein, a chitin-binding protein of 43 amino acids found in rubber tree (*Hevea brasiliensis)* (Lee et al. 1991). Members of this AMP family contain a conserved chitin-binding motif involved peptide-carbohydrate interactions with different pathogens, mainly chitin-containing fungi. Most of them possess eight cysteine residues forming four disulfide bridges, but variants with six and ten cysteine residues also occur (Odintsova and Egorov 2012). We found eight transcripts with similarity with pro-hevein, whereas five of them have a hevein motif (with eight cysteines) (Fig. 2), while the remaining three have only the C-terminal domain named Barwin domain. These last have similarities with proteins annotated as PR-4 proteins or hevein-like. PR-4 proteins are classified as chitinases type I and II (Ali et al. 2018). The three identified transcripts could codify class II chitinases since this type of chitinases does not have the amino-terminal hevein motif (Araki and Torikata 1995).

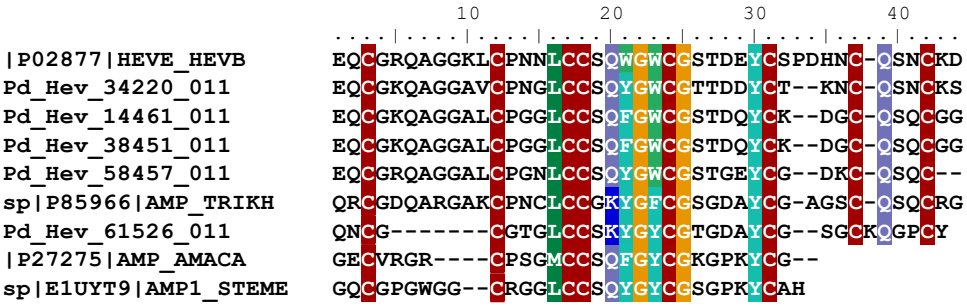


Fig. 2 Sequence comparison of putative *P. dubium* hevein-like peptides with corresponding Hevein-like members from other well-studied plants. Hevein motif alignment of *P. dubium* predicted hevein-like peptides with known hevein-like peptides. Sequence names to known hevein-like correspond to its UniProt Entry identifier. *P. dubium* hevein-like were named Pd_Hev and the number of the transcript. HEVE_HEVB: hevein from *H. brasiliensis*; AMP_TRIKH: hevein-like from *Triticum kiharae*; AMP_AMACA: hevein-like from *Amaranthus caudatus*; AMP1: hevein-like from *S. media*. Threshold (%) for Identity/Similarity shading was 90 %

LIPID-TRANSFER PROTEINS (LTPs)

The name of this group of peptides is due to its ability to facilitate the transfer of phospholipids between a donor and an acceptor membrane (Kader 1996). LTPs are subdivided into two families, which present molecular masses of around 10 (LTP1) and seven (LTP2) kDa. Both subfamilies include proteins with eight cysteines conserved at specific positions (Carvalho and Gomes 2007). Due to their possible role in plant defense, LTPs are recognized as PR proteins and are classified in the PR-14 family. Their ability to bind a wide range of lipids could explain their inhibitory activity against fungi and some bacterial pathogens (Sels et al. 2008). This

antimicrobial activity could result from the interaction with biological membranes, probably leading to membrane permeabilization (Kader 1996). We found 28 transcripts with LTP cysteine motif, 18 with similarity with subfamily 1 (Fig. 3a) and 10 with subfamily 2 (Fig. 3b). We observed three groups within subfamily 1, the first (I) presents similarities with typical LTP1 (like A0AT29 from *Lens culinaris* and Q42589 from *A. thaliana*) since 13 (group II) have similarities with EARLI1, a putative Arabidopsis LTP. This protein has a putative signal peptide of 25 amino acids at the N-terminus, a hydrophilic proline-rich domain in the middle, and a hydrophobic C-terminus (with eight Cys and high similarity to plant LTPs) (Li et al. 2012). In group III there is only one protein which, like the others has 8 Cys, where the third and fourth Cys are consecutive in the polypeptide chain, and the fifth and sixth Cys are separated by only one residue, but the number of residues among the remaining cysteines differs from the other two groups. This protein has a high similarity with a putative LTP (A0A1S3TFR4) described in *Vigna radiata*. Within subfamily 2 we detected a typical LTP2 (group I) with five members and another (group II), that exhibits a variable number of residues in the C-terminus (not shown in the alignment; see Supplementary Table S2), following the conserved motif of 8 Cys. The number of putative LTP transcripts in seedlings of *P. dubium* (28) was similar to that found from the weed *Stellaria media* (31) (Slavokhotova et al. 2017). LTPs are widely distributed in the plant kingdom and seem to be abundant in plant genomes forming multigenic families of related proteins. Genome-wide analysis of the LTP gene family identified 52 members in rice and 49 in Arabidopsis, many of them arranged in tandem duplication repeats (Boutrot et al. 2008). If the number of LTP genes was comparable in the genome of *P. dubium*, then, half of the genes would be expressed in seedlings.

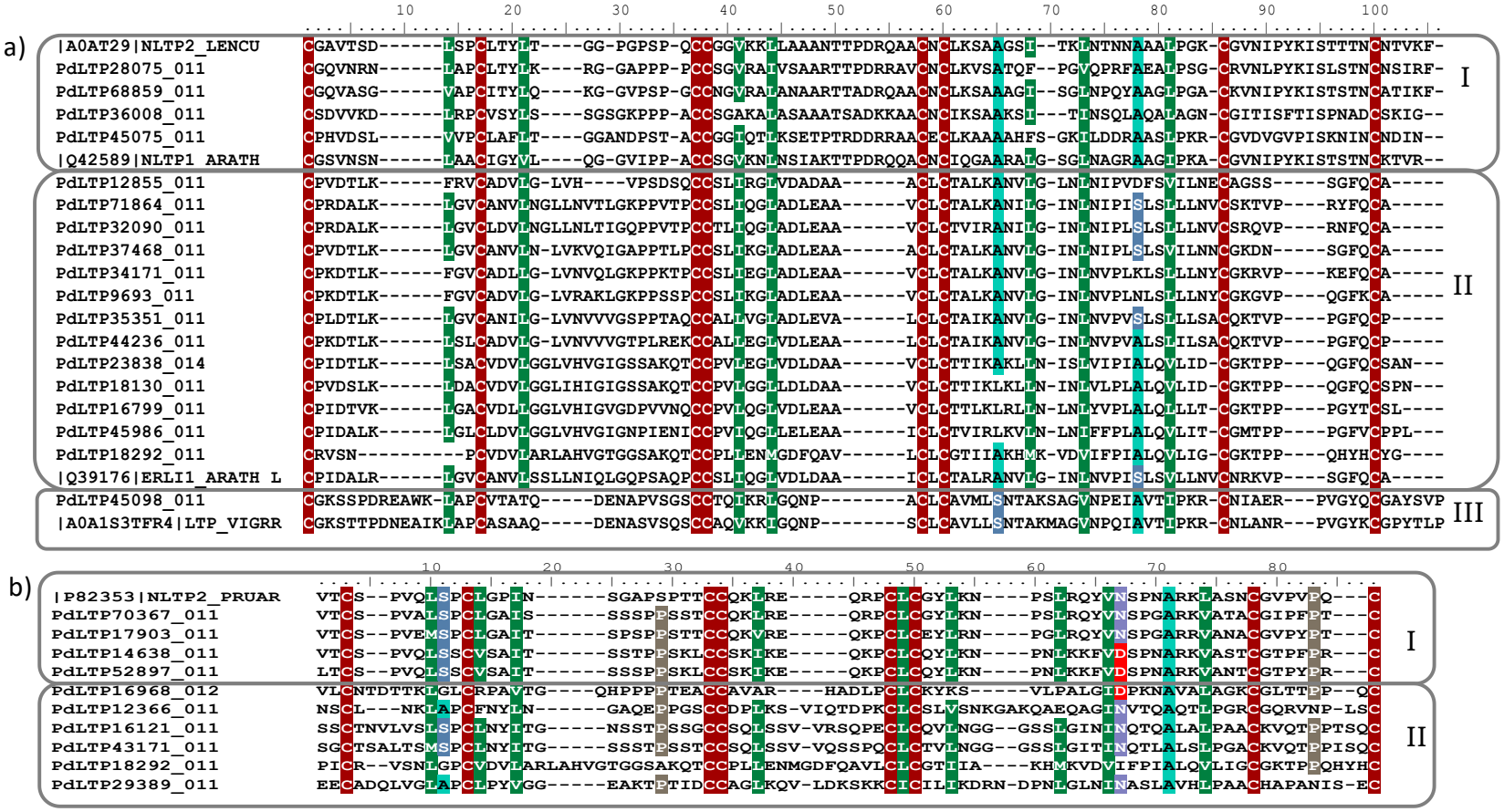


Fig. 3 Sequence comparison of putative *P. dubium* LTP peptides with corresponding LTP members from other well-studied and model plants.

a) Alignment of *P. dubium* predicted mature LTP1 peptides and known LTP subfamily 1 peptides. Three groups were identified within this subfamily showing similarity to different LTP1 members from other plants. b) Alignment of *P. dubium* predicted mature LTP2 and known LTP subfamily 2 peptides. Two groups were identified within this subfamily. Sequence names to known LTP correspond to its UniProt Entry identifier. *P. dubium* LTPs were named PdLTP and the number of the transcript. LENCU: *Lens culinaris*; ARATH: *A. thaliana*; VIGRR: *Vigna radiata*; PRUAR: *Prunus armeniaca*. Threshold (%) for Identity/Similarity shading was 75 %

DEFENSINS

Defensins seem to be the only class of AMP that are conserved across the eukaryotic lineage and have a widespread distribution throughout the plant kingdom (Thomma et al. 2002). The constitutive expression of defensin genes constitutes a primary plant defense against microorganisms colonization (Broekaert et al. 1995), but the production of these molecules can also be induced in response to pathogen attack, wounding and some abiotic stresses (Lay and Anderson 2005). The primary structure of these peptides generally consists of a signal peptide at the amino-terminal, and a basic, cysteine-rich mature peptide (45 to 54 amino acids). However, defensin-like peptides with unusual structures, have been identified (De Coninck et al. 2013). For several plant defensins, their three-dimensional structure has been elucidated by NMR spectroscopy. Despite limited amino acid sequence identities, these defensins have similar overall fold features with one α-helix and three antiparallel β-sheets that are stabilized by four intramolecular disulfide bonds formed by eight conservative cysteine residues (Lay and Anderson 2005; Carvalho and Gomes 2009). This cysteine-stabilized αβ motif [CSαβ; $CX_nCXXXCX_nCX_nCXC$, where C=Cys, X=any amino acid, and n indicates a non-conserved number of amino acids] (Silverstein et al. 2005) includes three disulfide bridges and is remarkably similar to those of insect defensins (Cornet et al. 1995). These peptides are also characterized by the occurrence of a γ-core motif $GXCX_{3-9}C$ (Yount and Yeaman 2004). Defensins from plants belong to the *cis*-defensin superfamily whereas mammalian defensins belong to the trans-defensin superfamily. This classification was based on their secondary structure orientation, disulfide topology, tertiary structure similarities, and precursor gene sequence. These two independent superfamilies (*cis* and *trans* defensins) with similar structures have arisen via an extreme case of convergent evolution (Shafee et al. 2017).

We found ~ 140 transcripts with defensin like cysteine motifs of which ~80 did not have annotation (maximum expected value of $10e^{-5}$). In several cases, the ORF including the cysteine motif was not the longest for the transcript and/or had an annotation in another reading frame, so they were discarded. In cases when the ORF with the cysteine motif was the longest and had a different annotation (not correspond with defensins, generally encoding proteins with more than 200 aa) the sequences were discarded. Remaining transcripts without annotation only those where the ORF with the cysteine motif was the longest, and had no other annotation in another reading frame were further analyzed. Many contained some kind of repeated sequence, like transposons. Some of these transcripts could be pseudogenes or assembly errors, so they were not taken into account.

A list of 14 defensin-like genes was finally considered. The deduced peptides were separated into three subgroups. The first (Fig. 4a) include 11 peptides with similarity with "true defensins", having eight Cys at conserved positions, with the classic cysteine-stabilized (CSαβ) motif that was common for most known defensins and a γ-core motif. Eight of them had other conserved amino acids: a glycine (G), a glutamic acid (E), and a serine (S). These residues are present in all defensins whose 3D structure was determined, except for a defensin from *Petunia hybrida*. Like the latter, four of the *P. dubium* deduced peptides do not present these residues in conserved positions.

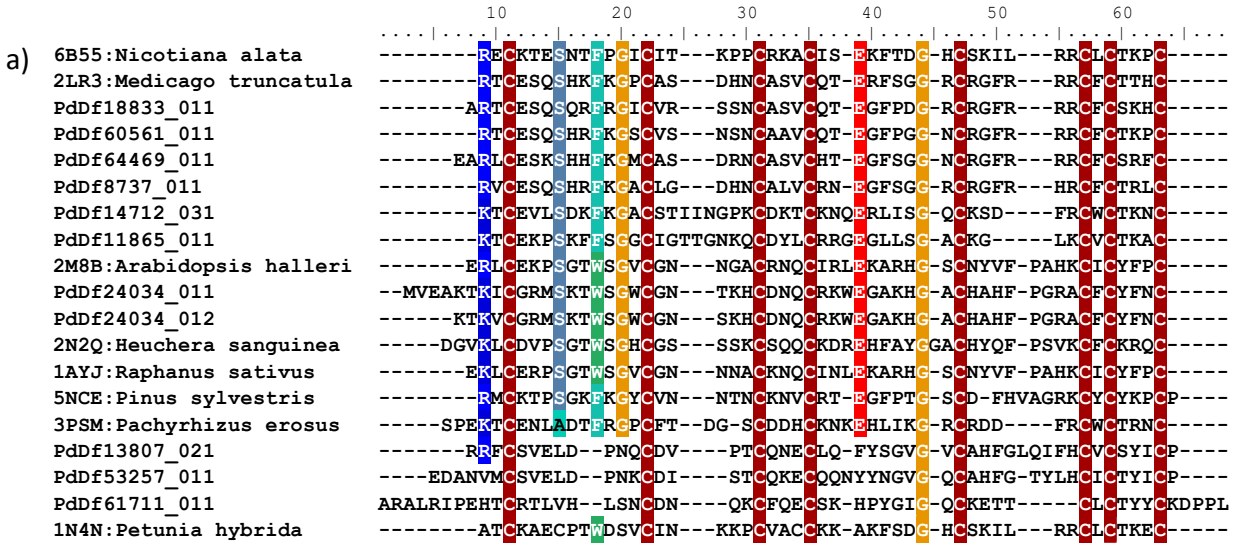
The second group includes two peptides with eight Cys that have the (CSαβ) motif, whereas de γ-core motif is absent. These proteins have similarities with uncharacterized proteins and with Kazal-type serine protease inhibitor. However, this kind of protein has six Cys with a cysteine motif $CX_{6\text{-}9}CX_7CX_6YX_3CX_{2\text{-}3}CX_{17}C$ (Pariani et al. 2016) where CXC is not present. This two putative defensins were aligned with floral defensins identified in *Nicotiana alata* and *Petunia hybrida* (Fig. 4b) These defensins are characterized by having a prodomain at the C-terminus (Lay et al. 2003), belonging to the class II defensins, according with its classification in solanaceous species. In the first class, the precursor protein is composed of an endoplasmic reticulum signal sequence and a mature defensin domain. These proteins enter the secretory pathway and have no obvious signals for post-translational modification or subcellular targeting. The second class is produced as larger precursors with a C-terminal prodomain of about 33 amino acids (Lay and Anderson 2005). Our two putative defensins could have a C-ter prodomain. However, only six cysteines exhibited conserved positions with respect to the mature peptide of *N. alata* and *P. hybrida* defensins. These *P. dubium* peptides have some characteristics in common with defensins, and some in common with Kazal-type serine protease inhibitors. Further studies are necessary to define their classification and function.

The third group includes only one transcript with the defensin cysteine motif $CX_{3\text{-}21}CX_{2,12}CX_{3,4}CX_{3,15}CX_{4,23}CCC$ (Fig 4c). The predicted peptide has similarity with some defensin-like proteins from *A. thaliana* that have the sixth, seventh and eighth cysteine located consecutively. This peptide presents a modified αβ motif but maintains de γ-core motif $GXCX_{3\text{-}9}C$. Of the 14 defensin-like transcripts, two are not complete at 5'; of these, no signal peptide was detected for Pd61711_011 predicted peptide. The remaining 13 defensin-like peptides appear to have signal peptide detected by SignalP tool (See Table S2), could be soluble and have extracellular localization, based on DeepLoc analysis.

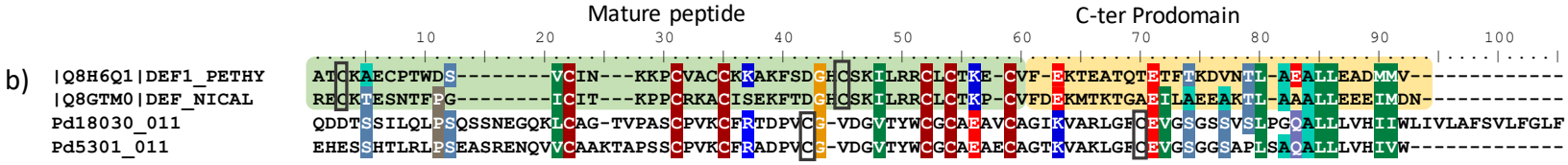
According to Silverstein et al. (2005), the number of defensins in plant genomes has been underestimated. These authors found more than 300 defensin-like genes in Arabidopsis. Of these, about 51 are expressed in seedlings while a similar number (56) was found to be

expressed in seeds of the *M. truncatula* legume (Tesfaye et al. 2013). The number of putative defensins genes found by our group in the *P. dubium* seedling transcriptome is a third less but is similar to that found from the weed *Stellaria media* seedling transcriptome (16) (Slavokhotova et al. 2017). *De novo* assembled strategy used in this work and in Slavokhotova et al. (2017) did not allow detecting a higher number of defensin-like genes. Alternatively, *P. dubium* seedlings expressed a lower number of defensin genes.

a)


6B55:Nicotiana alata
2LR3:Medicago truncatula
PdDf18833_011
PdDf60561_011
PdDf64469_011
PdDf8737_011
PdDf14712_031
PdDf11865_011
2M8B:Arabidopsis halleri
PdDf24034_011
PdDf24034_012
2N2Q:Heuchera sanguinea
1AYJ:Raphanus sativus
5NCE:Pinus sylvestris
3PSM:Pachyrhizus erosus
PdDf13807_021
PdDf53257_011
PdDf61711_011
1N4N:Petunia hybrida

b)


Mature peptide
C-ter Prodomain
|Q8H6Q1|DEF1_PETHY
|Q8GTM0|DEF_NICAL
Pd18030_011
Pd5301_011

c)

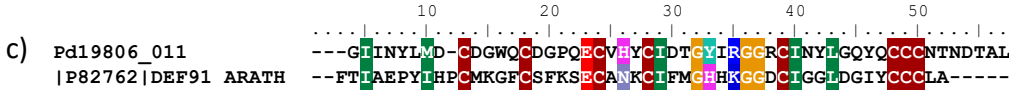

Pd19806_011
|P82762|DEF91 ARATH

Fig. 3 Sequence comparison of putative *P. dubium* defensin-like peptides with corresponding members of other well-studied and model plants. a) Alignment of *P. dubium* predicted "true" mature defensin peptides (named PdDf and the number of the transcript) and known mature defensin peptides. All the sequences used in this alignment correspond to defensins whose 3D structure has been determined. The name of each sequence corresponds to the PDB identifier. b) Alignment of *P. dubium* predicted defensin-like peptides and known defensin peptides with a C-ter Prodomain. that differ in location with respect to known defensins are highlighted with gray rectangles. Sequence names to known defensins correspond to its UniProt Entry identifier. (Q8H6Q1=1N4N; Q8GTMD=6B55 at the alignment in a). PETHY: *P. hybrida*; NICAL: *N. alata*. c) Alignment of a *P. dubium* predicted mature defensin-like peptide and a defensin-like sequence from *A. thaliana* that have the sixth, seventh and eighth cysteine located consecutively. Sequence name to *A. thaliana* defensin corresponds to its UniProt Entry identifier. Threshold (%) for Identity/Similarity shading was 75 %

SNAKINS (Snakin/GASA)

Snakins belong to a group of proteins encoded by Snakin/GASA genes. This family involves a group of widely distributed peptides among higher plants that are characterized by having a GASA (Gibberellic Acid Stimulated in Arabidopsis) domain of approximately 60 amino acids, with 12 cysteine residues that may be involved in the formation of up to six disulfide bonds (Oliveira-Lima et al. 2017). Snakin/GASA proteins contain three distinct domains: a signal peptide with 18-29 residues, a variable region that is highly divergent between family members, both in sequence length and amino acid composition, and a conserved C-terminal region: the GASA domain (Nahirñak et al. 2012). This cysteine-rich C-terminal domain contains 12 cysteine amino acids, in highly conserved positions ($XCX_3CX_3CX_8CX_3CX_2CCX_2CXCX_{11}CXCX_{12}CX$). Constitute a family of AMPs defined from potato Snakin-1 (StSN1, Segura et al. 1999) and Snakin-2 (StSN2, Berrocal-Lobo et al. 2002), the first peptides isolated of this family. Most Snakin/GASA genes are regulated by plant hormones, and they seem to play an essential role in plant development, plant responses to biotic or abiotic stress and in redox homeostasis (Nahirñak et al. 2012).

Our group has previously isolated a snakin gene (PdSN1) from *P. dubium* genome, similar to StSN1. PdSN1 is expressed in leaves and seedlings and presents *in vitro* antimicrobial activity when produced in *E. coli* (Rodríguez-Decuadro et al. 2018). In this work, we found 18 transcripts with snakin cysteine motif, including PdSN1. Alignment of mature predicted peptides is shown in Fig. 4, following their classification in subfamilies I, II and III (according to Berrocal-Lobo et al. 2002). Two predicted peptides have higher similarity with StSN1, one of them is PdSN1 (PdSN_16208_011). Ten peptides have similarities with StSN2 and six with StSN3 (M1B38_SOLTU). Some conserved residues within each subfamily were shaded in gray. Two

members of subfamily II (PdSN_11759_011 and 16337_011) have a proline-rich variable region like the Arabidopsis protein GASAE (Roxrud et al. 2007).

All predicted SN/GASA peptides have a peptide at the N-terminus (with 19 to 27 amino acids) according to to SignalP software and are predicted to be soluble and with extracellular location according to DeepLoc. However, the peptides with a proline-rich domain could be important components of cell wall proteins. Proline-rich proteins (PRPs) seem to play an important role in cell wall signal transduction cascades, plant development and stress tolerance, contributing to cell wall modification under stress conditions (Kavi Kishor et al. 2015). A subclass of PRPs is the hybrid PRPs, that represent putative cell wall proteins consisting of a repetitive proline-rich N-terminal domain and a conserved C-terminal domain (Kavi Kishor et al. 2015), like PdSN_11759_011 and PdSN16337_011. These putative SN/GASA peptides could fulfill their function at the cell wall level.

Alignment analysis of *P. dubium* Sn/GASA mature sequences and other plant mature proteins with GASA domain revealed three groups that coincide with the three subfamilies described for Berrocal-Lobo et al. (2002) (Fig. 5). For this analysis, sequences with reported experimental evidence such as antimicrobial activity were used, including a snakin-2 from tomato (|E5KBY0|E5KBY0_SOLLC; Herbel et al. 2015), a snakin-1 from alfalfa (|H9D2D5|H9D2D5_MEDSA; García et al. 2014), a snakin-1 from pepper (|B2ZAW4|B2ZAW4_CAPAN; Mao et al. 2011), a snakin-3 from potato (|M1BA38|M1BA38_SOLTU; Nahirñak et al. 2016), and sequences obtained from the curated UniProt database. A genome-wide search for new Snakin/GASA members in potato uncovered 16 Snakin/GASA genes, in addition to StSN1 and StSN2, a number equal to that found by our group in the *P. dubium* seedling transcriptome; however the number within each subfamily was different (subfamily I: four genes; subfamily II: seven genes; subfamily III: seven genes) (Nahirñak et al. 2016). Fourteen GASA proteins have been described in Arabidopsis (Roxrud et al. 2007). If genomes of different plant species have a similar number of Snakin/GASA genes, then the seedlings of *P. dubium* could be expressing all or almost all of the Snakin/GASA genes. Moreover, the number of *P. dubium* Snakin/GASA transcripts was similar to that found from the weed *Stellaria media* seedling transcriptome (16) (Slavokhotova et al. 2017). The identification of a large number of these genes in several species with a highly conserved domain suggests that they contribute to an important biological function.

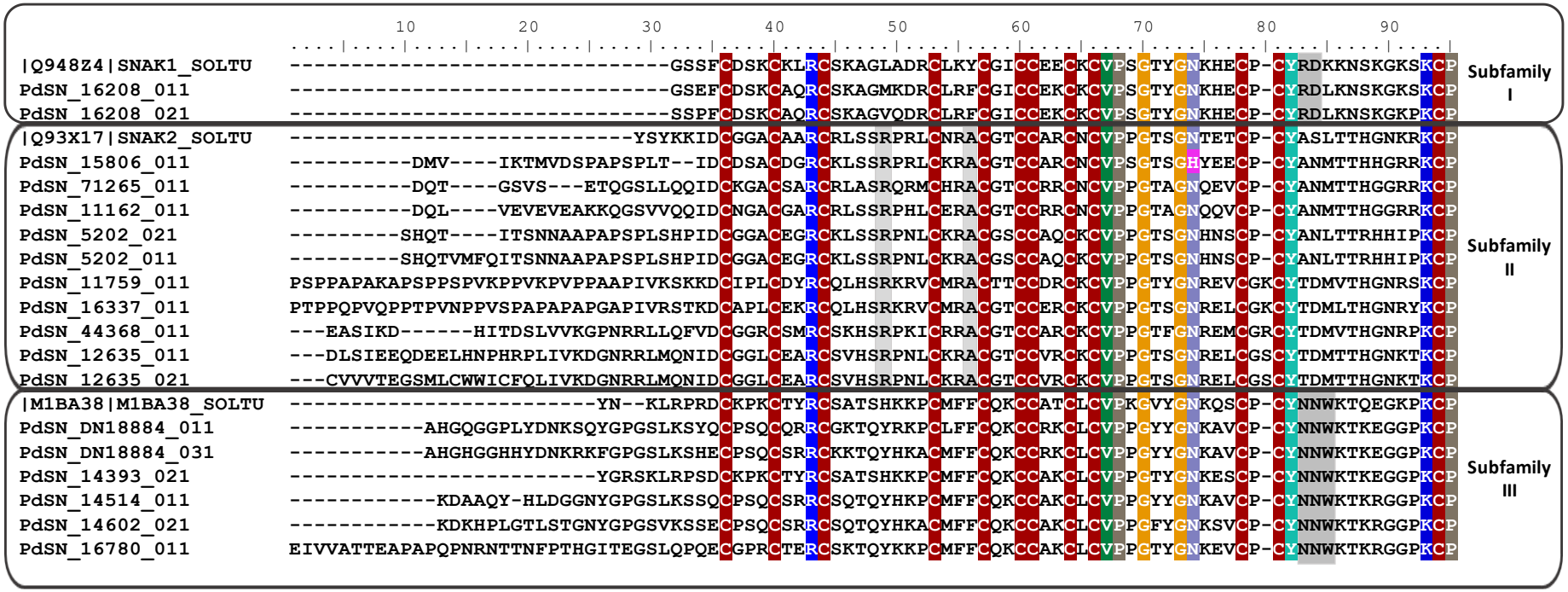


Fig. 4 Sequence comparison of putative *P. dubium* Snakin/GASA peptides with corresponding Snakin/GASA members from *Solanum tuberosum*. Alignment of *P. dubium* predicted mature Snakin/GASA peptides and snakin peptides from potato. Alignment revealed three groups that coincide with the three subfamilies described for Berrocal-Lobo et al. (2002). Conserved residues exclusively within each subfamily were shaded in gray. *S. tuberosum* sequence names correspond to its UniProt Entry identifier. Threshold (%) for Identity/Similarity shading was 100 %

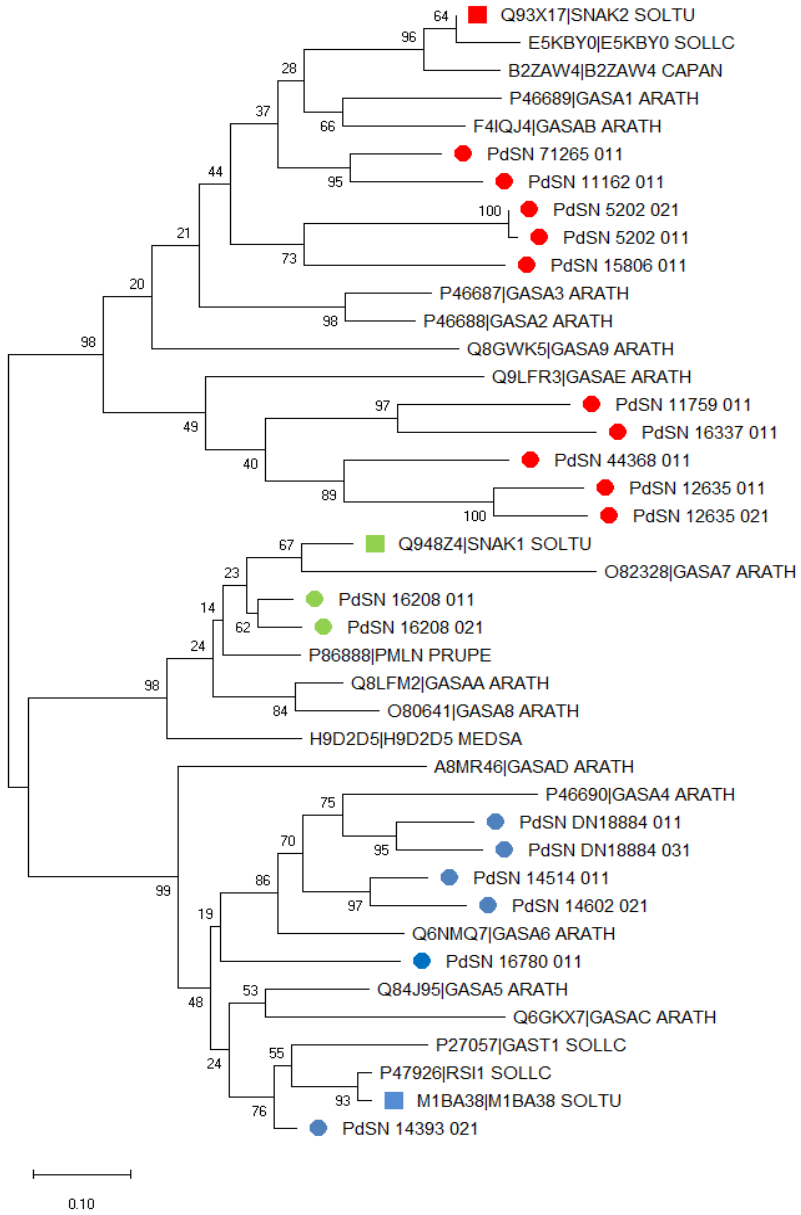


Fig. 5 Neighbor-Joining unrooted tree of predicted *P. dubium* Snakin/GASA mature peptides and members of the plant Snakin/GASA protein family. The deduced amino acid sequences of the predicted genes were aligned with a set of mature plant proteins with GASA domains. Sequences used were manually annotated and retrieved from the UniProt database (http://www.uniprot.org/). Four snakins with reported experimental evidence were also included (|E5KBY0|E5KBY0_SOLLC, |H9D2D5|H9D2D5_MEDSA, |B2ZAW4|B2ZAW4_CAPAN, |M1BA38|M1BA38_SOLTU). Squares indicate snakin-1 snakin-2 and snakin-3 (M1BA38) from *S. tuberosum* and circles indicate snakin/GASA predicted peptides from *P. dubium*. In green, members from subfamily I; in red, members of subfamily II, and in blue members from subfamily III. Values in the nodes regard bootstrap values (5000 replicates). Each sequence was named according to its UniProt Entry identifier, followed by a Mnemonic identifier of a UniProtKB entry. ARATH: *A. thaliana*, SOLTU: *S. tuberosum*, SOLLC: *Solanum lycopersicum*, PRUPE: *Prunus persica*. MEDSA: *Medicago sativa*, CAPAN: *Capsicum annuum*

Defensin and Snakin/GASA genes validation by PCR

To confirm the reliability of the transcriptome assembly, 18 transcripts encoding six putative defensin and 12 Snakin/GASA genes were selected for PCR amplification from genomic DNA, also allowing the description of their exons-introns structure. Of the six defensin genes examined, we obtained four genes from the start codon to stop codon (Fig 6). For the remaining two (PdDf14712_011 and PdDf18833_011) only the part corresponding to the mature peptide was amplified. As for other reported defensins (Carvalho and Gomes 2009), PdDf gene structure is comprised of two exons and one intron where the first exon entirely encodes the signal peptide, and the second exon encodes the last amino acids of the signal peptide and the mature defensin peptide. We believe that for this reason, we could not detect the presence of an intron in the genes PdDf14712_011 and PdDf18833_011, only verifying part of the second exon. The intron is variable in size (Carvalho and Gomes 2009), finding three defensin genes with a similar size while the four has ~200 nt more (Fig. 6).

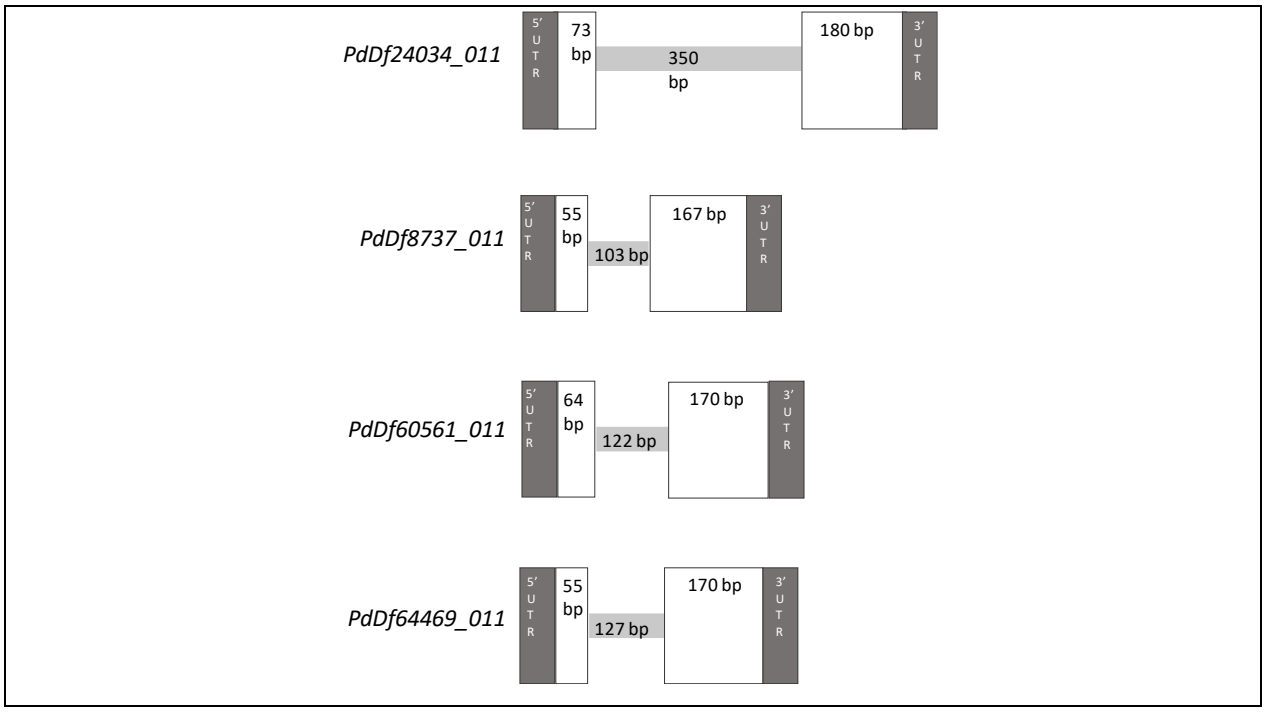


Fig. 6 Schematic representation of the structure of four *P. dubium* complete defensin genes. The corresponding exons and introns and their respective sizes are shown

For Snakin/GASA family, we verified two genes from subfamily I (Pd16208_011 was previously verified and named PdSN1; Rodríguez-Decuadro et al. 2018), seven from subfamily II and three from subfamily III (Fig. 7). The comparison of the *de novo* assembled transcript sequences with the corresponding genomic DNA sequences showed that the coding sequences of all Snakin/GASA genes verified are interrupted by one to three introns. According to the number of introns, subfamily II seems to be the most variable with one two or three introns while subfamily I seems to have only one intron. These results are similar to those found by Nahirñak et al. (2016) in potato but with the difference that the most variable subfamily was III. The number of introns within subfamily I was one for all de genes reported in potato. All verified snakin/GASA genes have a very conserved size in regard to the last exon, with 182 or 185 nucleotides; this exon encodes the GASA domain of these proteins. According to the SignalP analysis, the first exon at subfamilies I and II encodes the signal peptide and a small fragment of the mature peptide; however, for family III, it encodes only the signal peptide. The length of the introns varied from ~100 to ~250 nucleotides.

Some of the Snakin/GASA predicted genes are very similar and may have arisen by duplication. PdSN16208_011 and 021 are structurally similar, with an ORF with five differences at signal peptide and five substitutions at mature peptide. PdSN5202_011 and 021 are highly similar even in the UTR regions, (differing in 12 nucleotides, at the ORF, that codes in 021 for 4 amino acids more) so we could not verify both genes, only being able to verify 5202_021. PdSN12635_011 and 021 would encode proteins of 115 and 111 amino acids with an identical GASA domain but flexible at the signal peptide and the variable region. The 3' UTR was very similar, so we could not design two R primers, being able to verify only PdSN5202_021.

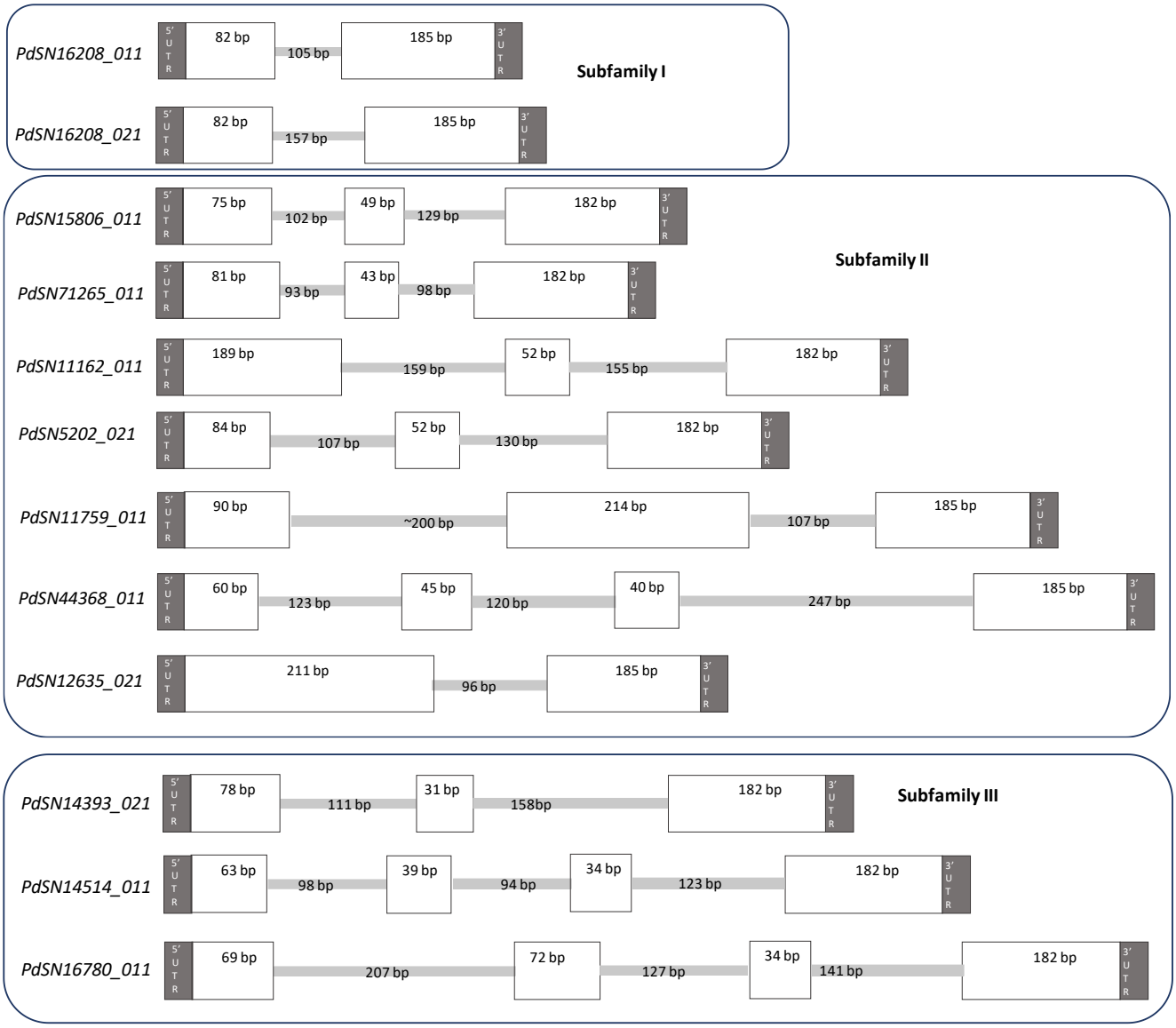


Fig. 7 Schematic representation of the structure of 12 *P. dubium* Snakin/GASA genes. The corresponding exons and introns and their respective sizes are shown. Genes were grouped according to the classification into subfamilies I, II and III

Conclusions

In this work, the first survey of *P. dubium* antimicrobial peptides were carried out from a *de novo* seedlings transcriptome. Considering that AMPs have low primary structure similarity, they could not be revealed by performing only homology-based search. However, they share conserved cysteine motifs so we could identify putative AMPs from *P. dubium* native legume by searching for those motifs, and found eight Hevein-like, 28 lipid transfer proteins, 14 defensins and 18 snakin/GASA putative genes. Although no reference genome is available for this South American legume, NGS-technologies allowed inferring on the abundance and diversity of AMPs

in *P. dubium* seedlings transcriptome. This is a good starting point to deepen the knowledge on their gene organization and evolution and their precise role in the plant defense system. Validated genes deserve more focused investigations (expression profiling and functional analysis), regarding the biological roles of defensin and Snakin/GASA genes during plant development and defense response. Also, these genes could be the first targets for production in heterologous systems to evaluate their potential as antimicrobial agents.

References

Ali S, Ganai BA, Kamili AN, Bhat AA, Mir ZA, Bhat JA, Tyagi A, Islam ST, Mushtaq M, Yadav P, Rawat S, Grover A (2018) Pathogenesis-related proteins and peptides as promising tools for engineering plants with multiple stress tolerance. Microbiol Res 212–213:29–37. doi: 10.1016/j.micres.2018.04.008

Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O (2017) Sequence analysis DeepLoc : prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387–3395. doi: 10.1093/bioinformatics/btx431

Almasia NI, Molinari MP, Maroniche GA, Nahirñak V, Barrios Barón MP, Taboga OA, Vazquez-Rovere C (2017) Successful production of the potato antimicrobial peptide Snakin-1 in baculovirus-infected insect cells and development of specific antibodies. BMC Biotechnol 17:1–11. doi: 10.1186/s12896-017-0401-2

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, Donovan CO, Redaschi N, Yeh LL (2004) UniProt : the Universal Protein knowledgebase. Nucleic Acids Res 32:D115–D119. doi: 10.1093/nar/gkh131

Araki T, Torikata T (1995) Structural classification of plant chitinases. two subclasses in Class I and Class II chitinases. Biosci Biotech Biochem 59:336–338 . doi: 10.1248/cpb.37.3229

Benko-Iseppon AM, Galdino SL, Calsa T, Kido EA, Tossi A, Belarmino LC, Crovella S (2010) Overview on plant antimicrobial peptides. Curr Protein Pept Sci 11:181–8. doi: 10.2174/138920310791112075

Berrocal-Lobo M, Segura A, Moreno M, López G, García-Olmedo F, Molina A (2002) Snakin-2 , an Antimicrobial Peptide from Potato Whose Gene Is Locally Induced by Wounding and Responds to Pathogen Infection 1. Plant Physiol 128:951–961. doi: 10.1104/pp.010685.1

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data Bioinformatics 30**:**2114-2120.

Bolson M, Hefler SR, Dall'Oglio Chaves EI, Gasparotto Junior A, Cardozo Junior EL (2015) Ethno-

medicinal study of plants used for treatment of human ailments, with residents of the surrounding region of forest fragments of Paraná, Brazil. J Ethnopharmacol 161:1–10. doi: 10.1016/j.jep.2014.11.045

Boutrot F, Chantret N, Gautier MF (2008) Genome-wide analysis of the rice and arabidopsis non-specific lipid transfer protein (nsLtp) gene families and identification of wheat nsLtp genes by EST data mining. BMC Genomics 9:1–19. doi: 10.1186/1471-2164-9-86

Broekaert WF, Terras FR, Cammue BP, Osborn RW (1995) Plant defensins: novel antimicrobial peptides as components of the host defense system. Plant Physiol 108:1353–1358. doi: 10.1104/pp.108.4.1353

Carvalho A de O, Gomes VM (2007) Role of plant lipid transfer proteins in plant cell physiology-A concise review. Peptides 28:1144–1153. doi: 10.1016/j.peptides.2007.03.004

Carvalho DO, Gomes VM (2009) Peptides Plant defensins — Prospects for the biological functions and biotechnological properties. Peptides 30:1007–1020. doi: 10.1016/j.peptides.2009.01.018

Cornet B, Bonmatin JM, Hetru C, Hoffmann JA, Ptak M, Vovelle, F (1995). Refined three-dimensional solution structure of insect defensin A Structure 3: 435-448.

Craik DJ, Daly NL, Bond T, Waine C (1999) Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J Mol Biol 294:1327–1336. doi: 10.1006/jmbi.1999.3383

De Coninck B, Cammue BPA, Thevissen K (2013) Modes of antifungal action and in planta functions of plant defensins and defensin-like peptides. Fungal Biol Rev 26:109–120. doi: 10.1016/j.fbr.2012.10.002

Doyle J (1991) DNA protocols for plants-CTAB total DNA isolation. In: Hewitt GM, Johnston A (eds) Molecular Techniques in Taxonomy, Springer, Berlin, pp. 283–293

García AN, Ayub ND, Fox AR, Gómez MC, Diéguez MJ, Pagano EM, Berini CA, Muschietti JP, Soto G (2014) Alfalfa snakin-1 prevents fungal colonization and probably coevolved with rhizobia. BMC Plant Biol 14:248

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011:29**:**644-652.

Graham M a, Silverstein K a T, Cannon SB, Vandenbosch K a (2004) Computational Identification and Characterization of Novel Genes from Legumes 1. Plant Physiol 135:1179–1197. doi: 10.1104/pp.104.037531.exchange

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98

Herbel V, Schäfer H, Wink M (2015) Recombinant production of snakin-2 (an antimicrobial peptide from tomato) in E. Coli and analysis of its bioactivity. Molecules 20:14889–14901. doi: 10.3390/molecules200814889

Kader J (1996) Lipid-Transfer Proteins in Plants. Annu Rev Plant Physiol Plant Mol Biol 47:627–654. doi: 10.1146/annurev.arplant.47.1.627

Kavi Kishor PB (2015) Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. Front Plant Sci 6:1–17. doi: 10.3389/fpls.2015.00544

Koehbach J, Attah AF, Berger A, Hellinger R, Kutchan TM, Carpenter EJ, Rolf M, Sonibare MA, Moody JO, Wong GKS, Dessein S, Greger H, Gruber CW (2013) Cyclotide discovery in Gentianales revisited--identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants. Biopolymers 100:438–452. doi: 10.1002/bip.22328

Lay FT, Anderson MA (2005) Defensins - Components of the Innate Immune System in Plants. Curr Protein Pept Sci 6:85–101. doi: 10.2174/1389203053027575

Lay FT, Brugliera F, Anderson MA (2003) Isolation and Properties of Floral Defensins from Ornamental Tobacco and Petunia 1. 131:1283–1293 . doi: 10.1104/pp.102.016626.)

Lee HI, Broekaert WF, Raikhel N V. (1991) Co- and post-translational processing of the hevein preproprotein of latex of the rubber tree (*Hevea brasiliensis*). J Biol Chem 266:15944–15948

Li L, Zhang C, Xu D, Schläppi M, Xu ZQ (2012) Expression of recombinant EARLI1, a hybrid proline-rich protein of Arabidopsis, in *Escherichia coli* and its inhibition effect to the growth of fungal pathogens and S*accharomyces cerevisiae.* Gene 506:50–61. doi: 10.1016/j.gene.2012.06.070

Macedo M, Freire M, Cristina E, Lı M, Toyama MH (2003) A trypsin inhibitor from Peltophorum dubium seeds active against pest proteases and its effect on the survival of *Anagasta kuehniella* (Lepidoptera:Pyralidae). Biochim Biophys Acta 1621 1621:170–182. doi: 10.1016/S0304-4165(03)00055-2

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

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671–682. doi: 10.1038/nrg3068

Montesinos E (2007) Antimicrobial peptides and plant disease control. FEMS Microbiol Lett 270:1–11. doi: 10.1111/j.1574-6968.2007.00683.x

Mylne JS, Colgrave ML, Daly NL, Chanson AH, Elliott AG, Mccallum EJ, Jones A, Craik DJ (2011) Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. Nat

Chem Biol 7:1–3. doi: 10.1038/nchembio.542

Nahirñak V, Almasia NI, Hopp HE, Vazquez-Rovere C (2012) Involvement in hormone crosstalk and redox homeostasis Snakin/GASA proteins. Plant Signal Behav 7:1004–1008

Nahirñak V, Rivarola M, Gonzalez de Urreta M, Paniego N, Hopp HE, Almasia NI, Vazquez-Rovere C (2016) Genome-wide Analysis of the Snakin/GASA Gene Family in *Solanum tuberosum* cv. Kennebec. Am J Potato Res doi: 10.1007/s12230-016-9494-8

Nawrot R, Barylski J, Nowicki, G, Broniarczyk J, Buchwald W, Goździcka-Józefiak A (2014) Plant antimicrobial peptides. Folia microbiol 59:181-196.

Nielsen H (2017) Predicting secretory proteins with SignalP. In: D. K (ed) Methods in Molecular Biology. New York, pp 59–73

Odintsova T, Egorov, T (2012) Plant antimicrobial peptides. In Plant signaling peptides, Springer, Berlin, Heidelberg pp. 107-13).

Oliveira-Lima M, Benko-Iseppon AM, Ferreira, Costa RJ, Rodríguez-Decuadro S, Kido EA, Crovella S, Pandolfi V (2017) Snakin : Structure , Roles and Applications of a Plant Antimicrobial Peptide. Curr Protein Pept Sci 18:1–7. doi: 10.2174/1389203717666160619 1

Oparin PB, Mineev KS, Dunaevsky YE, Arseniev AS, Belozersky MA, Grishin EV, Egorov TA, Vassilevski AA (2012) Buckwheat trypsin inhibitor with helical hairpin structure belongs to a new family of plant defence peptides. Biochem J 446:69–77. doi: 10.1042/BJ20120548

Pariani S, Contreras M, Rossi FR, Sander V, Corigliano MG, Simón F, Busi M V., Gomez-Casati DF, Pieckenstain FL, Duschak VG, Clemente M (2016) Characterization of a novel Kazal-type serine proteinase inhibitor of *Arabidopsis thaliana*. Biochimie 123:85–94. doi: 10.1016/j.biochi.2016.02.002

Pestana-Calsa MC, Ribeiro ILAC, Jr TC (2010) Bioinformatics-Coupled Molecular Approaches for Unravelling Potential Antimicrobial Peptides Coding Genes in Brazilian Native and Crop Plant Species. Curr Protein Pept Sci 199–209

Poth AG, Colgrave ML, Lyons RE, Daly NL, Craik DJ (2011) Discovery of an unusual biosynthetic origin for circular proteins in legumes. Proc Natl Acad Sci 108:10127–10132. doi: 10.1073/pnas.1103660108

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16**:**276-277.

Richero M, Barraco-Vega M, Cerdeiras MP, Cecchetto G (2013) Development of SCAR molecular markers for early and late differentiation. Trees 27:249–257. doi: 10.1007/s00468-012-0792-6.

Roxrud I, Lid SE, Fletcher JC, Schmidt EDL, Opsahl-Sorteberg H-G (2007) GASA4 , One of the 14-Member Arabidopsis GASA Family of Small Polypeptides , Regulates Flowering and Seed

Development. Plant Cell Physiol 48:471–483. doi: 10.1093/pcp/pcm016

Segura A, Moreno M, Madueño F, Molina A, García-Olmedo F (1999) Snakin-1, a peptide from potato that is active against plant pathogens. Mol Plant Microbe Interact 12:16–23. doi: 10.1094/MPMI.1999.12.1.16

Sels J, Mathys J, De Coninck BMA, Cammue BPA, De Bolle MFC (2008) Plant pathogenesis-related (PR) proteins: A focus on PR peptides. Plant Physiol Biochem 46:941–950. doi: 10.1016/j.plaphy.2008.06.011

Shafee TMA, Lay FT, Phan TK, Anderson MA, Hulett MD (2017) Convergent evolution of defensin sequence, structure and function. Cell Mol Life Sci 74:663–682. doi: 10.1007/s00018-016-2344-5

Silverstein K a T, Graham M a, Paape TD, VandenBosch K a (2005) Genome organization of more than 300 defensin-like genes in Arabidopsis. Plant Physiol 138:600–610. doi: 10.1104/pp.105.060079

Silverstein KAT, Moskal WA, Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J 51:262–280. doi: 10.1111/j.1365-313X.2007.03136.x

Slavokhotova AA, Rogozhin EA, Musolyamov AK, Andreev YA, Oparin PB, Berkut AA, Vassilevski AA, Egorov TA, Grishin E V., Odintsova TI (2014) Novel antifungal α-hairpinin peptide from *Stellaria media* seeds: Structure, biosynthesis, gene structure and evolution. Plant Mol Biol 84:189–202. doi: 10.1007/s11103-013-0127-z

Slavokhotova AA, Shelenkov AA, Korostyleva T V., Rogozhin EA, Melnikova N V., Kudryavtseva A V., Odintsova TI (2017) Defense peptide repertoire of *Stellaria media* predicted by high throughput next generation sequencing. Biochimie 135:15–27. doi: 10.1016/j.biochi.2016.12.017

Slavokhotova AA, Shelenkov AA, Odintsova TI (2015) Prediction of *Leymus arenarius* (L.) antimicrobial peptides based on de novo transcriptome assembly. Plant Mol Biol 89:203–214. doi: 10.1007/s11103-015-0346-6

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S (2016) TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res 26**:**1134-1144.

Stec B (2006) Plant thionins – the structural perspective. Cell Mol Life Sci 63:1370–1385. doi: 10.1007/s00018-005-5574-5

Tesfaye M, Silverstein KAT, Nallu S, Wang L, Botanga CJ, Gomez SK, Costa LM, Harrison MJ, Samac DA, Glazebrook J, Katagiri F, Gutierrez-Marcos JF, VandenBosch KA (2013) Spatio-temporal expression patterns of *Arabidopsis thaliana* and *Medicago truncatula* defensin-like genes. PLoS One 8: doi: 10.1371/journal.pone.0058992

Thevissen K, Kristensen H, Thomma BPHJ, Cammue BPA, Francois IEJA (2007) Therapeutic potential of antifungal plant and insect defensins. Drug Discov Today 12:966–971. doi: 10.1016/j.drudis.2007.07.016

Thomma BPHJ, Cammue BPA, Thevissen K (2002) Plant defensins. Planta 216:193–202. doi: 10.1007/s00425-002-0902-6

Troncoso MF, Zolezzi P, Hellman U, Wolfenstein-todel C (2003) A novel trypsin inhibitor from *Peltophorum dubium* seeds, with lectin-like properties, triggers rat lymphoma cell apoptosis. Arch Biochem Biophys 411:93–104. doi: 10.1016/S0003-9861(02)00726-9

Yount NY, Yeaman MR (2004) Multidimensional signatures in antimicrobial peptides. Proc Natl Acad Sci 101:7363–7368. doi: 10.1073/pnas.0401567101

Zou C, Lehti-shiu MD, Prakash T (2009) Evolutionary and expression signatures of pseudogenes. Plant Physiol 151:3–15. doi: 10.1104/pp.109.140632

Supplementary Information

Table S1. Summary of the assembly results

Total Paired reads	Filtered paired reads	percent	Total contigs	Filtered contigs	re maping percent	N50
26079982	23835863	91.4	127142	108151	95	1347

Table S2. Predicted *P. dubium* AMPs from seedling transcriptome

Family name	**Gene name**	**Sequence[a]**
cyclotide	PdCy67497_011	DAQVHITLCTVCETSGFTGTELTPGCTCGAFVLASTSKLCHHFINHGSLLFLRDRIAAFHVS
cyclotide	Pd2S albumin20440_021	**MAKLAILVFLAALLFVAVTASA**YKTTVTTVEIDGEEVNPRRKPGEGECRQKLQQVNLRQCERYLMEQC RDRGQDVIAFRGLRRNPSQALQQCCAGLRELDDEKCQCEALRRIMEKQREELEGKEILEMVRELRRLPEK CGMETPRCEIHRMGF
hevein-like	PdHev_10927_011	**MGRKQQLGMGSVVVVWWVGLSLSLSLIAIAEG**QSGSNVRSTYHLYNPQQHGWDLKAVSAFCAT WDADKPLAWRRKYGWTAFCGPAGPHGRDACGKCLRVTNTETGAEETVRIVDQCANGGLDLDVNVFQTL DTNGRGNALGHLTVNYRFVNCGD
hevein-like	PdHev_55563_011	**MAKLSLWVLCLVSAVALAYA**QSATNVRATYHMYQPEQHNWDLLAVSAFCATWDASKPLSWRNKYG WTAFCGPVGPQGQAACGKCLRVTNTRTGTQATVRIVDQCSNGGLDLDVGVFQKLDTDGSGNAQGHLTV NYDFVDCGD
hevein-like	PdHev_34220_011	**MKLGTNMFALAIALLLLKASA**EQCGKQAGGAVCPNGLCCSQYGWCGTTDDYCTKNCQSNCKSGGSP TPTPPSSGGDVASIISSSLFEQMLKHRNDAACQGKGFYSYDAFIKAARSFGGFGTTGDDATRKKELAAFFG QTSHETTGGWDTAPDGRYAWGYCYVREGSNTASSGYYGRGPIQLTHDYNYEQAGKAIGVDLLKNPDLV ATDAVISFKTAIWFWMTPQGNKPSSHDVITNRWTPSSADTSAGRVPGYGVITNIINGGLECGKGPNDKVA DRIGFYKRYCDIFGIGYGNNLDCYNQTPFA
hevein-like	PdHev_14461_011	**MSQILRNRKMDNLNKMRLGVLAWLIVGGLA**EQCGKQAGGALCPGGLCCSQFGWCGSTDQYCKD GCQSQCGGGGGGGGGDLGGLISRDTFNQMLKHRDDGGCPAKGFYTYDAFIQAAKAFPAFGNTGDTATR KREIAAFFGQTSHETTGGWATAPDGPYAWGYCFNKEQNPGSAYCASSPQWPCAPGKQYYGRGPIQISW NYNYGQCGRAIGADLLNNPDLVTSDPVISFKTALWFWMTAQSPKPSCHDVMTGRWSPSGADSAAGRVP GYGAVTNIINGGLECGRGQDARVADRIGFYKRYCDLLGVGYGNNLDCYGQRPFGSGLLVDTM
evein-like	PdHev_38451_011	**MSQILRNRKMDNLNKMRLGVLAWLIVGGL**AEQCGKQAGGALCPGGLCCSQFGWCGSTDQYCKD GCQSQCGGGGGGGGGDLGGLISRDTFNQMLKHRDDGGCPAKGFYTYDAFIQAAKAFPAFGNTGDTATR KREIAAFFGQTSHETSGGWATAPDGPYAWGYCFNKEQNPGSAYCASSPQWPCAPGKQYYGRGPIQISW YIFLYIYIYINIIVAMLN
hevein-like	PdHev_61526_011	**MVKLSIVVVGIIALVVAKNVSA**QNCGCGTGLCCSKYGYCGTGDAYCGSGCKQGPCYSSSASTTPAAT ADVSVADIVTPEFFNSIIDEADASCAGKNFYTRAAFLDALNSFDQFGRLGSVDDSKREIAAAFAHFTHETG HFCYIEEIDGASKDYCDESNTQYPCAPNKGYYGRGPIQLSWNFNYGPAGESIGFDGLNSPETVANDVVVS FKTALWFWMQNVRPVVNQGFGATIRAINGQLECDGANPNTVQARVNYYTQYCSQLGVAPGDNLTC
hevein-like	PdHev_58457_011	HHWAILVIGISGKDSLSIKTTSGGIQEIQKLGLKMAKLVGCLVVGLSVCVLGRARAEQCGRQAGGALCPGN LCCSQYGWCGSTGEYCGDKCQ
hevein-like	PdHev_70134_011	SHGQAACGKCLRVTNAGTGAQETVRIVDQCSNGGLDLDVGVFNRLDTDGSGYQQGHLTVNYQFIDCGN ELDSFNPLLSILDH

LTP	PdLTP71864_011	**MASKTSSSLALFLALNLLFFTFANA**CGSCPGYPKPNPRPRPNPNPSPSTGTCPRDALKLGVCANVLNG LLNVTLGKPPVTPCCSLIQGLADLEAAVCLCTALKANILGINLNIPISLSLLLNVCSKTVPRYFQCA
LTP	PdLTP32090_011	**LFLVINLLFFALVSA**WGTYPIPNPNPRPSPNNPGPSPNNPSPNPNNPSPSSGGTSCPRDALKLGVCLDVL NGLLNLTIGQPPVTPCCTLIQGLADLEAAVCLCTVIRANILGINLNIPLSLSLLLNVCSRQVPRNFQCA
LTP	PdLTP37468_011	**MASKALATCALLLTLNLLFSTMVSS**TYAPCPPTPKSHPNTPSTPSSSNGSCPVDTLKLGVCANVLNLVK VQIGAPPTLPCCSLIKGLADLEAAACLCTALKANVLGINLNIPLSLSVILNNCGKDNSGFQCA
LTP	PdLTP34171_011	**MASKCALLLCLSLVFFTVVSS**TYVPCPPPPHKAPKHPYVPKPPSSGSGMQPSCPKDTLKFGVCADLLGL VNVQLGKPPKTPCCSLIEGLADLEAAVCLCTALKANVLGINLNVPLKLSLLLNYCGKRVPKEFQCA
LTP	PdLTP9693_011	**MASKAALLLCFNLVLFTVASS**TYVPCPPGPKAPTPKHPHVPSKQPGCPKDTLKFGVCADVLGLVRAKLG KPPSSPCCSLIKGLADLEAAVCLCTALKANVLGINLNVPLNLSLLLNYCGKGVPQGFKCA
LTP	PdLTP35351_011	**MASNKKHLCATIFLLSLLARSAIT**EGCCSCRPKPRPCSPPAVPQCPLDTLKLGVCANILGLVNVVVGSP PTAQCCALLVGLADLEVALCLCTAIKANVLGINLNVPVSLSLLLSACQKTVPPGFQCP
LTP	PdLTP44236_011	**MASIFNIKLRAIILVLSLLAYSTFTNA**NGCPTTPATPPPSPSTERCPKDTLKLSLCADVLGLVNVVVGTP LREKCCALLEGLVDLEAALCLCTAIKANVLGINLNVPVALSLILSACQKTVPPGFQCP
LTP	PdLTP23838_014	VKPPPYVPKPPIVKPPPYVPKPPVVKPPPYVPKPPIVKPPPYVPKPPVVKPPPYAPKPPVVKPPPYVPKPPVVK PPPYVPKPPVVSPPYVPTPPVVTPPYVPKPPFVSPPYVPTPPVVTPPPYVPQPPPVVTPTPPVVTPPSPYVPK PPPAVTPCPPPPAQPTCPIDTLKLSACVDVLGGLVHVGIGSSAKQTCCPVLEGLVDLDAAVCLCTTIKAKLL NISLVIPIALQVLIDCGKTPPQGFQCSAN
LTP	PdLTP18130_011	VTPPYVPHPPVVNPPYVPKPPVTPPSVPNPPVVNPPYVPKPPVTPPVVPKPPVVAPPIVKPPPIVRPPYVPIPP PTVIPPPVVTPPTPPAPPTTPCPPPPPPAQPTCPVDSLKLDACVDVLGGLIHIGIGSSAKQTCCPVLGGLLDL DAAVCLCTTIKLKLLNINLVLPLALQVLIDCGKTPPQGFQCSPN
LTP	PdLTP16799_011	**METSSKIHAFLIICMLLLSSATPTLG**CGHCKKPGPPKKHKPGRKPIPFPPVVKPPPPGPATCPIDTVKLG ACVDLLGGLVHIGVGDPVVNQCCPVLQGLVDLEAAVCLCTTLKLRLLNLNLYVPLALQLLLTCGKTPPPGYT CSL
LTP	PdLTP45986_011	**MSSKATASFFIFMIFTLNVLPPTYA**CGSCTQPHPPYYRPGHPKTRHPKPHPPHHGGHPKVLPPLPPVI VPPIIVTPPPLSPPVIYPPPTAPSPPYWGVPPPFPPVTCPIDALKLGLCLDVLGGLVHVGIGNPIENICCPVIQ GLLELEAAICLCTVIRLKVLNLNIFFPLALQVLITCGMTPPPGFVCPPL
LTP	PdLTP12855_011	**MASKAAFLLSLNLLLFIMVSFTFVPSSEA**KCPVDTLKFRVCADVLGLVHVPSDSQCCSLIRGLVDADA AACLCTALKANVLGLNLNIPVDFSVILNECAGSSSGFQCA
LTP	PdLTP18292_011	RIRSPPYAPKPPVASPPYVATPPVVTPPYVPKLPVGTPPCVPQARPPPPPPKAPICRVSNLGPCVDVLARLAH VGTGGSAKQTCCPLLENMGDFQAVLCLCGTIIAKHMKVDVIFPIALQVLIGCGKTPPQHYHCYG
LTP	PdLTP65143_011	**MDCSRSSSLALFLCVNLLFFVLTSG**CYTCTQPKPNKPNPNSFPFPNPSPAKSCPRDALKLGVCAKLLNG PVG
LTP	PdLTP45098_011	**MIRIEETQAPMKYISFLVLLLVLGLINHNEVYG**AGECGKSSPDREAWKLAPCVTATQDENAPVSGS CCTQIKRLGQNPACLCAVMLSNTAKSAGVNPEIAVTIPKRCNIAERPVGYQCGAYSVP
LTP	PdLTP28075_011	**MASLLKLACLLVVCVAVVVG**APMAQAITCGQVNRNLAPCLTYLKRGGAPPPPCCSGVRAIVSAARTTP DRRAVCNCLKVSATQFPGVQPRFAEALPSGCRVNLPYKISLSTNCNSIRF
LTP	PdLTP68859_011	**MAGLMKLACVLVLCVAVVGAPMAEA**ITCGQVASGVAPCITYLQKGGVPSPGCCNGVRALANAARTT ADRQAACNCLKSAAAGISGLNPQYAAGLPGACKVNIPYKISTSTNCATIKF
LTP	PdLTP36008_011	**MRTSSVAASGVLLLLLLFASTSEA**VISCSDVVKDLRPCVSYLSSGSGKPPPACCSGAKALASAAATSAD KKAACNCIKSAAKSITINSQLAQALAGNCGITISFTISPNADCSKIG
LTP	PdLTP45075_011	**MGVAVVAGLAIMMVMVEPGEA**IDCPHVDSLVVPCLAFLTGGANDPSTACCGGIQTLKSETPTRDDRR AACECLKAAAAHFSGKILDDRAASLPKRCGVDVGVPISKNINCNDIN

LTP	PdLTP70367_011	**MTKKMWLCAMVVVMLVVVEVPHTAEG**VTCSPVALSPCLGAISSSSPPSSTCCQKLREQRPCLCGYLK NPSLRQYVNSPGARKVATACGIPFPTC
LTP	PdLTP17903_011	**MTTMKEKRVCASVAVLAIAIMLLVEVPRMAEG**VTCSPVEMSPCLGAITSPSPPSTTCCQKVREQKPC LCEYLRNPGLRQYVNSPGARRVANACGVPYPTC
LTP	PdLTP14638_011	**MKVSHIALCILVVLLLAEVQVSMA**VTCSPVQLSSCVSAITSSTPPSKLCCSKIKEQKPCLCQYLKNPNLK KFVDSPNARKVASTCGTPFPRC
LTP	PdLTP52897_011	**MKASHIALCALLVLLLAEA**QVSMALTCSPVQLSSCVSAITSSSPPSKLCCSKIKEQKPCLCQYLKNPNLK KFVDSPNARKVANTCGTPYPRC
LTP	PdLTP16121_011	**MAHLRMDLVGVVVVMAVMYAGAAA**QSSCTNVLVSLSPCLNYITGNSSTPSSGCCSQLSSVVRSQPE CLCQVLNGGGSSLGININQTQALALPAACKVQTPPTSQCKAASPGNSPAAESPNSVPSGTGSKSVPSTDNG SSDGNSIKLSIPLLFIGFAAIYASTFRTY
LTP	PdLTP12366_011	**METKSSALLCMAFLLLALAATWA**QDNSCLNKLAPCFNYLNGAQEPPGSCCDPLKSVIQTDPKCLCSLV SNKGAKQAEQAGINVTQAQTLPGRCGQRVNPLSCLSSSPDSNNTAQNSAAKFMNVSRGVIVLLVLLNVSA QMLWV
LTP	PdLTP43171_011	**MAFRGMVLGLAVVLVATLWSQSAA**QSGCTSALTSMSPCLNYITGSSSTPSSTCCSQLSSVVQSSPQC LCTVLNGGGSSLGITINQTLALSLPGACKVQTPPISQCKGAGNTPTTPSTAPVGSPADSSSESPQGSISPSA SDVPNAGAGSKTVPSMDGGSSDGSTIKISFHSMISLLFVASFVSAATHF
LTP	PdLTP29389_011	**MGLTDLAFTCGFLFLLLAN**FAASDLAKDREECADQLVGLAPCLPYVGGEAKTPTIDCCAGLKQVLDKSK KCICILIKDRNDPNLGLNINASLAVHLPAACHAPANISECVDLLHLKPNSAEAKVFEGFDKATQTNSSTPVA PSGKGSSPSVSANVQSDGWSVRGKRWLVGEVVCVIVPFVFNFLV
LTP	PdLTP16968_012	**MGSKLKALVQCVIAAAILVIGLLCGGAKG**VVLCNTDTTKLGLCRPAVTGQHPPPPTEACCAVARHAD LPCLCKYKSVLPALGIDPKNAVALAGKCGLTTPPQCKGN
LTP	PdLTP18292_011	RIRSPPYAPKPPVASPPYVATPPVVTPPYVPKLPVGTPPCVPQARPPPPPPKAPICRVSNLGPCVDVLARLAH VGTGGSAKQTCCPLLENMGDFQAVLCLCGTIIAKHMKVDVIFPIALQVLIGCGKTPPQHYHCYG
defensin	PdDf18833_011	**MARSLYQFVSTMFVVLLLLVASGMGPTMVAEA**RTCESQSQRFRGICVRSSNCASVCQTEGFPDGR CRGFRRRCFCSKHC
defensin	PdDf14712_031	**MAALCFFLLFLLAAHEIGEAEA**KTCEVLSDKFKGACSTIINGPKCDKTCKNQERLISGQCKSDFRCWCT KNC
defensin	PdDf24034_011	**MAKFLRLSTTFTLLLCVCFLIFSNEAAG**MVEAKTKICGRMSKTWSGWCGNTKHCDNQCRKWEGAK HGACHAHFPGRACFCYFNC
defensin	PdDf24034_012	**MAKFLRLSATFTLLLCICLLIFSNEAGMVEA**KTKVCGRMSKTWSGWCGNSKHCDNQCRKWEGAKH GACHAHFPGRACFCYFNC
defensin	PdDf60561_011	**MGRSVPMVSTIFVLLLLLVATEMAPMVAEA**RTCESQSHRFKGSCVSNSNCAAVCQTEGFPGGNCRG FRRRCFCTKPC
defensin	PdDf64469_011	**MERARFGVLFLLLVILASEEMVVQS**EARLCESKSHHFKGMCASDRNCASVCHTEGFSGGNCRGFRRR CFCSRFC
defensin	PdDf8737_011	**MERKSYGLFLLLLIVLAAEVAVQTEG**RVCESQSHRFKGACLGDHNCALVCRNEGFSGGRCRGFRHRC FCTRLC
defensin	PdDf53257_011	**MAKFTFFYLFLLVLVFAVSIGDG**EDANVMCSVELDPNKCDISTCQKECQQNYYNGVGQCAHFGTYLH CICTYICP
defensin	PdDf13807_021	**MAKFPFPHLFLLFLFVSVVLMLGGTTA**RRFCSVELDPNQCDVPTCQNECLQFYSGVGVCAHFGLQIF HCVCSYICP
defensin	PdDf61711_011	ARALRIPEHTCRTLVHLSNCDNQKCFQECSKHPYGIGQCKETTCLCTYYCKDPPL

defensin	PdDf18030_011	**MAGISRTPFPCFLGVAIIVFALCLFPMARA**QDDTSSILQLPSQSSNEGQKLCAGTVPASCPVKCFRTD PVCGVDGVTYWCGCAEAVCAGIKVARLGFCEVGSGSSVSLPGQALLLVHIIWLIVLAFSVLFGLF
defensin	PdDf5301_011	**MAGIYQTKHLHRFFSSSSAAVAAAIVVVLALSLCPVVLA**EHESSHTLRLPSEASRENQVVCAAKTAP SSCPVKCFRADPVCGVDGVTYWCGCAEAECAGTKVAKLGFCEVGSGGSAPLSAQALLLVHIVW
defensin	PdDf19806_011	**SQTYLFAGFLICLVLASGTTS**GIINYLMDCDGWQCDGPQECVHYCIDTGYIRGGRCINYLGQYQCCCN TNDTALP
snakin/GASA	PdSN12635_011	**MALRMLLVLGMLLLMASLPQISS**DLSIEEQDEELHNPHRPLIVKDGNRRLMQNIDCGGLCEARCSVH SRPNLCKRACGTCCVRCKCVPPGTSGNRELCGSCYTDMTTHGNKTKCP
snakin/GASA	PdSN12635_021	**MALFIIIIFSLLIKKYLES**CVVVTEGSMLCWWICFQLIVKDGNRRLMQNIDCGGLCEARCSVHSRPNLC KRACGTCCVRCKCVPPGTSGNRELCGSCYTDMTTHGNKTKCP
snakin/GASA	PdSN18884_011	**MAKFLAALILALIAISMLQTVVTA**AHGQGGPLYDNKSQYGPGSLKSYQCPSQCQRRCGKTQYRKPCL FFCQKCCRKCLCVPPGYYGNKAVCPCYNNWKTKEGGPKCP
snakin/GASA	PdSN18884_031	**MARVLAALILALIVISVLHTTVMA**AHGHGGHHYDNKRKFGPGSLKSHECPSQCSRRCKKTQYHKACM FFCQKCCRKCLCVPPGYYGNKAVCPCYNNWKTKEGGPKCP
snakin/GASA	PdSN16337_011	**MASKSTLLLLGIFVLFTTNVSS**NDEELAMKVKYAKPPVSAVAPVVAPKPAPPSKPLTPPPPAVKAPTPPQ PVQPPTPVNPPVSPAPAPAPGAPIVRSTKDCAPLCEKRCQLHSRKRVCMRACGTCCERCKCVPPGTSGNRE LCGKCYTDMLTHGNRYKCP
snakin/GASA	PdSN11162_011	**MAMASSKLLVAYLLISVLLLHDHIVDA**DQLVEVEVEAKKQGSVVQQIDCNGACGARCRLSSRPHLCE RACGTCCRRCNCVPPGTAGNQQVCPCYANMTTHGGRRKCP
snakin/GASA	PdSN15806_011	**MAISKTLVASILLSMLILHLAES**DMVIKTMVDSPAPSPLTIDCDSACDGRCKLSSRPRLCKRACGTCCA RCNCVPSGTSGHYEECPCYANMTTHHG
snakin/GASA	PdSN5202_011	**MAFSRILAASLVVTLLLLLHLTEA**SHQTVMFQITSNNAAPAPSPLSHPIDCGGACEGRCKLSSRPNLCK RACGSCCAQCKCVPPGTSGNHNSCPCYANLTTRHHIPKCP
snakin/GASA	PdSN5202_021	**MAFSRILAASLVVTLLLLLHLTEA**SHQTITSNNAAPAPSPLSHPIDCGGACEGRCKLSSRPNLCKRACG SCCAQCKCVPPGTSGNHNSCPCYANLTTRHHIPKCP
snakin/GASA	PdSN14393_021	**MAAPPYRSIFIILSLLLLITCSEVAEA**YGRSKLRPSDCKPKCTYRCSATSHKKPCMFFCQKCCAKCLCV PPGTYGNKESCPCYNNWKTKEGGPKCP
snakin/GASA	PdSN14514_011	**MAKLVCVLLLALLGISMIATQAMA**KDAAQYHLDGGNYGPGSLKSSQCPSQCSRRCSQTQYHKPCMF FCQKCCAKCLCVPPGYYGNKAVCPCYNNWKTKRGGPKCP
snakin/GASA	PdSN44368_011	**MASRFLLVMALLLFCMALSQVSS**EASIKDHITDSLVVKGPNRRLLQFVDCGGRCSMRCSKHSRPKICR RACGTCCARCKCVPPGTFGNREMCGRCYTDMVTHGNRPKCP
snakin/GASA	PdSN11759_011	**MASKSTIILVFLFLQVATKVYS**SDEELKIRLNYAKPPVTSPAAPPVVAHVPPPPLKPPTPPPPVEKAPSPP APAKAPSPPSPVKPPVKPVPPAAPIVKSKKDCIPLCDYRCQLHSRKRVCMRACTTCCDRCKCVPPGTYGNR EVCGKCYTDMVTHGNRSKCP
snakin/GASA	PdSN16208_011	**MKLVFATLLLVCLVLSASFVEVTLA**GSEFCDSKCAQRCSKAGMKDRCLRFCGICCEKCKCVPSGTYGN KHECPCYRDLKNSKGKSKCP
snakin/GASA	PdSN16208_021	**MKLGFATFLLVCLVLSSSLFEVTLA**SSPFCDSKCAQRCSKAGVQDRCLRFCGICCEKCKCVPSGTYGNK HECPCYRDLKNSKGKPKCP
snakin/GASA	PdSN71265_011	**MALSKLLIASLLMSLLLLHHLVDA**DQTGSVSETQGSLLQQIDCKGACSARCRLASRQRMCHRACGTC CRRCNCVPPGTAGNQEVCPCYANMTTHGGRRKCP
snakin/GASA	PdSN16780_011	**MAGKLSIIMLFLVTMLLFLVEHHA**EIVVATTEAPAPQPNRNTTNFPTHGITEGSLQPQECGPRCTERC SKTQYKKPCMFFCQKCCAKCLCVPPGTYGNKEVCPCYNNWKTKRGGPKCP

snakin/GASA	PdSN14602_021	**MAMAKLVCLLFLSLLAILMVPNQVMA**KDKHPLGTLSTGNYGPGSVKSSECPSQCSRRCSQTQYHKA CMFFCQKCCAKCLCVPPGFYG

[a] In bold predicted signal peptides

Table S3. Representative members of AMP families from well-studied or model plants

Family name	Protein name	Sequence
cyclotide	|C0HJS5|CYC18_CLITE	GLPICGETCFTGTCYTPGCTCSYPVCKKN
cyclotide	|A0A0S1RQH5|CLITE Albumin1	MAYLRLAPLAVIFLFAVMFAVEKTEGGLPICGETCFTGTCYTPGCTCSYPVCKKNHIIAIAAQAVD QHRLLCESHEDCFKKGTGNYCASFPDSDIHFGWCFHAESEGYLLKDFLKMSKDDLKMPTEITN
cyclic peptide	|Q4GWU5|SFTI1_HELAN	MAKLITLVVLAILAFVEVSVSGYKTSISTITTIEDNGRCTKSIPPICFPDGRP
Cyclic peptide precursor	|A0A023I288|Albumin_HEL AN	MAKLITLVVLAILAFVEVSVSGYKTSISTITTIEDNGRCTKSIPPICFPDGLDNPRGCQIRIQQLNHC QMHLTSFNYKLRSMAVENPKQQQHLSLCCNQLQEVEKQCQCEAIKQVVEQAQKQLEQGQGGQ QQVQQMVKKAQMLPNQCNLQCSI
albumin 2S	|P19594|2 2S albumin_SOYBN	MTKFTILLISLLFCIAHTCSASKWQHQQDSCRKQLQGVNLTPCEKHIMEKIQGRGDDDDDDDD DNHILRTMRGRINYIRRNEGKDEDEEEEGHMQKCCTEMSELRSPKCQCKALQKIMENQSEELEE KQKKKMEKELINLATMCRFGPMIQCDLSSDD
hevein-like	|P02877|HEVE_HEVB	EQCGRQAGGKLCPNNLCCSQWGWCGSTDEYCSPDHNCQSNCKD
hevein-like	|P27275|AMP_AMACA	MVNMKCVALIVIVMMAFMMVDPSMGVGECVRGRCPSGMCCSQFGYCGKGPKYCGRASTTVDH QADVAATKTAKNPTDAKLAGAGSP
hevein-like	|E1UYT9|AMP1_STEME	MLNMKSFALVMLFATLVGVTIASGPNGQCGPGWGGCRGGLCCSQYGYCGSGPKYCAHNTPLSE IEPTDAGRCSGRGTCSGGRCCSKYGYCGTGPAYCGLGMCQGSCLPDMPNHPAQIQARTEAAQA EAQAEAYNQANEAAQVEAYYQATQAQTQAQPQVEPAVTKAP
hevein-like	|P85966|AMP_TRIKH	AQRCGDQARGAKCPNCLCCGKYGFCGSGDAYCGAGSCQSQCRGCR
LTP	|Q9SU35|ERLL1_ARATH	MASKNSASLALFFALNILFFTLTVATNCNCKPSPKPKPVPSPKPKPVQCPPPPRPSVPSPNPRPVTP PRTPGSSGNSCPIDALKLGVCANVLSSLLNIQLGQPSSQQCCSLIQGLVDVDAAICLCTALRANVL GINLNVPISLSVLLNVCNRKLPSGFQCA
LTP	|A0AT29|NLTP2_LENCU	MARGMKLACVVLVICMVVIAPMAEGAISCGAVTSDLSPCLTYLTGGPGPSPQCCGGVKKLLAAAN TTPDRQAACNCLKSAAGSITKLNTNNAAALPGKCGVNIPYKISTTTNCNTVKF
LTP	|Q42589|NLTP1_ARATH	MAGVMKLACLLLACMIVAGPITSNAALSCGSVNSNLAACIGYVLQGGVIPPACCSGVKNLNSIAK TTPDRQQACNCIQGAARALGSGLNAGRAAGIPKACGVNIPYKISTSTNCKTVR
LTP	|A0A1S3TFR4|_VIGRR	MEGVVKFACLVGFVVLVSVAKVDSAGECGKSTTPDNEAIKLAPCASAAQDENASVSQSCCAQVK KIGQNPSCLCAVLLSNTAKMAGVNPQIAVTIPKRCNLANRPVGYKCGPYTLP
LTP	|P82353|NLTP2_PRUAR	VTCSPVQLSPCLGPINSGAPSPTTCCQKLREQRPCLCGYLKNPSLRQYVNSPNARKLASNCGVPV PQC
defensin	3PSM:Pachyrhizus erosus	SPE10KTCENLADTFRGPCFTDGSCDDHCKNKEHLIKGRCRDDFRCWCTRNC
defensin	1N4N:Petunia hybrida	ATCKAECPTWDSVCINKKPCVACCKKAKFSDGHCSKILRRCLCTKEC
defensin	6B55:Nicotiana alata	RECKTESNTFPGICITKPPCRKACISEKFTDGHCSKILRRCLCTKPC
defensin	5NCE:Pinus sylvestris	RMCKTPSGKFKGYCVNNTNCKNVCRTEGFPTGSCDFHVAGRKCYCYKPCP
defensin	2LR3:Medicago truncatula Def4	RTCESQSHKFKGPCASDHNCASVCQTERFSGGRCRGFRRRCFCTTHC
defensin	2M8B:Arabidopsis halleri	ERLCEKPSGTWSGVCGNNGACRNQCIRLEKARHGSCNYVFPAHKCICYFPC
defensin	2N2Q:Heuchera sanguinea	DGVKLCDVPSGTWSGHCGSSSKCSQQCKDREHFAYGGACHYQFPSVKCFCKRQC
defensin	1AYJ:Raphanus sativus	EKLCERPSGTWSGVCGNNNACKNQCINLEKARHGSCNYVFPAHKCICYFPC
defensin	|Q8GTM0|DEF_NICAL	RECKTESNTFPGICITKPPCRKACISEKFTDGHCSKILRRCLCTKPCVFDEKMTKTGAEILAEEAKT LAAALLEEEIMDN

defensin	|Q8H6Q1|DEF1_PETHY	ATCKAECPTWDSVCINKKPCVACCKKAKFSDGHCSKILRRCLCTKECVFEKTEATQTETFTKDVN TLAEALLEADMMV
defensin	|P82762|DEF91 ARATH	FTIAEPYIHPCMKGFCSFKSECANKCIFMGHHKGGDCIGGLDGIYCCCLA
snakin/GASA	|Q93X17|SNAK2_SOLTU	YSYKKIDCGGACAARCRLSSRPRLCNRACGTCCARCNCVPPGTSGNTETCPCYASLTTHGNKRK CP
snakin/GASA	|Q948Z4|SNAK1_SOLTU	GSSFCDSKCKLRCSKAGLADRCLKYCGICCEECKCVPSGTYGNKHECPCYRDKKNSKGKSKCP
snakin/GASA	|P46690|GASA4_ARATH	SSGSNVKWSQKRYGPGSLKRTQCPSECDRRCKKTQYHKACITFCNKCCRKCLCVPPGYYGNKQ VCSCYNNWKTQEGGPKCP
snakin/GASA	|P46689|GASA1_ARATH	DVENSQKKNGYAKKIDCGSACVARCRLSRRPRLCHRACGTCCYRCNCVPPGTYGNYDKCQCYA SLTTHGGRRKCP
snakin/GASA	|Q8LFM2|GASAA_ARATH	DSSPCGGKCNVRCSKAGRQDRCLKYCNICCEKCNYCVPSGTYGNKDECPCYRDMKNSKGTSKC P
snakin/GASA	|P46687|GASA3_ARATH	AEDSQVGEGVVKIDCGGRCKGRCSKSSRPNLCLRACNSCCYRCNCVPPGTAGNHHLCPCYASIT TRGGRLKCP
snakin/GASA	|Q84J95|GASA5_ARATH	ARGGGKLKPQQCNSKCSFRCSATSHKKPCMFFCLKCCKKCLCVPPGTFGNKQTCPCYNNWKTK EGRPKCP
snakin/GASA	|P46688|GASA2_ARATH	ADGAKVGEGVVKIDCGGRCKDRCSKSSRTKLCLRACNSCCSRCNCVPPGTSGNTHLCPCYASIT THGGRLKCP
snakin/GASA	|Q8GWK5|GASA9_ARATH	ELSSSSNNETSSVSQTNDENQTAAFKRTYHHRPRINCGHACARRCSKTSRKKVCHRACGSCCAK CQCVPPGTSGNTASCPCYASIRTHGNKLKCP
snakin/GASA	|O82328|GASA7_ARATH	TISDAFGSGAVAPAPQSKDGPALEKWCGQKCEGRCKEAGMKDRCLKYCGICCKDCQCVPSGTY GNKHECACYRDKLSSKGTPKCP
snakin/GASA	|F4IQJ4|GASAB_ARATH	DMVTSNDAPKIDCNSRCQERCSLSSRPNLCHRACGTCCARCNCVAPGTSGNYDKCPCYGSLTT HGGRRKCP
snakin/GASA	|Q6GKX7|GASAC_ARATH	DELESQAQAPAIHKNGGEGSLKPEECPKACEYRCSATSHRKPCLFFCNKCCNKCLCVPSGTYGHK EECPCYNNWTTKEGGPKCP
snakin/GASA	|Q6NMQ7|GASA6_ARATH	KEAEYHPESYGPGSLKSYQCGGQCTRRCSNTKYHKPCMFFCQKCCAKCLCVPPGTYGNKQVCP CYNNWKTQQGGPKCP
snakin/GASA	|A8MR46|GASAD_ARATH	AHMHFLINVCAECETKSAIPPLLECGPRCGDRCSNTQYKKPCLFFCNKCCNKCLCVPPGTYGNKQ VCPCYNNWKTKSGGPKCP
snakin/GASA	|O80641|GASA8_ARATH	DSSCGGKCNVRCSKAGQHEECLKYCNICCQKCNCVPSGTFGHKDECPCYRDMKNSKGGSKCP
snakin/GASA	|P27057|GAST1_SOLLC	NIMRDEQQQQQRNNQLYGVSEGRLHPQDCQPKCTYRCSKTSYKKPCMFFCQKCCAKCLCVPA GTYGNKQSCPCYNNWKTKRGGPKCP
snakin/GASA	|P47926|RSI1_SOLLC	YNKLRPTDCKPRCTYRCSATSHKKPCMFFCQKCCATCLCVPKGVYGNKQSCPCYNNWKTQEGK PKCP
snakin/GASA	|Q9LFR3|GASAE_ARATH	ASNEESNALVSLPTPTLPSPSPATKPPSPALKPPTPSYKPPTLPTTPIKPPTTKPPVKPPTIPVTPVK PPVSTPPIKLPPVQPPTYKPPTPTVKPPSVQPPTYKPPTPTVKPPTTSPVKPPTTPPVQSPPVQPPT YKPPTSPVKPPTTTPPVKPPTTTPPVQPPTYNPPTTPVKPPTAPPVKPPTPPPVRTRIDCVPLCGT RCGQHSRKNVCMRACVTCCYRCKCVPPGTYGNKEKCGSCYANMKTRGGKSKCP
snakin/GASA	|P86888|PMLN_PRUPE	GSSFCDSKCGVRCSKAGYQERCLKYCGICCEKCHCVPSGTYGNKDECPCYRDLKNSKGNPKCP
snakin/GASA	|E5KBY0|E5KBY0_SOLLC	IQTDQVSSNAISEGADSYKKIDCGGACAARCRLSSRPRLCHRACGTCCARCNCVPPGTSGNTET CPCYASLTTHGNKRKCP
snakin/GASA	|H9D2D5|H9D2D5_MEDSA	GTDSGRFCSSICGQRCSKAGMKDRCMKFCGICCGKCKCVPSGTYGNKHECPCYRDMKNSKGKP KCP
snakin/GASA	|B2ZAW4|B2ZAW4_CAPAN	IQTDHVASNAISEAAYSYKKIDCGGKCSARCRLSSRPRLCKRACGTCCARCNCVPPGTSGNTQT CPCYANMTTHGNRRKCP
snakin/GASA	|M1BA38|M1BA38_SOLTU	YNKLRPRDCKPKCTYRCSATSHKKPCMFFCQKCCATCLCVPKGVYGNKQSCPCYNNWKTQEGK PKCP

Table S4. Primers designed for Snakin/GASA and defensin PCR validation

Primer name	Primer sequence
PdSN16208_F:	CCATTGCCTCCAACTCTTGC
PdSN16208_R:	GCTAAGAAGGGTCGTCCAGG
PdSN16780_F:	AGGCCACACTTCCGCAAATA
PdSN16780_R:	AGCATTGAGAGTTGAGAGTGAGA
PdSN14393_F:	TTCCCCTGACTTCACTGCAC
PdSN14393_R:	TGCCCTTCAAGCACCAAATT
PdSN14514_F:	GCCCCCTCACTCTTCTCTCT
PdSN14514_R:	TGAGTGGATCGTGGGAAGTG
PdSN71265_F:	ACCCACTCCATCCCCTCTTT
PdSN71265_R:	TGGAAGAAGACGTGGTTGCT
PdSN11162_F:	AGTCACCAACGTTACTCGGC
PdSN11162_R:	ATTAAGGGCACTTGCGTCGG
PdSN11759_F:	TCTTTCTCGTTCTGCTCCACC
PdSN11759_R:	AGGACACTGGCAAAACCACT
PdSN44368_F:	CGAGGGGAGAAGAAGATGGC
PdSN44368_R:	ATGGGGATGGATGCAAAGGG
PdSN5202_021_F:	CACACCCAAATCCCACTCAGA
PdSN5202_021_R:	AGGGCATGACGGATGAAAGA
PdSN12635_011_F:	TCTTCATACATCTGCTCTGCCA
PdSN12635_121_F	GCCCAGTCAGTACCATTCCC
PdSN12635_R:	AGAGGCATGAGAAGGAAGCC
PdSN15806_F:	TTCCCCACTTTCAGCCTCAG
PdSN15806_R:	CCATGGTGGGTGGTCATGTT
PdDf24034_F	ATCGATCGGAGCACAAACA
PdDf24034_R	ATCAGCAACACCCTGAATTTATG
PdDf8737_F	TGCTCATAATTGAACTCCATTTC
PdDf8737_R	ACTGCTTTTATTTGGGCTCTG
PdDf60561_F	CGCTTTGCATCTTGAAACG
PdDf60561_R	ATGTGGACCCCACTCGATC
PdDf64469_F	AGATAGGAGGTGGTATAGACATCG
PdDf64469_R	CTCTGTAGGGGTGGCGAG
PdDf14712_F	GAAGCGAAGACTTGTGAGGTG
PdDf14712_R	AGTTCTTGGTGCACCAGCAG
PdDf18833_F1	TCACTGATCTGTCCGTACCAC
PdDf18833_R1	TCTTCCCTGACTCAGAACCA
PdDf18833_F2	ATGGCTCGCTCTTTGTACCA
PdDf18833_F mature	AGGACATGCGAGTCTCAGAG
PdDf18833_R2	TCAACAGTGCTTGCTGCAG

Ensamblado *de novo* de un *pool* de transcriptomas de congorosa, para la predicción de AMPs

La búsqueda de transcriptos que codifiquen AMPs putativos en un *pool* de transcriptomas (hojas/brotes/flores) de congorosa, se llevó a cabo de la misma forma que el realizado para el transcriptoma de brotes de ibirapitá. La secuenciación (plataforma Illumina HiSeq2000) de la biblioteca de ADNc obtenida del *pool* de ARNm de *M. ilicifolia* generó ~53 millones de lecturas de 100 pb (Tabla 1). Las lecturas de buena calidad se ensamblaron con Trinity (Grabherr et al. 2011) y este ensamblado se evaluó y filtró con la herramienta Transrate (Smith-Unna et al 2016), estimando la calidad de *contigs* individuales. Finalmente, el número final de transcriptos no redundantes fue aproximadamente 102K con una estadística N50 de 1093.

Tabla 1. Estadística de datos crudos de secuenciación

Bases totales en *Reads* (pb)	*Reads* totales	GC(%)	AT(%)	Q20(%)	Q30(%)
5,341,331,066	52,884,466	45.68	54.32	96.29	91.49

Para la predicción de AMPs putativos, se usaron los motivos cisteína descritos por Slavokhotova et al. (2015). A partir de esta búsqueda de motivos, se obtuvieron 2456 transcriptos con uno o más motivos cisteína. De éstos, únicamente 409 presentaron anotación contra las bases de datos SwissProt (https://www.uniprot.org/uniprot/?query=reviewed:yes) y TrEMBL (https://www.uniprot.org/uniprot/?query=reviewed:no) con valor esperado máximo (e-value) de $10e\text{-}0^5$. En sólo 52 casos, la anotación corresponde a AMPs, 80 a proteínas no caracterizadas y los restantes corresponden a transcriptos con similitud con otro tipo de proteínas. No se descartaron los transcriptos con ORF incompletos (desde Metionina inicial hasta el codón stop), si contenían el motivo cisteína completo. Para el análisis de cada familia de péptidos se tuvieron en cuenta los 2456 transcriptos (y en algunos casos el transcriptoma completo (102k) para una búsqueda más profunda en base a BLAST con *e-values* menores, dependiendo del péptido).

Los resultados fueron similares a los encontrados en el transcriptoma de brotes de *P. dubium*, detectando 32 transcriptos con similitud con proteínas de transferencia de lípidos, 10 con proteínas *hevein-like* y uno con tioninas. En el caso del péptido deducido tipo tionina, éste no presentó un motivo tionina (de los presentes en la lista obtenida de Slavokhotova et al. 2015) pero tiene un 44 % de identidad con una tionina de *A. thaliana*. Al igual que para ibirapitá, tampoco se detectaron genes que codifiquen ciclótidos ni *α-hairpinin*.

DEFENSINAS

Con respecto a transcriptos que codifiquen defensinas putativas, encontramos 100 transcriptos con motivo cisteína tipo defensina, de los cuales sólo 10 tienen anotación (valor esperado máximo: $10e\text{-}0^5$). Los criterios para la selección final de un transcripto como defensina-*like* fueron los siguientes: a) cuando el ORF que incluye el motivo cisteína no es el más largo y/o tienen anotación en otro marco de lectura fueron descartados; b) cuando el ORF con el motivo cisteína es el más largo pero tiene anotación que no corresponde con defensinas (en general estos transcriptos codifican proteínas de más de 200 aa) fueron descartados; c) de los transcriptos sin anotación, sólo aquellos donde el ORF con el motivo cisteína fue el más largo y no tenía otra anotación en otro marco de lectura, fueron tenidos en cuenta. Aplicando estos criterios, se obtuvo una lista de 12 defensinas putativas. Excepto por dos de ellas, en el que el transcripto está incompleto en 5', las restantes 10 presentan péptido señal de acuerdo con SignalP (http://www.cbs.dtu.dk/services/SignalP/).

Los péptidos deducidos se podrían separar en 4 grupos. El primer grupo (con siete péptidos) presenta mayor similitud con "defensinas verdaderas" como las que se encuentran en la base de datos PDB, con un motivo clásico (CSαβ) y un motivo γ-core. Dos de ellas presentan dos cisteínas adicionales, al igual que NaD1 (Lay et al. 2003) pero en donde las dos cisteínas adicionales se encuentran en posiciones diferentes a NaD1. Un segundo grupo contiene tres péptidos con sólo seis cisteínas; uno de ellos con los motivos CSαβ ($CX_nCX_3CX_nCX_nCXC$) y γ-*core* ($GXCX_{3\text{-}9}C$) y las otras dos con un motivo CSαβ modificado ($CX_5CX_{7,8}CX_{4,5}CX_4CXC$) ya que entre las cisteínas 3 y 4 en lugar de tres aminoácidos hay siete u ocho. Los últimos dos grupos están representados por un péptido cada uno. Uno de ellos presenta similitud con algunas proteínas de *A. thaliana* anotadas como *defensin-like* (inferidas por homología de secuencias, sin datos funcionales) que tienen tres cisteínas localizadas consecutivamente (Silverstein et al. 2005). La última tiene el motivo CSαβ pero el γ-*core* no está presente. Parece tener un prodominio, al igual que las defensinas florales de *N. alata* y *P. hybrida* (Lay et al. 2003). La anotación de este péptido es con una proteína no caracterizada y presenta, al igual que dos péptidos encontrados en *P. dubium* algunas características de defensinas y otras de inhibidores de serin-proteasa tipo Kazal.

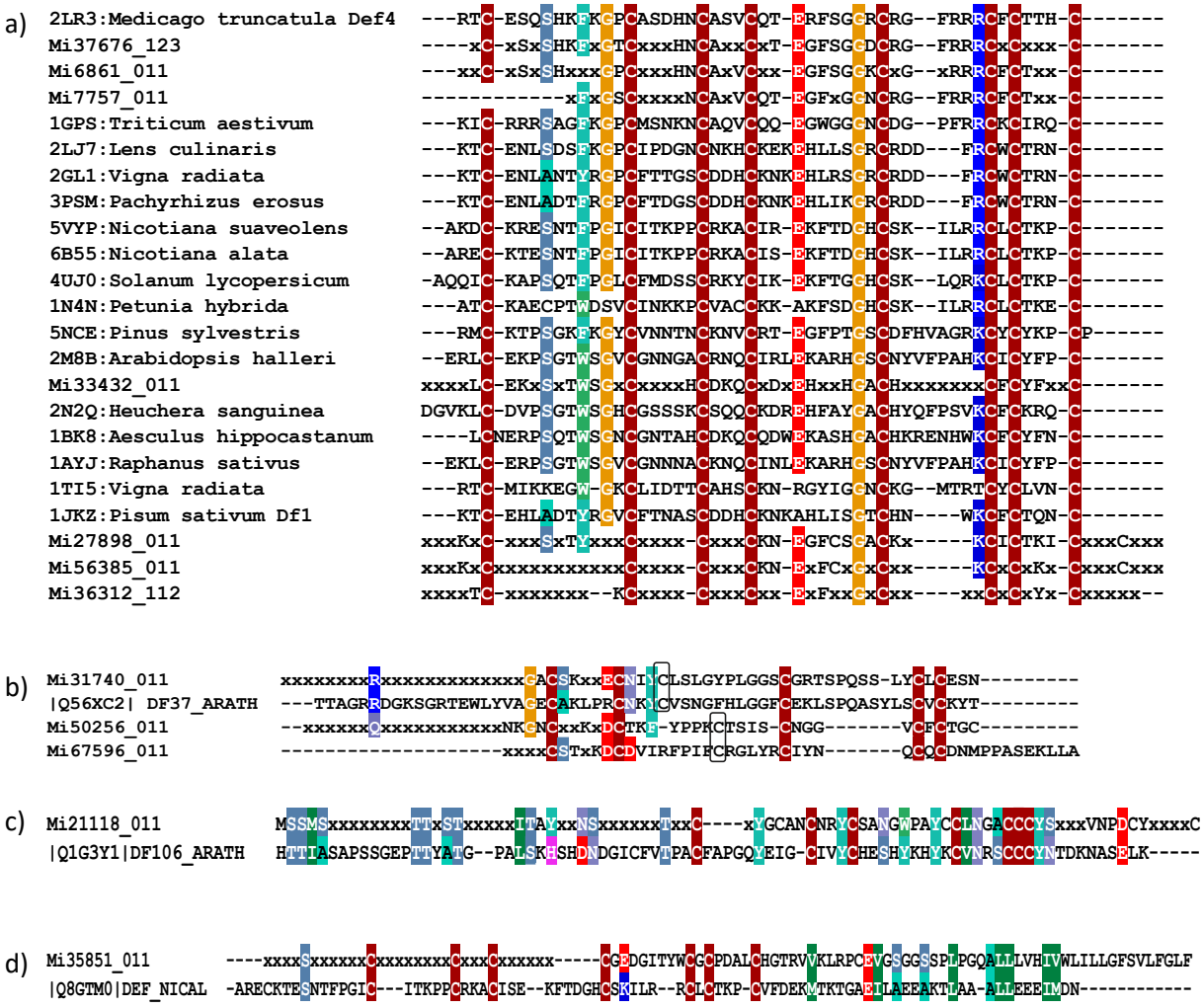


Figura 1. Comparación de secuencias de péptidos tipo defensina putativos de *M. ilicifolia* con defensinas de otras plantas. a) Alineamiento de "verdaderos" péptidos de tipo defensina de *M. ilicifolia* y péptidos maduros de defensinas conocidas. Todas las secuencias usadas en este alineamiento corresponden a defensinas cuya estructura 3D ha sido determinada. El nombre de cada secuencia corresponde al identificador de PDB. B) alineamiento de péptidos con motivo defensina de seis cisteínas y *defensin-like* proteína de *A. thaliana* (ARATH) con seis cisteínas. Las cisteínas que difieren del motivo αβ se marcan con rectángulos c) alineamiento de péptido tip defensina con una *defensin-like* de *A. thaliana* que posee las cisteínas 6,7 y 8 ubicadas de forma consecutiva. d) alineamiento de péptidos con motivo defensina y defensina con Prodominio C-terminal de *N. alata* (NICAL). Los péptidos de *M. ilicifolia* se denominaron Mi y a continuación el número del transcripto). Las X en los péptidos de congorosa representan algunos aminoácidos no conservados. No se incluyeron en las secuencias debido a que presentan variaciones de secuencia interesantes, por lo que se podrían considerar para licenciamiento, etc.

ESNAQUINAS

En relación a los genes Snakin/GASA, se encontraron 15 transcriptos en el *pool* de transcriptomas de *M. ilicifolia*. Todos los transcriptos con motivo esnaquina, presentaron anotación correspondiente a genes Snakin/GASA y todos estaban completos, presentando péptido señal en el extremo N-terminal, de acuerdo con la predicción de SignalP. Las secuencias deducidas de los 15 transcriptos, correspondientes a los péptidos maduros, fueron alineadas y se construyó una matriz de distancia genética, detectando tres grupos que coincidirían con las tres subfamilias típicas de proteínas Snakin/GASA de acuerdo a la clasificación de Berrocal-Lobo (2002) (Figura 2). Para este transcriptoma, sólo un péptido de esta familia tiene mayor similitud con StSN1 (subfamilia I), 10 agrupan con StSN2 (Subfamilia II) y cuatro con la esnaquina 3 de papa (Subfamilia III). Tres miembros de la subfamilia II tienen una región variable, entre el péptido señal y el dominio GASA, rica en prolina, como la proteína GASAE de *A. thaliana* (Roxrud et al. 2007). De éstos tres péptidos dos son idénticos en la región del dominio GASA y del péptido señal, variando sólo en la zona rica en prolinas, por lo que parecen haber surgido por duplicación. El tercero varía levemente en el péptido señal, además de la zona rica en prolinas, por lo que podría tener otra localización subcelular o podría provenir de un tejido diferente al de los otros dos. Con respecto al número de genes Snakin/GASA en los genomas y transcriptomas de diferentes especies, en el genoma de *A. thaliana* se han detectado hasta ahora 14 genes de esta familia (Roxrud et al. 2007) y en el de papa 18 (Nahirñak et al. 2016). El número de genes Snakin/GASA fue similar en el trancriptoma de brotes de ibirapitá (18) y el transcriptoma de brotes de *Stellaria media* (16) (Slavokhotova et al. 2017). Si el número de genes Snakin/GASA se conserva en las diferentes especies vegetales, entonces, en brotes se podrían estar expresando todos o casi todos los genes de esta familia. En congorosa, como se analizó un *pool* de distintos tejidos, no podemos saber cuántos genes Snakin/GASA se están expresando en cada uno de ellos. Si consideráramos, que en brotes se podrían estar expresando los 15 genes, parecería que el número y tipo de genes Snakin/GASA que se expresa en los tres tejidos (brotes, flores, hojas) es el mismo. Alternativamente, en flores y hojas podrían expresarse un menor número de genes Snakin/GASA. En papa por ejemplo, para la esnaquina-3 no se detectó expresión ni en hojas ni en flores (Nahirñak et al. 2016)

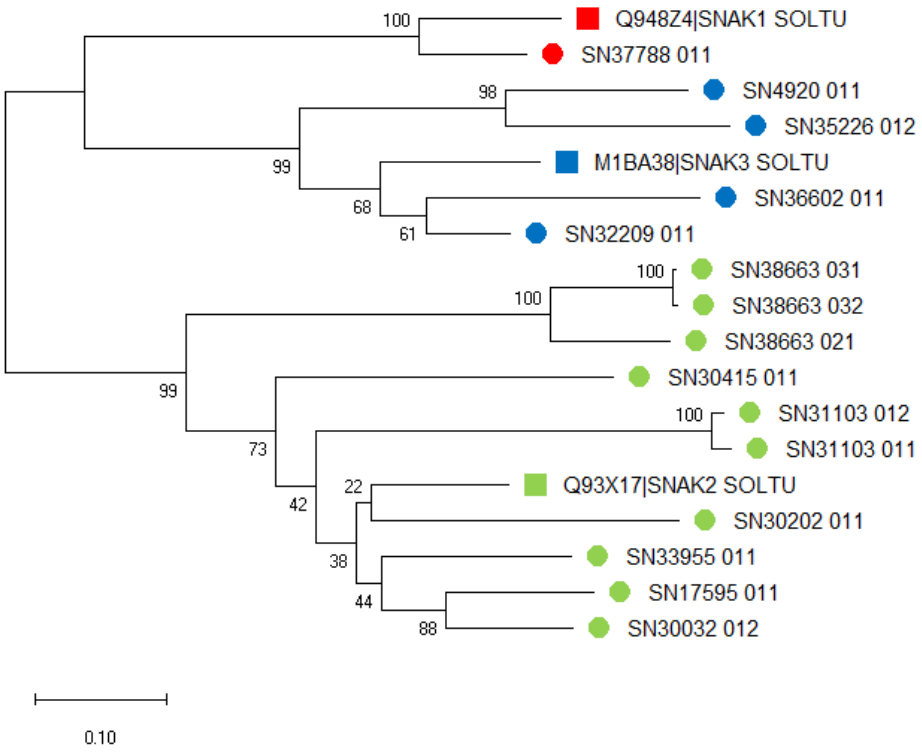


Figura 2. Árbol Neighbor Joining de distancia genética de péptidos Snakin/GASA putativos encontrados en el transcriptoma de *M. ilicifolia* y esnaquinas 1, 2 y 3 de papa. La secuencia aminoacídica deducida de los transcriptos fue alineada con StSN1, StSN2 y StSN3, agrupándose en las tres subfamilias propuestas por Berrocal Lobo etr al. (2002). Los cuadrados representan a las esnaquinas de papa y los círculos a las proteínas Snakin/GASA de M. ilicifolia. En rojo se muestra la subfamilia I, en azul la subfamilia III y en verde la subfamilia II. Los valores en los nodos muestran los valores de bootstrap (5000 réplicas)

Los genes que codifican péptidos putativos de tipo defensina o Snakin/GASA serán validados mediante amplificación usando ADN genómico. Algunos de ellos podrán ser caracterizados funcionalmente. Debido a que varios de los genes encontrados en *M. ilicifolia* presentan variaciones de secuencia interesantes, estamos considerando que, una vez validados, podrían ser considerados para patentamiento, etc.

Consideraciones Finales

En este trabajo nos planteamos como objetivo el aislamiento de genes que codifican defensinas y esnaquinas en plantas nativas y su caracterización estructural y funcional. Se seleccionaron tres especies nativas con usos en la medicina popular. Para la identificación y el aislamiento de genes se utilizó en un principio la amplificación por PCR, con cebadores degenerados diseñados a partir de secuencias conocidas de otras especies. Con esta estrategia se obtuvo el gen que denominamos *PdSN1*, que codifica una esnaquina de tipo 1, aislado de ibirapitá. En paralelo, se probaron varios cebadores específicos diseñados para otras especies. Usando *primers* diseñados por Padovan et al. 2010 para *V. unguiculata*, se aisló un gen del genoma de ceibo, *EcgDf1*, que codifica una defensina. Con esta metodología esperábamos aislar varios genes de defensinas y esnaquinas, seleccionando los más variables (diversos) para su posterior producción en un sistema heterólogo. Sin embargo, los resultados no fueron tan satisfactorios, posiblemente debido a la baja similitud en la estructura primaria de estos péptidos, que hace difícil su búsqueda sólo basada en homología de secuencias.

Posteriormente incorporamos la estrategia de secuenciado masivo de ARN, ahora más accesible, y la metodología de ensamblado *de novo,* que nos permitieron obtener un catálogo de AMPs pertenecientes a plantas de las que se sabe muy poco y para las que no se cuenta con un genoma de referencia propio. Varios de los genes de tipo defensina y snakin/GASA encontrados fueron verificados en ADN genómico y caracterizados estructuralmente. Tanto PdSN1, EcgDf1, como los péptidos deducidos de los transcriptos presentan características comunes a defensinas y esnaquinas. En relación a nuestra primera hipótesis, podemos decir que los genes y los péptidos de las especies estudiadas presentan patrones estructurales comunes como la presencia de un intrón en genes de defensinas y uno, dos o tres intrones en genes Snakin/GASA. Además, presentan un péptido señal de secreción y una distribución de cisteínas típica de cada familia de AMP, con la excepción de algunas defensinas putativas (tanto en congorosa como en ibirapitá) que tienen el motivo "αβ motif" modificado o el motivo γ-core no está presente. En congorosa además, se encontraron dos defensinas con el motivo "αβ motif" y el motivo γ-core pero con dos cisteínas adicionales. Los transcriptos de estos genes "no típicos" no fueron aún verificados; este tipo de análisis y otros como su caracterización estructural y funcional serían necesarios para su mejor clasificación. La predicción de la estructura 3D se realizó para los péptidos PdSN1 y EcgDf11 y ambos parecen seguir los mismos patrones estructurales que los de otros péptidos de tipo defensina y esnaquina cristalizados. Para PdSN1, nuestros resultados

mostraron además que el motivo *hélix-turn-helix* (HTH) estaría presente independientemente del patrón de conectividad específico de los enlaces disulfuro. Incluso, la ausencia completa de puentes disulfuro, también conduciría a un motivo HTH estable. Estos resultados se discuten más adelante en relación a la producción del péptido en sistemas heterólogos y a su actividad antimicrobiana. EcgDf1 podría dimerizar, al igual que otras defensinas vegetales, futuros estudios son necesarios para confirmar dicha hipótesis. Con respecto a la mayor diversidad de secuencia esperada para éstas moléculas en plantas nativas podemos decir que PdSN1 (84 % de similitud de secuencia con StSN1) y EcgDf1 (81 % de identidad con defensina de *M. truncatula*) no son tan variables en relación a otros péptidos de su tipo en otras plantas. Sin embargo, varias defensinas putativas encontradas en el transcriptoma de *P. dubium* y de *M. ilicifolia* difieren en varios residuos incluso aquellos que son altamente conservados en la mayoría de las defensinas depositadas en PDB.

De acuerdo con los antecedentes, consideramos que la producción recombinante de péptidos de tipo defensina y esnaquina podría ser exitosa en un sistema procariota, obteniendo péptidos biológicamente activos. Por las características de estos péptidos, se seleccionaron en un principio, para el péptido PdSN1, sistemas *E. coli*/plásmido en los que la bacteria pudiera formar puentes S-S en el citoplasma (cepas Shuffle y Rosetta-gami) y que el vector (pET102/D) tuviera una proteína que fusionada al péptido facilitara su solubilidad y evitara la posible toxicidad del péptido hacia la bacteria productora. Con ambas cepas se obtuvo el péptido recombinante en forma soluble, pero con mejores resultados con la cepa Rosetta-gami. Este péptido pudo ser purificado y mostró actividad antimicrobiana contra varios patógenos. Las predicciones de la estructura 3D de PdSN1, en la que parecería que la ausencia completa de puentes disulfuro también conduciría a un péptido con un motivo HTH estable, además del hecho de que otra esnaquina StSN1 presentó actividad aún sin puentes disulfuro y de que varios péptidos recombinantes de tipo defensina con actividad han sido producidos en la cepa BL21(DE3) nos animó a probar la producción de la defensina EcgDf1 también dicha cepa, además de Rosetta-gami. Los resultados fueron similares en ambas cepas en cuanto a solubilidad con un leve aumento de la forma soluble en Rosetta-gami. Elegimos esta última para continuar con los ensayos, obteniendo un péptido biológicamente activo contra un grupo de microorganismos. Por razones de tiempo, no seguimos los ensayos con el péptido producido por la cepa BL21(DE3), sin embargo, sería interesante comparar los resultados de actividad antimicrobiana obtenidos con el péptido producido por esta cepa, con los del procedente de Rosetta-gami. Con ambos péptidos los rendimientos fueron bajos (~1 mg/L cultivo), lo que limitó la cantidad de ensayos de actividad realizados. Futuros ensayos deberían realizarse para mejorar dichos

rendimientos, por ejemplo, uso de medio autoinductor, incorporación de otros vectores que permitan mejorar el porcentaje de péptido en las fracciones solubles, etc.

Para contrastar nuestra tercera hipótesis, la actividad antimicrobiana de PdSN1 y EcgDf1 fue evaluada *in vitro* contra varios microorganismos patógenos humanos y de plantas, incluyendo hongos filamentosos, levaduras y bacterias. Ambos péptidos presentaron actividad, en general con una mejor performance para hongos, especialmente en el caso de la defensina. Ambos péptidos mostraron actividad contra *A. niger, C. albicans, B. cinerea, C. michiganensis ssp. michiganensis* and *A. altenata.* Los valores IC50 fueron mejores para PdSN1 en el caso de *A. alternata* y similares en el caso de *B. cinerea*. Para *C. albicans* y *A. niger* los IC50 fueron menores con EcgDf1. Las concentraciones requeridas para la capacidad inhibitoria de estos AMPs parecen depender del microorganismo y del péptido analizado. Como se ha discutido en el capítulo 2, este fenómeno puede deberse al hecho de que un organismo produce un conjunto de AMP que pueden actuar de forma cooperativa, con espectros superpuestos de actividades antimicrobianas, pero con especificidad contra grupos particulares de microorganismos. Destacamos que la defensina presentó actividad contra *P. expansum*, un patógeno de amplio rango de huéspedes, que causa enfermedades importantes enfermedades pos-cosecha. En el caso de la esnaquina de ibirapitá, destacamos una potente actividad contra la bacteria *S. scabies*, un importante patógeno de la papa. Es interesante notar también que ambos péptidos presentaron actividad contra patógenos oportunistas del ser humano, como *C. albicans* y *A. niger*, produciendo infecciones en pacientes inmunodeprimidos. Si bien ya se han reportado AMPs con actividad contra estos patógenos, nuestros péptidos ya están siendo considerados como candidatos para ser utilizado como péptidos radiomarcados en proyectos enfocados en el diagnóstico de infecciones producidas por estos patógenos, en colaboración con la cátedra de Radioquímica de Facultad de Química.

Finalmente, respondiendo nuestra pregunta final, consideramos que algunos de los AMP de tipo defensina o esnaquina encontrados en este trabajo presentan características distintivas en su secuencia, especialmente algunos de los deducidos de los transcriptomas de *P. dubium* y *M. ilicifolia*. Queda pendiente la confirmación biológica de algunos de ellos, y su expresión en sistemas heterólogos para evaluar su potencial como agentes antimicrobianos. Además, como describimos en el párrafo anterior, tanto PdSN1 como EcgDf1 presentan características novedosas a nivel de su perfil de actividad antimicrobiana, ya que hasta el momento no hemos encontrado reportes de otros AMPs vegetales con actividad para patógenos de importancia comercial como *S. scabies* y *P. expansum*. Varios de estos péptidos podrían ser considerados entonces como buenos candidatos para el desarrollo de productos útiles para el diagnóstico o

control eficaz de enfermedades. En agricultura, se podrían usar en la protección de plantas tanto en el campo como después de la cosecha, como alternativa al uso de agroquímicos. Además, dado que EcgDf1 y PdSN1 muestran actividad contra patógenos fúngicos humanos, en el futuro, si se demuestra que no presentan toxicidad, podrían ser un sustituto de los antibióticos clásicos. Para ello, es necesario continuar los estudios en estos péptidos, con el objetivo de comprender los mecanismos de acción que utilizan y sus relaciones estructura-función.

Más allá de los objetivos planteados en este trabajo, la metodología de RNA-seq permitió detectar otros tipos de AMPs, como *hevein-like* y LTPs, no encontrando péptidos con motivo ciclótido o *γ-hairpinin*. Un único péptido con similitud con tioninas fue encontrado en el transcriptoma de congorosa. Además, los datos provenientes de los transcriptomas ensamblados nos permite acceder a la secuencia de genes que podrían ser usados como genes de referencia en estudios de expresión relativa.

Perspectivas

El desarrollo exitoso de AMPs con actividad, para el diagnóstico o el control eficaz y a largo plazo de enfermedades, implica comprender mejor los mecanismos de acción que utilizan estas moléculas. Por esta razón nos hemos propuesto continuar con los estudios en estos péptidos de la siguiente manera:

a) Con respecto al péptido PdSN1, los resultados encontrados con respecto a la predicción de su estructura 3D nos lleva a preguntarnos sobre la importancia del motivo HTH y de los puentes disulfuro para la estructura del péptido y la función del mismo. Entonces, con la finalidad de saber más sobre la relación entre estructura 3D y función de PdSN1 nos hemos propuesto: 1) analizar la incidencia de los puentes disulfuro, mediante seguimiento de la actividad antimicrobiana de mutantes PdSN1 deficientes en cisteínas; 2) evaluar el efecto de PdSN1 sobre la membrana celular de los microorganismos para los que presentó actividad, utilizando diferentes técnicas (microscopía de fluorescencia, citometría de flujo) que permitan visualizar la viabilidad celular y la localización del péptido en las células blanco y 3) analizar mediante simulación y retardos en gel la posible unión de PdSN1 a través del motivo HTH al ADN de los microorganismos.

b) Para la defensina EcgDf1, pensamos analizar su posible modo de acción mediante 1) evaluación de su efecto sobre la membrana celular de los microorganismos para los mostró actividad, utilizando diferentes técnicas (microscopía de fluorescencia, citometría de flujo) que permitan visualizar la viabilidad celular y la localización del péptido en las células blanco; 2) analizar su posible dimerización u oligomerización mediante simulación y ensayos *in vitro*: western blots, análisis de actividad de mutantes en sitios claves para la posible dimerización, etc.

c) En relación a los transcriptomas ensamblados *de novo* de ibirapitá y congorosa 1) Continuar con la confirmación de la fiabilidad del ensamblado, mediante la validación por PCR a partir de ADN genómico de los restantes genes de defensinas y esnaquinas. Son particularmente interesantes aquellos genes con mayor variabilidad en su secuencia, por ejemplo, defensinas putativas con motivos cisteínas que difieren de los típicos de las "verdaderas" defensinas. Estos péptidos podrían estudiarse más profundamente en cuanto a su organización génica y su estructura 3D, para una mejor clasificación. 2) En relación a su potencial actividad antimicrobiana, se podría realizar un rastreo primario de la actividad de varias secuencias expresadas en *E. coli*, enfrentando

las células productoras a diferentes patógenos, sorteando el paso de purificación de cada péptido. Por otro lado, nos interesa conocer las posibles funciones de estos genes en la planta por lo que podemos analizar si los mismos se expresan de forma constitutiva o se activan frente a estrés biótico por lo que nos planteamos 3) Determinar la expresión relativa de genes Snakin/GASA y defensinas en brotes de ibirapitá enfrentados a patógenos comunes del suelo como *A. niger* y *S. scabies.* En este sentido, ya contamos con varios genes probados como housekeeping en hojas y brotes de ibirapitá. El contar con los transriptomas nos permite además tener la secuencia de otros posibles genes de referencia.

En principio pensamos que algunos genes y péptidos deducidos del transcriptoma de congorosa podrían ser candidatos a patentamiento, licenciamiento, etc., por lo que serán estudiados con mayor profundidad.

Bibliografía general

Ali S, Ganai BA, Kamili AN, Bhat AA, Mir ZA, Bhat JA, Tyagi A, Islam ST, Mushtaq M, Yadav P, Rawat S, Grover A (2018) Pathogenesis-related proteins and peptides as promising tools for engineering plants with multiple stress tolerance. Microbiol Res 212–213:29–37. doi: 10.1016/j.micres.2018.04.008

Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O (2017) Sequence analysis DeepLoc : prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387–3395. doi: 10.1093/bioinformatics/btx431

Almasia NI, Bazzini AA, Hopp HE, Vazquez-Rovere C (2008) Overexpression of snakin-1 gene enhances resistance to *Rhizoctonia solani* and *Erwinia carotovora* in transgenic potato plants. Mol Plant Pathol 9:329–338. doi: 10.1111/J.1364-3703.2008.00469.X

Almasia NI, Molinari MP, Maroniche GA, Nahirñak V, Barrios Barón MP, Taboga OA, Vazquez-Rovere C (2017) Successful production of the potato antimicrobial peptide Snakin-1 in baculovirus-infected insect cells and development of specific antibodies. BMC Biotechnol 17:1–11. doi: 10.1186/s12896-017-0401-2

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, Donovan CO, Redaschi N, Yeh LL (2004) UniProt: The Universal Protein knowledgebase. Nucleic Acids Res 32: D115–D119. doi: 10.1093/nar/gkh131

Araki T, Torikata T (1995) Structural classification of plant chitinases. two subclasses in Class I and Class II chitinases. Biosci Biotech Biochem 59:336–338. doi: 10.1248/cpb.37.3229

Balaji V, Sessa G, Smart CD (2011) Silencing of host basal defense response-related gene expression increases susceptibility of *Nicotiana benthamiana* to *Clavibacter michiganensis* subsp. michiganensis. Phytopathology 101:349–57. doi: 10.1094/PHYTO-05-10-0132

Balaji V, Smart CD (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 Res 21:23–37. doi: 10.1007/s11248-011-9506-x

Benko-Iseppon AM, Galdino SL, Calsa T, Kido EA, Tossi A, Belarmino LC, Crovella S (2010) Overview on plant antimicrobial peptides. Curr Protein Pept Sci 11:181–8. doi: 10.2174/138920310791112075

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684-3688. doi: 10.1063/1.448118

Berrocal-Lobo M, Segura A, Moreno M, López G, García-Olmedo F, Molina A (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. doi: 10.1104/pp.010685.1

Blaser MJ (2002) The genetic gymnastics of our indigenous microbes. N Engl J Med 346:2083-2085.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data Bioinformatics 30:2114-2120.

Bolson M, Hefler SR, Dall'Oglio Chaves EI, Gasparotto Junior A, Cardozo Junior EL (2015) Ethno-medicinal study of plants used for treatment of human ailments, with residents of the surrounding region of forest fragments of Paraná, Brazil. J Ethnopharmacol 161:1–10. doi: 10.1016/j.jep.2014.11.045

Boutrot F, Chantret N, Gautier MF (2008) Genome-wide analysis of the rice and arabidopsis non-specific lipid transfer protein (nsLtp) gene families and identification of wheat nsLtp genes by EST data mining. BMC Genomics 9:1–19. doi: 10.1186/1471-2164-9-86

Bradford MM (1976) A rapid and sensitive method for the quantitation microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254 doi: 10.1016/0003-2697(76)90527-3

Broekaert WF, Terras F, Cammue BP, Vandedeyden J (1990) An automated quantitative assay for fungal growth inhibition. FEMS Microbiol Lett 69:55–60. doi: 10.1016/0378-1097(90)90412-J

Broekaert WF, Terras FR, Cammue BP, Osborn RW (1995) Plant defensins: novel antimicrobial peptides as components of the host defense system. Plant Physiol 108:1353–1358. doi: 10.1104/pp.108.4.1353

Brogden KA (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3:238–250. doi: 10.1038/nrmicro1098

Cabral KM., Almeida MS, Valente AP, Almeida FC., Kurtenbach E (2003) Production of the active antifungal *Pisum sativum* defensin 1 (Psd1) in *Pichia pastoris*: overcoming the inefficiency of the STE13 protease. Protein Expr Purif 31:115–122. doi: 10.1016/S1046-5928(03)00136-0

Cantrell, CL, Dayan, FE, Duke, SO 2012. Natural products as sources for new pesticides. J. Natl Products 75: 1231-1242

Carrasco L, Vazquez D, Hernandez-Lucas C, Carbonero P Garcia-Olmedo F (1981) Thionins: plant peptides that modify membrane permeability in cultured mammalian cells. Europ. J. Biochem 116:185–189.

Carvalho DO, Gomes VM (2007) Role of plant lipid transfer proteins in plant cell physiology-A concise review. Peptides 28:1144–1153. doi: 10.1016/j.peptides.2007.03.004

Carvalho DO, Gomes VM (2009) Peptides Plant defensins- Prospects for the biological functions and biotechnological properties. Peptides 30:1007–1020. doi: 10.1016/j.peptides.2009.01.018

Carvalho DO, Gomes VM (2011) Plant Defensins and Defensin-Like Peptides - Biological Activities and Biotechnological Applications Curr Pharm Des 17:4270–4293. doi: 10.2174/138161211798999447

Case DA, Cerutti DS, Cheatham III TE, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, Izadi S, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein D, Merz KM, Monard G, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Simmerling CL, Botello-Smith WM, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Xiao L, York DM, Kollman PA (2017) AMBER 2017, University of California, San Francisco.

Ceroni A, Passerini A, Vullo A Frasconi P (2006) DISULFIND: A disulfide bonding state and cysteine connectivity prediction server. Nucl Acids Res 34: W177-81. doi:10.1093/nar/gkl266

Chakraborty S, Newton AC (2011) Climate change, plant diseases and food security: An overview (Review). Plant Pathol 60:2-14

Charlet M, Chernysh S, Philippe H, Hetru C, Hoffmann JA. Bulet P (1996) Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, *Mytilus edulis*. J. Biol. Chem 271: 21808–817.

Cheng J, Saigo H, Baldi P (2006) Large-scale prediction of disulphide bridges using kernel methods, two-dimensional recursive neural networks, and weighted graph matching. proteins: structure, function. Bioinformatics 62:617-629. DOI: 10.1002/prot.20787

Cornet B, Bonmatin JM, Hetru C, Hoffmann JA, Ptak M, Vovelle, F (1995). Refined three-dimensional solution structure of insect defensin A Structure 3: 435-448.

Craik DJ, Daly NL, Bond T, Waine C (1999) Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J Mol Biol 294:1327–1336. doi: 10.1006/jmbi.1999.3383

Cuervo A, Dans PD, Carrascosa JL, Orozco M, Gomila G, Fumagalli L (2014) Direct measurement of the dielectric polarization properties of DNA. Proc Natl Acad Sci 111: E3624–E3630. doi: 10.1073/pnas.1405702111

Dalla Rizza M, Dellavalle P, Narancio R, Cabrera A, Ferreira F (2008) Biomolecules as Host Defense Weapons Against Microbial Pathogens. Recent patents on DNA & gene sequences 2:82-96.

Dans PD, Danilāne L, Ivani I, Dršata T, Lankaš F, Walther J, Illa Pujagut R, Battistini F, Gelpí JL, Lavery R, Orozco M (2016) Long-timescale dynamics of the Drew-Dickerson dodecamer. Nucl Acids Res 44:4052-4066. doi: 10.1093/nar/gkw264

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. J Chem Phys 98:10089-10095. doi: 10.1063/1.464397

de Beer A, Vivier MA (2011) Four plant defensins from an indigenous South African Brassicaceae species display divergent activities against two test pathogens despite high sequence similarity in the encoding genes. BMC Res Notes 4:459. doi: 10.1186/1756-0500-4-459

De Coninck B, Cammue BPA, Thevissen K (2013) Modes of antifungal action and in planta functions of plant defensins and defensin-like peptides. Fungal Biol Rev 26:109–120. doi: 10.1016/j.fbr.2012.10.002

Díaz-Dellavalle P, Cabrera A, Alem D, Larrañaga P, Ferreira F, Dalla-Rizza M. (2011). Antifungal activity of medicinal plant extracts against phytopathogenic fungi *Alternaria* sp. Chil J Agr Res, 71: 231-239.

Do HM, Lee SC, Jung HW, Sohn KH, Hwang BK (2004) 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 166:1297–1305. doi: 10.1016/j.plantsci.2004.01.008

Doyle J (1991) DNA protocols for plants-CTAB total DNA isolation. In: Hewitt GM, Johnston A (eds) Molecular Techniques in Taxonomy, Springer, Berlin, pp. 283–293

El-Mounadi K, Islam KT, Hernández-Ortiz P, Read ND, Shah DM (2016) Antifungal mechanisms of a plant defensin MtDef4 are not conserved between the ascomycete fungi *Neurospora crassa* and *Fusarium graminearum*. Mol Microbiol 100:542-559. doi: 10.1111/mmi.13333

Estrada G, Garcia B, Schiavon E, Ortiz E, Cestèle S, Wanke E, Possani L, Corzo G (2007) Four disulfide-bridged scorpion beta neurotoxin CssII: heterologous expression and proper folding in vitro. Biochim. Biophys. Acta 1770, 1161–1168.

Ferrè F, Clote P (2006) DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. Nucl Acids Res 34: W182-5. doi:10.1093/nar/gkl189

Ferreira Lacerda A, Perseghini Del Sarto R, Santos Silva M, Rosas de Vasconcelos E, Ramos Coelho R, Olinto dos Santos V, Vieira Godoy C, Santos Seixas C, Mattar da Silva M, Grossi-de-Sa M (2016) The recombinant pea defensin Drr230a is active against impacting soybean and cotton pathogenic fungi from the genera Fusarium, Colletotrichum and Phakopsora. 3 Biotech 6:59. doi: 10.1007/s13205-015-0320-7

Frohman MA, Dush, MK, Martin, GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci 85:8998-9002.

Games PD, dos Santos IS, Mello ÉO, Diz MSS, Carvalho AO, de Souza-Filho GA, Da Cunha M, Vasconcelos IM, Ferreira B dos S, Gomes VM (2008) Isolation, characterization and cloning of a cDNA encoding a new antifungal defensin from *Phaseolus vulgaris* L. seeds. Peptides 29:2090–2100. doi: 10.1016/j.peptides.2008.08.008

Gao AG, Hakimi SM, Mittanck CA, Wu Y, Woerner M, Stark DM, Rommens CM (2000) Fungal pathogen protection in potato by expression of a plant defensin peptide. Nature Biotechnol 18:1307-1310.

García AN, Ayub ND, Fox AR, Gómez MC, Diéguez MJ, Pagano EM, Berini CA, Muschietti JP, Soto G (2014) Alfalfa snakin-1 prevents fungal colonization and probably coevolved with rhizobia. BMC Plant Biol 14:248. doi: 10.1186/s12870-014-0248-9

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, New York, pp 571–607

Gentile F, Amodeo P, Febbraio F, Picaro F, Motta A, Formisano S, Nucci R (2002) SDS-resistant active and thermostable dimers are obtained from the dissociation of homotetrameric β-glycosidase from hyperthermophilic *Sulfolobus solfataricus* in SDS: Stabilizing role of the A-C intermonomeric interface. J Biol Chem 277:44050–44060. doi: 10.1074/jbc.M206761200

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011:29:644-652.

Graham M a, Silverstein K a T, Cannon SB, Vandenbosch K a (2004) Computational Identification and Characterization of Novel Genes from Legumes 1. Plant Physiol 135:1179–1197. doi: 10.1104/pp.104.037531.exchange

Guillen-Chable F, Arenas-Sosa I, Islas-flores I, Corzo G, Martinez-liu C, Estrada G (2017) Antibacterial activity and phospholipid recognition of the recombinant defensin J1-1 from Capsicum genus. Protein Expr Purif 136:45–51. doi: 10.1016/j.pep.2017.06.007

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98

Harris PWR, Yang SH, Molina A, Lopez G, Middleditch M, Brimble MA (2014) Plant antimicrobial peptides snakin-1 and snakin-2: chemical synthesis and insights into the disulfide connectivity. Chemistry 20:5102–5110. doi: 10.1002/chem.201303207

Hayes BM, Bleackley MR, Wiltshire JL, Anderson MA, Traven A, van der Weerden NL (2013) Identification and mechanism of action of the plant defensin NaD1 as a new member of the antifungal drug arsenal against *Candida albicans*. Antimicrob. Agents Chemother 57:3667–3675

Hegedüs N, Marx F (2013) Antifungal proteins: More than antimicrobials? Fungal Biol Rev 26:132–145. doi: 10.1016/j.fbr.2012.07.002

Herbel V, Schäfer H, Wink M (2015) Recombinant production of snakin-2 (an antimicrobial peptide from tomato) in *E. coli* and analysis of its bioactivity. Molecules 20:14889–14901. doi: 10.3390/molecules200814889

Herbel V, Sieber-frank J, Wink M (2017) The antimicrobial peptide snakin-2 is upregulated in the defense response of tomatoes (*Solanum lycopersicum*) as part of the jasmonate-dependent signaling pathway. J Plant Physiol 208:1–6. doi: 10.1016/j.jplph.2016.10.006

Hughes AL (1999) Evolutionary diversification of the mammalian defensin. Cell. Mol. Life Sci. 56: 94-103.

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38. doi: 10.1016/0263-7855(96)00018-5

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–35. doi: 10.1063/1.445869

Kader J (1996) Lipid-Transfer Proteins in Plants. Annu Rev Plant Physiol Plant Mol Biol 47:627–654. doi: 10.1146/annurev.arplant.47.1.627

Kant P, Liu WZ, Pauls KP (2009) PDC1, a corn defensin peptide expressed in *Escherichia coli* and *Pichia pastoris* inhibits growth of Fusarium graminearum. Peptides 30:1593–1599. doi: 10.1016/j.peptides.2009.05.024

Kavi Kishor PB (2015) Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. Front Plant Sci 6:1–17. doi: 10.3389/fpls.2015.00544

Koehbach J, Attah AF, Berger A, Hellinger R, Kutchan TM, Carpenter EJ, Rolf M, Sonibare MA, Moody JO, Wong GKS, Dessein S, Greger H, Gruber CW (2013) Cyclotide discovery in Gentianales revisited--identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants. Biopolymers 100:438–452. doi: 10.1002/bip.22328

Koike M, Okamoto T, Tsuda S, Imai R (2002) A novel plant defensin-like gene of winter wheat is specifically induced during cold acclimation Biochem Bioph Res Co 298: 46–53.

Kovaleva V, Krynytskyy H, Gout I, Gout R (2011) Recombinant expression, affinity purification and functional characterization of Scots pine defensin 1. Appl Microbiol Biotechnol 89:1093–1101. doi: 10.1007/s00253-010-2935-2

Kovalskaya N, Hammond RW (2009) Expression and functional characterization of the plant antimicrobial snakin-1 and defensin recombinant proteins. Protein Expr Purif 63:12–7. doi: 10.1016/j.pep.2008.08.013

Kuddus MR, Rumi F, Tsutsumi M, Takahashi R, Yamano M, Kamiya M, Kikukawa T, Demura M, Aizawa T (2016) Expression, purification and characterization of the recombinant cysteine-rich antimicrobial peptide snakin-1 in *Pichia pastoris*. Protein Expr Purif 122:15–22. doi: 10.1016/j.pep.2016.02.002

Laemmli UK (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature 227:680–685

Larrañaga P, Díaz-Dellavalle P, Cabrera A, Alem D, Leoni C, Almeida-Souza AL, Dalla-Rizza M (2012) Activity of Naturally Derived Antimicrobial Peptides against Filamentous Fungi Relevant for Agriculture. Sustainable Agriculture Research, 1: p211-221

Lay FT, Anderson MA (2005) Defensins-components of the innate immune system in plants. Curr Protein Pept Sci 6:85–101. doi: 10.2174/1389203053027575

Lay FT, Brugliera F, Anderson MA (2003) Isolation and Properties of Floral Defensins from Ornamental Tobacco and Petunia 1. 131:1283–1293. doi: 10.1104/pp.102.016626.)

Lay FT, Mills GD, Poon IKH, Cowieson NP, Kirby N, Baxter AA, Van Der Weerden NL, Dogovski C, Perugini MA, Anderson MA, Kvansakul M, Hulett MD (2012) Dimerization of plant defensin NaD1 enhances its antifungal activity. J Biol Chem 287:19961–19972. doi: 10.1074/jbc.M111.331009

Lee HI, Broekaert WF, Raikhel N V. (1991) Co- and post-translational processing of the hevein preproprotein of latex of the rubber tree (*Hevea brasiliensis*). J Biol Chem 266:15944–15948

Li L, Zhang C, Xu D, Schläppi M, Xu ZQ (2012) Expression of recombinant EARLI1, a hybrid proline-rich protein of Arabidopsis, in Escherichia coli and its inhibition effect to the growth of fungal pathogens and *Saccharomyces cerevisiae*. Gene 506:50–61. doi: 10.1016/j.gene.2012.06.070

Lister R, Gregory BD, Ecker JR (2009) Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Curr Opin Plant Biol 12: 107-118.

Lobo DS, Pereira IB, Fragel-Madeira L, Medeiros LN, Cabral LM, Faria J (2007) Antifungal *Pisum sativum* defensin 1 interacts with *Neurospora crassa* cyclin F related to the cell cycle. Biochemistry-US 46: 987–96.

Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M (2012) SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact 11:753. doi: 10.1186/1475-2859-11-56

Macedo M, Freire M, Cristina E, Lı M, Toyama MH (2003) A trypsin inhibitor from *Peltophorum dubium* seeds active against pest proteases and its effect on the survival of *Anagasta kuehniella* (Lepidoptera:Pyralidae). Biochim Biophys Acta 1621 1621:170–182. doi: 10.1016/S0304-4165(03)00055-2

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99sb. J Chem Theor Comput 11:3696-3713. doi: 10.1021/acs.jctc.5b00255

Manning M, Colón W (2004) Structural basis of protein kinetic stability: Resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward β-sheet structure. Biochemistry 43:11248–11254. doi: 10.1021/bi0491898

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

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671–682. doi: 10.1038/nrg3068

Meiyalaghan S, Thomson SJ, Fiers MW, Barrell PJ, Latimer JM, Mohan S, Jones EE, Conner AJ, Jacobs JM (2014) 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 15:2. doi: 10.1186/1471-2164-15-2

Mello ÉDO, Santos IS, Carvalho ADO, de Souza LS, de Souza-filho GA, do Nascimento V V, Machado OL, Zottich U, Gomes VM (2014) Functional expression and activity of the recombinant antifungal defensin PvD1r from *Phaseolus vulgaris* L. (common bean) seeds. BMC Biochem 15:1–13. doi: 10.1186/1471-2091-15-7

Mendez E, Moreno A, Colilla F, Pelaez F, Limas GG, Mendez R (1990) Primary structure and inhibition of protein synthesis in eukaryotic cell-free system of a novel thionin, g-hordothionin, from barley endosperm. Eur J Biochem 194: 533–539.

Mohan S, Meiyalaghan S, Latimer JM, Gatehouse ML, Monaghan KS, Vanga BR, Pitman AR, Jones EE, Conner AJ, Jacobs JME (2014) GSL2 over-expression confers resistance to *Pectobacterium atrosepticum* in potato. Theor Appl Genet 127:677–689. doi: 10.1007/s00122-013-2250-2

Montesinos E (2007) Antimicrobial peptides and plant disease control. FEMS Microbiol Lett 270:1–11. doi: 10.1111/j.1574-6968.2007.00683.x

Mylne JS, Colgrave ML, Daly NL, Chanson AH, Elliott AG, Mccallum EJ, Jones A, Craik DJ (2011) Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. Nat Chem Biol 7:1–3. doi: 10.1038/nchembio.542

Nahirñak V, Almasia NI, Fernandez PV, Hopp HE, Estevez JM, Carrari F, Vazquez-Rovere C (2012a) Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. Plant Physiol 158:252–63. doi: 10.1104/pp.111.186544

Nahirñak V, Almasia NI, Hopp HE, Vazquez-Rovere C (2012b) Involvement in hormone crosstalk and redox homeostasis Snakin/GASA proteins. Plant Signal Behav 7:1004–1008. doi: 10.4161/psb.20813

Nahirñak V, Rivarola M, Gonzalez de Urreta M, Paniego N, Hopp HE, Almasia NI, Vazquez-Rovere C (2016) Genome-wide Analysis of the Snakin/GASA Gene Family in *Solanum tuberosum* cv. Kennebec. Am J Potato Res. doi: 10.1007/s12230-016-9494-8

Nawrot R, Barylski J, Nowicki, G, Broniarczyk J, Buchwald W, Goździcka-Józefiak A (2014) Plant antimicrobial peptides. Folia microbiol 59:181-196.

Neill DA (1988) Experimental Studies on Species Relationships in Erythrina (Leguminosae: Papilionoideae). Ann Missouri Bot Gard 75:886–969

Nielsen H (2017) Predicting secretory proteins with SignalP. In: D. K (ed) Methods in Molecular Biology. New York, pp 59–73

Oard S, Karki B (2006) Mechanism of b-purothionin antimicrobial peptide inhibition by metal ions: molecular dynamics simulation study. Biophysical Chemistry, 121: 30–43.

Odintsova T, Egorov, T (2012) Plant antimicrobial peptides. In Plant signaling peptides, Springer, Berlin, Heidelberg pp. 107-13).

Oliveira-Lima M, Benko-Iseppon AM, Ferreira, Costa RJ, Rodríguez-Decuadro S, Kido EA, Crovella S, Pandolfi V (2017) Snakin: structure, roles and applications of a plant antimicrobial peptide. Curr Protein Pept Sci 18:1–7. doi: 10.2174/1389203717666160619 1

Oparin PB, Mineev KS, Dunaevsky YE, Arseniev AS, Belozersky MA, Grishin EV, Egorov TA, Vassilevski AA (2012) Buckwheat trypsin inhibitor with helical hairpin structure belongs to a new family of plant defence peptides. Biochem J 446:69–77. doi: 10.1042/BJ20120548

Oren Z, Shai Y 1998 Mode of action of linear amphipathic α-helical antimicrobial peptides. Biopolymers 47:451-463

Osborn RW, De Samblanx GW, Thevissen K, Goderis I, Torrekens S, Van Leuven F, Attenborough S, Rees SB, Broekaert WF (1995) Isolation and characterisation of plant defensins from seeds of Asteraceae, Fabaceae, Hippocastanaceae and Saxifragaceae. FEBS Lett 368:257–262. doi: 10.1016/0014-5793(95)00666-W

Padovan L, Crovella S, Tossi A, Segat L (2010a) Techniques for Plant Defensin Production. Curr Protein Pept Sci 11:231–235

Padovan L, Segat L, Tossi A, Calsa T, Ederson AK, Brandao L, Guimarães RL, Pandolfi V, Pestana-Calsa MC, Belarmino LC, Benko-Iseppon AM, Crovella S (2010b) Characterization of a new defensin from cowpea (*Vigna unguiculata* (L.) Walp.). Protein Pept Lett 17:297–304. doi: 10.2174/092986610790780350

Pariani S, Contreras M, Rossi FR, Sander V, Corigliano MG, Simón F, Busi M V., Gomez-Casati DF, Pieckenstain FL, Duschak VG, Clemente M (2016) Characterization of a novel Kazal-type serine proteinase inhibitor of *Arabidopsis thaliana*. Biochimie 123:85–94. doi: 10.1016/j.biochi.2016.02.002

Parisi K, Shafee TMA, Quimbar P, Weerden NL Van Der, Bleackley MR, Anderson MA (2018) The evolution, function and mechanisms of action for plant defensins. Semin Cell Dev Biol. doi: 10.1016/j.semcdb.2018.02.004

Pérez A, Luque FJ, Orozco M (2007) Dynamics of B-DNA on the microsecond time scale. J Am Chem Soc 129: 14739-14745.doi: 10.1021/ja0753546

Pestana-Calsa MC, Ribeiro ILAC, Jr TC (2010) Bioinformatics-coupled molecular approaches for unravelling potential antimicrobial peptides coding genes in brazilian native and crop plant species. Curr Protein Pept Sci 199–209. doi: 10.2174/138920310791112138

Peters BM, Shirtliff ME, Jabra-Rizk MA (2010) Antimicrobial Peptides: Primeval Molecules or Future Drugs? PLoS Pathog 6: e1001067. doi: 10.1371/journal.ppat.1001067

Portieles R, Rodriguez M, Castillo J, Pujol M, Enriquez G, Borroto C, Trujillo L, Thomma BPHJ, Borras-Hidalgo O (2010) NmDef02, a novel antimicrobial gene isolated from *Nicotiana megalosiphon* confers high-level pathogen resistance under greenhouse and field conditions. Plant Biotechnol J 678–690. doi: 10.1111/j.1467-7652.2010.00501.x

Porto WF, Franco OL (2013) Theoretical structural insights into the snakin/GASA family. Peptides 44:163–7. doi: 10.1016/j.peptides.2013.03.014

Poth AG, Colgrave ML, Lyons RE, Daly NL, Craik DJ (2011) Discovery of an unusual biosynthetic origin for circular proteins in legumes. Proc Natl Acad Sci 108:10127–10132. doi: 10.1073/pnas.1103660108

Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci 106:1760–1765. doi: 10.1073/pnas.0813167106

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16:276-277.

Richero M, Barraco-Vega M, Cerdeiras MP, Cecchetto G (2013) Development of SCAR molecular markers for early and late differentiation of. Trees 27:249–257. doi: 10.1007/s00468-012-0792-6

Roe DR, Cheatham III TE (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 9:3084–3095. doi: 10.1021/ct400341p

Rong W, Qi L, Wang J, Du L, Xu H, Wang A, Zhang Z (2013) Expression of a potato antimicrobial peptide SN1 increases resistance to take-all pathogen *Gaeumannomyces graminis* var. tritici in transgenic wheat. Funct Integr Genomics 13:403–409. doi: 10.1007/s10142-013-0332-5

Roxrud I, Lid SE, Fletcher JC, Schmidt EDL, Opsahl-Sorteberg H-G (2007) GASA4, one of the 14-Member Arabidopsis GASA Family of Small Polypeptides, Regulates Flowering and Seed Development. Plant Cell Physiol 48:471–483. doi: 10.1093/pcp/pcm016

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (Eds.) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ. 365-386.

Ryckaert J-P, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341. doi: 10.1016/0021-9991(77)90098-5

Sagaram US, El-Mounadi K, Buchko GW, Berg HR, Kaur J, Pandurangi RS, Smith TJ, Shah DM (2013) Structural and functional studies of a phosphatidic acid-binding antifungal plant defensin MtDef4: Identification of an RGFRRR motif governing fungal cell entry. PLoS One 8:1–22. doi: 10.1371/journal.pone.0082485

Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh Ewald. J Chem Theory Comput 9:3878–3888. doi:10.1021/ct400314y

Savojardo C, Fariselli P, Alhamdoosh M, Martelli PL, Pierleoni A, Casadio R (2011) Improving the prediction of disulfide bonds in Eukaryotes with machine learning methods and protein subcellular localization. Bioinformatics 27:2224-2230. doi:10.1093/bioinformatics/btr387

Schägger G, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 379:368–379. doi: 10.1016/0003-2697(87)90587-2

Segura A, Moreno M, Madueño F, Molina A, García-Olmedo F (1999) Snakin-1, a peptide from potato that is active against plant pathogens. Mol Plant Microbe Interact 12:16–23. doi: 10.1094/MPMI.1999.12.1.16

Sels J, Mathys J, De Coninck BMA, Cammue BPA, De Bolle MFC (2008) Plant pathogenesis-related (PR) proteins: A focus on PR peptides. Plant Physiol Biochem 46:941–950. doi: 10.1016/j.plaphy.2008.06.011

Shafee TMA, Lay FT, Phan TK, Anderson MA, Hulett MD (2017) Convergent evolution of defensin sequence, structure and function. Cell Mol Life Sci 74:663–682. doi: 10.1007/s00018-016-2344-5

Shah DM, Read ND 2013 Antifungal peptides come of age. Fungal Biol Rev 26:107-108.

Shenkarev ZO, Gizatullina AK, Finkina EI, Alekseeva EA, Balandin S V., Mineev Konstantin S. Arseniev AS, Ovchinnikova T V (2014) Heterologous Expression and Solution Structure of Defensin from *Lentil lens*. Biochem Biophys Res Commun 451:252–257. doi: 10.1016/j.bbrc.2014.07.104

Shi Y, Mowery RA, Ashley J, Hentz M, Ramirez AJ, Bilgicer B, Slunt-brown H, Borchelt DR, Shaw BF (2012) Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. Protein Sci 21:1197–1209. doi: 10.1002/pro.2107

Silva ON, Mulder KCL, Barbosa AEAD, Otero-Gonzalez AJ, Lopez-Abarrategui C, Rezende TMB, Dias SC, Franco OL (2011) Exploring the pharmacological potential of promiscuous host-defense peptides: from natural screenings to biotechnological applications. Front Microbiol 2:1–14. doi: 10.3389/fmicb.2011.00232

Silverstein K a T, Graham M a, Paape TD, VandenBosch K a (2005) Genome organization of more than 300 defensin-like genes in Arabidopsis. Plant Physiol 138:600–610. doi: 10.1104/pp.105.060079

Silverstein KAT, Moskal WA, Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J 51:262–280. doi: 10.1111/j.1365-313X.2007.03136.x

Simoes CMO, Faikenberg M, Mentz LA, Schenkef EP, Amoros M, Girre L (1999) Antiviral activity of South Brazilian medicinal plant extracts. Phytomedicine 6:205–214. doi: 10.1016/S0944-7113(99)80010-5

Slavokhotova AA, Odintsova TI, Rogozhin EA, Musolyamov AK, Andreev YA, Grishin E V., Egorov TA (2011) Isolation, molecular cloning and antimicrobial activity of novel defensins from common chickweed (*Stellaria media* L.) seeds. Biochimie 93:450–456. doi: 10.1016/j.biochi.2010.10.019

Slavokhotova AA, Rogozhin EA, Musolyamov AK, Andreev YA, Oparin PB, Berkut AA, Vassilevski AA, Egorov TA, Grishin E V., Odintsova TI (2014) Novel antifungal α-hairpinin peptide from *Stellaria media* seeds: Structure, biosynthesis, gene structure and evolution. Plant Mol Biol 84:189–202. doi: 10.1007/s11103-013-0127-z

Slavokhotova AA, Shelenkov AA, Korostyleva T V., Rogozhin EA, Melnikova N V., Kudryavtseva A V., Odintsova TI (2017) Defense peptide repertoire of *Stellaria media* predicted by high throughput next generation sequencing. Biochimie 135:15–27. doi: 10.1016/j.biochi.2016.12.017

Slavokhotova AA, Shelenkov AA, Odintsova TI (2015) Prediction of *Leymus arenarius* (L.) antimicrobial peptides based on de novo transcriptome assembly. Plant Mol Biol 89:203–214. doi: 10.1007/s11103-015-0346-6

Smith DE, Dang LX (1994) Computer simulations of NaCl association in polarizable water. J Chem Phys 100:3757-3762. doi: 10.1063/1.466363

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S (2016) TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res 26:1134-1144.

Song X, Zhang M, Zhou Z, Gong W (2011) Ultra-high resolution crystal structure of a dimeric defensin SPE10. FEBS Lett 585:300–306. doi: 10.1016/j.febslet.2010.12.039

Spelbrink RG, Dilmac N, Allen A, Smith TJ, Shah DM, Hockerman GH (2004) Differential antifungal and calcium channel-blocking activity among structurally related plant defensins. Plant Physiol 135:2055–6.

Stec B (2006) Plant thionins-the structural perspective. Cell Mol Life Sci 63:1370–1385. doi: 10.1007/s00018-005-5574-5

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods research resource. Mol Biol Evol 28:2731–2739. doi: 10.1093/molbev/msr121

Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 60:523–533. doi: 10.1007/s00253-002-1158-6

Tesfaye M, Silverstein KAT, Nallu S, Wang L, Botanga CJ, Gomez SK, Costa LM, Harrison MJ, Samac DA, Glazebrook J, Katagiri F, Gutierrez-Marcos JF, VandenBosch KA (2013) Spatio-temporal expression patterns of *Arabidopsis thaliana* and *Medicago truncatula* defensin-like genes. PLoS One 8: doi: 10.1371/journal.pone.0058992

Thevissen K, de Mello Tavares P, Xu D, Blankenship J, Vandenbosch D, Idkowiak-Baldys J, Govaert G, Bink A, Rozental S, de Groot PW, Davis TR, Kumamoto CA, Vargas G, Nimrichter L, Coenye T, Mitchell A, Roemer T, Hannun YA, Cammue BP (2012) The plant defensin RsAFP2 induces cell wall stress, septin mislocalization and accumulation of ceramides in *Candida albicans*. Mol Microbiol. 84:166-80.

Thevissen K, Kristensen H, Thomma BPHJ, Cammue BPA, Francois IEJA (2007) Therapeutic potential of antifungal plant and insect defensins. Drug Discov Today 12:966–971. doi: 10.1016/j.drudis.2007.07.016

Thevissen K, Warnecke DC, Francois IE, Leipelt M, Heinz E, Ott C, Zahringer U, Thomma BP, Ferket KK, Cammue BP (2004) Defensins from insects and plants interact with fungal glucosylceramides. J Biol Chem 279: 3900-3905.

Thomma BP, Penninckx IA, Broekaert WF, Cammue BP (2001) The complexity of disease signaling in *Arabidopsis*. Curr Opin Immunol 13: 63–68.

Thomma BPHJ, Cammue BPA, Thevissen K (2002) Plant defensins. Planta 216:193–202. doi: 10.1007/s00425-002-0902-6

Tossi A, Sandri L (2002) Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. Curr Pharm Des 8:743–61. doi: 10.2174/1381612023395475

Trapalis M, Li SF, Parish RW (2017) The Arabidopsis GASA10 gene encodes a cell wall protein strongly expressed in developing anthers and seeds. Plant Sci 260:71–79. doi: 10.1016/j.plantsci.2017.04.003

Troncoso MF, Zolezzi P, Hellman U, Wolfenstein-todel C (2003) A novel trypsin inhibitor from *Peltophorum dubium* seeds, with lectin-like properties, triggers rat lymphoma cell apoptosis. Arch Biochem Biophys 411:93–104. doi: 10.1016/S0003-9861(02)00726-9

Tuppo L, Alessandri C, Pomponi D, Picone D, Tamburrini M, Ferrara R, Petriccione M, Mangone I, Palazzo P, Liso M, Giangrieco I, Crescenzo R, Bernardi M, Zennaro D, Helmer-Cittrerich M, Mari A, Ciardiello M. (2013) Peamaclein-a new peach allergenic protein: similarities, differences and misleading features compared to Pru p 3. Clin Exp Allergy 43:128–140. doi: 10.1111/cea.12028

Van Den Ent F, Löwe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. J Biochem Bioph Meth 67: 67-74.

Van der Weerden NL, Bleackley MR, Anderson MA (2013) Properties and mechanisms of action of naturally occurring antifungal peptides. Cell Mol Life Sci 70:3545-3570

Vandesompele J, De Preter K, Pattyn F, Poppe B, Roy N Van, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: Research 0034. doi: 10.1186/gb-2002-3-7-research0034

Vijayan S, Guruprasad L, Kirti PB (2008) Prokaryotic expression of a constitutively expressed *Tephrosia villosa* defensin and its potent antifungal activity. Appl Microbiol Biotechnol 80:1023–1032. doi: 10.1007/s00253-008-1648-2

Vila-perello M, Sanchez-Vallet A, Garcıa-Olmedo F, Molina A, Andreu D (2005) Structural dissection of a highly knotted peptide reveals minimal motif with antimicrobial activity. J Biol Chem 280:1661–1668. doi: 10.1074/jbc.M410577200

Vriens K, Cammue BP, Thevissen K (2014) Antifungal plant defensins: mechanisms of action and production. Molecules 19:12280-303

Vriens K, Cools TL, Harvey PJ, Craik DJ, Spincemaille P, Cassiman D, Braem A, Vleugels J, De Coninck B, Cammue BPA, Thevissen K (2015) Synergistic activity of the plant defensin HsAFP1 and caspofungin against *Candida albicans* biofilms and planktonic cultures. PloS one 10,e0132701.

Wang S, Rao P, Ye X (2009) Isolation and biochemical characterization of a novel leguminous defense peptide with antifungal and antiproliferative potency. Appl Microbiol Biotechnol 82:79–86. doi: 10.1007/s00253-008-1729-2

Webb B, Sali A (2014) Comparative protein structure modeling using modeller. Curr Protoc Bioinformatics, John Wiley & Sons, Inc., 5.6.1-5.6.32. doi:10.1002/0471250953.bi0506s15

Wintjens R, Rooman M (1996) Structural classification of hth dna-binding domains and protein–DNA interaction modes. J Mol Biol 262:294–313. doi: 10.1006/jmbi.1996.0514

Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80:1715-35. doi: 10.1002/prot.24065

Yang J, He BJ, Jang R, Zhang Y, Shen HB (2015) Accurate disulfide-bonding network predictions improve ab initio structure prediction of cysteine-rich proteins. Bioinformatics 31:3773-3781. doi: 10.1093/bioinformatics/btv459

Yeaman MR, Yount NY (2003) Mechanisms of antimicrobial peptide action and resistance. Pharmacol Rev 55:27–55. doi: 10.1124/pr.55.1.2

Yeung H, Squire CJ, Yosaatmadja Y, Panjikar S, López G, Molina A, Baker EN, Harris PWR, Brimble MA (2016) Radiation damage and racemic protein crystallography reveal the unique structure of the GASA/snakin protein superfamily. Angew Chemie Int Ed 55:7930–7933. doi: 10.1002/ange.201602719

Yount NY, Yeaman MR (2004) Multidimensional signatures in antimicrobial peptides. Proc Natl Acad Sci 101:7363–7368. doi: 10.1073/pnas.0401567101

Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature 415:389–395. doi: 10.1038/415389a

Zhang L, Simmons M (2006) Phylogeny and delimitation of the Celastrales inferred from nuclear and plastid genes. Syst Bot 31:122–137. doi: 10.1600/036364406775971778

Zhu J, Zhang L, Li W, Han S, Yang W, Qi L (2013) Reference gene selection for quantitative real-time pcr normalization in Caragana intermedia under different abiotic stress conditions. PLoS One 8:e53196. doi: 10.1371/journal.pone.0053196

Zou C, Lehti-shiu MD, Prakash T (2009) Evolutionary and expression signatures of pseudogenes. Plant Physiol 151:3–15. doi: 10.1104/pp.109.140632

Cysteine motifs used in Slavokhotova et al. (2015) for AMP prediction

Family name	Motif name	Motif formula[a]
Cys-rich	#N01	C(X{3,30})CX{3}C(X{3,30})C(X{3,30})CXC
cyclotide	CYC01	CX{3}CX{4}CX{6}CX{1}CX{4}C
cyclotide	CYC02	CX{3}CX{4,5}CX{4,6}CX{1}CX{4,5}C
Cys-rich	#N02	CX{8,11}CX{3,8}CX{3}CX{9,11}CX{4,8}CX{1}CX{3}C
Cys-rich	#N09	CX{9}CX{1}CX{25}CX{14}CX{11}C
Cys-rich	#N19	CX{3,5}CX{8,17}CX{4,6}C
Cys-rich	#N20	CX{4,6}CX{6}CX{3}CX{3,31}CX{3}CX{2,4}CX{3}
Cys-rich	#N21	CX{3}CX{11}CX{8}CX{3}CX{12}CX{33,37}CX{23,27}C
Cys-rich	#N22	CX{6,9}CX{22,33}CX{3}CX{9}CX{26,27}C
Cys-rich	#N23	CX{12}CX{21}CXCX{8}CX{21}C
Cys-rich	#N26	CX{6}CX{3}CX{12,14}CX{8}CXCX{19}CX{16,17}CX{2,3}CX{5}C
Cys-rich	#N27	CX{3,14}CX{4,5}CX{3}CX{8,11}CX{5,10}C
Cys-rich	#N29	CX{8}CX{12,13}CX{3}CX{3}CX{5}CX{8,9}C
Cys-rich	OUT01	CX{3}CX{17}CX{5}CX{6}C
Cys-rich	OUT02	CX{16}CX{3}CX{14}CX{5}CX{6}
Cys-rich	OUT03	CX{2}CXCX{2}CX{5}CX{3}CX{5}CX{3}CXCX{6}C
Cys-rich	OUT04	CX{24}CX{9}CXCX{3}C{3}CX{9}CX{28}CXC
Cys-rich	OUT05	CX{2}CXCX{11}CX{5}CX{2}CX{3}CX{11}CX{7}CXC
Cys-rich	OUT06	CX{5}CX{13}CX{6}CX{2}CX{4}CX{3}CX{7}C
Cys-rich	OUT07	CXCX{3}CX{13}CX{4}C
Cys-rich	OUT08	CX{2}CX{3}CX{24}CX{4} CX{14}CX{4}CX{5}C
Cys-rich	OUT09	CX{6}CX{2}CX{5}CX{15}CX{5}CX{2}C
Cys-rich	OUT10	CX{3}CX{5}CX{11}CX{5}CX{15}CX{11}C
Cys-rich	OUT11	CX{3}CX{17}CX{2}CX{4}CX{16}CX{6}C
Cys-rich	OUT12	CX{4}CX{5}CX{13}CX{4}CX{5}C
Cys-rich	OUT13	CX{7}CX{21}CX{3}CX{15}CX{4}CX{22}C
Cys-rich	OUT14	CX{3}CX{20}CX{3}CX{7}CX{3}C{4}CX{12}CX{3}CX{3}C
Cys-rich	OUT15	CX{3}CX{2}CX{3}CX{17}CX{2}C
Cys-rich	OUT16	CX{9}CX{3}CX{2}CX{3}C
Cys-rich	OUT17	CX{2}CX{8}CX{3}CX{4}CX{3}C
Cys-rich	OUT18	CX{3}CX{3}CX{4}CX{3}CX{4}CX{4}C
Cys-rich	OUT19	CX{3}CX{7}CX{3}CX{2}CX{9}C
Cys-rich	OUT20	CX{9}CX{15}CX{5}CX{10}CX{4}CX{12}CX{2}C
Cys-rich	OUT21	CX{5}CX{3}CX{7}CX{17}CX{1}C
Cys-rich	OUT22	CX{8}CX{13}CX{3}CX{10}CX{3}C
Cys-rich	OUT23	CX{1}CX{4}CX{3}CX{14}CX{6}CX{22}C
Cys-rich	OUT24	CX{1}CX{3}CX{12}CX{3}C
Cys-rich	OUT25	CX{5}CX{8}CX{5}CX{12}C
Cys-rich	OUT26	CX{4}CX{13}CX{5}CX{7}C

Cys-rich	OUT27	CX{5}CX{15}CX{5}CX{7}C
Cys-rich	OUT28	CX{4}CX{15}CX{5}CX{4}C
Cys-rich	OUT29	CX{4}CX{9}CX{4}CX{9}C
Cys-rich	OUT30	CX{2}CX{3}CX{24}CX{4}CX{14}CX{4}CX{5}C
Cys-rich	OUT31	CX{19}CX{19}CX{5}CX{1}CX{4}CX{10}CX{1}CX{1}CX{3}CX{3}CX{1}CX{3}CX{1}CX{5}CX{5}C
Cys-rich	OUT32	CX{3}CX{1}CX{3}CX{2}CX{9}CX{3}CX{7}CX{32}C
Cys-rich	OUT33	CX{19}CX{5}CX{1}CX{4}CX{10}CX{1}CX{1}CX{3}CX{3}CX{1}CX{3}CX{1}CX{5}CX{5}C
Cys-rich	OUT34	CX{3}CX{2}CX{3}CX{1}CX{2}CX{1}CX{10}CX{3}CX{5}C
Cys-rich	OUT35	CX{3}CX{20}CX{3}CX{7}CX{3}CX{4}CX{12}CX{3}CX{3}C
Cys-rich	OUT36	CX{4}CX{3}CX{30}CX{3}CXC
DEFL	DEF01	CX{8}CX{13}CX{3}CX{5}CX{3}CX{11}CX{6}CCC
DEFL	DEF02	CX{17}CX{11}CX{5}CX{3}CX{9}CX{8}CCCX{3}CX{12}C
DEFL	DEF03	CX{8}CX{3}CX{3}CX{3}CX{9}CCC
DEFL	DEF04	CX{7}CX{4}CX{3}CX{12}CX{5}CCC
DEFL	DEF05	CX{10}CX{5}CX{3}CX{9}CX{8}CX{1}CX{3}C
DEFL	DEF06	CX{10}CX{5}CX{3}CX{9}CX{6}CX{1}CX{3}C
DEFL	DEF07	CX{10}CX{5}CX{3}CX{9}CX{3}CX{4}CX{1}CX{3}C
DEFL	DEF08	CX{10}CX{9}CX{4}CX{3}CX{15}CX{10}CX{1}CX{3}CX{7}C
DEFL	DEF09	CX{11}CX{4}CX{3}CX{9}CX{3}CXCXCX{3}C
DEFL	DEF10	CX{10}CX{4}CX{3}CX{11}CX{4}CXCX{2}C
DEFL	DEF11	CX{10}CX{4}CX{3}CX{11}CX{4}CXC
DEFL	DEF12	CX{10}CX{4}CX{3}CX{11}CX{4}CX{1}CX{2}C
DEFL	DEF13	CX{8}CX{4}CX{3}CX{11}CX{5}CX{1}CX{2}C
DEFL	DEF14	CX{4}CX{3}CX{12}CX{7}CXCX{2}C
DEFL	DEF15	CX{8}CV{4}CX{3}CX{17}CXCX{2}C
DEFL	DEF16	CX{5}CX{9}CX{5}CX{1}CX{2}C
DEFL	DEF17	CX{3}CXCX{2}CX{5}CX{3}CX{5}CX{3}CXCX{6}C
DEFL	DEF18	CX{3}CX{2}CX{3}CX{1}CC
DEFL	DEF19	CX{3}CX{2}CX{3}CX{6}CX{9}CC
DEFL	DEF20	CX{4}CX{11}CX{5}CX{8}CX{5}CX{1}CC
DEFL	DEF21	CX{4}CX{11}CX{5}CX{8}CX{4}CXCC
DEFL	DEF22	CX{3}CX{10}CX{3}CC
DEFL	DEF23	CX{2}CX{3}CX{6}CX{10}CXCCX{3}C
DEFL	DEF24	CX{2}CX{7}CXCX{X}CX{5}CX{10}CX{6}CX{4}CX{2}CXCC
DEFL	DEF25	CX{1}CX{3}CX{16}CX{6}CX{1}C
DEFL	DEF26	CXCX{3}CX{15}CX{6}CXC
DEFL	DEF27	CX{3}CX{6}CX{3}CX{8}CX{1}C
DEFL	DEF28	CX{4}CX{13}CX{4}CX{2}CX{3}CX{11}CX{7}CXC
DEFL	DEF29	CX{9}CX{4}CX{9}CX{4}CX{4}CX{5}CXC
DEFL	DEF30	CX{5}CX{3}CX{1}CX{3}CX{1}CX{5}CCCX{1}CX{3}CX{5}CX{1}CX{1}CX{3}CX{1}CX{3}C
DEFL	DEF31	CX{8,11}CX{3,8}CX{3}CX{9,11CX{4,8}CX{1}C{3}C
DEFL	DEF32	CX{4,25}CX{2,12}CX{3,4}CX{3,17}CX{4,32}CXCX{1,6}C
DEFL	DEF33	CX{3,21}CX{2,12}CX{3,4}CX{3,15}CX{4,23}CCC
DEFL	DEF34	CX{2,14}CX{3,5}CX{3,16}CX{4,28}CXC
DEFL	DEF35	CX{5,10}CX{4,6}CX{3}CX{9,15}CX{5,12}CX{1}CX{3}C
DEFL	DEF36	CX{3,14}CX{4,5}CX{3}CX{8,11}CX{5,10}CCC
hairpin	4C01	CX{3}CX{6,13}CX{3}C

hairpin	4C02	CX{3}CX{6,13}CX{3}CX{3}C
hairpin	4C03	CX{3}CX{1,20}CX{3}C
hairpin	4C04	CX{3}CX{1,20}CX{3}CX{3}C
hevein-like	HEV001	CX{12}CCX{5}CX{6}CX{3}CXCX{3}C
hevein-like	HEVHIP01	CX{4,5}CX{4}CCX{5}CX{6}C
hevein-like	HEVHIP02	CX{3,8}CX{4}CCX{5}CX{6}CX{3,5}CX{1,3}C
hevein-like	HEVHIP03	CX{6}CX{8}CCX{3}CX{10}C
hevein-like	HEVHIP04	CX{6}CX{7,8}CCX{3}CX{10}C
hevein-like	HEVHIP05	CX{12}CCX{5}CX{6}CX{3}C
LTP	LTP001	CX{9}CX{16}CCX{12}CXC{25}CX{9}C
LTP	LTP002	CX{9}CX{14}CCX{12}CXC{29}CX{9}C
LTP	LTP003	CX{9}CX{16}CCX{12}CXC{26}CX{8}C
LTP	LTP004	CX{9}CX{14}CCX{19}CXC{22}CX{13}C
LTP	LTP005	CX{9}CX{16}CCX{14}CXC{23}CX{9}C
LTP	LTP006	CX{9}CX{16}CCX{14}CXC{22}CX{9}C
LTP	LTP007	CX{9}CX{19}CCX{9}CXC{24}CX{7}C
LTP	LTP008	CX{12}CX{9}CX{20}CCX{13}CXC{24}CX{9}C
LTP	LTP009	CX{15}CX{16}CCX{9}CX{21}C
LTP	LTP010	CX{10}CX{3}CX{12}CCX{9}CX{23}C
LTP	LTP011	CX{14}CX{14}CCX{3}CX{11}CXCX{24}C{10}C
LTP	LTP012	CX{9}CX{14}CX{19}CXC{22}C{13}C
LTP	LTP013	CX{9}CX{14}CX{19}CCX{19}CXC{22}CX{14}C
LTP	LTP014	CX{9}CX{16}CCX{12}CXC{24}CX{9}C
LTP	LTP015	CX{7}CX{12}CCX{8}CXC{23}CX{6}C
LTP	LTP016	CX{7,9}CX{12,14}CCX{8,19}CX{1}CX{19,23}CX{13,15}C
LTP	LTP017	CX{10}CX{3}CX{12}CCX{9}CX{1}CX{21}CX{6}C
LTP	LTP018	CX{14}CX{14}CCX{11}CX{1}CX{24}CX{10}C
LTP	LTP019	CX{9}CX{8}CCX{8}CXC{9}CX{5}CX{15}CX{5}C
LTP	LTP020	CX{9}CX{13}CCX{19}CX{1}CX{23}CX{13}C
LTP	LTP021	CX{9}CX{8}CCX{8}CX{1}CX{9}CX{5}CX{15}CX{5}C
LTP	LTP022	CX{9}CX{8}CCX{7}CX{1}CX{9}CX{5}CX{17}CX{5}C
LTP	LTP023	CX{4}CX{3}CX{3}CX{3}CCX{13}C
LTP	LTP024	CX{3}CCX{4}CXCX{9}CX{5}C
LTP	LTP025	CX{8}CX{2}CCX{7}CX{2}CX{8}CX{3}CX{9}CX{4}CX{13}C
LTP	LTP026	CX{4}CCX{3}CX{9}CX{4}C
LTP	LTP027	CX{18}CCX{3}CX{14}CX{4}C
LTP	LTP028	CX{5}CCX{3}CX{4}CX{3}C
LTP	LTP029	CX{2}CCX{3}CX{2}CX{6}CX{3}CX{7}CX{2}CX{7}CX{3}CX{3}CX{3}CX{4}C
snakin	SNA01	CX{3}CX{3}CX{9}CX{3}CX{2}CCX{2}CX{1}CX{11}CX{1}CX{14}C
snakin	SNA02	CX{3}CX{3}CX{8}CX{3}CX{2}CCX{2}CX{1}CX{11}CX{1}CX{12}C
thionin	THI001	CCCX{6}CX{3}CX{9}CX{3}CXCX{5}CX{3}CX{17}CX{3}CX{4}C
thionin	THI002	CCCX{6}CX{3}CX{9}CX{3}CXCX{5}CX{3}CX{17}CX{3}CX{4}CX{5}C
thionin	THI003	CCCX{6}CX{3}CX{9}CX{3}CXCX{5}CX{3}CX{17}CX{3}CX{4}CX{18}CX{3}CX{3}C
thionin	THI004	CCX{7}CX{3}CX{9}CX{3}CXCX{7}CX{21}CX{3}CX{4}CX{18}CX{3}CX{3}C
thionin	THI005	CCX{7}CX{3}CX{9}CX{3}CXCX{7}CX{21}CX{3}CX{4}C
thionin	THI006	CCX{5}CX{12}CX{5}CXC
thionin	THI007	CCX{3}CX{10}CX{6}C

thionin	THI008	CC{11}CX{13}C{3}CX{2}CX{19}CXC
thionin	THI009	CCX{16}CX{18}C{3}CX{11}CX{3}CXC
thionin	THI010	CCX{5}CX{14}CX{5}C
thionin	THI011	CCX{5}CX{14}CX{6}CX{36}C
thionin	THI012	CCX{16}CX{18}CX{3}CX{11}CX{3}CX{1}C
thionin	THI013	CCX{2}CX{4}CX{2}CX{3}CX{8}CX{3}CCC
thionin	THI014	CCX{31}CX{4}CX{15}CX{5}CX{10}C
thionin	THI015	CCX{4}CX{20}CX{5}CX{13}CX{6}CX{2}CX{4}CX{3}CX{7}C
thionin	THI016	CCX{6}CCX{3}CX{6}CX{3}CX{27}CX{5}CX{5}C
thionin	THI017	CCX{7,8}CX{3}CX{8,14}CX{3}CX{1}CX{6,8}C
thionin	THI018	CCX{11}CX{9,15}CX{5}CX{6,11}C
thionin	THI019	CCX{8}CX{3}C
thionin	THI020	CCX{8}CX{12,13}CX{3}CX{3}CX{5}CX{8,9}CHCYC

[a] X is any amino acid except cysteine and numbers in curly brackets reflect an allowed range for the number of amino acids