



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Trabajo de Tesis Doctoral

***Estudio de peptidasas aspárticas de flores de Asteraceae
con potencial aplicación biotecnológica.***

Tesista María Laura Colombo

Directora: Sandra E. Vairo Cavalli

Año 2016



El presente trabajo de tesis para optar por el grado de Doctor de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata, ha sido realizado bajo la dirección de la Dra. Sandra E. Vairo Cavalli en el Centro de Investigaciones de Proteínas Vegetales (CIProVe) de la Facultad de Ciencias Exactas, UNLP. El mismo fue financiado por el Consejo Nacional de Investigaciones Científicas y Técnicas y por el Programa Erasmus Mundus.

La Plata, abril de 2016

“Siempre hay que saber cuándo una etapa llega a su fin. Cerrando ciclos, cerrando puertas, terminando capítulos; no importa el nombre que le demos, lo que importa es dejar en el pasado los momentos de la vida que ya se han acabado”

—Paulo Coelho

A mis amores, Salvador y Gonzalo

Agradecimientos



Este trabajo de Tesis ha significado mucho para mí, no sólo en el aspecto profesional sino también por el crecimiento personal por el cual he transitado con el mismo. Por ello, no quiero dejar de agradecer a cada una de las personas que han sido parte de estos años, de este trabajo...

Mi mayor agradecimiento a la Dra. Sandra Vairo Cavalli, no solo por la oportunidad de realizar la tesis, por confiar en mí durante todos estos años, por guiarme y aconsejarme, mostrando siempre disposición y entusiasmo, por su dirección y apoyo incondicional. Por la tolerancia y la comprensión, por enseñarme y ayudarme, por su optimismo, tranquilidad y paciencia (¡¡¡sobre todo en estos últimos días!!!).

A mis queridos compañeros CIPROVIANOS con quienes en el día a día hemos compartido alegrías y nuestras penas, nuestros logros y nuestras frustraciones, almuerzos, meriendas, cumpleaños, casamientos y porque no asados en la puerta del laboratorio... pero sobretodo por haber compartido gran parte del día, por hacer del laboratorio una segunda casa, ya que nunca fueron 8 hs de trabajo sino empezar y no saber cuándo se termina, gracias por compartir no sólo los días semanales sino también fines y porque no feriados. Pero también quiero darles las gracias a todos los que compartieron un mate, una charla, un consejo profesional como personal, por continuarme un ensayo cuando no llegaba porque me tenía que ir corriendo a buscar a Salvi. Por la paciencia que me tuvieron en estos últimos meses, por escucharme, por ayudarme en la mesada y a no bajar los brazos. Gracia al “profe” Caffini y a Nora por brindarme su apoyo durante todos estos años.

Pero en este GRACIAS, quiero agradecer a especialmente a Constanza, Agustina y a Mariela que en estos últimos meses me ayudaron a la par mía. Por nunca decirme que no, porque siempre estuvieron cuando las necesite, porque cuando les decía “...este es el último ensayo” y no lo era siempre estuvieron dispuestas a hacerlo cuantas veces más fuera necesario.

No quiero olvidarme de todos aquellos que ayudaron a remar desde afuera, por las tantas veces que aportaron reactivos, plantas, equipos, experiencias, y hasta alegrías o tristezas por los resultados obtenidos, muchas veces mates por medio. La gente del IBBM Gabriel; Belén, Matías, Agustín; del CIDCA, Vanesa, Flavia, Carolina; del INFIVE Dana, Lucía, Mario, Pedro y gracias a tantos otros que seguramente me han visto dando vueltas por los pasillos. Al Dr. Pablo Tornero por guiarme, enseñarme y hacer comfortable mi estadía en el IBMCP.

Pero quiero darle las gracias a alguien muy especial en mi vida, porque sin su ayuda y sin su apoyo nada de esto hubiera sido posible. Gracias Gonza, por ayudarme a no bajar los brazos, porque cuando llegaba triste porque los ensayos no daban o porque tenía una discusión o algún problema en el laboratorio siempre estuviste dispuesto a escucharme. Gracias por apoyarme y aconsejarme del mejor modo, por enseñarme “a ver la mitad del vaso lleno y no la mitad del vaso vacío”. Gracias por haberme acompañado a España, cuando tenías muchos compromisos en Argentina. Gracias por acompañarme un sábado, un domingo o feriado a continuar un ensayo, por esperarme aún cuando le decía que eran 5 minutos y pasaban horas, por ayudarme a cuidar a Salvi cuando trabajaba, por dejar de trabajar para que yo pueda terminar mis ensayos. Por tratar de comprender porque mi trabajo no se terminaba cuando cerraba en la puerta del laboratorio, sino que continuaba en casa con la computadora... ¡¡¡gracias por estar presente incondicionalmente en los momentos más difíciles y por enseñarme entre tantas cosas, lo básico de una vida, que no todo es trabajo!!!...¡¡¡gracias por estar, por quererme y cuidarme. Gracias a mi solcito, Salvador, también por hacerme tan feliz. Gracias a los dos por la paciencia que me tuvieron todo este tiempo!!!!

Finalmente, mi reconocimiento al CIProVe y sus autoridades, por brindarme el ámbito para realizar parte de este trabajo, al CONICET y sus autoridades por otorgarme las becas para el desarrollo de esta tesis doctoral y para la futura formación posdoctoral, y a la Facultad de Ciencias Exactas y la Universidad Nacional de La Plata, por brindarme la formación de grado y de posgrado.

Índice



Índice

| | |
|--|----|
| 1. INTRODUCCIÓN | 1 |
| 1.1. PEPTIDASAS..... | 1 |
| 1.1.1. Importancia..... | 1 |
| 1.1.2. Nomenclatura y clasificación de las peptidasas | 2 |
| 1.1.3. Especificidad de secuencia mediada por especificidad de subsitios | 11 |
| 1.1.4. Mecanismo catalítico de las peptidasas aspárticas..... | 13 |
| 1.1.5. Distribución de peptidasas | 14 |
| 1.1.6. Peptidasas de plantas..... | 15 |
| 1.1.6.1. Peptidasas aspárticas de plantas | 16 |
| 1.1.7. Aplicaciones biotecnológicas de las PAs | 18 |
| 1.1.7.1. Aplicaciones de las peptidasas aspárticas de plantas | 20 |
| 1.1.8. Organización estructural de las PAs vegetales..... | 21 |
| 1.1.8.1. Organización de la estructura primaria | 21 |
| 1.1.8.2. Organización de la estructura terciaria..... | 25 |
| 1.1.9. Rol fisiológico de las peptidasas vegetales..... | 29 |
| 1.1.10. Mecanismos de procesamiento de peptidasas aspárticas | 29 |
| 1.1.11. Peptidasas aspárticas en flores de cardo | 33 |
| 1.2. DESCRIPCIÓN GENERAL DE LA VÍA SECRETORIA EN PLANTAS..... | 34 |
| 1.2.1. Componentes de la vía secretoria | 35 |
| 1.2.2. Tipos de vacuolas..... | 39 |
| 1.2.3. Señales de direccionamiento vacuolar..... | 40 |
| 2. OBJETIVOS | 42 |
| 3. MATERIALES | 45 |
| 3.1. CEPAS BACTERIANAS..... | 46 |
| 3.1.1. Cepas de <i>Escherichia coli</i> | 46 |
| 3.1.2. Cepas de <i>Agrobacterium tumefaciens</i> | 47 |
| 3.2. MEDIOS DE CULTIVO..... | 47 |
| 3.3. SOLUCIONES BUFFERS | 48 |
| 3.4. SOLUCIONES DE ANTIBIÓTICO | 49 |
| 3.5. VECTORES | 50 |
| 3.5.1. Clonación | 50 |
| 3.5.1.1. pGEM®-T Easy | 50 |
| 3.5.1.2. pDONR 222..... | 51 |
| 3.5.2. Expresión heteróloga en sistema eucariota | 52 |
| 3.5.2.1. pB7FWG2..... | 53 |
| 3.5.2.2. pB7FRWG2..... | 54 |
| 3.6. CEBADORES | 55 |

| | | |
|-----------|--|----|
| 3.7. | ENZIMAS..... | 55 |
| 3.7.1. | Enzimas de restricción..... | 56 |
| 3.8. | MATERIAL VEGETAL..... | 56 |
| 3.9. | HERRAMIENTAS BIOINFORMÁTICAS..... | 57 |
| 3.9.1. | Herramientas generales..... | 57 |
| 3.9.2. | Herramientas empleadas para el modelado molecular por homología..... | 59 |
| 3.9.3. | Herramientas empleadas para el modelado molecular por homología análisis filogenético..... | 60 |
| 4. | MÉTODOS..... | 62 |
| 4.1. | CLONACIÓN DE PEPTIDASAS DE FLORES..... | 62 |
| 4.1.1. | Extracción de ARN total de plantas..... | 62 |
| 4.1.2. | Determinación de la concentración, pureza del ARN total y ADN..... | 63 |
| 4.1.3. | Electroforesis en geles de agarosa..... | 64 |
| 4.1.4. | Purificación de ADN a partir de gel usando un kit comercial..... | 64 |
| 4.1.5. | Amplificación de ácidos nucleicos..... | 65 |
| 4.1.5.1. | Síntesis de ADNc..... | 65 |
| 4.1.5.2. | Amplificación por PCR..... | 66 |
| 4.1.6. | Ligación..... | 67 |
| 4.1.7. | Digestión enzimática de ácidos nucleicos..... | 68 |
| 4.1.8. | Secuenciación y análisis de datos..... | 69 |
| 4.1.9. | Crecimiento y mantenimiento de cultivos bacterianos..... | 69 |
| 4.1.10. | Preparación de células competentes químicas..... | 70 |
| 4.1.10.1. | Método de cloruro de calcio..... | 70 |
| 4.1.10.2. | Método de cloruro de rubidio..... | 70 |
| 4.1.11. | Preparación de células electrocompetentes..... | 71 |
| 4.1.12. | Transformación de bacterias competentes químicas por shock térmico..... | 71 |
| 4.1.13. | Transformación de bacterias electrocompetentes por electroporación..... | 72 |
| 4.1.14. | Amplificación y preparación de los plásmidos..... | 72 |
| 4.1.15. | Extracción de ADN plasmídico..... | 73 |
| 4.1.15.1. | Método de lisis alcalina..... | 73 |
| 4.1.15.2. | Tecnología de membranas a base de sílice..... | 73 |
| 4.2. | ESTUDIO DE LOCALIZACIÓN DE PEPTIDASAS DE <i>Silybum marianum</i> y SUS MUTANTES POR EXPRESIÓN TRANSIENTE EN <i>Nicotiana benthamiana</i> | 74 |
| 4.2.1. | Obtención de construcciones para expresión de proteínas de fusión wt y mutantes..... | 74 |
| 4.2.2. | Reacción de recombinación con BP clonasa™ (vía lisogénica)..... | 74 |
| 4.2.3. | Obtención de los mutantes de preprosilpepsinas..... | 76 |
| 4.2.4. | Reacción de recombinación con LR clonasa™ (vía lítica)..... | 80 |
| 4.2.5. | Obtención de cepas <i>Agrobacterium tumefaciens</i> recombinantes..... | 81 |
| 4.2.5.1. | Obtención de <i>Agrobacterium tumefaciens</i> competentes y electrocompetentes..... | 81 |

| | | |
|----------|--|-----|
| 4.2.5.2. | Transformación <i>A. tumefaciens</i> por shock térmico..... | 82 |
| 4.2.5.3. | Transformación <i>A. tumefaciens</i> por electroporación..... | 82 |
| 4.2.6. | Germinación y mantenimiento de plantas de <i>Nicotiana benthamiana</i> | 83 |
| 4.2.7. | Infiltración con <i>A. tumefaciens</i> y preparación de las muestras vegetales para microscopía..... | 83 |
| 4.2.8. | Microscopía confocal..... | 85 |
| 4.2.9. | Estudio bioinformático de peptidasas clonadas de <i>S. marianum</i> | 86 |
| 4.2.9.1. | Búsqueda en base de datos..... | 86 |
| 4.2.9.2. | Análisis de los motivos conservados..... | 86 |
| 4.2.9.3. | Análisis estructural y funcional..... | 86 |
| 4.2.9.4. | Modelado molecular por homología: predicción de estructura terciaria..... | 87 |
| 4.2.9.5. | Evaluación y validación del modelo obtenido..... | 88 |
| 5. | RESULTADOS..... | 92 |
| 5.1. | FUENTE DE PEPTIDASAS..... | 92 |
| 5.2. | CLONACIÓN DE ASPARTILPEPTIDASAS DE FLORES..... | 93 |
| 5.3. | ANÁLISIS BIOINFORMÁTICO..... | 95 |
| 5.3.1. | Búsqueda de motivos conservados..... | 101 |
| 5.3.2. | Búsqueda de dominios conservados..... | 107 |
| 5.3.3. | Búsqueda de las peptidasas homólogas con mayor grado de similitud con las preprosilpepsinas..... | 110 |
| 5.3.3.1. | Comparación de los precursores de peptidasas clonadas con sus homólogos más cercanas..... | 112 |
| 5.3.4. | Organización de la estructura primaria..... | 112 |
| 5.3.4.1. | Comparación de los precursores de las peptidasas clonadas entre sí..... | 116 |
| 5.3.5. | Predicción de la estructura secundaria..... | 117 |
| 5.3.6. | Predicción de estructura terciaria, evaluación y validación del modelo..... | 118 |
| 5.3.7. | Evaluación del modelo tridimensional obtenido..... | 128 |
| 5.3.8. | Predicción funcional..... | 133 |
| 5.4. | ANÁLISIS FILOGENÉTICO..... | 136 |
| 5.5. | Localización de una peptidasa de <i>S. marianum</i> en <i>Nicotiana benthamiana</i> | 140 |
| 5.5.1. | Obtención de las construcciones para la expresión de preprosilpepsina 2..... | 143 |
| 5.5.2. | Agroinfiltración de hojas de <i>N. benthamiana</i> , expresión transitoria y microscopía confocal..... | 152 |
| 6. | DISCUSIÓN Y CONCLUSIONES..... | 170 |
| 7.1. | ANEXO I..... | 185 |
| 7.2. | ANEXO II..... | 200 |
| 8. | ABREVIATURAS..... | 207 |
| 9. | BIBLIOGRAFÍA..... | 209 |



Introducción

1.1. PEPTIDASAS

1.1.1. Importancia

Las peptidasas son proteínas catalíticamente activas que escinden enlaces peptídicos por hidrólisis de proteínas y péptidos (Rawlings & Barrett, 2014).

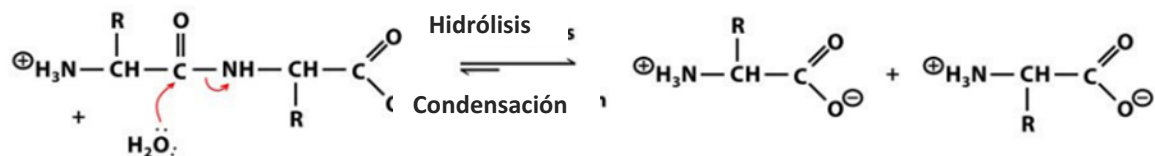


Figura 1. Mecanismo hidrolítico de las peptidasas

Las peptidasas son probablemente una de las más antiguas familias de enzimas. Casi todos los organismos, incluyendo virus, tiene al menos una peptidasa (Salvesen, Hempel, & Coll, 2015). Estos catalizadores biológicos llevan a cabo el desensamblaje de las moléculas de proteínas necesarias para crear el pool de aminoácidos libres para la síntesis de nuevas proteínas. Esta actividad es tal vez más familiar en el contexto de la digestión de las proteínas de los alimentos en el tracto digestivo de los animales, pero de hecho las peptidasas han evolucionado para realizar innumerables funciones vitales dentro y fuera de las células. Tal vez la más importante de estas funciones es la de procesar proteínas haciendo unos pocos clivajes específicos, con frecuencia sólo uno. De este modo por proteólisis algunas proteínas son enviadas a la ubicación subcelular correcta o son secretadas; otras proteínas se activan a partir de su precursor latente, lo que es especialmente importante para las enzimas; se liberan uno o más péptidos biológicamente activos a partir de algunas proteínas; mientras que otras proteínas se desactivan. Esta diversificación de funciones ha llevado a una situación en la que los genes para peptidasas representan entre el 2% y el 5% del total de genes en todo tipo de organismos (Rawlings & Barrett, 2014)

El genoma humano contiene más de 550 genes de peptidasas, los más abundantes de los cuales corresponden a metalo, cisteín y serinproteasas representados por 191, 178 y 161 genes, respectivamente. Las proteasas treonínicas y aspárticas son de abundancia relativamente baja con sólo 27 y 21 genes, respectivamente. Dada la diversidad estructural y funcional, las proteasas llevan a cabo una amplia gama de funciones críticas. La diversificación de las funciones biológicas de las proteasas deriva de la evolución estructural, que converge con geometrías de sitios activos similares y con variados motivos de reconocimiento de sustrato. Existe una gran variedad de proteasas utilizadas comercialmente que van desde aditivos en detergentes hasta eficaces terapéuticos. Desde el punto de vista de la industria, las proteasas constituyen el grupo de enzimas con mayor relevancia en el mercado mundial, representando cerca del 60% del total de las enzimas comercializadas (Qing, Yi, Marek, & Iverson, 2013). Debido a que la proteólisis cambia las propiedades químicas, físicas, biológicas e inmunológicas de las proteínas, las enzimas proteolíticas cumplen un rol muy importante en biotecnología (González-Rábade, Badillo-Corona, Aranda-Barradas, & Oliver-Salvador, 2011), siendo muy diversos los campos de aplicación: tecnología alimentaria, industria farmacéutica y manufacturas de detergente, entre otras (Feijoo-Siota & Villa, 2011). La gran diversidad de los sitios activos de las peptidasas permite que presenten diferentes grados de selectividad y especificidad por el sustrato, por lo tanto estas enzimas son un marco prometedor para la ingeniería de nuevas actividades (Qing et al., 2013).

1.1.2. Nomenclatura y clasificación de las peptidasas

Existe una confusa redundancia en la terminología utilizada en esta área de la bioquímica. Los términos "proteasa" y "peptidasa" son esencialmente sinónimos. Proteasa es el término más familiar, pero peptidasa es estrictamente más correcto

y es el término recomendado en la nomenclatura enzimática de la Comisión de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (NC-IUBMB), y por lo tanto es el término que se utilizará en este trabajo de tesis. Una peptidasa es también una "enzima proteolítica", aunque los términos no son sinónimos porque hay enzimas proteolíticas que no son hidrolasas, como aquellas proteínas que se auto-procesan utilizando la capacidad de la asparagina o glutamina para ciclarse y formar una succinimida que conduce a la ruptura de un enlace peptídico (Rawlings & Barrett, 2014).

Debido a la variedad en su comportamiento catalítico, las peptidasas son difíciles de clasificar. Siguiendo las reglas establecidas para el resto de las enzimas se utilizan tres criterios para su clasificación:

a) *Naturaleza del sitio activo.* Hartley (1960) estableció una clasificación de peptidasas de acuerdo con el tipo de sitio catalítico. En esta clasificación las peptidasas fueron divididas en seis clases mecánicas: serin-endopeptidasas (EC 3.4.21); cistein-endopeptidasas, anteriormente llamadas tiol-proteasas (EC 3.4.22); aspartil-endopeptidasas, conocidos como proteasas ácidas; endopeptidasas glutámicas (EC 3.4.23); metaloendopeptidasas (EC 3.4.24); y treonin-endopeptidasas (EC 3.4.25). Un quinto grupo incluye peptidasas con un modo de acción no identificado (EC 3.4.99). Serin, treonin, y cisteín-peptidasas son catalíticamente muy diferentes a las aspartil y metalopeptidasas. En el primer grupo el nucleófilo del sitio catalítico es parte de un aminoácido, mientras que para el segundo grupo lo es una molécula de agua activada.

b) *Reacción catalizada.* En agosto de 1955 el NC-IUBMB crea la Comisión Internacional de Enzimas (Enzymes Commission, EC) que aborda el estudio, los criterios y reglas para la clasificación de enzimas. Desde 1964 la EC clasifica los diversos tipos de enzimas en seis grandes grupos: EC 1: Oxidorreductasas,

EC 2: Transferasas, EC 3: Hidrolasas, EC 4: Liasas, EC 5: Isomerasas y EC 6: Ligasas.

Dentro de este sistema el número 3.4 le pertenece a las peptidasas que actúan sobre enlaces peptídicos, dentro de la clase 3 de las enzimas hidrolíticas o hidrolasas. Las peptidasas en la lista EC se dividen en 14 subclases (Tabla 1).

| Subclase | Tipo de peptidasa |
|----------|--|
| 3.4.11 | Aminopeptidasas |
| 3.4.13 | Dipeptidasas |
| 3.4.14 | Dipeptidil-Peptidasas y Tripeptidil-Peptidasas |
| 3.4.15 | Peptidil-Dipeptidasas |
| 3.4.16 | Carboxipeptidasas tipo serínico |
| 3.4.17 | Metalocarboxipeptidasas |
| 3.4.18 | Carboxipeptidasas tipo cisteínico |
| 3.4.19 | Omega peptidasas |
| 3.4.21 | Endopeptidasas serínicas |
| 3.4.22 | Endopeptidasas cisteínicas |
| 3.4.23 | Endopeptidasas aspárticas |
| 3.4.24 | Endopeptidasas metálicas |
| 3.4.25 | Endopeptidasas treonínicas |
| 3.4.99 | Endopeptidasas de tipo no conocido |

Tabla 1. Sistema EC de clasificación de peptidasas

Se reconocen dos grupos principales en función de su capacidad de hidrolizar enlaces peptídicos internos o situados en los extremos. Las exopeptidasas (3.4.11-19) son enzimas que hidrolizan enlaces peptídicos preferentemente o sólo si existe un grupo terminal libre, mientras que las endopeptidasas (3.4.21-25

y 3.4.99) actúan sólo lejos del final de las cadenas polipeptídicas (Fig. 2, URL: <http://www.chem.qmul.ac.uk/iubmb/enzyme/EC34>).

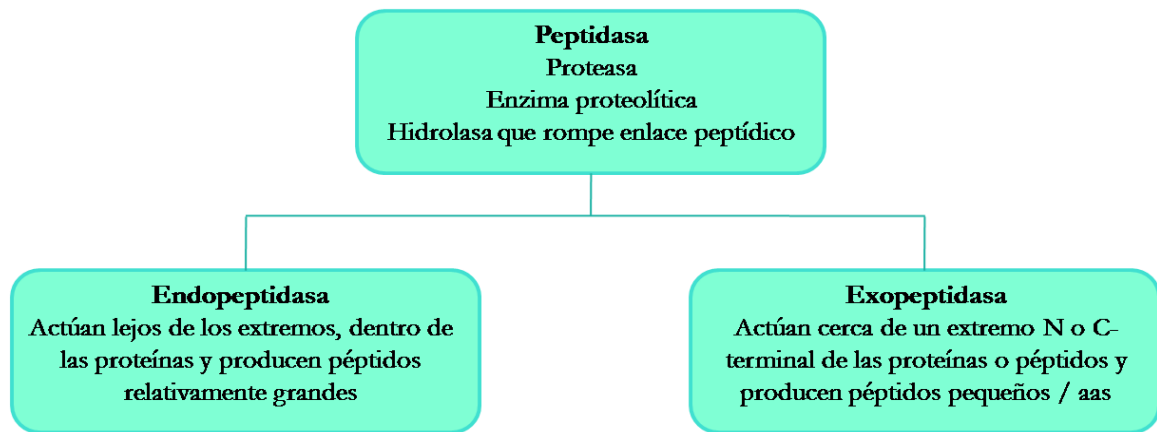


Figura 2. Diferentes tipos de actividad peptidásica

c) *Relaciones evolutivas basadas en la estructura.* Este sistema es el más moderno y organizado, correlaciona la entidad individual con la jerarquía evolutiva, clasificando las peptidasas en "familias" de proteínas con secuencias similares (Barrett & Rawlings, 1991). Este esquema fue desarrollado para dar estructura a la base de datos MEROPS que utiliza una clasificación basada en la estructura jerárquica de las peptidasas que el sistema EC no tiene en cuenta (Barrett, Rawlings, & O'Brien, 2001; Rawlings & Barrett, 1999; Rawlings, Tolle, & Barrett, 2004). El sistema MEROPS de clasificación tiene en cuenta los conceptos de tipo catalítico, clan, familia y peptidasa (Fig. 3). Cada peptidasa se asigna a una familia en base de similitudes en la secuencia de aminoácidos y las familias que son homólogas se agrupan luego en clanes.

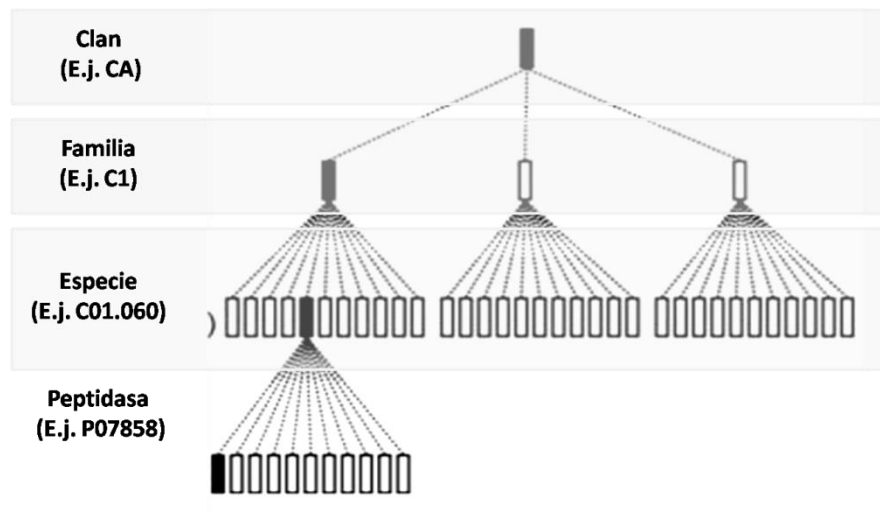


Figura 3. Sistema MEROPS de clasificación de peptidasas (Barrett & Rawlings, 2007)

I. Clan: Conjunto de familias en las que todas las peptidasas han evolucionado a partir de un único ancestro. Las familias en el mismo clan tienen en común que las peptidasas que las integran exhiben tipos de plegamiento similares. Cada clan se identifica por dos letras mayúsculas, la primera especifica el tipo catalítico y la segunda es exclusiva del clan. Las primeras letras usadas en el clan son **A**-aspártico, **C**-cisteína, **G**-glutámico, **M**-metalopeptidasa, **S**-serina, **T**-treonina, **U**-desconocido y **P**-tipo catalítico mixto o compuesto del clan peptidasa.

Las peptidasas de tipo aspártico (PA) se asignan a los clanes **AA**, **AC**, **AD**, **AE** y **AF** (Tabla 3). Las estructuras terciarias de los miembros de clanes **AA**, **AD**, **AE** y **AF** muestran un tipo de plegamiento único, no relacionado con ninguna otra peptidasa. Para el clan **AC**, no hay hasta ahora ninguna estructura cristalina, pero otros criterios apuntan a su carácter distintivo.

| Clan AA | Residuos catalíticos Asp, Asp (o His) | Plegamiento doble barril β (o dímeros de barriles β simples) |
|----------------|--|--|
| Familias | A1A | Pepsina A (<i>Homo sapiens</i>) |
| | A1B | Nepentesina (<i>Nepenthes gracilis</i>) |
| | A2A | Retropepsina de HIV-1 |
| | A2B | Ty3 transposón peptidasa (<i>Saccharomyces cerevisiae</i>) |
| | A2C | Gypsy transposón peptidasa (<i>Drosophila melanogaster</i>) |
| | A2D | Osvaldo transposón peptidasa (<i>Drosophila buzzatii</i>) |
| | A2E | Retrotransposón peptidasa (<i>Schizosaccharomyces pombe</i>) |
| | A2G | Retrotransposón peptidasa 17.6 (<i>Drosophila melanogaster</i>) |
| | A2H | Retropepsina (virus del sarcoma dérmico de Alaska) |
| | A3A | Endopeptidasa del virus del mosaico del coliflor (<i>cauliflower mosaic virus</i>) |
| | A3B | Proteasa putativa del virus baciliforme |
| | A9 | Spumapepsina (<i>spumaretrovirus humano</i>) |
| | A11A | Transposón copia (<i>Drosophila melanogaster</i>) |
| | A11B | Ty1 transposón peptidasa (<i>Saccharomyces cerevisiae</i>) |
| | A28 | Proteína-1 inducible por daño en el ADN (<i>Saccharomyces cerevisiae</i>) |
| | A32 | Peptidasa PerP (<i>Caulobacter crescentus</i>) |
| | A33 | Proteasa aspártica skin (<i>Mus musculus</i>) |
| Clan AC | Residuos catalíticos Asp, Asp | <i>Plegamiento</i> Desconocido |
| Familia | A8 | Peptidasa señal II (<i>Escherichia coli</i>) |
| Clan AD | Residuos catalíticos Asp, Asp | <i>Plegamiento</i> Desconocido |
| Familias | A22A | Presenilina 1 (<i>Homo sapiens</i>) |
| | A22B | Peptidasa impas 1 (<i>Homo sapiens</i>) |
| | A24A | Peptidasa tipo M1 prepilina tipo IV (<i>Pseudomonas aeruginosa</i>) |
| | A24B | Peptidasa preflagelínica (<i>Methanococcus maripauldis</i>) |
| Clan AE | Residuos catalíticos Asp, Asp, His (o Lys) | <i>Plegamiento</i> hoja beta central rodeado de hélices |
| Familias | A25 | peptidasa gpr (<i>Bacillus megaterium</i>) |
| | A31 | peptidasa HybD (<i>Escherichia coli</i>) |

| Clan AF | Residuos catalíticos Asp, Asp, Asp, His | Plegamiento cilindro β |
|---------------------------------|--|--|
| Familia | A26 | Omptina (<i>Escherichia coli</i>) |
| Otras familias (clan AX) | | |
| | A5 | Thermopepsina (<i>Sulfolobus acidocaldarius</i>) |
| | A36 | Factor de esporulación SpoIIIGA (<i>Bacillus subtilis</i>) |
| Clan NA | | |
| Familias | N1 | Endopeptidase nodavirus |
| | N2 | Endopeptidase tetravirus (<i>Nudaurelia capensis</i> omega virus) |
| | N8 | Poteína de la cápside de auto-escisión de tipo VP0 (poliovirus humano 1) |

Tabla 3. Clanes y familias de las peptidasas aspárticas. Para cada familia la peptidasa citada es la peptidasa ejemplo típica usada para crear la familia en la base de datos MEROPS

II. Familia: Una familia incluye peptidasas que presentan homología en la secuencia aminoacídica de la región encargada de la actividad enzimática

El clan **AA** contiene la gran familia **A1**, siendo la pepsina A (código PDB: 1PSO) el ejemplo típico. Las estructuras cristalinas de un número de enzimas de la familia A1 muestran que son moléculas bilobuladas, con el sitio activo situado entre los lóbulos. Cada lóbulo lleva uno de los residuos Asp de la díada catalítica. Muchas peptidasas son moléculas bilobuladas, pero en la familia A1, se cree que un lóbulo ha evolucionado a partir del otro por duplicación génica (Tang, James, Hsu, Jenkins, & Blundell, 1978). Evidencias estructurales sugieren que las enzimas tipo pepsina evolucionaron a través de una duplicación de genes y eventos de fusión a partir de un gen ancestral tipo retropepsina, retrotransposones y peptidasas de virus lamda (Pfam PF00077). Éstas se caracterizan por ser mucho más pequeñas y parecen ser homólogos a cualquiera de los dos lóbulos de las PAs eucariotas (Tang, 2004). La retropepsina de HIV-1 (código PDB: 5HVP, Fig. 4) de la familia A2 presenta un único lóbulo con un Asp catalítico, muy similar en estructura a un lóbulo de la pepsina. Las PAs tipo

retrovirales son homodímeros obligados con el sitio catalítico ubicado entre las dos moléculas. Otros autores sugieren que las PAs pueden ser descritas como una serie de elementos de estructura secundaria que se repiten (Dunn, 2002; Rawlings et al., 2004; Simões, Faro, Bur, Kay, & Faro, 2011). En las enzimas actuales, se mantiene una muy limitada similitud en las secuencias de aminoácidos de los dos lóbulos, pero el plegamiento es muy similar.

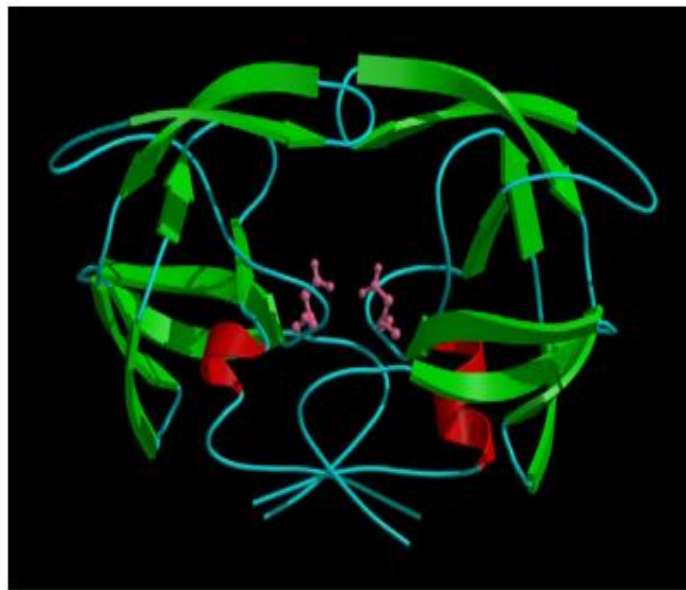


Figura 4. Diagrama de cintas del dímero de la retropepsina del HIV-1 (código PDB: 5HVP). El residuo catalítico Asp25 se muestra como barras y esferas en color rosa (uno por cada monómero). Tomado de. Rawlings & Salvesen (2013).

El hecho de que las endopeptidasas en las familias A1 y A2 presenten plegamientos similares es totalmente compatible con la idea de que han evolucionado a partir de un ancestro común, por esto ambas familias están agrupadas en el clan AA (Rawlings & Salvesen, 2013).

Cada Asp del sitio activo está presente en el motivo distintivo Asp-Thr/Ser-Gly seguido por un motivo de secuencia hidrofóbico-hidrofóbico-Gly, que se conserva en las familias A1, A2 y en algunas otras familias del clan AA. Dos de estos motivos están conservados en todas las secuencias conocidas o que se predicen como activas dentro de la familia A1 (Rawlings & Bateman, 2009). El

motivo hidrofóbico-hidrofóbico-Gly, en la familia A2, precede inmediatamente el residuo invariante Arg 123 (numeración de prepropepsinógeno A) que se cree es importante para la dimerización (Rawlings & Barrett, 2014). Juntos los motivos catalítico e hidrofóbico-hidrofóbico-Gly constituyen una característica estructural conocida como *loop-ψ*, este *loop* sirve para ubicar los dos residuos Asp necesarios para el funcionamiento de la maquinaria catalítica (Simões et al., 2011).

Las enzimas con la organización arquetípica de pepsina (peptidasas de la familia A1) son las PAs más numerosas y mejor caracterizadas que se encuentran en eucariotas con una pequeña cantidad de homólogos conocidos en bacterias marinas y bacterias de plantas (Simões et al., 2011). La gran mayoría son activas a pH ácido, pero algunas muestran actividad en condiciones neutras. Esta familia incluye a enzimas del tracto digestivo animal, como la pepsina, gastrisina y quimosina, enzimas lisosomales como la catepsina D y E y enzimas involucradas en procesos postraduccionales, como la renina producida en el riñón, que procesa angiotensinógeno en el plasma. Las aspartil-peptidasas fúngicas homólogas de la pepsina presentan diversas estructuras y algunas de ellas como la PA de *Rhizopus chinensis* (rizopuspepsina) intervienen en el proceso de esporulación. Las plasmepsinas I y II aisladas del protozoo *Plasmodium falciparum*, también son PAs homólogas de la pepsina (Rawlings, N.D., Barrett, 2004) participan en la degradación de la hemoglobina, su inhibición es el blanco de potenciales drogas antimaláricas (Bernstein & James, 1999; Gupta, Yedidi, Varghese, Kovari, & Woster, 2010). Homólogos de la pepsina se conocen en plantas, como la fitepsina de cebada (Glathe et al., 1998; Törmäkangas et al., 2001).

1.1.3. Especificidad de secuencia mediada por especificidad de subsitios

Tanto las exopeptidasas como las endopeptidasas exhiben “especificidad de secuencia”, es decir, muestran distintas preferencias por determinados aminoácidos cerca de los enlaces peptídicos que clivan. El sitio catalítico de una peptidasa, que es directamente responsable de la hidrólisis del enlace peptídico, se encuentra comúnmente en una ranura en la superficie de la molécula. La especificidad de la hidrólisis está dictada por las propiedades de los sitios de unión dispuestos a lo largo de la ranura en uno o ambos lados del sitio catalítico (Rawlings & Barrett, 2014). La especificidad usualmente es descrita mediante el modelo propuesto por Schechter y Berger en el que cada subsitio específico es capaz de acomodar la cadena lateral de un solo residuo de aminoácido. Por convención, los residuos de aminoácidos de sustrato se denominan P (por péptido) y los subsitios en la peptidasa que interactúan con el sustrato se denominan S (por subsitio). Los sitios se numeran desde el sitio catalítico, S1, S2 y así sucesivamente hacia el extremo N-terminal del sustrato y S1', S2' y así sucesivamente hacia el C-terminal (Fig.5). Mientras que los aminoácidos del sustrato que se acomodan en estos subsitios se numeran P1, P2, etc. y P1', P2', etc., respectivamente (Schechter & Berger, 1967). Diferentes peptidasas tienen diferentes requisitos para las interacciones que se establecen en cada subsitio y esto determina la especificidad de la escisión. Por ejemplo, el subsitio S1 de tripsina tiene una marcada preferencia por la unión de residuos de aminoácidos básicos (arginina y lisina), mientras que las interacciones con varios de los subsitios más distantes del enlace escindible son críticos para la unión del sustrato en la renina, y las caspasas (Hooper, 2002).

Como se observa en la Fig. 5 B los residuos de sustrato se representan a partir de una única cadena polipeptídica continua, mientras que los residuos de enzima pueden provenir de diferentes pliegues de la enzima, pero que tienen que

disponerse en una adecuada posición espacial para permitir que los residuos de sustrato entren en el bolsillo de especificidad e interactúen con los residuos del sitio activo durante la escisión (Timmer & Salvesen, 2007).

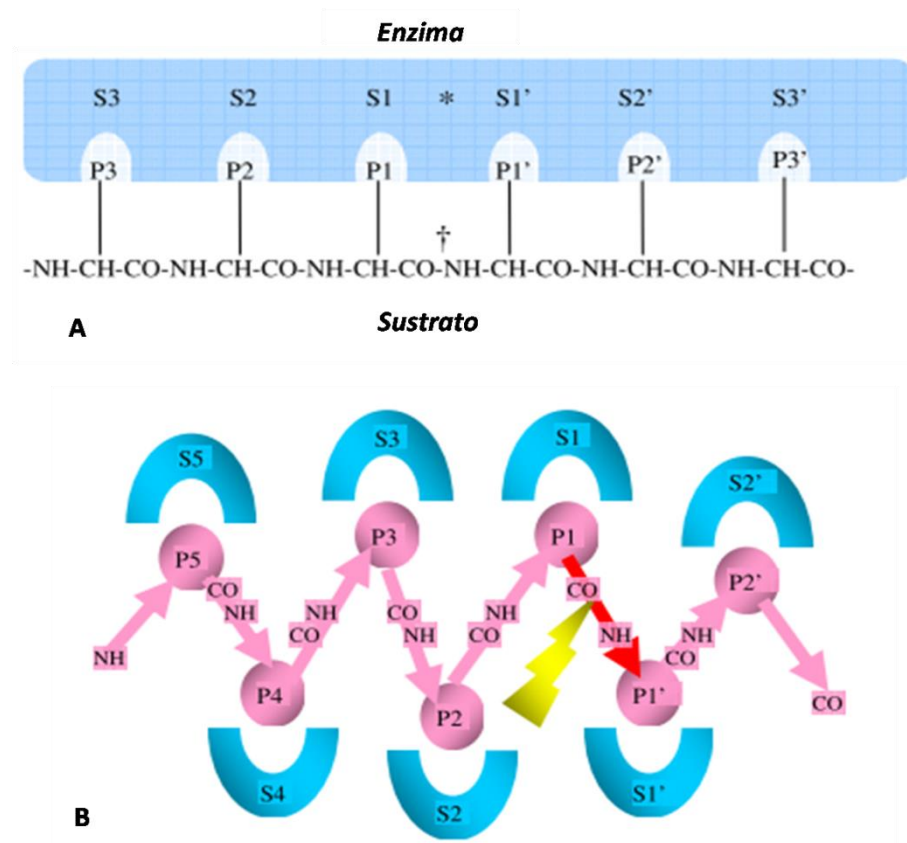


Figura 5. A: Nomenclatura propuesta por Schechter y Berger para la unión de un sustrato peptídico a una peptidasa. El área sombreada de azul representa la peptidasa. P1, P1', etc. son las cadenas laterales de los seis aminoácidos que rodean el enlace peptídico escindible (indicado por la flecha) en el sustrato. S1, S1', etc., son los subsitios correspondientes de la peptidasa. **B. Representación 2D de especificidad del subsitio de la peptidasa.** Se muestra en azul la peptidasa y en rosa el sustrato. Tomado de Timmer & Salvesen, 2007.

Las PAs muestran especificidad por aquellos residuos que presentan grandes cadenas laterales hidrófobas a cada lado de la unión escindible (por ej: Phe, Val, Ile, Leu-en P1 y P10, Dunn & Hung, 2000; Rawlings, Barrett, & Bateman, 2012).

1.1.4. Mecanismo catalítico de las peptidasas aspárticas

En la gran mayoría de las PAs son dos los residuos de ácido aspártico (Asp32 y Asp215 en la numeración de la pepsina) que actúan juntos para unir y activar la molécula de agua (díada catalítica), pero en algunas enzimas, otro residuo reemplaza al segundo Asp (Rawlings, N.D., Barrett, 2004). El mecanismo de la hidrólisis del enlace peptídico (Fig. 9) ha sido establecido mediante estudios estructurales de estas enzimas unidas a compuestos que mimetizan estados de transición al unirse al sitio activo. El nucleófilo que escinde la unión peptídica es una molécula de agua activada a un ion hidróxido por el carboxilato, base general, del Asp215. El carboxilo del Asp33 protonado presta asistencia electrofílica a través de un puente de hidrógeno. Los grupos carboxílicos de los dos residuos aspartato se encuentran aproximadamente coplanares y la molécula de agua unida yace, entre los dos, en ese plano común. Como consecuencia del ataque nucleofílico se forma un intermediario tetraédrico. A continuación se protona el átomo de nitrógeno del compuesto -NH₂ saliente conjuntamente con la regeneración del estado protonado del Asp33, esto está acoplado a la liberación de los nuevos grupos amino y carboxilo. Finalmente se reasocian los grupos carboxilo del sitio activo con una nueva molécula de agua. En las peptidasas de la familia A1 del clan AA la hendidura de unión al sustrato abarca 7 u 8 aminoácidos (P5 a P3') con los Asp en posición central. Esta hendidura se forma donde se encuentran los dominios N y C terminales. Los dos residuos de Asp comúnmente se encuentran en el motivo Asp-Thr-Gly-Thr/Ser, lo cual es importante para establecer y mantener el ambiente del grupo carboxílico (James, 2004)

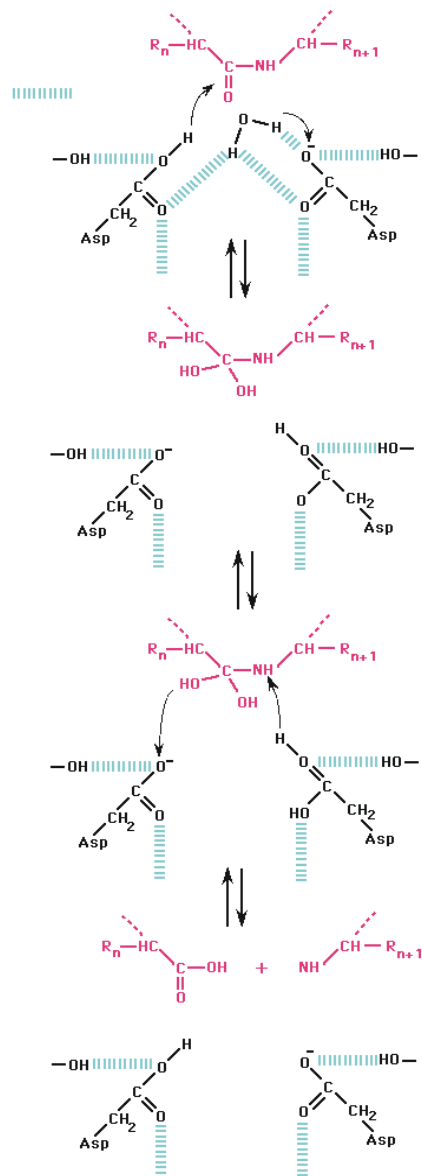


Fig. 9. Mecanismo catalítico de peptidasas aspárticas

1.1.5. Distribución de peptidasas

En los organismos vivos, la amplia distribución entre plantas, animales y microorganismos demuestra que las peptidasas son esenciales para todos ellos, llevando a cabo importantes roles fisiológicos en numerosos procesos. Las peptidasas fueron inicialmente reconocidas como enzimas proteolíticas del jugo gástrico involucradas en la degradación inespecífica de las proteínas dietarias.

Sin embargo, avances posteriores han provisto una nueva visión del mundo proteolítico debido a que juegan un rol fundamental en la maduración de las proteínas. Las proteínas sintetizadas como un precursor se someten a escisión enzimática selectiva de ciertos enlaces peptídicos. Este proceso es crucial para la formación de la estructura tridimensional activa y/o correcta localización celular. Estas moléculas también juegan un rol en la degradación de proteínas durante el mantenimiento normal de células y la restructuración de los tejidos (Mutlu & Gal, 1999). Actúan reciclando polipéptidos en sus aminoácidos constitutivos, mientras que otras, actúan escindiendo enlaces peptídicos covalentes de proteínas o péptidos, una modificación post-traducciona esencial que altera las vidas medias, el tráfico subcelular y las actividades de una amplia gama de proteínas. En consecuencia, las peptidasas están potencialmente implicadas en una multitud de procesos biológicos que van desde la simple digestión de las proteínas a las cascadas de señalización altamente reguladas (Olivares et al., 2011).

1.1.6. Peptidasas de plantas

Una amplia variedad de peptidasas están asociadas con procesos biológicos vitales, pero el origen y la evolución de su enorme diversidad no están bien definidos (Page & Di Cera, 2008). Los genomas de plantas codifican cientos de peptidasas, que representan a decenas de familias no relacionadas. Así el genoma de *Arabidopsis* (*Arabidopsis thaliana*) codifica más de 800 peptidasas que están distribuidas en más de 62 familias de acuerdo a su homología de secuencia aminoacídica, las cuales pertenecen a 30 clanes diferentes de los cuales el más grande es el de las peptidasas serínicas, seguido de las aspárticas, cisteínicas, metalo-peptidasas y treónicas (Faro & Gal, 2005; Tsiatsiani, Gevaert, & Van Breusegem, 2011; van der Hoorn, 2008).

1.1.6.1. Peptidasas aspárticas de plantas

Las PAs de plantas están presentes en una vasta variedad de especies. En gimnospermas por ejemplo se las ha estudiado en semillas de pino (*Pinus sylvestris*, Salmia, 1981). También se las ha encontrado en especies de angiospermas. En monocotiledóneas se puede ejemplificar su presencia en cebada (*Hordeum vulgare*), arroz (*Oryza sativa*), trigo (*Triticum aestivum*), sorgo (*Sorghum bicolor*) y maíz (*Zea mays*, Asakura et al., 1995; Asakura, Watanabe, Abe, & Arai, 1997; Sarkkinen et al., 1992). Mientras que, en eudicotiledóneas, se las pueden encontrar en girasol (*Helianthus annuus*), cacao (*Theobroma cacao*), Arabidopsis (*A. thaliana*), repollo (*Brassica oleracea*), espinaca (*Spinacia oleracea*), papa (*Solanum tuberosum*), tabaco (*Nicotina tabacum*), tomate (*Solanum lycopersicum*), cardo de Castilla (*Cynara cardunculus*), abrepuños (*Centaurea calcitrapa*) y plantas carnívoras como *Nepenthes spp* (Domingos et al., 2000), entre otros ejemplos.

Los miembros de la familia A1 están presentes en una variedad de tejidos, su diversidad ha provocado la redefinición de la clasificación de las PAs de plantas de la familia A1 en aspartilpeptidasas típicas, tipo nucelina y atípicas (Lufrano et al., 2012). Con respecto a su localización dentro de la planta las PAs se han aislado, caracterizado y purificado a partir de diferentes órganos tales como semillas, flores y hojas (Lufrano et al., 2012). Las peptidasas se expresan específicamente en tiempo y espacio y se acumulan en diferentes compartimentos subcelulares (Pesquet, 2012; Schaller, 2004; Tsiatsiani et al., 2011; van der Hoorn, 2008). En cuanto a su localización subcelular, las PAs de plantas de la familia A1 que han sido caracterizadas, se dirigen principalmente a las vacuolas líticas o al compartimento vacuolar de almacenamiento. Sin embargo, algunas de ellas, se dirigen al retículo endoplasmático o se secretan a la pared celular o al apoplasto (Almeida et al., 2012; González-Rábade et al., 2011; Mazorra-Manzano et al., 2013; van der Hoorn, 2008). Algunos ejemplos de PAs secretadas a la matriz extracelular lo constituyen las PAs de hojas de tabaco

(*Nicotina tabacum*, Rodrigo, Vera, van Loon, & Conejero, 1991), tomate (*Solanum lycopersicum*, Rodrigo, Vera, & Conejero, 1989) y cardosina B que se encuentra en la matriz extracelular del tejido floral de transmisión en el cardo de Castilla (*C. cardunculus*, Vieira et al., 2001). Mendieta y colaboradores demostraron que las PAs de papa (StAPs, *Solanum tuberosum*), inducidas por estrés biótico y abiótico, poseen localización extracelular. Una diferente localización celular incluye las membranas del fotosistema II en los plástidos de espinaca (*Spinacia oleracea*, Kuwabara & Suzuki, 1995). El análisis *in silico* de las señales de direccionamiento para la mayoría de las PAs putativas del genoma de *Arabidopsis* (*A. thaliana*) está de acuerdo con la localización esperada en la vía secretoria, pero también predice otros sitios, como cloroplasto o mitocondria, con varias PAs que se indican como unidas a membranas (Faro & Gal, 2005). La Tabla 1 resume nombres de PAs de diferentes organismos de la tribu Cardue que han sido estudiadas con sus número acceso (tomado de Vairo Cavalli, Lufrano, Colombo, & Priolo, 2013).

| Especie | Enzima | Referencia |
|-----------------------------|---|---|
| <i>Centaurea calcitrapa</i> | Proteinasa aspártica Cyprosina | Q96383 Domingos et al. (1998) |
| <i>Cynara cardunculus</i> | Cyprosina A X69193 Cyprosina B (aka cyprosin 3) Q39476/X81984 Cardosina A Q9XFX3/AJ132884 Cardosina B Q9XFX4/AJ237674 Cardosina C Cardosina D Cardosina E P85136 Cardosina F P85137 | Cordeiro et al. 1994 White et al. (1999) Faro et al. (1995) Faro et al. (1995) Pimentel et al. (2007) Pimentel et al. (2007) Sarmiento et al. (2009) Sarmiento et al. (2009) Sarmiento et al. (2009) Sarmiento et al. (2009) |

| | | |
|----------------------------|--|---|
| | Cardosina G P85138 | |
| | Cardosina H P85139 | |
| <i>Cynara humilis</i> | Cardosina A | Esteves (1995) |
| <i>Cynara scolymus</i> | Cynarasa A Cynarasa B Cynarasa C | Llorente et al. (2004) Sidrach et al. (2004) |
| <i>Cirsium vulgare</i> | Cirsina JN703462 | Lufrano et al. (2012) |
| <i>Onopordum turcicum</i> | Proteinasa aspártica | Tamer (1993) |
| <i>Onopordum acanthium</i> | Onopordosina | Brutti et al. (2012) |

Tabla 4: PAs presentes en varias especies de la tribu *Cardueae*Cass. (Cynareae Less.). Son miembros de la familia A1 (base de datos MEROPS). En la segunda columna se indican los números de acceso de secuencia de las bases de datos UniProtKB / TrEMBL o de GenBank. Cuando es posible se indica el nombre asignado por los autores citados. Tomado de Vairo Cavalli (2013).

1.1.7. Aplicaciones biotecnológicas de las PAs

En la actualidad, se conocen cerca de 4000 enzimas, y de éstas, aproximadamente 200 tipos se utilizan comercialmente. Sin embargo, sólo alrededor de 20 enzimas se producen en escala verdaderamente industrial. Casi el 75% de todas las enzimas industriales son enzimas hidrolíticas, amilasas, peptidasas y lipasas dominan el mercado de enzimas, lo que representa más del 70% del total de ventas de enzimas (Shuang, Yang, Yang, Zhu, & Wang, 2012). El mercado de enzima industrial se valoró en 4,2 mil millones de dólares en el 2014 y se prevé que crezca a una tasa anual del 7,0% entre 2015 y 2020.

Las peptidasas constituyen el grupo más importante de enzimas industriales actualmente en uso, contribuyendo en gran medida a la industria química

tradicional y moderna, mejorando los procesos existentes. Son utilizadas en variados procesos tecnológicos que incluyen la tiernización de carnes, la elaboración de cerveza y de quesos, la panificación, la industria de polvos detergentes y el procesado de fibras textiles y cueros, además se utilizan en la industria farmacéutica y en el tratamiento de efluentes industriales, en cosméticos, medicamentos, y como herramientas en investigación y desarrollo (González-Rábade et al., 2011).

La medición de la actividad hidrolítica sobre sustratos sintéticos es una forma sencilla de conocer la especificidad de clivaje de estas enzimas, proporcionando información importante para aplicaciones biotecnológicas, como la producción de péptidos bioactivos a partir de proteínas de alimentos (Cimino, Colombo, Liggieri, Bruno, & Vairo-Cavalli, 2015; Silva & Malcata, 2005). Algunos péptidos están ocultos e inactivos en las proteínas originales, pero cuando son liberados pueden tener diversas aplicaciones biomédicas, como agentes antihipertensivos o antioxidantes, entre otros (Bertucci, Liggieri, Colombo, Vairo Cavalli, & Bruno, 2015; Perpetuo, Juliano, & Lebrun, 2003). Dado que la proteólisis cambia las propiedades químicas, físicas, biológicas, e inmunológicas de las proteínas, juega un papel importante en la biotecnología. La hidrólisis de las proteínas de los alimentos, mejora las características nutricionales, retarda el deterioro, modifica diferentes propiedades funcionales (solubilidad, formación de espuma, coagulación, y capacidades emulsionantes), previene interacciones no deseadas, modifica el sabor y olor, y elimina factores tóxicos o inhibidores, entre otros (González-Rábade et al., 2011). Una de las aplicaciones más importantes de las peptidasas en la industria alimentaria es el uso como cuajo en la fabricación de queso. PAs tales como quimosina, pepsina y algunas enzimas fúngicas han sido utilizadas en el proceso de fermentación y en la producción de una variedad de productos, incluyendo la producción de queso, formación de sabor y producción de péptidos bioactivos como nutracéuticos (Horimoto, Dee, & Yada, 2009). Las

PAs también han despertado un gran interés como dianas para el tratamiento farmacológico de varias condiciones fisiológica y patológicas humanas. Estas incluyen el uso de la pepsina en la enfermedad de úlcera péptica, la renina en la hipertensión, la plasmepsina en la malaria, la catepsina D (CatD) en la metástasis en cánceres, la catepsina E (CatE) en el sistema inmunológico, peptidasa del virus de la inmunodeficiencia humana (VIH) en el síndrome de inmunodeficiencia adquirida (SIDA) y BACE (enzima que cliva la proteína precursora β -amiloide) en la enfermedad de Alzheimer (Horimoto et al., 2009).

1.1.7.1. Aplicaciones de las peptidasas aspárticas de plantas

Se han estudiado diversas aplicaciones biotecnológicas y biomédicas de PAs de flores de cardo. Estas aplicaciones incluyen la manufactura de quesos –un proceso que data del período de dominación romana sobre la península Ibérica– junto con otros usos, algunos potenciales o en proceso de evaluación, absolutamente contemporáneos (Vairo Cavalli et al., 2013). Cardosina A (*C. cardunculus*) muestra una acción específica sobre colágeno, por lo que esta enzima ha sido propuesta como herramienta en procedimientos médicos que requieran la remodelación de la matriz extracelular. Además, esta peptidasa ha sido evaluada para disociar la corteza de embriones con el fin de establecer cultivos neuronales primarios (Duarte et al., 2005; Duarte, Rosa, Duarte, Pires, & Barros, 2007). Las neuronas tratadas con cardosina exhiben morfología y función normal cuando se comparan con cultivos preparados con el método de disociación de tripsina ampliamente utilizado (Duarte et al., 2007). Además, mejorando la regeneración neuronal en una fase inicial del cultivo después de la disociación celular con cardosinas (Duarte, Duarte, Correia, Pires, & Barros, 2009).

La inmovilización de cardosina A en esponjas de quitosano no indujo alteraciones en la especificidad de la enzima, manteniendo así la capacidad de hidrolizar el colágeno de tipo I. Por su biocompatibilidad esta forma de

cardosina A inmovilizada resulta una herramienta promisorio en la prevención de la formación y reducción de adherencias en cirugías abdominales (Pereira et al., 2005).

Otra aplicación desarrollada tiene que ver con la estabilidad en sistemas bifásicos acuosos-orgánicos de las cardosinas A y B que fueron capaces de catalizar la síntesis enzimática de dipéptidos y tripéptidos en dichos medios (Sarmiento, Silvestre, Barros, & Pires, 1998; Sarmiento, Oliveira, Pires, Amado, & Barros, 2004; Shmyrova et al., 2006).

1.1.8. Organización estructural de las PAs vegetales

1.1.8.1. Organización de la estructura primaria

Existen tres grupos bien diferenciados de PAs de plantas: PAs típicas, tipo nucelina y atípicas (Chen, Ouyang, Wang, Xie, & Zhang, 2009; Faro & Gal, 2005). La Fig 6 muestra la organización de dominios de cada tipo.

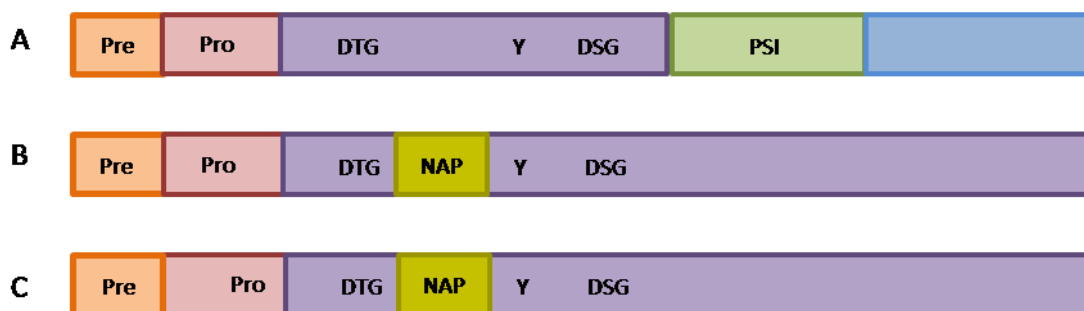


Figura 6. Aspectos distintivos de las PAs, típica (A) tipo nucelina (B) y atípica (C). Características distintivas de las PAs de plantas: péptido señal (**Pre**), prosegmento (**Pro**), primera tríada catalítica (**DTG**); inserto específico de la proteasa aspártica tipo nepentesina (**NAP**); Tyr conservada (**Y**); segunda tríada catalítica (**DSG**). Tomado de Faro & Gal, 2005.

La estructura primaria es bien conservada entre PAs vegetales típicas (Dunn, 2002; Simões & Faro, 2004). Se sintetizan como precursores inactivos, proenzimas (zimógenos), estos son posteriormente convertidos en enzimas maduras, ya sea monoméricas o heterodiméricas. Los precursores de PAs tienden a tener un alto grado de similitud en términos de estructura primaria y se

caracterizan por la presencia de una secuencia señal N-terminal hidrófoba (Pre), responsable de la translocación al retículo endoplasmático (RE) de la célula vegetal, seguida por un prosegmento (Pro) o propéptido presente en todas las PAs de alrededor de 40 aminoácidos. El Pro es responsable de la inactivación del zimógeno durante su tránsito a través del sistema endomembranas, el plegamiento, la estabilidad o el direccionamiento intracelular de las proenzimas (Kervinen et al., 1999; Simões & Faro, 2004). Por último, entre el dominio N-terminal y el dominio C-terminal, se puede encontrar un fragmento conocido como *inserto específico de planta* (PSI) situado entre los extremos amino y carboxilo terminales de la proteína madura, el cual abarca 100 aminoácidos (Guevara, Veríssimo, Pires, Faro, & Daleo, 2004; Fig.6). La mayoría de las PAs estudiadas y caracterizadas hasta el momento pertenecen a este grupo.

Las PAs atípicas muestran características intermedias entre las secuencias típicas y tipo nucelina. Las PAs tipo nucelina codifican una proteína similar a la nucelina (número de acceso: U87148), que fue detectada en células nucleares de cebada y no contienen el dominio PSI (Faro & Gal, 2005). Las PAs atípicas de planta caracterizadas contienen un elevado número de cisteínas y muestran una localización diferente al de las PAs típicas, sin embargo, tienen una baja identidad de secuencia entre ellas y variadas funciones. Así, la proteína de unión a ADN en el nucleoide cloroplástico (CND41, número de acceso BAA22813.1), aislada de cloroplastos de células de tabaco (Kato et al., 2004), está involucrada en la movilización de nitrógeno. Mientras que las PAs nepentesina I y nepentesina II (número de acceso:BAD07474 y número de acceso: BAD07475, respectivamente) aisladas de *Nepenthes alata* (Athauda et al., 2004) son secretadas a los reservorios de la planta insectívora, y en *Arabidopsis* la proteasa aspártica PSC1 (número de acceso: NP_195839) está involucrada en la regulación de la muerte celular y se localiza en el RE (Olivares et al., 2011). Con la excepción de CND41 y PCS1 (promotor de la supervivencia celular 1) todas las PAs atípicas

caracterizadas hasta el momento tienen una arquitectura estructural característica con algunas variaciones: el péptido señal, seguido de un prosegmento y un dominio de codificación de la proteína madura. CND41 tiene un péptido de 120 aminoácidos que participa en el direccionamiento de la proteasa al cloroplasto (Nakano et al., 1997). Además, la PA PCS1 parece carecer del prosegmento clásico. Después de la secuencia de señal, presenta una secuencia rica en serinas, cuya función no ha sido descrita (Faro & Gal, 2005).

Si bien el PSI sólo se identificó en PAs típicas de plantas, su estructura es semejante a saposinas y proteínas tipo saposinas (SAPLIPs). La familia de las SAPLIPs es diversa, se trata de proteínas que interactúan con lípidos (Muñoz et al., 2010; Muñoz, Palomares-Jerez, Daleo, Villalaín, & Guevara, 2014) y que normalmente se elimina durante la maduración de las PAs (Dunn, 2002; Simões & Faro, 2004). Existe una similitud significativa (alrededor del 40% de identidad) entre el PSI y las saposinas. Las saposinas, se generan a partir del precursor prosaposina, son glicoproteínas estables a altas temperaturas con un peso molecular de 12-15 kDa que funcionan como activadores de hidrolasas lisosomales en el metabolismo de diversos esfingolípidos en seres humanos y animales (Kishimoto, Hiraiwa, & O'Brien, 1992; Schuette, Pierstorff, Huettler, & Sandhoff, 2001). Se han identificado cuatro clases de saposina (A, B, C y D) que tienen la capacidad de unirse a una variedad de esfingolípidos y gangliósidos. La especificidad y el modo de activación de las hidrolasas de esfingolípidos difieren entre las saposinas individuales. La prosaposina se encuentra en diversos fluidos corporales y membranas neuronales. Además de la función de precursor, las prosaposinas poseen una variedad de actividades neurotróficas. La secuencia neurotrófica fue localizada a la porción N-terminal del dominio de saposina C. Las cuatro saposinas son estructuralmente similares entre sí en términos de la ubicación de seis cisteínas, un sitio de glicosilación conservado, y la presencia de residuos de prolina y una tirosina conservados. Estas características estructurales

también se encuentran en los PSIs (Brodelius et al., 2005; Faro & Gal, 2005; Simões & Faro, 2004). El PSI se pliega como un dominio independiente. La caracterización estructural del PSI de profitepsina y de otras proteínas de la familia SAPLIP muestra que comparten una estructura globular compacta, cinco alfa-hélices, unidos por tres enlaces disulfuro (Bryksa et al., 2011; Simões & Faro, 2004). Egas y colaboradores (2000) han demostrado la importancia de estas cisteínas y sus puentes, que se asocia con el mantenimiento de la estructura tridimensional del PSI en cardosina A, la unión del PSI a lípidos y las actividades de permeabilización de membranas.

En los PSIs la secuencia de aminoácidos revela una inversión entre los dominios amino y carboxilo terminal con respecto a las saposinas, dando lugar a la expresión “swaposin” (Ponting & Russel, 1995). Esta permutación resulta en una estructura tridimensional muy similar, con un patrón de conectividad algo diferente entre sus regiones (Bliven & Prlic, 2012).

Algunas proteínas tipo saposina (SAPLIPs) como NK-lisina y granulisin presentan actividad antimicrobiana, ambas proteínas están clasificadas como antimicrobianos (<http://aps.unmc.edu/AP/main.php>). Al igual que SAPLIPs, el dominio StAsp-PSI (PSI de PA de *Solanum tuberosum*) es capaz de interactuar con la membrana plasmática de células de mamíferos, aumentando la permeabilidad de la célula y produciendo muerte celular. Esta actividad citotóxica de StAsp-PSI es selectiva. Considerando que estas proteínas son tóxicas para microorganismos patógenos humanos y de plantas, hongos, oomicetos, bacterias Gram - y Gram +, así como células cancerígenas, pero no son capaces de matar las células T humanas, los glóbulos rojos humanos y células vegetales. Egas y colaboradores (2000) encontraron que el PSI recombinante es capaz de interactuar con vesículas fosfolipídicas y de inducir la fusión de su contenido. Aún así, el rol del PSI no ha sido completamente esclarecido, Simões y Faro (2004) propusieron que participaría en mecanismos de defensa contra patógenos y/o en la muerte

celular, en base a su capacidad de interactuar con membranas lipídicas. Sin embargo, White y colaboradores (1999) encontraron en la procyprosina recombinante, que el PSI tiene como función asegurar que el polipéptido naciente de la proenzima esté correctamente plegado y sea capaz de activarse hasta ser una enzima madura.

El PSI ha sido propuesto también como una señal de direccionamiento de proteínas. Originalmente fitepsina (*Hordeum vulgare*) se acumula en la vacuola, con la delección de su dominio PSI la fitepsina truncada es secretada, lo que lleva a pensar en un rol del PSI en el direccionamiento de la proteína (Törmäkangas et al., 2001). Los resultados con otras peptidasas, como soyAP2, una PA de soja, revelan información contradictoria. Esto puede deberse al hecho que diferentes dominios PSI podrían tener diferentes funciones fisiológicas (Terauchi et al., 2005).

En el caso de cardosina A, se sabe que su dominio PSI no sólo posee información de direccionamiento hacia la vacuola, sino también que este direccionamiento se lleva a cabo por una vía independiente al aparato de Golgi. Pereira (2012) planteó la hipótesis de que la ausencia de residuos glicosilados en este dominio de alguna manera podría estar involucrada en esta nueva función.

1.1.8.2. Organización de la estructura terciaria

La estructura terciaria de varios miembros de la familia A1 ha sido determinada, pero de estas sólo dos estructuras cristalinas corresponden a PAs vegetales, ambas pertenecientes al grupo de PAs típicas. Estas son la estructura de cardosina A madura (código de PDB: 1B5F, Frazão et al., 1999) y profitepsina, precursor de la PA de cebada conteniendo el prosegmento junto con el dominio PSI (código de PDB: 1QDM, Kervinen et al., 1999).

La homología de los dos lóbulos de las moléculas de peptidasas tipo pepsina se refleja por el hecho de que los dos residuos Aspárticos, Asp94 y Asp277

(numeración de acuerdo al alineamiento de la familia A1, <http://merops.sanger.ac.uk>) se encuentran en motivos de secuencias muy similares. El motivo central se resume comúnmente como Asp-Thr-Gly (el motivo DTG), aunque una descripción más completa sería: Xaa-Xaa-**Asp**-Gly-Xbb, en la que Xaa es un residuo hidrofóbico y Xbb es Ser o Thr. En la familia A1, hay un tercer residuo importante para la catálisis además de los dos residuos de ácido aspártico; este residuo se encuentra en una estructura flexible conocida como “flap”, ubicada a lo largo de la hendidura catalítica y juega un papel importante en la especificidad de la PA. El *flap* se proyecta hacia afuera sobre la hendidura del sitio activo y encierra a los sustratos y a los inhibidores en el sitio activo (Rawlings & Barrett, 1995; Simões & Faro, 2004; Vairo Cavalli et al., 2013). El residuo es una Tyr conservada (Tyr 75, numeración de pepsina) que interactúa con los residuos del sustrato en el bolsillo catalítico (Andreeva & Rumsh, 2001).

La estructura cristalográfica de cardosina A madura de *C. cardunculus* fue determinada con alta resolución 1,7 Å (Frazão et al., 1999). La molécula está compuesta por dos cadenas polipeptídicas glicosiladas de 31 y 15 kDa, respectivamente y un plegamiento típico dentro de la familia de las PAs. Esencialmente está formada por la duplicación de un motivo de cuatro cadenas β anti-paralelas y un α -hélice, lo que se repite dos veces en cada uno de los dominios (1B5F; Fig. 7). El sitio activo, como ya ha sido mencionado precedentemente, se encuentra entre los dos dominios en la parte inferior de una gran hendidura. Las dos mitades de la enzima actúan como estructuras plegadas independientes que se mueven una con relación a la otra y contribuyendo cada una con un de los aspártico catalítico en una disposición coplanar de los carboxilatos (Frazão et al., 1999).

Una característica única de cardosina A, entre las peptidasas de plantas, es la presencia de dos motivos aminoacídicos Arg-Gly-Asp (RGD, residuos 246-248,

(Frazão et al., 1999) y Lys-Gly-Glu (KGE, residuos 455-457, Simões et al., 2005). RGD es una secuencia conocida de unión a integrinas en tejidos de mamíferos en donde facilita muchas funciones de reconocimiento celular tales como adhesión, migración, señalización, diferenciación y crecimiento. Es posible que cardosina A participe en mecanismos proteolíticos mediados por adhesión asociados con el crecimiento del tubo polínico, en forma similar a las peptidasas unidas a integrina que actúan durante la proliferación e invasión celular. Así, de acuerdo a la estructura cristalina de cardosina A, la secuencia RGD localizada en un giro que conecta dos cadenas β y proyectada por encima de la superficie molecular actuaría en la interacción polen-pistilo (Faro et al., 1999). Se ha identificado a la fosfolipasa D (FLD) como la proteína de unión a cardosina A mediante la interacción de su dominio C2 con los motivos RGD y KGE de cardosina A siendo los residuos de aminoácidos básicos de RGD/KGE los preponderantes en el reconocimiento del dominio C2. Ambas proteínas, cardosina A y FLD, forman un complejo lo que sugiere que las mismas podrían llevar a cabo acciones concertadas y/o sinérgicas en procesos degenerativos, tales como los observados durante las respuestas al estrés, la senescencia de las plantas, y/o las interacciones polen-pistilo. A su vez, la asociación de ambas proteínas podría facilitar la disgregación de las vacuolas en la fase de desmantelamiento que ocurre en el tipo de muerte celular dada por colapso vacuolar en las plantas.

A pesar de la alta similitud existente entre las dos peptidasas aspárticas de pistilos, cardosina A y B, ninguna de estas dos secuencias se encuentran conservadas en cardosina B, la que contiene por su parte los motivos RGN y EGE (Simões et al., 2005).



Figura 7. Estructura de cardosina A madura (*C. cardunculus* L). La molécula contiene dos cadenas peptídicas glicosiladas (31 kDa en color morado oscuro y 15 kDa en gris claro). La proteína es bilobulada con dos dominios separados por la gran hendidura donde se encuentra el sitio activo (indicado por una flecha); el Asp35 y Asp218, dentro del sitio activo, se muestran en naranja, se muestran como bolas y palos. Los glúcidos, están representados también como barras y esferas, son N-glicanos unidos a Asn67 y Asn178, en color cian, rojo y azul que representan átomos de C, O y N, respectivamente. Los enlaces disulfuro, así como las cisteínas implicadas en los puentes, se muestran en amarillo. Tomado de Vairo Cavalli et al., 2013.

1.1.9. Rol fisiológico de las peptidasas vegetales

Las peptidasas vegetales están involucradas en muchos aspectos fisiológicos y en el desarrollo de la planta. Desempeñan un papel fundamental en procesos tales como, recambio proteico, degradación de proteínas mal plegadas, senescencia y en la vía ubiquitina/proteasoma. Las peptidasas son también responsables de la modificación post-traducciona de proteínas por proteólisis limitada en sitios muy específicos. Están implicadas en una gran diversidad de procesos celulares, incluyendo la fotoinhibición en el cloroplasto, mecanismos de defensa, muerte celular programada, fotomorfogénesis en la plántula en desarrollo, reproducción sexual de plantas, meiosis, supervivencia del gametofito, embriogénesis, formación de la cubierta de semillas y maduración de las mismas, formación del xilema, desarrollo del estoma, biogénesis de cloroplastos, movilización de nutrientes, respuestas a estímulos ambientales y factores de virulencia y senescencia (Mendieta, Pagano, Muñoz, Daleo, & Guevara, 2006; Simões & Faro, 2004; van der Hoorn, 2008). De esta manera, las peptidasas vegetales son reguladores claves de una enorme variedad de procesos biológicos que están involucrados en todos los aspectos del ciclo de vida de las plantas, desde la movilización de proteínas de reserva durante la germinación hasta el inicio de la muerte celular y senescencia programadas (González-Rábade et al., 2011; Schaller, 2004; van der Hoorn, 2008).

1.1.10. Mecanismos de procesamiento de peptidasas aspárticas

Tres tipos distintos de procesamiento han sido observados en los zimógenos de las PAs: autoprosesamiento total (intra o intermolecula); autoprosesamiento parcial asistido por diferentes peptidasas y procesamiento totalmente asistido. El pepsinógeno se incluye en la primer categoría, la procatepsina D en la segunda (requiere la acción de proteinasas cisteínicas lisosomales), y la prorrenina en la tercera (Koelsch, Mares, Metcalf, & Fusek, 1994). Sin embargo aún no se ha

identificado ninguna enzima particular como responsable del inicio de los eventos de procesamiento de las PAs de plantas. Evidencias obtenidas mediante experimentos *in vitro* sugieren que la maduración completa de las PAs podría requerir la acción de otra(s) endopeptidasa(s) y exopeptidasas en adición al procesamiento autocatalítico –que requieren un descenso en el pH– (Castanheira et al., 2005; Domingos et al., 2000; Sampaio, Fortes, Cabral, Pais, & Fonseca, 2008; Simões & Faro, 2004; White et al., 1999).

El procesamiento de las PAs de planta en enzimas plenamente activas implica varios pasos proteolíticos que incluyen la eliminación del péptido señal, propéptido y en la mayoría de los casos el PSI, originando formas heterodiméricas o monoméricas (Mutlu & Gal, 1999). La gran mayoría de las PAs de plantas son proteínas heterodiméricas con una subunidad mayor de 28-35 kDa, y una subunidad menor de 11-16 kDa y sólo un número reducido son proteínas monoméricas, que conservan el PSI aun en su forma madura, y presentan una masa molecular de 36-65 kDa (Balandin et al., 2005; Mutlu & Gal, 1999). Tal es el caso de las PAs aisladas de tomate (Rodrigo et al., 1989), tabaco (Rodrigo et al., 1991) y papa (Guevara, Oliva, Machinandiarena, & Daleo, 1999). En las PAs vegetales heterodiméricas, el PSI se elimina por proteólisis (Glathe et al., 1998; Ramalho-Santos, Verissimo, Cortes, Samyn, & Faro, 1998). En general, el dominio PSI es escindido del precursor en forma parcial o total, no obstante, existen algunas variaciones en el mecanismo y el orden en que el prosegmento y este inserto son eliminados durante la conversión hacia la forma activa (Simões & Faro, 2004). Se ha sugerido que el procesamiento de las preproenzimas involucra una serie de clivajes proteolíticos secuenciales durante el tráfico hacia el Golgi, lo que da lugar a una mezcla compleja de formas heterogéneas de peptidasas aspárticas maduras dentro de los tejidos de la planta (P. Duarte, Pissarra, & Moore, 2008). El mecanismo de procesamiento no es necesariamente siempre el mismo y no necesariamente sigue el mismo orden (Simões & Faro,

2004). En este sentido, se ha encontrado microheterogeneidad en ambas subunidades de cardosinas A, E, F, G y H y en cyprosinas (Cordeiro, Pais, & Brodelius, 1994). Aunque la diversidad y heterogeneidad no solo se debe a modificaciones postraduccionales, sino también a la naturaleza multigénica de esta familia de enzimas (Pimentel et al., 2007; Sarmiento et al., 2009).

Procardosina A sufre procesamiento proteolítico durante la maduración de la flor, el PSI es removido completamente, probablemente por una PA, luego de lo cual el prosegmento es removido. El primer sitio de escisión para el procesamiento ocurre entre el fragmento de 31 KDa y el PSI. El procesamiento sería secuencial, con la acumulación de intermediarios parcialmente procesados (Fig. 8, Ramalho-Santos et al., 1998). En el procesamiento de procathepsina D o pepsinógeno, por ejemplo, se han identificado pseudoformas de la remoción parcial del prosegmento (Conner & Richo, 1992; Dykes & Kay, 1976).

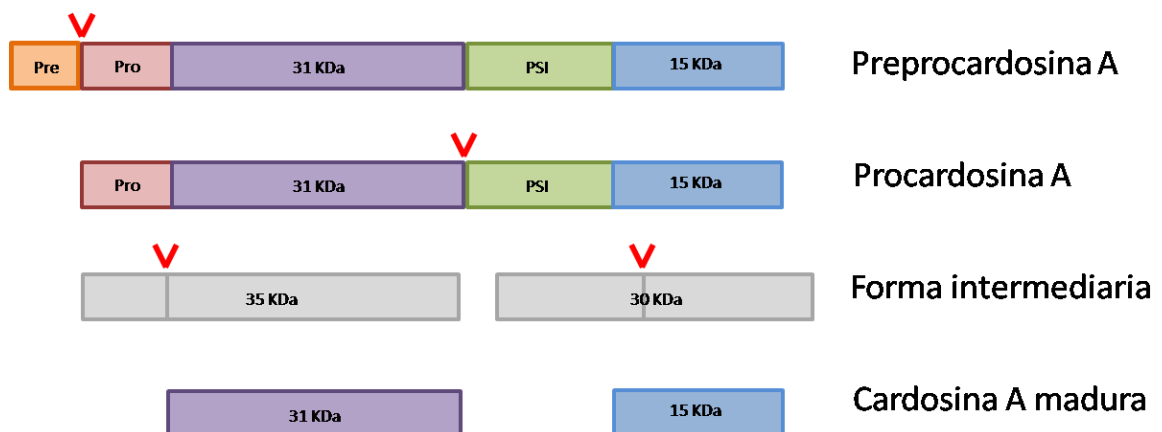


Figura 8. Representación esquemática de los pasos de procesamiento de cardosina A, como ejemplo ilustrativo del procesamiento de las PAs. Durante el direccionamiento, el precursor inicial zimógeno se expone a diferentes valores de pH y a enzimas proteolíticas, resultando en la escisión del péptido señal (Pre), prosegmento (Pro) y dominio PSI. Los sitios de escisión entre el Pro y el fragmento de 31-kDa, el fragmento de 31 kDa y el PSI, y el PSI y el fragmento de 15 kDa se indican con flechas. La eliminación de PSI parece ocurrir en el orden secuencial indicado, con la existencia de una forma intermediaria. La forma final, madura está presente en la vacuola y presenta actividad proteolítica. Tomado de Ramalho-Santos et al., 1998.

El mecanismo por el cual se elimina el PSI en procardosina A no está claro, pero involucra una conformación específica del precursor. Es posible que procardosina A durante su transporte intracelular interactúe directamente con la membrana a través del dominio PSI, limitando así la correcta conformación para su procesamiento en los sitios de escisión del PSI. Tras la eliminación del PSI, procardosina A adquiere una estructura típica de PA similar a las de mamíferos o microbianas (Ramalho-Santos et al., 1998). Un mecanismo de procesamiento autocatalítico similar al de cardosina A ha sido propuesto para cenprosina, PA de *Centaurea calcitrapa* (Domingos et al., 2000) y para oryzasina 1 recombinante, PA de arroz (Asakura, Matsumoto, Funaki, Arai, & Abe, 2000).

Otro esquema de procesamiento fue descrito para profitepsina, donde el propéptido es escindido previamente a la remoción incompleta del dominio PSI (Glathe et al., 1998; Törmäkangas et al., 2001). Las diferencias observadas entre los sitios de procesamiento *in vivo* e *in vitro* hacen pensar en la participación de otras endopeptidasas y/o exopeptidasas en el mecanismo de activación además del mecanismo de autoprosesamiento (Glathe et al., 1998). Para procyprosina recombinante producida en *Pichia pastoris* el procesamiento se asemeja al de profitepsina (White et al., 1999). Sin embargo, el procesamiento de procyprosina parece involucrar la acción de peptidasas de la célula huésped y no autoactivación (White et al., 1999). También ha sido propuesto un mecanismo de procesamiento similar para una de las PAs que se expresa en las semillas de girasol (Park, Kusakabe, Sakakibara, & Kobayashi, 2001). En resumen, la mayoría de los casos, el procesamiento de las PAs típicas de planta origina una enzima heterodimérica que carece tanto del prosegmento como del PSI, con una organización de los dominios semejante a la de sus homólogos de origen no-vegetal de la familia A1.

1.1.11. Peptidasas aspárticas en flores de cardo

Estudios de localización *in situ* han revelado que las PAs no se distribuyen al azar dentro de la planta, sino que están muy específicamente localizadas en diferentes tejidos según la especie (Simões & Faro, 2004).

Cynara cardunculus L. (cardo de Castilla) en las flores maduras produce dos grupos de peptidasas, cardosinas y cyprosinas, identificadas como PAs típicas de plantas. Incluso cuando la actividad aspartilendopeptidasa se encuentra en niveles relativamente bajos en la planta, estas flores poseen actividad extremadamente alta de PA en ciertos tejidos (Vairo Cavalli, Lufrano, Colombo & Priolo, 2013)

Cordeiro y colaboradores (1994) demostraron la presencia de tres peptidasas glicosiladas con actividad coagulante de leche en extractos de flores secas de *C. cardunculus*, obtenidos a pH alcalino (8.3). Fueron llamadas cynarasas 1, 2, y 3. El nombre cynarasa más tarde fue reemplazada por cyprosina según las convenciones de nomenclatura de PAs (Cordeiro et al., 1994).

Las cardosinas se han purificado y caracterizado a partir de estigmas frescos de *C. cardunculus* (Verissimo, Esteves, Faro, & Pires, 1995), sin embargo, cardosinas y cyprosinas nunca han sido co-purificadas (Pimentel et al., 2007).

Las cyprosinas están específicamente localizadas en las células epidérmicas de los estilos de las flores de *C. cardunculus*. Los genes que codifican tanto para cyprosinas como para cardosinas se organizan como una familia de multigenes que se expresan en las etapas tempranas del desarrollo floral (Pimentel et al., 2007). El ARNm de las cardosinas se detecta en inflorescencias jóvenes pero no en los pistilos de inflorescencias completamente abiertas (Faro et al., 1999; Vieira et al., 2001). La expresión de cardosina A ocurre esencialmente en los pistilos, específicamente en las vacuolas de almacenamiento de proteínas de las papilas estigmáticas así como también en la gran vacuola central de las células del estilo desde el inicio del desarrollo floral (Ramalho-Santos et al., 1998). Cardosina B se acumula en los pistilos hasta las últimas etapas del desarrollo floral localizándose

en las paredes de las células y en la matriz extracelular del tejido de transmisión floral (Vieira et al., 2001).

Las diferencias en actividad y localización, sugieren que las dos cardosinas cumplen funciones biológicas diferentes, para cardosina B se ha propuesto un rol en defensa o en la interacción polen pistilo, cardosina A intervendría en mecanismos proteolíticos mediados por adhesión probablemente asociados con el crecimiento del tubo polínico.

Recientemente se ha propuesto que cardosina A sería una enzima bifuncional, ya que procardosina A tendría un rol en eventos que ocurren durante y luego de la germinación de las semillas de cardo como la desestabilización de la membrana durante el intercambio de solutos y agua en estructuras en crecimiento activo. En los embriones, este zimógeno, se acumula en cuerpos proteicos y paredes celulares. En cambio, cardosina B solo es detectable en las primeras etapas luego de la imbibición de semillas, por lo que le cabría un rol en la pérdida de estructuras limitantes (Pereira et al., 2008).

En estudios posteriores, Sarmiento (2009) purificó y caracterizó cuatro nuevas PAs de pistilos de *C. cardunculus*, elevando así el número de PAs de esta especie a nueve. Estas cuatro enzimas: cardosina E, F, G, y H, son diméricas, glicosiladas, y con actividad máxima a alrededor de pH 4,3. Estas enzimas se asemejan a cardosin A más de lo que se asemejan a cardosina B o a cyprosina.

1.2. DESCRIPCIÓN GENERAL DE LA VÍA SECRETORIA EN PLANTAS.

Una típica célula animal o vegetal contiene hasta 10.000 diferentes tipos de proteínas, mientras que las células de levaduras contienen aproximadamente unas 5000 proteínas. Para un apropiado funcionamiento, cada una de estas numerosas proteínas debe localizarse en un compartimento intracelular preciso, membrana celular u organela, o dirigirse al exterior celular. Tanto las proteínas como otras

cargas se sintetizan y se programan para seguir una determinada vía hasta alcanzar su destino final (Xiang, Etxeberria, & Van Den Ende, 2013) .

Como se indicó en el apartado 1.1.8.1 una secuencia señal Pre dirige los precursores de PAs de planta, siguiendo la vía secretoria hacia las membranas del RE rugoso, una vez translocada al RE la proteína termina de ser sintetizada. En el siguiente apartado se describirá esta vía en forma general.

1.2.1. Componentes de la vía secretoria

En las células eucariotas, el sistema de membranas internas que delimitan distintos compartimientos dentro de la célula, excluyendo plástidos y mitocondrias, a menudo es denominado “sistema de endomembranas”. El mismo está compuesto por organelas específicas, las cuales están principalmente especializadas en la producción y maduración de macromoléculas, su almacenamiento, degradación y reciclado. En términos generales las organelas que constituyen el sistema de endomembranas son: el RE, el aparato de Golgi (AG), las vesículas secretorias (VS), distintas Vacuolas (Vs) y la membrana plasmática (Morré & Mollenhauer, 1974). Generalmente, la vía secretoria de proteínas se inicia en el ER, sigue a través del AG, y finalmente alcanza la vacuola, otros compartimientos o la superficie celular.

Las proteínas no citosólicas, como las proteínas solubles en agua, cuyo destino final es el lumen de un organela, se suelen importar al interior del RE. Esta importación está mediada por una secuencia señal del RE, responsable de la iniciación de su translocación hacia la membrana del RE, que generalmente se asocia con la región N-terminal compuesta por residuos hidrofóbicos. (Alberts et al., 2008). El RE es también el primer lugar para la N-glicosilación de proteínas, así como el control de calidad del plegamiento de las proteínas. De hecho, las proteínas solubles cuyo plegado es incorrecto son exportadas de vuelta al citoplasma donde son degradadas (Stahelin & Driouich, 1997).

En plantas, el AG presenta muchas funciones y rasgos que son similares a lo observado en mamíferos, aunque también existen algunas características únicas (Alberts et al., 2008). Entre otras funciones, es el responsable direccionamiento y correcta entrega de la carga de proteína a una amplia gama de diferentes destinos (Jürgens, 2004). El AG recibe la mayor parte de las proteínas del RE procesadas, a pesar de que la exportación del RE podrá, en situaciones muy particulares, evitar el paso por esta organela por una serie de mecanismos. Una vez dentro del AG las proteínas son adicionalmente procesadas, son glicosiladas o sufren modificaciones de lípidos (Hawes & Satiat-Jeunemaitre, 2005).

La mayor diferencia entre animales y plantas es en cuanto a su estructura, mientras que, en células de mamíferos el AG tiene una posición perinuclear casi estacionaria, en las plantas, está dividido en apilamientos individuales. La parte más próxima al RE se denomina *cis*-Golgi, seguido del Golgi-medio y el más alejado se conoce como *trans*-Golgi; al final se encuentra una red compleja de cisternas la cual se conoce como red del *trans*-Golgi (TGN, *trans golgi network*). Todas estas cisternas, en general, son consideradas funcionalmente independientes (Stachelin & Moore, 1995), ya que cada una puede describirse como una estructura polarizada en términos de morfología cisternal y actividad enzimática, cambiando gradualmente desde el *cis*- al *trans*-Golgi (Fitchette-Lainé, Gomord et al., 1997). Por otro lado, de manera más específica, en el sistema de endomembranas de células de planta encontramos el compartimiento prevacuolar (PVC, *prevacuolar compartment*) que está definido como una organela que recibe las proteínas desde las vesículas secretorias y posteriormente las transporta a las vacuolas (Bethke & Jones, 2000). Basado en el funcionamiento de los sistemas de endomembranas de levaduras o mamíferos, el PVC funciona además como organela intermediaria en la vía endocítica desde la membrana plasmática al Golgi (Van Deurs, Holm, Kayser, Sandvig, & Hansen, 1993).

Las vacuolas, junto con la membrana plasmática, son el punto más distal de la vía secretoria, muchas proteínas vacuolares son transportadas desde el RE a través de compartimentos intermedios. Las vacuolas de planta son organelas multifuncionales, cumplen funciones tanto físicas como metabólicas cruciales para la célula. Estas funciones vitales y esenciales para la planta son por ejemplo el mantenimiento de la turgencia, la homeostasis protoplasmática, almacenamiento tanto de metabolitos secundarios como de proteínas que luego son degradadas para proveer carbono, nitrógeno y fuente de azufre implicados en el crecimiento y el desarrollo de la planta así como en respuesta celular a factores ambientales y bióticos (Viotti, 2014; Xiang et al., 2013). El número y tamaño de las vacuolas dependen del tipo de célula y de la etapa de desarrollo. Las vacuolas tienen diferentes formas, tamaño, contenido y función; una vacuola central única puede ocupar hasta el 90% del volumen de una célula (Viotti, 2014). Las plantas presentan al menos dos tipos de vacuolas funcionalmente distintas (Fig. 9): vacuolas líticas (LV) y vacuolas de almacenamiento de proteínas (PSV) (Herman & Larkins, 1999; Hoh, Hinz, Jeong, & Robinson, 1995; Paris, Stanley, Jones, & Rogers, 1996; Alessandro Vitale & Raikhel, 1999).

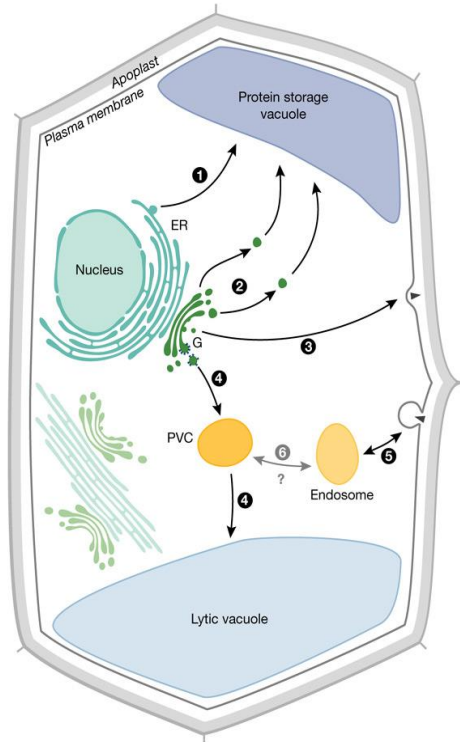


Figura 9. Mapa simplificado de la vía secretora de las plantas superiores. Las proteínas vacuolares se sintetizan y se procesan en el RE y se transfieren a vacuolas a través de diversas rutas, por ejemplo, se puede producir la transferencia directa de algunas proteínas de almacenamiento a la vacuola (ruta 1). Las proteínas destinadas a la vacuola de almacenamiento de proteínas pueden dejar el AG en compartimentos *cis* y *trans*-Golgi (ruta 2), mientras que las proteínas destinadas a la exocitosis lo hacen predominantemente por el *trans*-Golgi (ruta 3). Los productos destinados a la vacuola lítica lo hacen por el *trans*-Golgi y se encaminan de manera dependiente del receptor clatrina a través del compartimento PVC (ruta 4). El material de endocitosis y las proteínas de la membrana plasmática recicladas se internalizan en un compartimento tipo endosoma (ruta 5), que puede transferir el material fusionándose o madurando en el PVC (ruta 6). La función exacta de las vesículas en la transferencia entre los endosomas putativos, los PVCs y una vacuola lítica no está claramente establecida. (Tomado de Brandizzi & Hawes, 2004)

El transporte de proteínas solubles y de membranas entre estas organelas está mediado por distintas vesículas de transporte, donde cada una mantiene una composición única de su membrana y de las proteínas que transporta. Entre ellas podemos mencionar, las vesículas que median el transporte entre el RE y el AG, denominadas COP's (COP-I y COP-II, *coat protein complexes I y II*, respectivamente) o también, las encargadas del transporte desde estas organelas y su destino final como las vesículas cubiertas de clatrina (CCV, *Clathrin-Coated Vesicles*), las vesículas PAC (*Precursor Accumulating vesicles*) descritas en zapallo (*Cucurbita maxima*) y las vesículas densas (DV, *Dense Vesicles*), entre otras (Hara-Nishimura, Shimada, Hatano, Takeuchi, & Nishimura, 1998; Hillmer, Movafeghi, Robinson, & Hinz, 2001; A Vitale & Pedrazzini, 2005).

Las proteínas vacuolares llegan a los diferentes tipos de vacuolas a través de una vía de tráfico y biosíntesis mediada por vesículas que incluye el RE, el aparato de Golgi, el TGN y endosomas/ PVC. En las células vegetales, se conocen dos tipos diferentes de receptores de membrana: la familia de receptores de tráfico vacuolar (VSR, *vacuolar sorting receptor*) y la familia de receptores de membrana anillo-H2 (RMR, *receptor membrane ring-H2*) (Viotti, 2014).

Con respecto al destino final de las PAs, se ha observado que diferentes PAs se acumulan en compartimentos distintos, y la misma proteína puede ser secretada al apoplasto o dirigida a vacuolas, dependiendo del tipo de célula, y estado de desarrollo, sugiriendo un mecanismo estricto en la regulación del tránsito (da Costa, Pereira, Moore, & Pissarra, 2010; Oliveira et al., 2010).

1.2.2. Tipos de vacuolas

Como ya se indicó dos tipos de vacuolas pueden ser encontrados en las células vegetales, la vacuola de almacenamiento de proteínas (PSV) y la vacuola lítica (LV). Por lo general, las proteínas de almacenamiento vacuolar se acumulan en PSVs. Las PSVs tienen pHs más altos y actividades hidrolíticas más bajas que las LVs, y predominan en los tejidos de almacenamiento (cotiledones, endosperma en semillas y tubérculos), así como en los tejidos vegetativos de las plantas adultas (corteza, hojas, vainas). Las proteínas almacenadas en PSV se utilizan principalmente como fuente de nitrógeno o de carbono durante la germinación de semillas y desarrollo de las plantas (Wang et al., 2007). Sin embargo, las PSVs también pueden contener grandes cantidades de proteínas tóxicas (lectinas, inhibidores de péptidasas y proteínas inhibidoras de ribosomas), que pueden ser considerados como el resultado de un proceso de desintoxicación citosólica o como una defensa contra depredadores. Las proteínas almacenadas en PSVs podrían ser procesados a través de la acción de enzimas de procesamiento vacuolar o peptidasas. Las LVs se encuentran generalmente en los tejidos

vegetativos, tienen un pH ácido y contienen una gran cantidad de enzimas hidrolíticas. Las LVs se utilizan tanto para el almacenamiento como para el depósito de materiales no deseados de las células vegetales. Reciben componentes extracelulares a través de endocitosis y fagocitosis, y material intracelular a través de autofagia, así como a través de la vía de tráfico biosintética y los sistemas de transporte de membrana. Las LVs modulan la degradación de una multitud de macromoléculas y otros compuestos. Como tales, se consideran reguladores clave en la homeostasis celular. Dependiendo de las condiciones de fisiológicas, las LVs se pueden transformar en PSVs y viceversa (Xiang et al., 2013).

1.2.3. Señales de direccionamiento vacuolar

La señal de direccionamiento es parte de la secuencia de la proteína. Hay tres grupos principales de determinantes de direccionamiento vacuolar (VSD) de plantas identificados hasta el momento. Los VSDs son necesarios para el correcto direccionamiento hacia la vacuola, sin estos motivos, las proteínas vacuolares siguen el camino por defecto y son secretadas hacia la superficie celular.

1) VSDs de secuencia específica (ssVSDs) consisten en motivos conservados, como NPIRL o similares, y no toleran grandes modificaciones. Por lo general, este tipo de señal se encuentra en el extremo N-terminal de una proteína (Nakamura & Matsuoka, 1993). Se considera que los ssVSDs son reconocidos por VSRs para su tráfico a las LV.

2) VSDs C-terminales (ctVSD) (Nakamura & Matsuoka, 1993; Neuhaus & Rogers, 1998). Estas señales no son conservadas, ni tienen un tamaño definido.

3) VSD dependiente de la estructura física de la proteína (psVSD). Estas señales se encuentran en el centro de la proteína e implican uno o más motivos,

los cuales se exponen sólo cuando la proteína adquiere su conformación plegada (Jolliffe, Ceriotti, Frigerio, & Roberts, 51AD; Neuhaus & Rogers, 1998)

En las plantas, los ctVSDs fueron descubiertos en lectina de cebada y quitinasa de tabaco. La longitud mínima para estos motivos es de cuatro aminoácidos. Muchos péptidos aleatorios C-terminal son suficientes para dirigir una proteína reportera a la vacuola. Por ejemplo, el ctVSD de quitinasa A de tabaco GLLVDTM y los motivos FAEAI y LVAE de lectina de cebada, son necesarios y suficientes para el direccionamiento vacuolar (Xiang et al., 2013). Los ctVSDs deben estar localizados estrictamente en el C-terminal de la proteína. Más aún, la introducción de un sitio de glicosilación C-terminal o la adición de alanina en el C-terminal lleva a la proteína a la superficie celular. Cabe destacar, que los ctVSDs son distintivos de las plantas. Algunas proteínas incluso combinan un ssVSD y un ctVSD para el direccionamiento dual a la matriz de PSV o de globoides (Xiang et al., 2013). Se demostró que la región C-terminal de la subunidad alfa de beta-conglicinina de soja tiene dos señales de direccionamiento, una del tipo ctVSD y otra del tipo ssVSD. Ambos funcionan para el tráfico hacia las PSVs, pero no está claro si ambos utilizan el mismo VSR (Pereira, Pereira, & Pissarra, 2014; C. Pereira, Pereira, Satiat-Jeunemaitre, & Pissarra, 2013). Otro estudio reciente demostró la existencia de dos VSDs en la misma proteína con propiedades diferentes: una ctVSD típica y un dominio de direccionamiento vacuolar no convencional, el PSI, ambos en la secuencia de cardosin A (Pereira et al., 2014).



Objetivos

Los genomas de plantas codifican cientos de peptidasas lo que representa decenas de familias no relacionadas entre sí de estas enzimas. Esta diversidad funcional se correlaciona con los datos moleculares: las peptidasas se expresan específicamente en tiempo y espacio y se acumulan en diferentes compartimentos subcelulares. La caracterización bioquímica, funcional y estructural de estas enzimas permitirá avanzar en el área del conocimiento de las relaciones estructura-función de las mismas.

Las peptidasas aspárticas han sido empleadas como herramientas en distintos procesos, en la fermentación y manufactura de una variedad de productos, incluyendo producción de quesos, péptidos bioactivos, nutraceuticos y formación del sabor. El conocimiento de los motivos estructurales y funcionales y la comparación con otras peptidasas ya estudiadas permitirá el diseño de nuevas peptidasas, mediante ingeniería de proteínas, que puedan resultar atractivas para su potencial aplicación en determinados procesos biotecnológicos.

Extractos de flores de *Silybum marianum*, una especie de la familia Asteraceae, fueron probados en ensayos de coagulación de leche mostrando capacidad coagulante debida a la presencia de PAs para ser potencialmente utilizados en la industria láctea. Esta especie comúnmente conocida como *cardo mariano* es considerada una maleza en nuestra región.

El presente trabajo intenta generar un aporte al conocimiento de las PAs de origen vegetal mediante el estudio de las peptidasas de flores de *S. marianum*.

El **objetivo general** de este trabajo de investigación es caracterizar peptidasas aspárticas y sus precursores a partir de flores de *Silybum marianum* (familia Asteraceae) que crecen en nuestro país.

Para llevar adelante este propósito, se plantearon los siguientes **objetivos específicos** que contribuyen a lograr el objetivo general:

1. *Clonar y secuenciar los ADNc de PAs de S. marianum.*
 - a. *Aislar el ARN total de flores de S. marianum.*
 - b. *Clonar los precursores de las aspartilproteasas (preproenzimas) a partir del ARNm*
 - c. *Obtener la secuencia nucleotídica y de aminoácidos de estos precursores.*
 - d. *Predecir sus propiedades fisicoquímicas.*
 - e. *Realizar un estudio comparativo con los precursores de otras peptidasas aspárticas conocidas.*
2. *Predecir la estructura molecular de las peptidasas clonadas por modelado por homología a partir de la secuencia aminoacídica y en base al modelo estructural, realizar una predicción funcional.*
 - a. *Predecir motivos estructurales y funcionales en las peptidasas clonadas.*
 - b. *Generar un modelo tridimensional de las peptidasas clonadas.*
 - c. *Validar el modelo tridimensional.*
3. *Estudiar la localización subcelular de las peptidasas por expresión transiente en Nicotiana benthamiana y microscopía confocal.*
 - a. *Construir vectores de expresión para uno de los precursores y su mutante fusionados a proteínas fluorescentes.*
 - b. *Expresar en forma transiente las proteínas de fusión en N. benthamiana*
 - c. *Determinar la localización subcelular de las proteínas por microscopía confocal.*

Materiales



3.1. CEPAS BACTERIANAS

3.1.1. Cepas de *Escherichia coli*

| Cepa | Genotipo/Descripción | Aplicación |
|--|---|---|
| XL1-Blue (Stratagene, <i>Agilent Technologies, Santa Clara, CA</i>) | <i>endA1 gyrA96 (nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB⁺ lacI^q Δ(lacZ) M15] hsd R17(r_K⁻ m_K⁺)</i> | |
| TOP10F (Invitrogen, <i>Invitrogen Corp., Carlsbad, CA</i>) | <i>F-mcrA Δ(mrr-bsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL (Str^R) endA1 λ</i> | |
| TOP10F' (Invitrogen) | <i>F' [lacI^q Tn10(tet^R)] mcrA Δ(mrr-bsdRMS-mcrBC) φ80 lac ZΔM15 ΔlacX74 deo RnupG recA1 araD139 Δ(ara-leu)7697 gal Ugal KrpsL (Str^R) endA1 λ</i> | Clonado general, propagación de plásmidos |
| DH5α (Invitrogen) | <i>F-endA1 glnV44 thi-1 recA1 relA1 gyrA96 deo RnupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK-mK+), λ-</i> | |
| GC10 (Sigma-Aldrich, <i>San Luis, MO</i>) | <i>F- mcrA Δ(mrr-bsdRMS-mcrBC) Φ80 dlacZ ΔM15 ΔlacX74 endA1 recA1 Δ(ara, leu)7697 araD139 gal Ugal Knup GrpsL λ- T1R</i> | |
| DB 3.1 (Invitrogen) | <i>F-gyrA462 endA1 glnV44 Δ(sr1-recA) mcr Bmrr hsdS20(r_B⁻, m_B⁻) ara14 galK2 lacY1 proA2 rpsL20 (Smⁱ) xyl5 Δleu ml1</i> | Propagación y mantenimiento de vectores Gateway |
| Arctic Express™ (DE3) (Stratagene) | <i>E. coli B F- omp Tbs dS (rB- mB-) dcm+ Tetr gal λ (DE3) endA Hte [cpn10cpn60 Gentr]</i> | Expresión en <i>E. coli</i> |

Tabla 1. Cepas de *E. coli* utilizadas en el clonado y expresión de peptidasas.

Las cepas utilizadas, así como su genotipo y aplicación se muestran en la Tabla 1.

3.1.2. Cepas de *Agrobacterium tumefaciens*

A. tumefaciens GV3101: contiene integrado en su cromosoma, secuencias derivadas del plásmido Ti desarmado pTiC58 y un gen de resistencia a rifampicina. Además posee el plásmido helper pMP90 (pTiC58 γ Δ T-DNA) que confiere resistencia a gentamicina y provee los genes *vir*, necesarios para la transformación del T-DNA dentro de la célula vegetal.

3.2. MEDIOS DE CULTIVO

Los distintos medios empleados para el crecimiento de cultivo de *E. coli* y *A. tumefaciens* se detalla en la Tabla 2.

| Medio | Componente | Concentración | Aplicación |
|---------------------------------|---|---------------|--|
| Medio Luria-Bertani (LB) | <i>Extracto de levadura</i> | 5 g/L | Crecimiento y mantenimiento de <i>E. coli</i> y <i>A. tumefaciens</i> |
| | <i>NaCl</i> | 10 g/L | |
| | <i>Triptona</i> | 10 g/L | Expresión de proteínas recombinantes |
| | Solo para medio sólido: <i>Agar bacteriológico</i> | | |
| SOC | <i>Bactotriptona</i> | 20 g/L | Crecimiento de células competentes transformadas mediante choque térmico |
| | <i>Extracto de levaduras</i> | 5 g/L | |
| | <i>NaCl</i> | 0,6 g/L | |
| | <i>KCl</i> | 1,86 g/L | |
| | <i>MgCl₂</i> | 2,03 g/L | |
| | <i>MgSO₄</i> | 1,2g/L | |
| | <i>Glucosa</i> | 3,6 g/L | |

| | | | |
|------------|-----------------------------|-------|---|
| YEB | <i>Peptona</i> | 5 g/L | Crecimiento y mantenimiento de cepas de <i>A. tumefaciens</i> |
| | <i>Extracto de levadura</i> | 1g/L | |
| | <i>Extracto de carne</i> | 5g/L | |
| | <i>MgCl₂</i> | 5 g/L | |
| | <i>Sacarosa</i> | 5 g/L | |

Tabla 2. Medios de cultivo empleados en el crecimiento y expresión de *E. coli* y *A. tumefaciens*

3.3. SOLUCIONES BUFFERS

Las soluciones tampón que se utilizaron en el presente trabajo se describen en la Tabla 3.

| Buffers | Descripción | Aplicación |
|-------------------------------------|--|--|
| TAE | Tris 40 mM pH 7,6; ácido acético 20 mM; EDTA 1mM. | Preparación de geles de agarosa y como buffer de corrida en las electroforesis de ADN. |
| TE | Tris-Cl 10 mM pH 7,5; EDTA 1mM. | Resuspensión de oligonucleótidos. |
| TBF I | RbCl 100 mM; MnCl ₂ 50 mM; Acetato de Potasio 30 mM; CaCl ₂ 10 mM; Glicerol 15%; pH 5,8. | Preparación de células competentes químicas. |
| TBF II | MOPS 10 mM; RbCl 10 mM; CaCl ₂ 75 mM; Glicerol 15%; pH 6,5. | Preparación de células competentes químicas. |
| Buffer de lisis | KH ₂ PO ₄ /Na ₂ HP0 ₄ 50mM; NaCl 80mM pH 7,4. | Expresión de PSI recombinante |
| Buffer de muestra 1 × | Tris 65mM, Glicerol 8ml, SDS 69,35 mM, β-mercaptoetanol 5 ml, Azul de bromofenol 30 mM, pH 6,8 (se llevó a pH con HCL 1N). | Preparación de muestras de electroforesis |
| Buffer del gel de resolución | Tris 3 M pH8,8 (se llevó a pH con HCL 1N). | Preparación de geles de electroforesis SDS-PAGE |
| Buffer del gel de stacking | Tris 499,4 mM pH 6,8 (se llevó a pH con HCL 1N). | Preparación de geles de electroforesis SDS-PAGE |

| | | |
|-----------------------------------|---|---|
| Buffer del Reservorio 10 × | Tris 247,65 mM , glicina 1,92 M, SDS 34,68 mM. | Buffer de corrida para electroforesis SDS-PAGE |
| Buffer de transferencia | Tris-HCl 25 mM pH 8,3; glicina 192 mM, metanol 20%. | Buffer de transferencia para Western blot |
| Buffer PBS | NaCl 136,99mM, KCL 2,68mM, Na ₂ HP0 ₄ .2 H ₂ O 4,77 mM, K ₂ HP0 ₄ 0,014 mM pH 7,4. | Preparación de anticuerpos para el Western blot |
| Buffer PBS-Tween 20 | Buffer PBS con Tween al 0,05%. | Buffer de lavado para el revelado de Western blot |
| Medio de Infiltración | MgCl ₂ 10 mM, ácido morfolinoetansulfónico (MES) 10 mM pH 5,7 | Agroinfiltración de <i>N. benthamiana</i> |

Tabla 3. Soluciones tampón utilizadas en las distintas técnicas empleadas en este trabajo de Tesis

3.4. SOLUCIONES DE ANTIBIÓTICO

Para la selección de clones transformados de *E. coli* y *A. tumefaciens* se prepararon soluciones stock con una concentración 1000 veces superior (1000 ×) a la de trabajo en agua excepto para los antibióticos tetraciclina y cloranfenicol para los que se empleó etanol 70%, como se muestran en la Tabla 4.

| Antibiótico | Stock (mg/ml) | Concentración final <i>E. coli</i> (µg/mL) | Concentración final <i>A. tumefaciens</i> (µg/mL) |
|--|------------------------|--|---|
| Ampicilina (Ap) (Bagó) | 100 (H ₂ O) | 100 | No se usa |
| Tetraciclina (Tc) (Bagó) | 5 (50%EtOH) | 5 | 5 |
| Kanamicina (Km) (MP Biomedicals) | 100 (H ₂ O) | 50 | 100 |
| Gentamicina (Gm) (Bagó) | 50 (H ₂ O) | 20 | 50 |
| Estreptomocina (St) | 100 (H ₂ O) | 25-100 | 300 |

| | | | |
|-----------------------------|-----------------------|-----|-----|
| <i>(Sigma-Aldrich)</i> | | | |
| Espectinomicina (Sp) | 50 (H ₂ O) | 50 | 100 |
| <i>(MP Biomedicals)</i> | | | |
| Cloranfenicol (Cm) | 30 (EtOH) | 30 | 50 |
| <i>(Sigma-Aldrich)</i> | | | |
| Rifampicina (Rf) | 100 (DMF) | 100 | 100 |
| <i>(Genbiotech SRL)</i> | | | |

Tabla 4. Antibióticos utilizados en clonado y expresión. En la columna de solución stock se indica entre paréntesis el solvente correspondiente

3.5. VECTORES

3.5.1. Clonación

3.5.1.1. pGEM®-T Easy

El vector de clonación pGEM-T Easy, *Promega Corporation, Madison, WI*, se encuentra linealizado y presenta una timina 3' terminal en ambos extremos del sitio de inserción (Fig. 1), esto previene la recircularización del vector y genera una unión compatible con cualquier producto de la reacción en cadena de la polimerasa (PCR) obtenido por empleo de polimerasa Taq. Esta característica da versatilidad a la vez que aumenta la eficiencia de ligación de los productos de PCR al vector.

Debido a que el sitio de inserción en el vector pGEM-T Easy interrumpe el marco de lectura abierto de la enzima β -galactosidasa, los clones que incorporaron el vector conteniendo el inserto a clonar, pueden ser identificados como colonias blancas resistentes a Apen la placa de crecimiento. Cuando el gen de la enzima β -galactosidasa no es interrumpido, las colonias resultan de color azul.

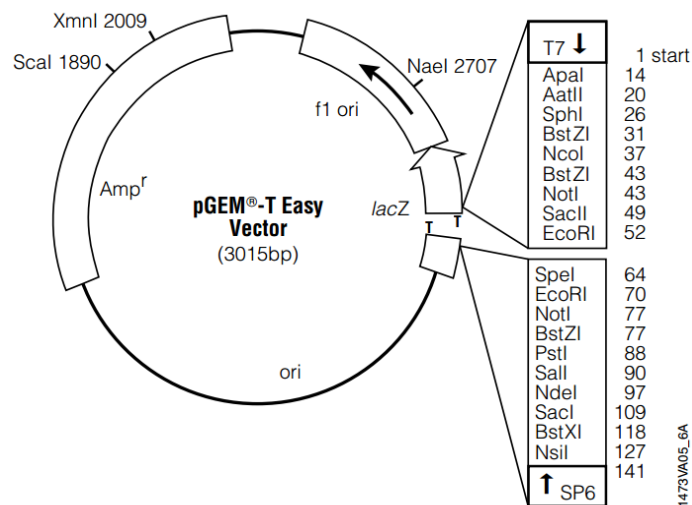


Figura 1. Mapa génico del vector de clonación pGEM-T Easy

3.5.1.2. pDONR 222

El vector de clonado pDORN222, Invitrogen, de tecnología de clonado Gateway contiene un origen de replicación y mantenimiento bacteriano de alta copia (ori pUC) y un marcador de resistencia a Km (Fig. 2). Además posee una región que contiene dos sitios de recombinación, attP1 y attP2, que facilitan la clonación por recombinación del gen de interés a partir de producto de PCR o del clon de expresión flanqueado por los sitios attB. Luego de la recombinación, el fragmento de ADN lineal se intercambia con la región del plásmido situada entre los sitio attP. Los vectores del sistema Gateway contienen el gen *ccdB*, entre los sitios attP, que codifica para una proteína que inhibe la enzima ADN girasa, una proteína esencial en la replicación del ADN. La presencia de *ccdB*, por tanto, inhibe el crecimiento en la mayoría de las cepas bacterianas. Los plásmidos que han perdido el gen *ccdB* a través de la recombinación con el fragmento de ADNc serán capaces de transformar las células y serán resistentes a Km. Los plásmidos que no se recombinan con el fragmento de ADNc retendrán el gen *ccdB*, el que será letal cuando se transforme en las bacterias.

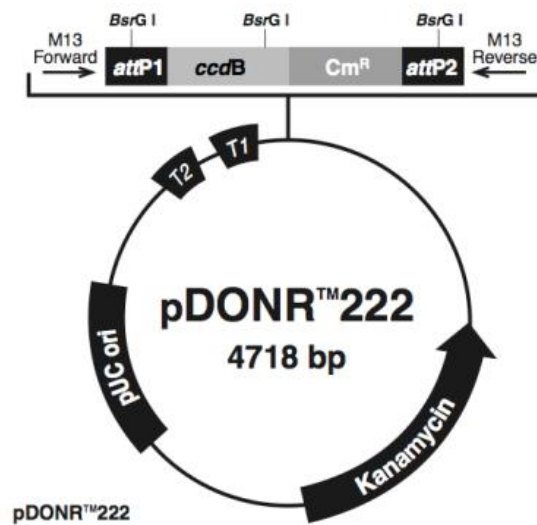


Figura 2. Mapa génico del vector de clonación pDONR222

3.5.2. Expresión heteróloga en sistema eucariota

Los vectores del sistema Gateway derivan del plásmido pPZP200. Este plásmido es relativamente pequeño (6,7 kb), contiene un origen de replicación para *E. coli* (ColE1) y otro para *A. tumefaciens* (pVS1), tiene un sitio para la movilización de *E. coli* a *A. tumefaciens* denominado bom (basis of mobility), proveniente del plásmido pBR322, posee el gen de resistencia a St/Sm. Las regiones fronterizas del ADN-T, derivan del gen nos que codifica para la enzima nopalina-sintasa (proveniente del plásmido Ti, pTiT37) y contiene una repetición de 25 bp y regiones fronterizas exteriores pero no las secuencias de ADN-T internas; por lo tanto, las secuencias de ADN-T naturales no se transfieren a las plantas transformadas. El vector de destino tiene además un marcador de selección para planta, el gen bialafosacetil-transferasa (*bar*), que confiere resistencia a glufosinato de amonio. El marcador está bajo la regulación transcripcional del promotor de nos y el terminador nos y se ubican en el lado izquierdo del ADN-T, mientras que el sitio de recombinación de Gateway se ubica del lado derecho del ADN-T. Para la sobreexpresión de una secuencia de ADN, cuenta con el

promotor y el terminador de la transcripción del virus mosaico de la coliflor (CaMV) 35S.

Para las transformaciones mediadas por *A. tumefaciens* se utilizaron los vectores binarios que se describen a continuación.

3.5.2.1. pB7FWG2

Estructura del T-ADN: (LB-[Tnos-BarR-Pnos]-[T35S-eGFP:(attR2, GW, attR1)-P35S]-RB) (VIB, University of Gent, Belgium). La Fig. 4 muestra el mapa génico del plásmido.

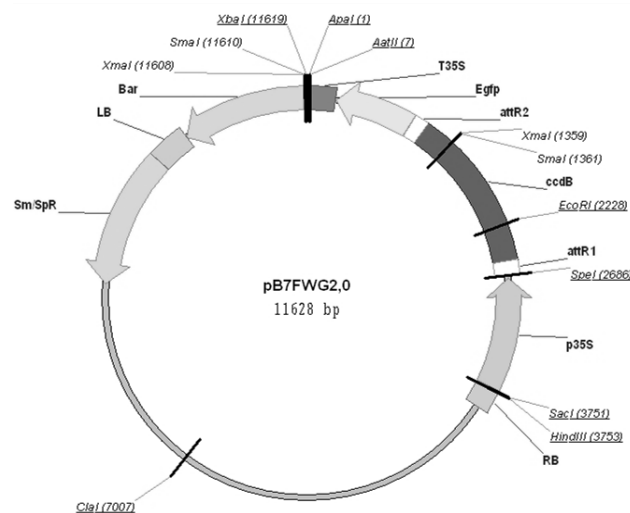


Figura 4. Mapa génico del vector binario pB7FWG2

El casete marcador contiene la secuencia codificante de la proteína fluorescente verde (GFP) (Egfp; Clontech Laboratories, Palo Alto, CA, EEUU). El eGFP es una versión mejorada de GFP por las siglas en inglés “enhanced Green fluorescent protein” (Heim, Cubitt, & Tsien, 1995), tiene la ventaja de ser una proteína compacta de pequeño tamaño, de este modo la fusión de la proteína GFP no modifica significativamente ni la función ni la localización subcelular de la proteína nativa a la que se ha unido.

3.5.2.2. pB7FRWG2

Estructura del T-ADN: (LB-[Tnos-BarR-Pnos]-[T35S-mRFP:(attR2, GW, attR1)-P35S]-RB) VIB, University of Gent, Belgium. En la Fig. 5 muestra el mapa génico del plásmido

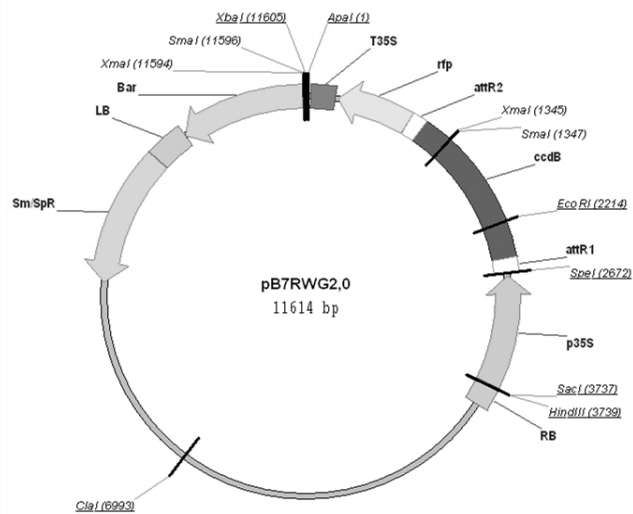


Figura 5. Mapa génico del vector binario pB7RFWG2

El casete marcador contiene la secuencia codificante de la proteína fluorescente roja RFP. El RFP es un análogo del GFP proveniente del coral *Discosoma*, capaz de emitir en el rango del rojo. La proteína nativa denominada DsRed (Campbell et al., 2002; Heikal, Hess, Baird, Tsien, & Webb, 2000), es de gran tamaño y tetramérica, razón por la cual se generó una versión monomérica de la DsRed. La proteína resultante se denominó mRFP por las siglas de *monomeric red fluorescente protein*.

3.6. CEBADORES

Se utilizaron distintos oligonucleótidos que se detallan en la Tabla 5.

| | | Cebador | Secuencia(5' -> -3') | Tm(°C) |
|--|-----|-----------|---|---------|
| Obtención delADNc | RT | OligodT | CCGGAATTCACTGCAGGG | 63,3 |
| | | | TACCCAATACGACTCACTA | |
| | | | TAGGGCTTTTTTTTTTTTTT | |
| | | | TTTT | |
| Clonación en pGEM-TEasy | PCR | Prepro_FW | atgggtaactcaatcaaagca | 56 |
| | | Prepro_Rw | tcaagctgcttctgcaaatcc | 56 |
| Clonación en pDONR222 | PCR | F4F | GGGGACAAGTTTGTACA AAAAAGCAGGCTTCGAA <u>GGAGATAGAACC</u> Catgggtaa ctcaatcaaagc | 52Δ30°C |
| | | F4R | GGGGACCACTTTGTACA AGAAAGCTGGGTG agctgct tctgcaaatccca | |
| Confirmación de clonado en vectores pDONR222 | PCR | 523 (F) | gtaaaagcagcgcagctctaagctcggg | 72 |
| | | 524 (R) | c cgggccagagctgccaggaacag | |
| Mutantes de sitio activo (<i>preprosilpepsina 2</i>) | PCR | 618 (R) | cccatagatcagaact tcctgtcg aaa aatcac | 68,9 |
| | | 619 (F) | tttcacagtgatttt gcgacagga agtt c | 67,3 |

Tabla 5. Secuencias de los oligonucleótidos utilizados. La Tm fue calculada por el programa Oligoanalyzer (<http://www.idtdna.com/analyzer/applications/oligoanalyzer/>) Δ: gradiente de temperatura. En mayúscula y negrita: GGGG-attB1 y GGGG-attB2; en mayúscula y subrayado con línea entera: secuencia de Shine-Dalgarno seguida de la secuencia Kozak (subrayado en línea punteada); en minúscula: secuencia específica del gen; en minúscula, negrita y subrayado: triada catalítica Asp-Thr-Gly (DTG); en minúscula y rojo: nucleótidos modificados para generar mutación

3.7. ENZIMAS

Transcriptasa reversa M-MuLV RT (virus de la leucemia murina de Moloney, Fermentas Life Sciences, Glen Burnie, MD, EEUU): La

transcriptasa reversa o retrotranscriptasa es una enzima de tipo ADN-polimerasa, que tiene como función sintetizar ADN de doble cadena utilizando como molde ARN monocatenario o ADN de cadena simple, es decir, catalizar la retrotranscripción o transcripción inversa. La transcriptasa inversa M-MuLV carece de actividad 3' → 5' exonucleasa

ADN polimerasa GoTaq (Promega): esta enzima posee además actividad transferasa terminal 3' → 5' incorporando en los extremos de las secuencias amplificadas bases nucleótidos de timidina adicionales.

BP Clonase®II (Invitrogen): es una mezcla de enzimas del sistema Gateway® que catalizan la recombinación *in vitro* de productos de PCR o segmentos de ADN subclonados a partir de clones (que contiene los sitios attB) y un vector donador (que contiene los sitios attP) para generar clones de entrada.

LR Clonase®II (Invitrogen): es una mezcla del sistema Gateway® que contiene enzimas (Integrasa, IHF, Integración Factor Host y X escisionasa) que catalizan la recombinación *in vitro* entre un clon de entrada, que contiene un gen de interés flanqueado por sitios attL y un vector de destino, que contiene los sitios attR para generar el clon de expresión.

3.7.1. Enzimas de restricción

EcoRI (*Fermentas Life Sciences*)

Bsp1407I (isoesquisómero de BsrGI)

SaI (*Promega Corporation*)

3.8. MATERIAL VEGETAL

Para la realización de este trabajo de tesis inflorescencias de *Silybum marianum* (L.) Gaertn. se recolectaron en los partidos de Tandil, Olavarría y La Plata (Pcia. de Buenos Aires) durante los meses de octubre y noviembre. Una muestra ha sido depositada en el herbario LPE de la Facultad de Ciencias Exactas de la

Universidad Nacional de La Plata, Argentina con el número de voucher LPE 1162.

Plantas de *Nicotiana benthamiana* wt jóvenes (4-5 semanas), crecidas en cámara de cultivo a 23° C con fotoperíodo 16:8 hs (luz: oscuridad) fueron empleadas en los ensayos de expresión transiente.

3.9. HERRAMIENTAS BIOINFORMÁTICAS

3.9.1. Herramientas generales

Oligo Analyzer 3.1 (<https://www.idtdna.com/calc/analyzer>, © Copyright 2016 Integrated DNA Technologies, Inc). Servidor empleado para calcular las temperaturas de *melting* de los oligonucleótidos y el contenido de guanina-citocina. Además, permite evaluar la formación de *hairpins* y dímeros de *primers*.

BioEdit Sequence Alignment Editor (copyright© 1997-2013 Tom Hall). Programa que permite analizar y editar secuencias de ADN.

DNASTartLasergene. Software que ofrece herramientas en los campos de biología molecular, genómica y biología estructural. La herramienta dedicada a la biología molecular proporciona aplicaciones que permiten la alineación de secuencias de ADN, ARN, y proteínas utilizando múltiples métodos de alineación que incluyen MUSCLE, Mauve, MAFFT, Clustal Omega, Jotun Hein, Wilbur-Lipman, Martinez Needleman-Wunsch, Lipman-Pearson y análisis Dotplot

T-coffee (<http://tcoffee.crg.cat/apps/tcoffee/do:regular>). Servidor utilizado para realizar alineamientos múltiples (Di Tommaso et al., 2011).

Restriction Mapper versión 3 (<http://www.restrictionmapper.org>). Servidor utilizado para llevar a cabo la digestión virtual a fin de establecer los patrones de restricción esperados.

PlasMapper versión 2.0 (<http://wishart.biology.ualberta.ca/PlasMapper/>). Servidor que permite generar mapas de vectores plasmídicos empleando sólo una secuencia de ADN como *input* (Dong, Stothard, Forsythe, & Wishart, 2004).

ProtParam (<http://web.expasy.org/protparam>). Servidor que permite el cálculo de diferentes parámetros físicoquímicos de una proteína. Los parámetros calculados incluyen peso molecular, pI teórico, composición de aminoácidos, composición atómica, coeficiente de extinción, índice de alifaticidad y promedio General de Hidropatía (GRAVY, (Kyte & Doolittle, 1982).

Signal IP 4.0 (www.cbs.dtu.dk/services/SignalP). Servidor que predice la presencia y ubicación de los sitios de corte del péptido señal en las secuencias de aminoácidos de diferentes organismos, procariontes Gram-positivos, procariontes Gram negativos y eucariotes (Petersen et al., 2011).

NetNGlyc 1.0 (www.cbs.dtu.dk/services/NetNGlyc). Servidor que predice los sitios de N-glicosilación en proteínas y determina el contexto de la secuencia Asn-Xaa-Ser/Thr (Gupta, Jung, & Brunak, 2004).

ProDom (Protein Domain Database) (<http://prodom.prabi.fr/prodom/current/html/form.php>). Base de datos de dominio de familias de proteínas generados a partir de la comparación global de todas las secuencias de proteínas disponibles (Kahn, Rezvoy, & Vivien, 2008).

Pfam (Protein Families Database) (<http://pfam.xfam.org>). Base de datos de familias de proteínas, cada una representada por alineamientos múltiples de secuencias y modelos HMMs (hidden Markovmodels) cubriendo dominios proteicos y familias comunes (Finn et al., 2014).

MEME suite 4.11.1 (Multiple Em for Motif Elicitation). Servidor que permite descubrir los motivos en un grupo de ADN o secuencias de proteínas relacionadas. MEME toma como entrada un grupo de secuencias de proteínas o

ADN y genera motivos con un umbral de confianza estadístico especificado por el usuario (Bailey & Elkan, 1994).

ELM server (<http://elm.eu.org>) es un recurso de biología computacional que permite investigar candidatos a sitios funcionales en proteínas eucariotas. La base de datos contiene más de 240 clases diferentes motivos con más de 2700 sitios comprobados experimentalmente, curada manualmente a partir de más de 2400 publicaciones científicas (Holger et al., 2016; Puntervoll et al., 2003).

3.9.2. Herramientas empleadas para el modelado molecular por homología

HHPred (<http://toolkit.tuebingen.mpg.de/hhpred>). Servidor utilizado para realizar la asignación de plegamiento (*fold assignment*). Permite la detección de proteínas homólogas remotas y la predicción estructural a partir de una secuencia problema empleando un método de comparación HMM-HMM (Söding, Biegert, & Lupas, 2005).

Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Servidor empleado para realizar el alineamiento entre la secuencia molde y la secuencia a modelar.

Gene Doc Multiple Sequence Alignment Editor and Shading Utility (versión 2.7.000). Este programa fue utilizado para analizar y editar alineamientos.

Modeller, versión 9.14 (*copyright*© 1989-2015 *Andrej Sali*). Programa que permite modelar por homología o comparación de proteínas, se utiliza para modelar una proteína cuya estructura molecular no es conocida empleando como molde una proteína homóloga cuya estructura ya ha sido resuelta (Sali & Blundell, 1993).

ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>). Servidor *online* que permite la evaluación energética de modelos tridimensionales de proteínas (Wiederstein & Sippl, 2007).

PROCHECK (<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>).

Programa empleado para evaluar la calidad estructural de los modelos tridimensionales (Laskowski, MacArthur, Moss, & Thornton, 1993).

ProFunc (<https://www.ebi.ac.uk/thornton-srv/databases/profunc/>). Servidor que permite la identificación de la función bioquímica más probable de una proteína a partir de las coordenadas moleculares de su estructura tridimensional.

3.9.3. Herramientas empleadas para el modelado molecular por homología análisis filogenético

BLAST—Basic Local Alignment Search Tool

(www.ncbi.nlm.nih.gov/BLAST/). Servidor que permite realizar búsquedas de similitud secuencial. Encuentra regiones de similitud entre las secuencias locales. El programa compara las secuencias de nucleótidos o de proteína de bases de datos de secuencia y calcula el significado estadístico de los matches. BLAST puede utilizarse para inferir las relaciones funcionales y evolutivas entre las secuencias, así como a identificar los miembros de las familias génicas.

ClustalX2 (Thompson, Gibson, Plewniak, Jeanmougin, & Higgins, 1997) es una versión de Clustal W con una interfaz gráfica (Larkin et al., 2007).

HyPhy. Es un paquete de software de código abierto para el análisis de secuencias genéticas que utilizan técnicas de filogenia, evolución molecular. Cuenta con una completa interfaz gráfica de usuario y un lenguaje de programación rico para la personalización ilimitada de análisis. (Kosakovsky Pond, Frost, & Muse, 2005).

PhyML (<http://atgc.lirmm.fr/phyml>). Software que permite estimar grandes filogenias en forma precisa, rápida y sencilla haciendo uso del método de máxima verosimilitud a partir de secuencias de ADN y proteínas. (Guindon et al., 2010).



Métodos

4.1. CLONACIÓN DE PEPTIDASAS DE FLORES

4.1.1. Extracción de ARN total de plantas

A partir de inflorescencias maduras e inmaduras de *Silybum marianum*, recolectadas en las inmediaciones de la ciudad de La Plata durante la primavera, se aisló el ARN total empleando el kit comercial RNA Kit RneasyPlant (*Qiagen*), basado en las propiedades de unión selectiva de una membrana a base de sílice y mediante la técnica del Trizol® (*Applied Biosystems, Thermo Fisher Scientific Inc.*) modificada para material con alto contenido de polisacáridos y proteoglicanos. En ambos casos el tejido se molió en un mortero agregando nitrógeno líquido hasta obtener un polvo muy fino. De este polvo se tomaron aproximadamente 100 mg. El proceso se continuó siguiendo las instrucciones proporcionadas por el fabricante para el caso de la extracción utilizando el kit, mientras que en la técnica del Trizol se continuó con el agregado de 1 ml de este reactivo. Luego se separaron las fracciones. Se homogeneizó e incubó 5 min a temperatura ambiente. Se centrifugó a $13,82 \times g$ durante 10 min a $4 \text{ }^\circ\text{C}$ y se extrajo el sobrenadante. Se adicionaron 200 μl de cloroformo por cada ml de Trizol. Se agitó vigorosamente durante 15 s y se incubó a temperatura ambiente durante 2 a 3 min. Se centrifugó a $13,82 \times g$, 15 min y $4 \text{ }^\circ\text{C}$ y se separaron las dos fases, la inferior rica en fenol/cloroformo de la superior rica en agua. El RNA presente en la fase superior se transfirió a un tubo sin arrastrar la interfase blanca que contiene las proteínas y se le agregaron 250 μl de isopropanol y 250 μl de la mezcla de NaCl 1,2 M (solución salina) y $\text{Na}_3\text{C}_6\text{H}_5\text{O}_7$ 0,8M. El tubo conteniendo la mezcla se invirtió suavemente y se incubó por 10 min a temperatura ambiente. Se centrifugó a $13,82 \times g$ durante 10 min y $4 \text{ }^\circ\text{C}$, luego de lo cual se removió el sobrenadante y en baño de hielo se agregó 1 ml de EtOH al 75% (preparado con H_2O libre de RNAsas). Se homogeneizó usando vórtex y se centrifugó a $10,94 \times g$ durante 5 min a $4 \text{ }^\circ\text{C}$. Se removió el EtOH con pipeta y

secó el pellet en estufa por 5 min aproximadamente, luego de lo cual se redisolvió en H₂O libre de RNAsas (aprox. 50 µl H₂O/100 mg tejido inicial de modo que quede una solución más o menos viscosa). El ARN obtenido en ambos casos se cuantificó por medida de la absorbancia a 260 nm y su integridad se evaluó en gel de agarosa al 1% (p/v) como se describe en el ítem 4.1.3. de esta sección

4.1.2. Determinación de la concentración, pureza del ARN total y ADN

La cuantificación y estimación de la pureza del ARN total y ADN de las distintas muestras utilizadas fue realizada utilizando un espectrofotómetro (UV-visible Agilent 8453E, *Agilent Technologies Inc.*, o NanoDrop, *NanoDrop™*, *Thermo Scientific*).

Para la cuantificación se midió la absorbancia a 260 y 280 nm y a partir de los valores obtenidos se calculó la concentración de ARN y ADN mediante las siguientes ecuaciones:

$$\text{Concentración de ARN } (\mu\text{g}/\mu\text{l}) = (\text{Absorbancia } 260\text{nm}) \times (\text{Dilución}) \times 40 / 1000$$

$$\text{Concentración de ADN } (\mu\text{g}/\mu\text{l}) = (\text{Absorbancia } 260\text{nm}) \times (\text{Dilución}) \times 50$$

Sabiendo que la densidad óptica a 260 nm (1 OD_{260}) = 40 µg para ARN y ADN de cadena simple y que el ADN de cadena doble presenta un máximo de absorbancia a 260 nm (1 OD_{260}) = 50 µg.

Para determinar la pureza, se calculó la relación A_{260}/A_{280} . Se consideró que las relaciones cercanas a 2,0 unidades de OD_{260} son óptimas para el ARN, mientras que entre 1,9-1,7 son óptimas para el ADN.

4.1.3. Electroforesis en geles de agarosa

La técnica de electroforesis se basa en el desplazamiento de los ácidos nucleicos a través de una matriz porosa de agarosa polimerizada, siendo el tamaño del poro determinado por la concentración de ésta en una solución de sales. Al aplicarse una corriente eléctrica a través del gel en condiciones de pH neutro los ácidos nucleicos que poseen una carga negativa conferida por los grupos fosfatos, migrarán hacia el ánodo (Sambrook, Fritsch, & Maniatis, 1989)

Las muestras de ADN plasmídico, productos de digestión enzimática y los fragmentos de amplificación de PCR fueron analizados en geles de agarosa en distintas concentraciones al 0,75-2,0% (p/v) según los tamaños de los fragmentos a separar. La suspensión de agarosa se realizó en buffer TAE 1×. Se mezclaron 5,0 µl de la muestra de ADN con 5 µl de colorante GelRed 10.000× (*Genbiotech, Buenos Aires*). La electroforesis se llevó a cabo a 70 V durante 30-60 min en un sistema Sub-Cell GT Cell o Mini-Sub Cell GT Cell (*Bio-Rad Laboratories, Inc., Hercules, CA, EEUU*). El tamaño del gel empleado 7,0 × 7,0cm con un espesor de 7,0 mm. Los ácidos nucleicos se observaron en un transiluminador (*Vilber Lourmat, Eberhardzell, Alemania*) con luz ultravioleta a 260nm y se tomaron fotografías para documentar los resultados. Para la realización de electroforesis de ADN la cuba y el peine fueron lavados previamente con SDS 0,5% (p/v) en agua tratada con dietilpírocarbonato (DEPC).

4.1.4. Purificación de ADN a partir de gel usando un kit comercial

Para la purificación de los fragmentos de ADN, luego de desarrollada la electroforesis la banda del peso molecular que se pretendía aislar se cortó del gel de agarosa y se purificó mediante el uso de kit comercial GeneJET Gel Extraction Kit (*Thermo Scientific Inc.*). Este kit se basa en la unión selectiva de ADN de más de 100 pb a columnas rellenas de lana de fibra de vidrio. Se siguió

el procedimiento indicado por el fabricante, estimándose la concentración del ADN eluído como se describió en 4.1.2. Las muestras de ADN recuperadas fueron, en todos los casos, almacenadas a -20°C.

4.1.5. Amplificación de ácidos nucleicos

4.1.5.1. Síntesis de ADNc

El ADNc se sintetizó a partir del ARN total obtenido de las inflorescencias maduras e inmaduras por medio de la transcripción reversa (RT). En esta reacción las moléculas de ARNm funcionan como molde y el oligo dT18 actúa como cebador en presencia de dNTPs que funcionan como sustrato para la transcriptasa reversa. La RT se realizó en dos etapas:

Etapas:

| Componente | Volúmen(μl) |
|--|-------------|
| ARN ([2,21 μg/μl], inflorescencia inmadura y [3.01 μg/μl] inflorescencia madura) | 3,0 |
| Primer Oligo(dT)18 (100mM) | 1,0 |
| Agua tratada con DEPC | c.s.p 12,5 |

Tabla 6. Composición de la mezcla de reacción de la RT

En la primera etapa, la mezcla (Tabla 6) se sometió a un ciclo de desnaturalización a 55-60 °C por 10 min en un termociclador (*Techne, Ramsey, MN, EEUU*), transcurrido este tiempo el tubo conteniendo la mezcla se puso en hielo para evitar su re-naturalización y se continuó con la Etapa 2 que consistió en agregar los reactivos enumerados en la Tabla 7.

Etapas 2:

| Componente | Volúmen(μl) |
|--|-------------|
| Buffer 5× (Tris- HCl 250 mM (pH 8.3 a 25°C), KCl 220 mM y MgCl ₂ 20mM, DTT 50 mM). | 4,0 |
| Mezcla de dinucleotidotrifosfato (dNTP) [10mM de cada uno, adeninatrifosfato (dATP), citosina trifosfato (dCTP), guanidinatrifosfato (dGTP), y timina trifosfato (dTTP) con NaOH 10 mM, pH 7,0 (<i>Fermentas Life Sciences</i>)] | 2,0 |
| Inhibidor RiboLock™ RNaseInhibitor (<i>ThermoScientific Inc.</i>) 20 u | 0,5 |
| Agua tratada con DEPC | 1,5 |

Tabla 7. Composición de la mezcla de reacción de PCR

La mezcla se sometió a un ciclo de calentamiento a 37° C durante 5 min. Luego de lo cual se adicionó 1,0 μl de Transcriptasa reversa M-MuLV RT (virus de la leucemia murina de Moloney, *Fermentas Life Sciences*). Esta mezcla se sometió a un ciclo de síntesis del ADNc a 42 °C por 60 min y a un ciclo final de desnaturalización de la enzima a 70 °C por 10 min en el termociclador. El ADNc se almacenó a -20 °C.

4.1.5.2. Amplificación por PCR

La amplificación de regiones determinadas de ADNc se llevó a cabo mediante la técnica de PCR bajo condiciones estándar (Mullis, 1990) en termociclador (*Techné*). La mezcla de reacción que se utilizó para la amplificación se describe en la Tabla 8.

| Componente | Volúmen(μ l) |
|-----------------------------------|-------------------|
| Buffer de reacción (10 \times) | 1,0 |
| Mezcla de dNTPs(10 mM) | 2,0 |
| Prepro_FW (10 μ M) | 2,0 |
| Prepro_Rw (10 μ M) | 2,0 |
| Taq polimerasa | 0,3 |
| ADNc | 1,5 |
| Agua Mili Q | c.s.p 50 |

Tabla 8. Composición de la mezcla de PCR para la amplificación de las secuencias codificantes de peptidasas a partir de ADNc

La reacción de PCR para las peptidasas se realizó con el programa de termociclado que se detallan en la Tabla 9.

| Segmento | Paso | Nº de repeticiones | Temperatura ($^{\circ}$ C) | Tiempo |
|------------|---------------------------|--------------------|-----------------------------|----------|
| I | Desnaturalización inicial | 1 | 95 | 5 min |
| II | Desnaturalización | 35 | 94 | 45seg |
| | Hibridación | | 56 | 45seg |
| | Extensión | | 72 | 4 min |
| III | Extensión final | 1 | 72 | 15 min |
| IV | Mantenimiento | 1 | 4 | ∞ |

Tabla 9. Programa de termociclado para la amplificación de ADN por PCR

Después de la amplificación, las muestras se almacenaron a -20° C. Para verificar el peso de los productos obtenidos por PCR estos se sometieron a electroforesis horizontal en un gel de agarosa al 1% (p/v).

4.1.6. Ligación

Las mezclas de ligación con el vector pGEM-T Easy se llevaron a cabo como se indican en la Tabla 10, incubando las reacciones durante la noche a 4° C.

| Componente | Reacción estándar (μ l) | Control Positivo (μ l) | Control de background (μ l) |
|----------------------------------|---------------------------------|-----------------------------------|--|
| Buffer de ligación 2X,T4 | 5,0 | 5,0 | 5,0 |
| DNA ligasa | | | |
| Vector pGEM-T Easy (50 ng) | 1,0 | 1,0 | 1,0 |
| Producto de PCR | variable(*) | - | - |
| Control de DNA insertado | - | 2,0 | - |
| T4 DNA ligasa (<i>Promega</i>) | 1,0 | 1,0 | 1,0 |
| 3U/ μ l | c.s.p 10,0 | c.s.p 10,0 | c.s.p 10,0 |
| Agua libre de nucleasas | | | |

Tabla 10. Composición de la mezcla utilizada en la ligación del pGEM-TEasy y el producto de PCR (*) La cantidad de ADN a ligar, se adecuaron a los requerimientos de las aplicaciones posteriores.

Se obtuvieron así las construcciones denominadas: pGEM-pPreproSm1-wt y pGEM-pPreproSm2-wt.

4.1.7. Digestión enzimática de ácidos nucleicos

Las digestiones de ADN se llevaron a cabo con endonucleasas de restricción que reconocen secuencias nucleotídicas específicas capaces de generar extremos cohesivos que posteriormente se ligaron entre sí. También se sometió a digestión por enzimas de restricción a aquellos plásmidos sobre los que se pretendía deducir la presencia/ausencia de una determinada secuencia mediante la escisión del fragmento de interés o por generación del perfil de bandas electroforéticas resultante (análisis de restricción).

La digestión de plásmidos aislados como minipreparaciones de ADN (*Métodos 4.1.15*) con una o más enzimas de restricción fue llevada a cabo durante 3-4 hs a 37 °C, la mezcla que se detalla en la Tabla 11. Se emplearon las enzimas *EcoRI*, *SaI**Bsp1407I*. Se utilizaron los buffers recomendados por la casa comercial para cada una de las endonucleasas.

| Componente | Volumen (µl) |
|--|--------------|
| Buffer de reacción compatible con la/s enzima/s usada/s (10 ×) (*) | 2,0 |
| Albumina Serica Bovina (BSA, 10 mg/ml)(*) | 0,2 0,5 |
| Enzima de restriccion | variable(**) |
| ADN plasmídico | variable(**) |
| Agua MilliQ | csp 20 |

Tabla 11. Composición de la mezcla utilizada en la digestión de plásmidos por enzimas de digestión. (*) El buffer apropiado así como el requerimiento de BSA dependen de la/s enzima/s de restricción usada/s. (**) en el caso de las enzimas *EcoRI* y *Bsp1407I* no fue necesario adicionar BSA, ya que el buffer de reacción ya lo posee La cantidad de ADN a digerir y en consecuencia la cantidad de enzima usada, se adecuaron a los requerimientos de las aplicaciones posteriores.

Transcurrido el tiempo de reacción, el perfil de bandas generado así como el grado de digestión del ADN fueron analizados en un gel de agarosa 1% (p/v).

4.1.8. Secuenciación y análisis de datos

Las reacciones de secuenciación se prepararon en un volumen final de 10 µL, incluyendo 50 ng/µl para producto de PCR o 100 ng/µl para plásmidos. Fueron realizados por el servicio de secuenciación de ADN de la Universidad Politécnica de Valencia– IBMCP-CSIC, Macrogen (Seul, Corea), o por el servicio de secuenciación de ADN del INTA-Castelar. Para el análisis de los cromatogramas y el procesamiento de las distintas secuencias se utilizaron a las herramientas DNASTart y BioEdit. Luego se llevó a cabo el análisis de la secuencia mediante programas de búsqueda en bases de datos como HHPred o BLASTn.

4.1.9. Crecimiento y mantenimiento de cultivos bacterianos

El crecimiento de las distintas cepas bacterianas se realizó, salvo excepciones, en medio LB o SOC líquido a 37°C y con agitación a 250 rpm en una incubadora orbital. Para la selección y mantenimiento de clones recombinantes se añadieron diferentes antibióticos (Tabla 4). El mantenimiento de las distintas cepas

bacterianas y clones recombinantes se realizó en medio LB o SOC, en presencia de glicerol 90%, manteniéndose a -80°C .

4.1.10. Preparación de células competentes químicas

4.1.10.1. Método de cloruro de calcio

Para la preparación de bacterias competentes para transformaciones de rutina se partió de un cultivo de bacterias de 2,0 ml en LB o SOC. Se dejó crecer *overnight* a 37°C y 250-280 rpm para utilizarlo como preinóculo. Se tomaron 500 μl del cultivo saturado, se realizó una dilución del mismo 1:100 en medio LB o SOC fresco (200 ml) y se incubó a 37°C con agitación (250-280 rpm) aproximadamente durante 2 hs hasta alcanzar una $\text{DO}_{600\text{nm}}$ de 0,6. Se dejó en hielo 10 min y luego se centrifugó a $3,45 \times g$, por 10 min a 4°C . Se eliminó el sobrenadante y se resuspendió en 100 ml de CaCl_2 50 mM (estéril y enfriado a 4°C). Se homogeneizó usando vórtex y se dejó reposar 10 min en baño de hielo. Se centrifugó nuevamente en las mismas condiciones y se resuspendió en 20 ml de CaCl_2 50 mM (estéril y enfriado a 4°C). Se agregaron 3,2 ml de glicerol y se homogeneizó usando mezclador de vórtice. Finalmente, se fraccionó en alícuotas de 50 μl en tubos estériles y se almacenó a -80°C .

4.1.10.2. Método de cloruro de rubidio

Para la preparación de bacterias competentes por este método se procedió de la misma manera que en el método de Cl_2Ca para preparar el preinóculo y el cultivo. Una vez alcanzada la $\text{DO}_{600\text{nm}}$ de 0,6 se dejó en hielo durante 10 min y se centrifugó a $6,9 \times g$ por 10 min y 4°C . Se descartó el sobrenadante y se resuspendió el precipitado en 10 ml de buffer TBFI. Se incubó en hielo por 5 min y se centrifugó a $6,9 \times g$ por 5 min. Se descartó el sobrenadante, se adicionó 1 ml de buffer TBFII y se incubó en hielo durante 60 min. Finalmente se

fraccionó de a 50 μ l en tubos Eppendorfs estériles y se almacenó inmediatamente a -80°C .

Este método resultó más eficiente que el de cloruro de calcio, y fue utilizado en aquellos casos en que no se obtuvieron transformantes.

4.1.11. Preparación de células electrocompetentes

Para la preparación de bacterias electrocompetentes se plaquéó la cepa bacteriana desde un stock glicerolado en una placa de agar LB sin antibiótico. Se incubó a 37°C *overnight*. Luego una colonia aislada se inoculó en 10ml de medio LB sin antibiótico. Se incubó a 37°C y 250 rpm *overnight* y 10 ml del cultivo saturado se usaron como preinóculo, se realizó una dilución 1:100 del mismo en medio LB o SOC fresco (1000 ml). Se incubó a 37°C y agitación (250-280 rpm). Cuando la $\text{DO}_{600\text{ nm}}$ alcanzó un valor de 0,6 se pusieron las células en hielo y se centrifugaron a $3,45 \times g$ por 10 min a 4°C . Se descartó el sobrenadante y resuspendió el pellet en 200 ml de agua destilada fría. Se homogeneizó usando vórtex, se centrifugó nuevamente en las mismas condiciones y se resuspendió en 100 ml de agua desionizada fría. Luego se centrifugó en las mismas condiciones y el pellet se resuspendió en frío con 40 ml de glicerol al 10%. Se fraccionó en alícuotas y se centrifugó a $3,45 \times g$ durante 10 min a 4°C . Finalmente, el pellet se resuspendió suavemente en frío con un 1 ml de glicerol 10% y las alícuotas se almacenaron a -80°C .

4.1.12. Transformación de bacterias competentes químicas por shock térmico

Para realizar la transformación por shock térmico de las cepas indicadas en la Tabla 1 una alícuota de 50 μ l de células competentes químicas se descongeló lentamente en baño de hielo se adicionaron 25 ng de ADN y 1 μ l de DMSO y se dejó en baño de hielo por 30 min. Luego se realizó un shock térmico durante 45

s a 42 °C en baño termostático y rápidamente se pasó el tubo al baño de hielo por otros 5 min. A la mezcla de bacterias con el plásmido se le adicionaron 250 µl de LB o SOC fresco a temperatura ambiente y se incubó a 37 °C durante 2 hs a 220-250 rpm. Luego con ayuda de una espátula de Drigalsky se sembró en superficie sobre una placa de Agar-LB o SOC suplementado con el antibiótico de selección correspondiente. Luego de incubar toda la noche a 37 °C se pudieron obtener las colonias de bacterias transformadas.

4.1.13. Transformación de bacterias electrocompetentes por electroporación

Para realizar la transformación por electroporación una alícuota de 50µl de células se descongeló lentamente en baño de hielo, se adicionaron 25 ng de ADN y se dejó en hielo por 30 min. Se colocaron en una celda de electroporación entre las dos protuberancias teniendo cuidado de que no queden burbujas. Luego se colocó la celda en el electroporador y se realizó un pulso de 5 milis de 1800V. Se adicionaron 250 µl de medio SOC o LB a la cubeta, se invirtió varias veces y se colocó en un tubo Eppendorf. Por último, se llevó 1 h a 37 °C para que se recuperen las células. Con ayuda de una espátula de Drigalsky se sembró en superficie sobre una placa de Agar-LB suplementado con el antibiótico correspondiente.

4.1.14. Amplificación y preparación de los plásmidos

Para la clonación y amplificación de todos los plásmidos mencionados en el apartado de *Materiales 3.5* se emplearon las cepas *E.coli DH5a*, *TOP 10*, *XL1-Blue* o *GC10*. Mientras que los vectores vacíos del sistema Gateway fueron amplificados en la cepa *DB3.1TM* que resiste el gen tóxico *ccdB*. Las cepas llevando los distintos plásmidos se crecieron a 37 °C con agitación (220-250 rpm) en medio líquido LB o SOC suplementado con el antibiótico de selección

correspondiente, *overnight*. Luego se centrifugaron ($10,43 \times g$) y a partir pellet obtenido se realizó la purificación empleando el método de lisis alcalina (Sambrook et al., 1989) o mediante el kit comercial de purificación de plásmidos GeneJET Plasmid Miniprep Kit (*Thermo Scientific*).

4.1.15. Extracción de ADN plasmídico

4.1.15.1. Método de lisis alcalina

Luego de centrifugar el cultivo obtenido como se indicó en el inciso 4.1.14 el sobrenadante fue descartado y el pellet resuspendido en 200 μ l de una mezcla de 0,2 M NaOH-SDS al 1% (preparado en el momento), se mezcló suavemente por inversión y se agregaron 200 μ l de solución de acetato de sodio 3,5 M. Se mezcló nuevamente por inversión y se centrifugó a $10,43 \times g$ durante 2 min. El sobrenadante fue transferido a un tubo nuevo conteniendo 600 μ l de isopropanol. Se mezcló nuevamente por inversión y centrifugó a $10,43 \times g$ durante 2 min. El sobrenadante fue descartado y el precipitado se lavó con 0,5 ml de EtOH al 70%, se mezcló suavemente y se centrifugó a $10,43 \times g$ durante 2 min. Se descartó el sobrenadante y se dejó secar. Por último se resuspendió en 50 μ l de agua libre de RNasas. Se incubó 10 min a 65°C y se realizó una última centrifugación.

4.1.15.2. Tecnología de membranas a base de sílice

El kit GeneJET Plasmid Miniprep Kit (*Thermo Scientific*) combina con el método de lisis alcalina de las bacterias (Sambrook et al., 1989), con la unión y la liberación selectiva del ADN plasmídico en columnas de sílice. El kit empleado permite recuperar hasta 20 μ g de ADN plásmido.

Para la obtención de las minipreparaciones se partió de un cultivo bacteriano proveniente de una única colonia de transformantes siguiendo las instrucciones

del fabricante. El ADN eluído fue posteriormente cuantificado en un equipo Nanodrop 1000 (*Thermo Fisher Scientific Inc.*)

4.2. ESTUDIO DE LOCALIZACIÓN DE PEPTIDASAS DE *Silybum marianum* Y SUS MUTANTES POR EXPRESIÓN TRANSIENTE EN *Nicotiana benthamiana*

4.2.1. Obtención de construcciones para expresión de proteínas de fusión wt y mutantes

Para la transformación de *A. tumefaciens* y posterior transformación de tejidos vegetales se emplearon vectores binarios conteniendo los insertos de interés. Se utilizaron los vectores binarios pB7FWG2 y pB7RFWG2 (*Materiales* 3.5.2). Estos vectores tienen la ventaja de expresar el gen de interés bajo el control del promotor constitutivo fuerte 35S del virus del mosaico de coliflor fusionado a un gen reportero.

4.2.2. Reacción de recombinación con BP clonasa™ (vía lisogénica)

A partir de la construcción de pGEM-pPreproSm2-wt se diseñaron dos nuevos cebadores (F4F y F4R) a fin de amplificar y clonar la peptidasa en el vector de clonado pDORN222 del sistema Gateway. Los cebadores se diseñaron flanqueando el ORF incluyendo los sitios *attB* (*attB1* y *attB2*) necesarios para la recombinación mediada por la mezcla enzimática BP *clonasa II*.

La amplificación se llevó a cabo mediante PCR en un termociclador (*Techné*). La mezcla de reacción que se utilizó se describe en la Tabla 12.

| Componente | Volúmen(μ l) |
|-----------------------------------|-------------------|
| Buffer de reacción (10 \times) | 2,0 |
| Mezcla de dNTPs(10 mM) | 2,0 |
| F4F (2 μ M) | 2,0 |
| F4R (2 μ M) | 2,0 |
| Taq polimerasa | 0,5 |
| Molde | 1,0 |
| Agua Mili Q | c.s.p 20 |

Tabla 12. Composición de la mezcla de reacción de la cadena polimerasa (PCR)

Los parámetros utilizados en el programa de termociclado se detallan en la Tabla 13.

| Segmento | Paso | Nº de repeticiones | Temperatura ($^{\circ}$ C) | Tiempo |
|------------|---------------------------|--------------------|-----------------------------|----------|
| I | Desnaturalización inicial | 1 | 96 | 3 min |
| | Desnaturalización | | 95 | 30 s |
| II | Hibridación | 20 | 52 Δ 30 $^{\circ}$ | 30 s |
| | Extensión | | 72 | 1,5 min |
| III | Extensión final | 1 | 72 | 15 min |
| IV | Mantenimiento | 1 | 4 | ∞ |

Tabla 13. Programa de termociclado para la amplificación de la secuencia de ADN por PCR. (Donde Δ representa al gradiente de temperatura: 46.3; 48.3; 50.8; 52.7; 56.7; 58.7; 61.8; 64.7)

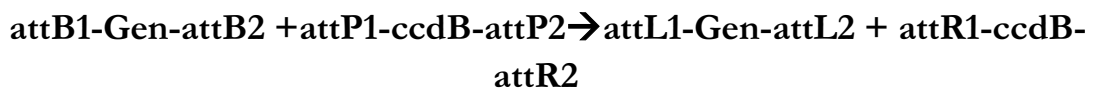
Después de la amplificación el productos de PCR se sometió a electroforesis en un gel de agarosa horizontal al 1% (p/v), a fin de chequear el tamaño del producto.

Las reacción de recombinación entre el fragmento de ADN flanqueado por los sitios *attB* (producto de purificación) y el vector pDONR222 conteniendo los sitios *attP* para generar un clon de entrada se indican en la Tabla 14.

| Componente | Volúmen(μl) |
|-----------------|-------------|
| Producto de PCR | 5,0 |
| Vector pDONR222 | 1,0 |
| BP Clonase™ | 1,5 |

Tabla 14. Composición de la mezcla de reacción BP clonasa

La reacción de recombinación puede resumirse del siguiente modo:



La mezcla de reacción, se agitó brevemente dos veces y se incubó a 25 °C *overnight*, luego de lo cual se utilizó para transformar *E. coli* GC10 competentes. Las colonias positivas de la construcción *pDONR222-pPreproSm2-wt* se seleccionaron por su resistencia a Km y fueron repicadas en medio líquido con el mismo antibiótico. Los cultivos crecidos durante la noche (SOC-Km^R) se cosecharon para el aislamiento de los plásmidos. Los vectores purificados fueron guardados a -20°C para su uso posterior. La presencia del inserto en el vector *pDONR222* fue confirmado por PCR utilizando cebadores específicos (523 y 524). Posteriormente, tres muestras provenientes de clones seleccionados al azar fueron remitidas al servicio de secuenciación para corroborar que la peptidasa se encontrara en marco de lectura.

4.2.3. Obtención de los mutantes de preprosilpepsinas

A partir de la peptidasa clonada en *pGEM-TEasy*, se generó una nueva construcción *pDONR222-pPreproSm2-D103A*. Para ello se diseñó dos nuevos oligonucleótidos: 618 y 619. El proceso se dividió en dos fases.

Para la obtención de la construcción mutante se realizó dos reacciones de PCR (Fase I). La primera utilizando el cebador original para el extremo N-terminal (F4F) y un cebador reverso para el mutante de sitio activo (618). La segunda utilizando el cebador original para el extremo C-terminal (F4R) más el cebador para generar el mutante de sitio activo (619). La mezcla de reacción utilizada se describe en las Tablas 15 y 16.

Fase I:

| Componente | Volúmen(μ l) |
|-----------------------------------|-------------------|
| Buffer de reacción (10 \times) | 2,0 |
| Mezcla de dNTPs(10 mM) | 2,0 |
| F4F (2 μ M) | 2,0 |
| 618 (2 μ M) | 2,0 |
| Taq polimerasa | 0,5 |
| Molde | 1,0 |
| Agua Mili Q | c.s.p 20 |

Tabla 15. Composición de la mezcla de reacción de la PCR

| Componente | Volúmen(μ l) |
|-----------------------------------|-------------------|
| Buffer de reacción (10 \times) | 2,0 |
| Mezcla de dNTPs(10 mM) | 2,0 |
| 619 (2 μ M) | 2,0 |
| F4R (2 μ M) | 2,0 |
| Taq polimerasa | 0,5 |
| Molde | 1,0 |
| Agua Mili Q | c.s.p 20 |

Tabla 16. Composición de la mezcla de reacción de la PCR

En la Tabla 19 se detallan los parámetros utilizados en el programa de termociclado.

| Segmento | Paso | Nº de repeticiones | Temperatura (°C) | Tiempo |
|----------|---------------------------|--------------------|------------------|-----------|
| I | Desnaturalización inicial | 1 | 96 | 3 min |
| II | Desnaturalización | 20 | 95 | 30s |
| | Hibridación | | 57 Δ 26° | 30s |
| III | Extensión | 1 | 72 | 1min 30 s |
| | Extensión final | | 72 | 15 min |
| IV | Mantenimiento | 1 | 4 | ∞ |

Tabla 19. Programa de termociclado para la amplificación de la secuencia de ADN por PCR Donde Δ representa al gradiente de temperatura: 53,8; 55,8; 55,9; 57,6; 61,6; 62,6; 65,5; 67,8; 68,8).

FaseII:

Luego se realizó una nueva PCR utilizando para ello los cebadores: F4F-618 y 619-F4R (Tabla 20, Fig. 6).

| Componente | Volúmen(μ l) |
|-----------------------------------|-------------------|
| Buffer de reacción (10 \times) | 1,0 |
| Mezcla de dNTP's(10 mM) | 1,0 |
| F4F(2 μ M) | 2,0 |
| F4R (2 μ M) | 2,0 |
| Producto de PCR juego F4F-618 | 5,0 |
| Producto de PCR juego 619-F4R | 3,0 |
| Taq polimerasa | 0,5 |
| Agua Mili Q | c.s.p 20 |

Tabla 20. Composición de la mezcla de reacción de la cadena polimerasa (PCR) para pGEM-pPreproSm1-D103A y pGEM-pPreproSm2-D103A

Se utilizó el siguiente ciclo:

| Segmento | Paso | Nº de repeticiones | Temperatura (°C) | Tiempo |
|----------|---------------------------|--------------------|------------------|----------|
| I | Desnaturalización inicial | 1 | 96 | 3 min |
| II | Desnaturalización | 20 | 95 | 30 s |
| | Hibridación | | 55,9° | 30 s |
| III | Extensión | 1 | 72 | 1.30 min |
| | Extensión final | | 72 | 15 min |
| IV | Mantenimiento | 1 | 4 | ∞ |

Tabla 22. Programa de termociclado para la amplificación de la secuencia de ADN por PCR Donde Δ representa al gradiente de temperatura: 53,8; 55,8; 55,9; 57,6; 61,6; 62,6; 65,5; 67,8; 68,8

El proceso puede resumirse del siguiente modo:

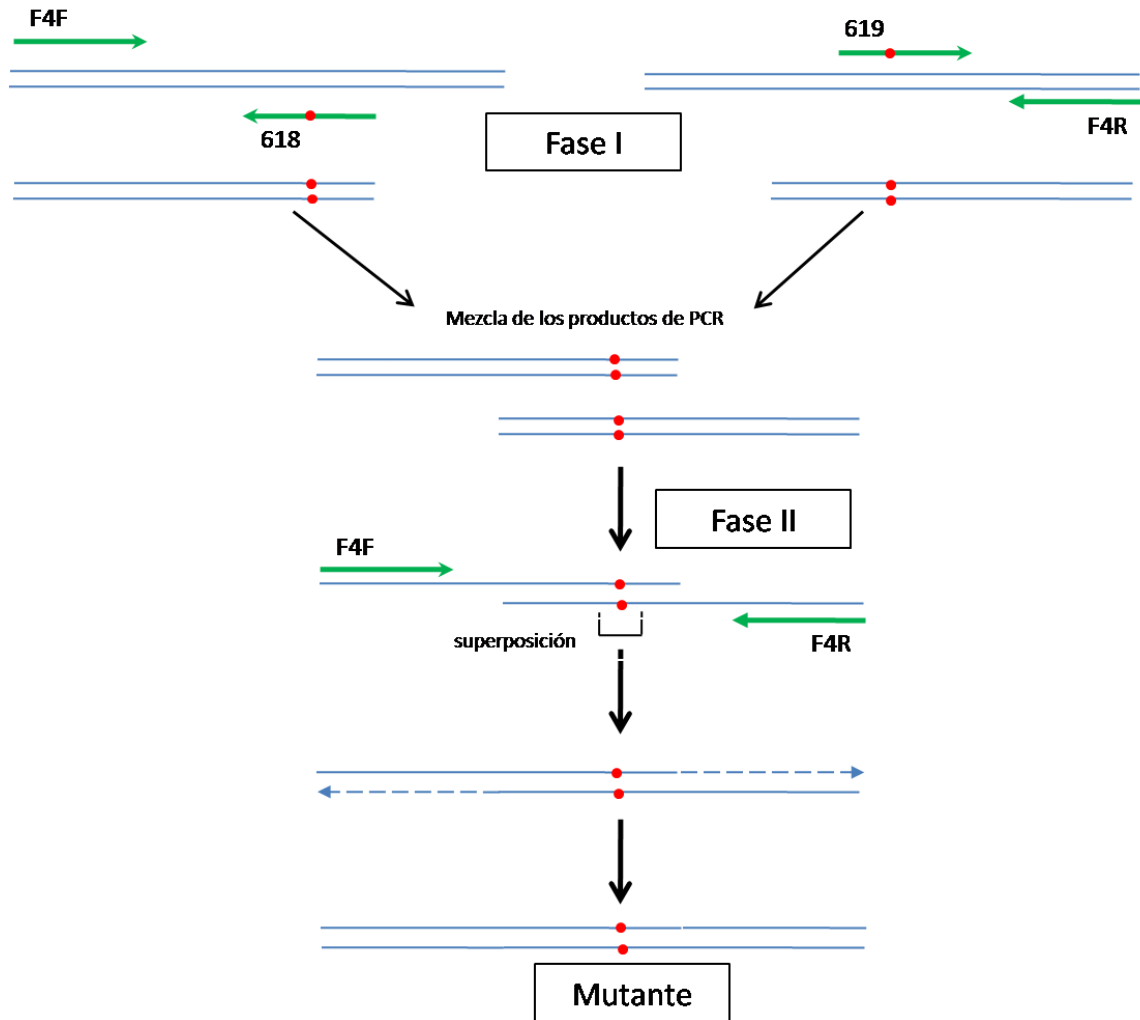
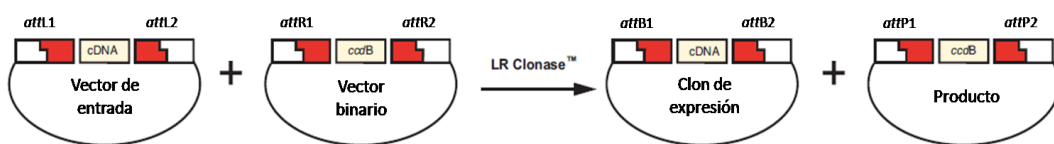
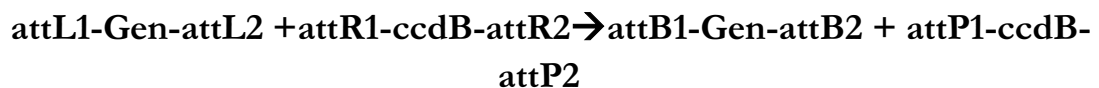


Figura 6. Representación de la mutagénesis dirigida por PCR.

Después de la amplificación, para verificar el peso de los productos de PCR obtenidos se realizó una electroforesis en gel de agarosa horizontal al 1% (p/v). A partir del producto de PCR se llevó a cabo la reacción de recombinación con la mezcla enzimática *BPclonase II*, para obtener así la construcción pDONR222-pPreproSm2-D103A, siguiendo la técnica descrita en el apartado 4.2.2. de esta sección.

4.2.4. Reacción de recombinación con LR clonasa™ (vía lítica)

Las construcciones pDONR222-pPreproSm1-D103A y pDONR222-pPreproSm2-D103A conteniendo el gen de interés en marco de lectura se recombinaron con el vector binario pB7FWG2. Además el vector pB7FRWG2 se utilizó para recombinar con la construcción pDONR222-pPreproSm2-wt. La reacción catalizada por la enzima *LR Clonase™* entre las construcciones flanqueadas por los sitios *attL* (vector de entrada) con los sitio *attR* del vector binario (vector destino: vector de expresión) puede resumirse del siguiente modo:



Para cada vector de expresión se realizó una mezcla de reacción descrita en la Tabla 23.

| Componente | Volúmen(μl) |
|------------------------|-------------|
| Molde | 3,5 |
| Vector pB7FWG2/pB7RWG2 | 0,5 |
| Enzimas de LR Clonase™ | 1,0 |

Tabla 23. Composición de la mezcla de reacción LR *clonase II*

Las mezclas de reacción se incubaron a 25 °C *overnight*, luego de lo cual se utilizaron para transformar *E.coli* GC10 (construcciones con el vector pB7FWG2) o *E.coli* XL1-blue (construcciones con el vector pB7FRWG2). Las colonias fueron seleccionadas (Sp o Sp/Tc) y fueron repicadas en medio líquido. Los cultivos crecidos durante la noche se utilizaron para aislar los plásmidos.

La presencia del gen que codifica para las peptidasas en el vector pB7FWG2 fue confirmado por PCR utilizando cebadores específicos (F4F y F4R) y utilizando el programa indicado en el apartado 4.2.2. de esta sección. Para el vector pB7RWG2 se confirmó la presencia del gen mediante digestión enzimática con la enzima Bsp1407I (*BsrGI*) y/o *SaI*. Las mezclas se incubaron durante 3 hs a 37 °C y luego se inactivaron por incubación a 65 °C durante 20 min. Los productos de las digestiones fueron analizados en geles de agarosa al 0,7%.

4.2.5. Obtención de cepas *Agrobacterium tumefaciens* recombinantes

4.2.5.1. Obtención de *Agrobacterium tumefaciens* competentes y electrocompetentes

Para la preparación *A. tumefaciens* electrocompetentes se partió de una placa crecida en YEB-Agar suplementado con el antibiótico de selección correspondiente durante dos días a 28 °C. Se tomó una colonia y se inoculó en 10 ml de YEB con antibiótico y se llevó a 28 °C *overnight* y 250-280 rpm. El preinóculo saturado, se utilizó para inocular 1000 ml de YEB con antibiótico. Se dejó crecer a 28 °C hasta que la DO_{600} fuera de aproximadamente 0,5 (más de 6 h). Luego se enfrió en un baño de hielo (0 °C) y se centrifugó durante 10 min a $9,21 \times g$ y 4 °C. Para lavar las células se eliminó el sobrenadante.

Para la preparación de competentes el sobrenadante se resuspendió en 100 ml de $CaCl_2$ 20 mM frío. Se centrifugó durante 5 min a $4,61 \times g$ y 4 °C. Se eliminó el sobrenadante y suavemente se resuspendió en 20 ml de $CaCl_2$ 20 mM frío. Se fraccionó en alícuotas de 200 μ l que se conservaron a -80 °C. Para las electrocompetentes al precipitado se le añadieron 500 ml de agua fría estéril y se centrifugó durante 10 min a $9,21 \times g$ y 4 °C. Se hicieron dos ciclos más de lavado, se resuspendió en 15 ml de glicerol frío al 10% y se centrifugó durante 10 min a $9,21 \times g$ y 4 °C. Por último, el precipitado se resuspendió en 3ml de glicerol frío al

10% y se fraccionó en alícuotas de 40 μ l, así se conservaron a -80 °C hasta su utilización.

4.2.5.2. Transformación *A. tumefaciens* por shock térmico

Para realizar la transformación se tomó una alícuota de células GV3101 competentes, se descongeló lentamente en baño de hielo y se adicionaron 500 ng de ADN plasmídico. La mezcla se mantuvo en hielo durante 15-30 min. Se colocó en N₂ líquido durante 5 min y se realizó un shock térmico en baño termostático a 37 °C durante 5 min, luego de lo cual se volvió a colocar en hielo durante 5 min. Finalmente se añadió 1,0 ml de LB y se incubó a 220-250 rpm y 28 °C durante 3-4 h. Luego se tomaron 50 μ l de la mezcla de transformación y se volcó sobre una placa de Agar-LB suplementado con el antibiótico de selección correspondiente. Se dispersó la mezcla sobre una placa con espátula de Drigalsky. Luego se incubó a 28 °C durante dos días.

4.2.5.3. Transformación *A.tumefaciens* por electroporación

Para realizar la transformación por electroporación de las células GV3101 electrocompetentes, se tomó una alícuota de células electrocompetentes y se descongelaron lentamente en baño de hielo. Paralelamente, la cubeta de electroporación de 2 mm (*Bio-Rad*) se colocó en hielo 10 min antes de llevar a cabo la transformación. Se mezclaron 1 μ l de preparación plasmídica con 40 μ l de células, se colocó la mezcla inmediatamente en la cubeta y se incubó 1 min en hielo. A continuación, se colocó la cubeta en el electroporador (*Bio-Rad Gene Pulser*) y se procedió a aplicar un pulso de 1800 V. Inmediatamente después de electroporar se adicionó 1 ml de medio YEB o LB y se volcó la mezcla en un tubo limpio y estéril. Las bacterias se recuperaron a 28 °C durante 3 h en agitación (200 rpm) y luego se dispersó en una placa con medio YEB o LB sólido en presencia del antibiótico correspondiente al vector en uso. Las colonias

deben ser visibles después de 2 días a 28 °C. Posteriormente se tomó una colonia y se volvió a repicar en medio YEB o LB sólido en presencia del antibiótico correspondiente para confirmar la transformación.

4.2.6. Germinación y mantenimiento de plantas de *Nicotiana benthamiana*

Las plantas de *N. benthamiana* usadas para los ensayos de agroinfiltración fueron germinadas en almácigos conteniendo perlita:tierra 1:3 embebidas con almacigol, un fertilizante líquido quelatinizado formulado para suministrar a la planta una relación balanceada de Fósforo, Zinc, Nitrógeno y Azufre. Las plantas se mantuvieron en cámara de cultivo con un fotoperíodo de 16 hs de luz/oscuridad, a 22-24 °C.

4.2.7. Infiltración con *A.tumefaciens* y preparación de las muestras vegetales para microscopía

Para los ensayos de agroinfiltración se utilizaron clones de *A. tumefaciens* GV3101 transformados con cada uno de los vectores binarios de expresión como se indicó anteriormente: pGEM-pPreproSm1-wt-eGFP, pGEM-pPreproSm2-wt-eGFP, pGEM-pPreproSm1-D103A-eGFP, pGEM-pPreproSm2-D103A-eGFP, pGEM-pPreproSm1-wt-mRFP y pGEM-pPreproSm2-wt-mRFP.

Se prepararon los cultivos de los clones con cada una de las construcciones generadas, se dejaron crecer *overnight* con agitación de 220-250 rpm y a 28°C en medio LB o YEB en presencia del antibiótico correspondiente a la cepa y al vector en uso. Estos inóculos se centrifugaron 15 min a 3,45 ×g y se resuspendieron en buffer MI en el mismo volumen inicial. Luego se centrifugó 10 min a 3,45 ×g, se eliminó el sobrenadante y se resuspendió en la mitad del volumen inicial. Por último se midió la DO_{600nm}. Para los ensayos de agroinfiltración con preprosilpepsina 2-eGFP se utilizó DO600 =0,1 y 1. Para

los ensayos de agroinfiltración con preprosilpepsina eGFP con el nanobody (Anti-GFP) se utilizó $DO_{600} = 0,1$ y 1. Para los ensayos de agroinfiltración con preprosilpepsina 2-mRFP con el nanobody (Anti-GFP) se utilizó $DO_{600} = 0,1, 0,3, 0,6$ y 1. En todos los casos la DO_{600nm} fue de 0,5 para la proteína P19 (Voinnet et al., 2003, supresor de la proteína viral de PTGS), 0,8 para el nanobody (anti-GFP, cedido por Petruccelli, Silvana) y 0,4 para pGDG (Goodin, Dietzgen, Schichnes, Ruzin, & Jackson, 2002). Se preparó la alícuota de agrobacterium de cada construcción con el agregado de acetosiringona (3',5'-dimetoxi 4'-hidroxiacetofenona) a una concentración final de 150 μ M. La acetosiringona es un producto natural fenólico relacionado con la acetofenona y el 2,6-dimetoxifenol. Es capaz de causar la activación de los genes de la región de virulencia del plásmido Ti, que son necesarios para iniciar la transferencia de la región T del ADN del plásmido Ti del *A. tumefaciens* al cromosoma de la planta. El tratamiento con acetosiringona sintética aumenta la velocidad con que *A. tumefaciens* infecta los tejidos vegetales.

Para la estabilización del GFP en el medio ácido de algunos compartimentos celulares se utilizó el plásmido pGreen II (Anexo II) que lleva al gen que codifica para un nanobody (anti-GFP) fusionado a la secuencia codificante de RFP. La proteína quimera no presenta señales de direccionamiento y posee una gran afinidad por GFP. pGreen II es un pequeño vector binario Ti capaz de replicarse en *E.coli*, pero incapaz de replicarse en *Agrobacterium* sin la presencia de otro plásmido binario, pSoup (Anexo II). pSoup co-reside en la misma cepa que el plásmido pGreen II, proporcionando las funciones de replicación en *trans* para pGreen II.

Para la infiltración se utilizó la proteína P19 como supresor post-trascricional del silenciamiento génico (PTGS), obtenida del virus TBSV (por *TomatoBushy-Stunt Virus*), presente en el vector pBIN61-p19 (Voinnet et al., 2003). El supresor funciona uniéndose a ARN de doble hebra generado por la maquinaria

molecular que dirige ese proceso inhibiendo así, el silenciamiento inducido por transgenes (Silhavy et al., 2002). Se sabe que su expresión es capaz de incrementar los niveles de acumulación de diversas proteínas recombinantes en 50 veces (Voinnet et al., 2003).

Las distintas muestras se incubaron durante 3 h a temperatura ambiente. Los inóculos se utilizaron para infectar la superficie abaxial de hojas jóvenes de *N. benthamiana* utilizando una jeringa de 1 ml sin aguja. Luego de 24, 48, 72 h y 9 días postinfiltración, se tomaron muestras de las hojas infectadas. Se cortaron trozos de hoja agroinfiltrada de 5 mm x 1 cm, los que se montaron entre portaobjetos y cubreobjetos para su observación.

4.2.8. Microscopía confocal

La expresión transitoria en hojas de *N. benthamiana* se observó por microscopía confocal de escaneo láser (LeicaDMI 6000 CS motorizado) con objetivo de 20 × (NA 0.70), 590 μm y de 63 × (NA 1.4) objetivo de inmersión, 100 μm. Para obtener las imágenes de GFP se utilizó un láser de excitación de 488 nm, la emisión se monitoreó con filtro de 510-560nm. Para los ensayos con RFP y en los ensayos de coinfiltración con el nanobody (anti-GFP) se excitó con láser de helio-neón 1.5 mW con líneas de excitación de 543 nm y de emisión de 690-630 nm. Para obtener imágenes de clorofila se utilizó un láser de helio-neón 1.5 mW con líneas de excitación de 543 nm.

Las imágenes mostradas son representativas de varias observaciones de la misma proteína en las mismas condiciones. Las imágenes fueron procesadas con Leica Confocal Software (LCS Lite 2.61.15)

4.2.9. Estudio bioinformático de peptidasas clonadas de *S. marianum*

4.2.9.1. Búsqueda en base de datos

Las secuencias de las PAs de 10 familias diferentes de plantas se obtuvieron del Centro Nacional de Información Biotecnológica (NCBI, <http://www.ncbi.nlm.nih.gov>).

4.2.9.2. Análisis de los motivos conservados

Los motivos de las secuencias de las peptidasas fueron analizados usando el programa *Multiple Em for Motif Elicitation* (MEME; version 4.9.1, <http://meme.nbcr.net/meme>, (Bailey & Elkan, 1994). MEME representa los motivos como matrices de probabilidad de letra, dependientes de la posición. Un motivo MEME es un patrón de secuencia que se produce repetidamente una o más veces en una secuencia. Describe el nivel de conservación en cada posición (letra) en el patrón. MEME busca los motivos más "significativos" en las secuencias de entrada, donde "significativo" es función de la longitud del patrón, el número de veces que se produce y el grado de similitud entre las ocurrencias. Informa un E-valor para cada motivo que encuentra, que es una estimación de la cantidad de motivos que uno esperaría encontrar por azar si se mezclaran las letras de las secuencias. Motivos con pequeños E-values son muy poco probables. Los parámetros de análisis para MEME utilizados fueron: ancho mínimo para cada motivo: 6; ancho máxima de cada motivo: 50; número máximo de motivos a encontrar: 10 y número de repeticiones: 0 o 1 por secuencia.

4.2.9.3. Análisis estructural y funcional

Varios servicios web *on-line* y software se utilizaron para el análisis de las PAs. Los análisis comparativos y bioinformáticos se llevaron a cabo *on-line* en los sitios web (<http://www.ncbi.nlm.nih.gov> y <http://expasy.org/tools>). Los parámetros

físico-químicos de las PAs fueron analizados por ProtParam. Para determinar el índice alifático de una proteína que se define como el volumen relativo ocupado por cadenas alifáticas laterales (alanina, valina, isoleucina y leucina), se utilizó la siguiente fórmula:

$$\text{Índice alifático} = X(\text{Ala}) + a * X(\text{Val}) + b * (X(\text{Ile}) + X(\text{Leu}))$$

donde X(Ala), X(Val), X(Ile), y X(Leu) moles por ciento (100 X fracción molar) de alanina, valina, isoleucina, y leucina. Los coeficientes a y b son el volumen relativo de la cadena lateral valina ($a = 2,9$) y de las cadenas laterales de Leu / Ile ($B = 3,9$) según la cadena lateral de alanina.

El valor GRAVY para un péptido o proteína se calcula como la suma de los valores de hidropatía de todos los aminoácidos, dividido por el número de residuos en la secuencia.

La predicción de estructura secundaria se analizó por ProFunc , la predicción de sitios de escisión del péptido señal se realizó por SignalP 4.1 server y los dominios funcionales de las proteínas AP fueron identificados por servidor ProDom

4.2.9.4. Modelado molecular por homología: predicción de estructura terciaria

El modelo tridimensional de prosilpepsina1 y prosilpepsina2 se obtuvo a partir de un método de modelado por homología. Los métodos incluidos dentro de esta categoría se basan en el hecho de que aquellas proteínas con un origen común (homólogas) tienden a plegarse en forma similar. Por lo tanto, empleando una proteína homóloga con estructura conocida (molde o template) es posible construir modelos estructurales utilizando esa información estructural. Para obtener el modelo tridimensional las peptidasas, en primer lugar, se realizó la asignación de plegamiento o *foldassignment* que permite encontrar la proteína homóloga (ya cristalizada) más adecuada para emplear como molde. El

foldassingment se llevó a cabo empleando el servidor HHPred(Söding et al., 2005).. Luego se realizó el alineamiento entre la secuencia de las peptidasas y el molde con el servidor ClustalW (EMBnet) y se editó con el programa GeneDocMultipleSequenceAlignment Editor and ShadingUtility. Este alineamiento, junto con el archivo PDB de la proteína elegida como molde así como también el *script*, es decir un archivo de comandos que indica las instrucciones para hacer el modelado, fueron los archivos que se utilizaron como *input* para ejecutar el cálculo con el programa Modeller, versión 9.14 (Sali & Blundell, 1993). Este programa constituye un método de modelado comparativo diseñado para encontrar la estructura más probable de una secuencia determinada, en función de su alineamiento con una proteína homóloga con estructura conocida. El modelo tridimensional se obtiene por satisfacción de restricciones espaciales. El programa toma las distancias entre residuos a partir del *template*, que se complementan con datos estereoquímicos, distancias de enlaces, ángulos diedros, distancias de interacciones débiles, etc; todos esos datos se combinan para dar una función objetivo (*objectivefunction*).El modelado consiste en optimizar esta *objective function*. El mejor modelo es aquel que viola el menor número de restricciones espaciales y será el que posea el menor valor de *objectivefunction*.

4.2.9.5. Evaluación y validación del modelo obtenido

La calidad del modelo estructural obtenido fue evaluada a través de dos tipos de criterios: energéticos y estructurales/geométricos.

La evaluación energética del modelo se llevó a cabo empleando el servidor ProSA-web(Wiederstein& Sippl 2007), que permitió realizar un análisis global y local. En el caso del análisis global, se comparó la energía del modelo obtenido con la de todas las proteínas ya cristalizadas (a través de un parámetro denominado *Z-score*); mientras que el análisis local permitió evaluar la energía para

cada residuo de la proteína en cuestión. En el caso de tratarse de un buen modelo, deben esperarse valores de energía negativos. También se usó como criterio energético el potencial estadístico DOPE (*DiscreteOptimizedProteinEnergy*) llevado a cabo a través del programa Modeller, que es similar a la evaluación local mencionada previamente, ya que permite analizar la energía de cada residuo, siendo un modelo de buena calidad aquel que presenta la mayor cantidad de residuos con un *DOPE score* menor a cero.

La calidad y la precisión del modelo estereoquímico se evaluó con el programa PROCHECK por análisis gráfico de Ramachandran. Este programa ubica a los residuos de la proteína modelada que estarán en distintas regiones dependiendo la calidad del modelo.

El gráfico Ramachandran muestra los ángulos de torsión phi-psi para todos los residuos en la estructura (excepto los que están en los extremos de la cadena). En la cadena polipeptídica los enlaces N-C α y C α -C tienen libre rotación. Estas rotaciones están representadas por los ángulos de torsión phi (Φ) y psi (Ψ), respectivamente, Fig.7.

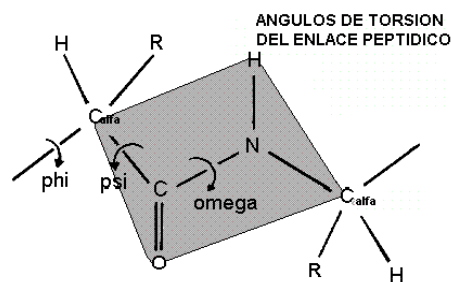


Figura 7. Ángulos de torsión principales del enlace peptídico o de un polipéptido: phi (Φ), psi (Ψ), y omega (Ω).

Se pueden distinguir cuatro regiones: las **zonas blancas** corresponden a conformaciones donde los átomos del polipéptido están más cercanos que sus radios de van der Waals. Estas regiones son estéricamente no permitidas para todos los aminoácidos, excepto para Glicina, el cual es el único aminoácido que

no presenta cadena lateral. Las **zonas cremas** corresponden a regiones generosamente permitidas. Las **regiones rojas** corresponden a conformaciones donde no hay impedimentos estéricos, es decir estas son zonas permitidas llamadas conformaciones α -hélices y Hojas- β . Las **zonas amarillas** muestran las regiones permitidas sí en los cálculos se usan radios van der Waals ligeramente más pequeños, es decir se permite que los átomos puedan estar un poco más cercanos. Esto origina una nueva región que corresponde a una hélice con giro hacia la izquierda. Los L-aminoácidos no pueden formar una hélice con giro a la izquierda pero ocasionalmente residuos individuales pueden adoptar esta conformación. Estos residuos generalmente son glicina pero también pueden ser aspargina y ácido aspártico en los que la cadena lateral forma un puente hidrógeno con la cadena principal y así estabiliza esta conformación desfavorable.

Las regiones no permitidas generalmente involucran impedimentos estéricos entre el carbono metilénico de la cadena lateral y los átomos de la cadena principal. La Glicina no tiene cadena lateral y por lo tanto puede adoptar ángulos phi y psi en los cuatro cuadrantes del gráfico de Ramachandran. Por esto ellas aparecen en regiones donde cualquier otro residuo estaría estéricamente impedido.



Resultados

5.1. FUENTE DE PEPTIDASAS

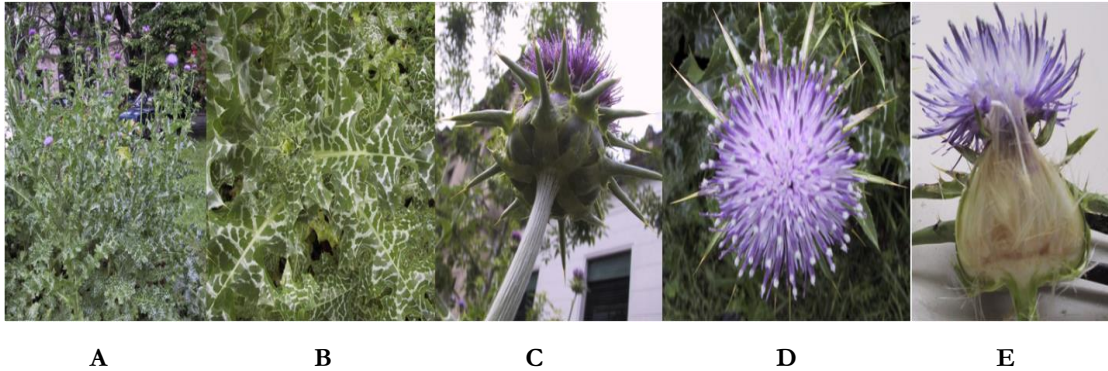


Figura 1. *Silybum marianum* (L.) Gaertn. **A.** Parte aérea de la planta florecida. **B.** Hojas basales. **C, D y E.** Inflorescencia

A fin de alcanzar los objetivos de la presente tesis doctoral se utilizaron como material de partida flores de la especie *Silybum marianum* (L.) Gaertn. perteneciente a la familia Asteraceae y dentro de la misma, a la Tribu Carduae.

S. marianum (Fig. 1) es conocida como cardo mariano, cardo asnal o cardo blanco. Es una hierba anual de 1,0 a 2,2 m de altura, con tallos robustos y hojas glabras o casi glabras de color verde marmoreado con blanco y espigas aceradas en los márgenes. Las hojas basales son subarrosetadas, grandes, lobuladas o pinatífidas, con lóbulos espinoso-dentados; mientras que las hojas caulinares son alternas, sésiles, oblongas, más o menos lobadas y con ápice y márgenes espinosos (Fig. 1 B). Las flores son violáceas, isomorfas, perfectas y se disponen en capítulos terminales, poco numerosos, largamente pedunculados, de 4 a 6 cm de diámetro (Fig. 1 C); el involucre es globoso, de unos 3 a 4 cm de altura, con brácteas terminadas en una larga punta espinosa y extendida o dirigida hacia abajo (Fig. 1 C). Las anteras tienen la base sagitada, los filamentos están engrosados y con pelos unidos entre sí. Las ramas estilares están unidas y tienen un anillo piloso por debajo del punto de bifurcación (Ariza Espinar & Delucchi, 1998). La especie es originaria del sur y sudoeste de Europa, oeste de Asia, norte

de África, Islas Canarias y Azores; naturalizada en Argentina (Jujuy, Salta, Chaco, Corrientes, Santiago del Estero, Santa Fe, Córdoba, Entre Ríos, La Pampa, Buenos Aires, Ciudad Autónoma de Buenos Aires, Río Negro y Tierra del Fuego). Es frecuente encontrarla en cultivos y ambientes modificados.

5.2. CLONACIÓN DE ASPARTILPEPTIDASAS DE FLORES

En primer lugar se extrajo el ARN total de flores inmaduras y maduras del cardo mariano utilizando dos métodos específicos para material vegetal. El rendimiento para la extracción a partir de flores inmaduras fue de 0,4 $\mu\text{g}/\mu\text{l}$ por el método de Trizol modificado para material vegetal y 2,2 $\mu\text{g}/\mu\text{l}$ por el kit comercial. Mientras que, para la muestra proveniente de flores maduras resultó de 0,1 $\mu\text{g}/\mu\text{l}$ por el método de método de Trizol modificado y 3,0 $\mu\text{g}/\mu\text{l}$ mediante el kit comercial. Por lo tanto en ambos materiales resultó más eficiente la extracción mediante kit comercial y estas muestras fueron las que se utilizaron en adelante.

La calidad del ARN se controló mediante electroforesis de agarosa (1%) como se muestra en la Fig. 2 A y entonces se sintetizó el ADNc por retrotranscripción. Luego, el ADNc se usó como molde en una reacción de PCR con cebadores diseñados específicamente a partir de regiones conservadas de los extremos 5' y 3' del marco de lectura abierto (*ORF*) de PAs de plantas (Fig. 2 B). Los productos de amplificación que resultaron del tamaño esperado de 1,5 kb se purificaron con un kit comercial desde un gel de agarosa (1%) y se ligaron al vector pGEM-T Easy. Con el producto de estas ligaciones se transformaron bacterias *E. coli* XL1-Blue químicamente competentes. Finalmente, se purificaron los plásmidos provenientes de varias colonias positivas para cada una de las construcciones. La Fig.2C muestra el análisis de restricción realizado con la enzima EcoRI que corrobora la presencia del inserto de 1,5 kb en el plásmido. Esquemas de estas construcciones se muestran en el Anexo 2.

Varios clones conteniendo el inserto se secuenciaron en forma automática. El análisis de las secuencias con el servidor BLAST, permitió la identificación de dos secuencias nucleotídicas diferentes. Estas secuencias exhiben altos niveles de identidad de secuencia entre ellas y con otras secuencias de PAs típicas de plantas disponibles en bases de datos públicas como *Cynara cardunculus*, *Centaurea calcitrapa*, *Cirsium vulgare* y *Helianthus annuus* pertenecientes a la familia Asteraceae, *Solanum tuberosum* y *Solanum lycopersicum*, entre otras.

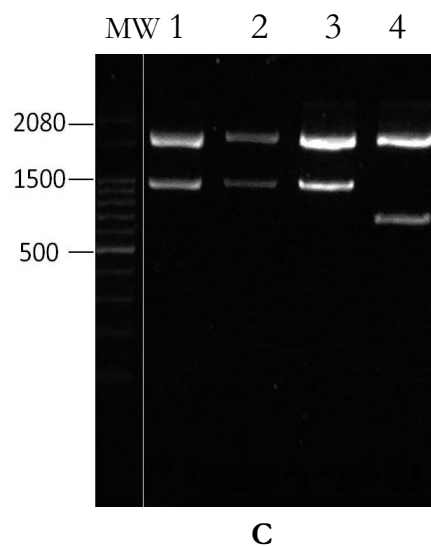
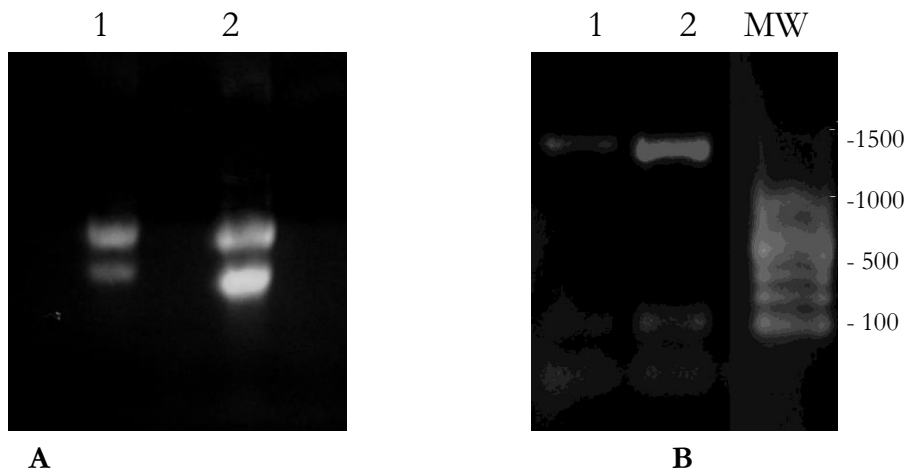


Figura 2. Amplificación y clonado de la secuencia correspondiente al marco de lectura abierto de las PAs de flores de *S. marianum* (página anterior). La primera cadena de ADNc se sintetizó a partir del ARN total aislado de flores *S. marianum* y se utilizó como molde en la reacción de PCR con cebadores específicamente diseñados para los extremos 5' y 3' altamente conservados de marcos de lectura abiertos que codifican PAs vegetales. El producto de amplificación obtenido se clonó y secuenció en forma automática. **A:** Análisis de la extracción de ARN total por electroforesis en gel de agarosa (1%): *Calles 1 y 2:* ARN total de *S. marianum* aislado de flores maduras y flores inmaduras, respectivamente. **B:** Análisis electroforético en gel de agarosa (1%) del producto amplificado mediante PCR: *Calle 1:* se tomó como molde el ADNc proveniente de flores maduras, *Calle 2:* se tomó como molde el ADNc proveniente de flores inmaduras. *MW:* marcadores de masas moleculares de ADN de 100 pb (*PB-L Productos Bio-Lógicos*). **C:** Análisis de restricción del plásmido pGEM-T Easy conteniendo el inserto de 1,5kb purificado a partir de clones positivos con la enzima *EcoRI* *Calles 1 y 2:* muestra original flores maduras, *Calles 3 y 4:* muestra original flores inmaduras, *MW:* marcadores de masas moleculares de ADN de 100 pb (*PB-L Productos Bio-Lógicos*).

Una de las secuencias se obtuvo a partir del ARN proveniente de ambos tipos de flores (maduras e inmaduras) mientras que la otra secuencia solo fue obtenida a partir del ARN de flores maduras.

5.3. ANÁLISIS BIOINFORMÁTICO

A fin de investigar las características de las dos proteínas clonadas de *S. marianum* se utilizaron diferentes métodos bioinformáticos. Las secuencias nucleotídicas y las secuencias aminoacídicas deducidas se muestran en las Fig. 3 y Fig. 4. Las secuencias nucleotídicas de 1511 y 1530 pares de base se depositaron en la base de datos GenBank con los números de acceso JX535387.1 y JX535388.1, respectivamente. La primera secuencia solo se encontró en flores maduras mientras que la segunda se encontró en ambos tipos de flores. Los productos proteicos codificados corresponden a dos precursores de PAs y fueron denominados **preprosilpepsina 1** y **preprosilpepsina 2**, respectivamente (números de acceso en la base de datos UniProtKB: W5QMD3 y W5QMH7). En este trabajo de tesis se emplearán los nombres preprosilpepsina1 o 2 en referencia a las secuencias completas de las

prepropeptidasas, prosilpepsina 1 o 2 en referencia a los zimógenos y silpepsina 1 y 2 para las PAs maduras.

Figura 3. ADNc que codifica para la forma completa de silpepsina 1 (preprosilpepsina 1) y su secuencia aminoacídica derivada (página siguiente). Se encuentran indicadas las regiones correspondientes al péptido señal N-terminal (flecha de puntos) y prosegmento (flecha discontinua). La secuencia del ADNc de preprosilpepsina1 ha sido incorporada en la base de datos GenBank bajo el número de acceso JX535387.1.

atgggtaactcaatcaaagcaagcctgcttgcttgtttttgttttttcttctatcacct
 M G N S I K A S L L A L F L F F L L S P
 actgcatttttcgggtctccaatggtagtttagagttggacttaaaaagaggaaggtg
 T A F S S N G G L V R V G L K K R K V
 gaccaaattaaccaacttagtggacatggtagctctatggaaggaaaggctagaaaagat
 D Q I N Q L S G H G A S M E G K A R K D
 tttggcttcggtagtactctgagggactcggacagtgcattattgcattaagaactac
 F G F G S T L R D S D S D I I A L K N Y
 atggatgctcagttattatggtagattggtagttggagaaccagctcaaaagttcacctg
 M D A Q Y Y G E I G I G E P A Q K F T V
 atttttgatactgggagttctaattctatgggtgccttctgcaaagtgctacttttcagta
 I F D T G S S N L W V P S A K C Y F S V
 gcttgcctttttcactcaaagtataagtcgagccattcaagtacctaagaataatgga
 A C L F H S K Y K S S H S S T Y K K N G
 acatctgctgctattcaatatggaactggatcaatatctggttttgtagccaagactct
 T S A A I Q Y G T G S I S G F V S Q D S
 gtcaaactcgggtgatcttgtttgtaaaagagcaggattttatagaggcaaccaagagcct
 V K L G D L V V K E Q D F I E A T K E P
 ggcatcactttcttggcagccaagtttgatggtagtttcttggccttggatttcaggagatc
 G I T F L A A K F D G I L G L G F Q E I
 tctgttgggaaactgtttcctgtctggtacaacatggtaatacaaggtcttgttcacgaa
 S V G K S V P V W Y N M V N Q G L V H E
 cctgtgttttcccttttggtttaatcgcaatgctgatgaggaagaagggggtgaacttgtg
 P V F S F W F N R N A D E E E G G E L V
 tttgggtggtgttgaccctaatacttttaagggtaggcacacatatgtccctgtgacccaa
 F G G V D P N H F K G R H T Y V P V T Q
 aagggtcatttggcagtttaataatgggtgacgttcttattgaagataaaaactaccggattt
 K G Y W Q F N M G D V L I E D K T T G F
 tgttctgatggttgtgctgcaattgccgactctggaacctcttggttggcaggtccaacg
 C S D G C A A I A D S G T S L L A G P T
 gctattattactcaaataatcatgcaattgggtgctaagggggtcatgagccagcaatgc
 A I I T Q I N H A I G A K G V M S Q Q C
 aaaacattgggtgatcagttatggaaagagtataattgagatgctcctgtctgaggctcaa
 K T L V D Q Y G K S I I E M L L S E A Q
 cctgataaaaactgtttctcaaataagttatgcaacttttaatgggtgctcgcgatgttagt
 P D K I C S Q M K L C T F N G A R D V S
 tcaataattgagagtggtggtgacaagaacaacggcaagtcttctggtaatgagatgtgt
 S I I E S V V D K N N G K S S G N E M C
 accttctgtgaaatggcagtcgtttggatgcaaaccaaatcaaacgaaaccagactcaa
 T F C E M A V V W M Q N Q I K R N Q T Q
 gataacataatcaactatgtcagtgagttgtgtgatcgttaccagtcctaatgggagaa
 D N I I N Y V S E L C D R L P S P M G E
 tcagcagtagactgcaacagttcttccctccatgcccaatatttcttcacaattgggtggc
 S A V D C N S L S S M P N I S F T I G G
 aaagtttttgagctctgcccagaacagtacatcctcaaaatcgggtgatggagaagcagca
 K V F E L C P E Q Y I L K I G D G E A A
 caatgcataagtgattcactgctatggatgtggcccctcctcgtggacctctatggatc
 Q C I S G F T A M D V A P P R G P L W I
 ttgggagatgtttttatgggtcgataccacacggtgttcgattatggcaagttacgagtg
 L G D V F M G R Y H T V F D Y G K L R V
 ggatttgcagaagcagcttga
 G F A E A A

```

atgggtaactcaatcaaagcaagcctgcttgcttgttcttgttttttcttctatcacca
M G N S I K A S L L A L F L F F L L S P
actgcatttttcggcctcgaatgggtggattgattcgcagtcggacttaaaaagaggaagggtg
T A F S A S N G G L I R V G L K K R K V
gaccgaagcgaccaacttcatggacttgggtgcatatcgaaacgaaatggtcgaaaagat
D R S D Q L H G L G V H T K R N G R K D
tttgcttccgtgggtacgctcagggactcggacaccggcattattgtactaacgaacgaa
F G F R G T L R D S D T G I I V L T N E
agagatactacgtatttcggtgagattgggtatcgggactccacttcaaatttcacagtg
R D T T Y F G E I G I G T P L Q N F T V
atTTTTgacacaggaagtTctgatctatgggtgccttcttcaaagtgtacagttcagta
I F D T G S S D L W V P S S K C Y S S V
tcgtgttgggttcattcaaggtataagtcgagcgattcaagtacctacaaggaaaacggg
S C W V H S R Y K S S D S S T Y K E N G
acatctgggtgctattcaatatggaactggatcaatcgctggtttttatagccaagactct
T S G A I Q Y G T G S I A G F Y S Q D S
gtcgaagtcgggtgatcttgttggtaaaagagcaggattttatagaggcaaccaagaggcc
V E V G D L V V K E Q D F I E A T K E A
gacaatgtatTTTTaaagagcccactttgatgggtataacttggccttggatttcaagagatc
D N V F L R A H F D G I L G L G F Q E I
tcggttcgggaagcgtgttctctgtctgtacaacatgggtgatcaaggtcttgttcaagaa
S V R K A V P V W Y N M V D Q G L V Q E
cctgtgttttcttttggcttaatcgcaatgttaatgaggaagaagggggggaacttgtg
P V F S F W L N R N V N E E E G G E L V
tttgggtgggggtgaccctaatacatttttaggggtaaccacacatatgtccctgtgattcaa
F G G V D P N H F R G N H T Y V P V I Q
aagggccattggcagtttgatattgggtgacgttcttattggagataaaagtacaggattt
K G H W Q F D M G D V L I G D K S T G F
tgtgccaatggttgtgcggaattgccgactctggaaccactttgttgacaggcccaacg
C A N G C A A I A D S G T T L L T G P T
gctattgttactcaaatacaatcaagcaattgggtgctaacaggggtcatgaaccaacaatgc
A I V T Q I N Q A I G A N R V M N Q Q C
aaaaccttgggttagtcgggtatggaaagaatataatgcagatgctcatgtctgaggcgcaa
K T L V S R Y G K N I M Q M L M S E A Q
cctgataaaaatttgttctcatgaacttatgcacttttgatgggtgctcacgatgttagt
P D K I C S H M N L C T F D G A H D V S
tcaataattgagagtggttgacatgaaaaacgacaagtcttcagggtggcgtacacgat
S I I E S V V D M K N D K S S G G V H D
gagatgtgtactctctgtgagatggcagtcgtttggatgcaaaaagaaatcaaacgaaac
E M C T F C E M A V V W M Q K E I K R N
gagattgaagataacataatacaactatgtcactgagctatgtgatcacttaccagttca
E I E D N I I N Y V T E L C D H L P S S
agtggagaatcagtagtagattgcaatagtccttctccatgccaataactgcctttaca
S G E S V V D C N S L S S M P N T A F T
attgggtggcaagatttttgagctcaccacagaacagtacattttcaaactcgggtgagggga
I G G K I F E L T P E Q Y I F K L G E G
gaagcagcacaatgcacatcagtggttactactatggatgtgccccctcctcgtgggtcct
E A A Q C I S G F T T M D V P P P R G P
ctatggatcctcggggatatttttatgggtccataccacacagttttcgattatggcaag
L W I L G D I F M G P Y H T V F D Y G K
ttaCgagtggtgatttgcagaagcagcttga
L R V G F A E A A

```

Figura 4. ADNc que codifica para la forma completa de silpepsina 2 (preprosilpepsina 2) y su secuencia aminoacídica derivada. Se encuentran indicadas las regiones correspondientes a: péptido señal N-terminal (flecha de puntos) y prosegmento (flecha discontinua). La secuencia del ADNc de preprosilpepsina 2 ha sido incorporada en la base de datos GenBank bajo el número de acceso JX535388.1.

Las estructuras primarias de las PAs de *S. marianum* fueron analizadas con el servidor ProtParam, que permitió calcular las propiedades físico-químicas a partir de la secuencia de las proteínas deducidas incluyendo peso molecular, punto isoeléctrico (pI) teórico, índice alifático y promedio general de hidropatía (GRAVY). Los resultados se muestran en la Tabla 1.

| Preproenzima de: | | silpepsina 1 | silpepsina 2 |
|---|-------------------------------------|---|--|
| Análisis de la estructura primaria | | | |
| ProtParam | Número de aminoácidos | 506 | 509 |
| | Peso molecular | 54988.5 | 55797.0 |
| | Punto isoeléctrico teórico | 5.60 | 5.17 |
| | Índice de alifaticidad | 79.76 | 79.23 |
| | GRAVY | -0.069 | -0.136 |
| Predicción de las modificaciones post-traduccionales | | | |
| Signal IP 4.0 | Sitio de escisión del péptido señal | A ₂₄ -A ₂₅ | A ₂₄ -A ₂₅ |
| NetNGlyc | Sitios de N-Glicosilación | N139: motivo NGT N397: motivo NQT N433: motivo NIS | N97: motivo NFT N139: motivo NGT N252: motivo NHT |

Tabla 1. Análisis de la estructura primaria, predicción propiedades fisicoquímicas, predicción de modificaciones postraduccionales y predicción de topología por diferentes herramientas bioinformáticas

Como se puede observar en la Tabla 1, no existen diferencias significativas entre ambas peptidasas en cuanto a peso molecular y pI, lo que haría dificultosa la purificación de estas peptidasas desde su fuente natural. El pI fue de 5,60 para preprosilpepsina 1 y 5,17 para preprosilpepsina 2, lo que indica su carácter ácido.

El índice alifático es el volumen relativo ocupado por las cadenas laterales alifáticas (Ikai, 1980). Un alto valor indica que la proteína tiene un alto contenido de cadenas alifáticas que le proveen termorresistencia y por lo tanto estabilidad en diferentes condiciones (Sivakumar, Balaji, & Gangaradhakrishnan, 2007). Para preprosilpepsina 1 el valor calculado fue de 79,76, mientras que para preprosilpepsina 2 fue de 79,23; estos valores son intermedios con respecto a los de otras PAs. La proteasa aspártica del plátano malayo (*Musa acuminata*) tiene un índice de 83,82 mientras que la PA de papa (*Solanum tuberosum*) tiene un índice de 74,81.

El índice de hidropatía es la variación de Energía Libre (ΔG) asociada a la transferencia de aminoácidos de un medio hidrofóbico a un medio hidrofílico (Muller-Esterl, 2008). Por lo tanto, un grado de hidrofobicidad (índice GRAVY) positivo, es indicativo de secuencias hidrofóbicas. Por el contrario, un índice GRAVY negativo es característico de secuencias hidrofílicas. La presencia de grupos polares en las cadenas laterales localizadas en la superficie de las proteínas, a diferencia de los grupos hidrofóbicos que se encuentran en el interior, es lo que favorece la solubilidad de las proteínas en el agua, y por lo tanto la que determina el valor GRAVY (Gasteiger et al., 2005). Ambas peptidasas presentaron valores bajos de GRAVY determinando que las proteínas son hidrofílicas. Preprosilpepsina 1 presenta un valor de GRAVY similar al de la PA 1 de células oclusivas de *A. thaliana* (-0,054; (Salanoubat et al., 2000), mientras que para preprosilpepsina 2 es similar al de la PA de *Oryza sativa* codificada por el gen Os07g0592200 (-0,164; (Ohyanagi et al., 2006).

Se analizaron las secuencias de los productos proteicos con el servidor SignalP 4.1 que predice la presencia y ubicación de los sitios de escisión del péptido señal en las secuencia de la proteína. El método incorpora una predicción de sitios de corte y una predicción del péptido señal/péptido no señal basadas en una combinación de redes neuronales artificiales. De acuerdo con los resultados del análisis SignalP, ambas PAs poseen un péptido señal que se escinde en la misma posición (Tabla 1; Fig. 5). En ambos casos se trata de proteínas de la vía secretoria.

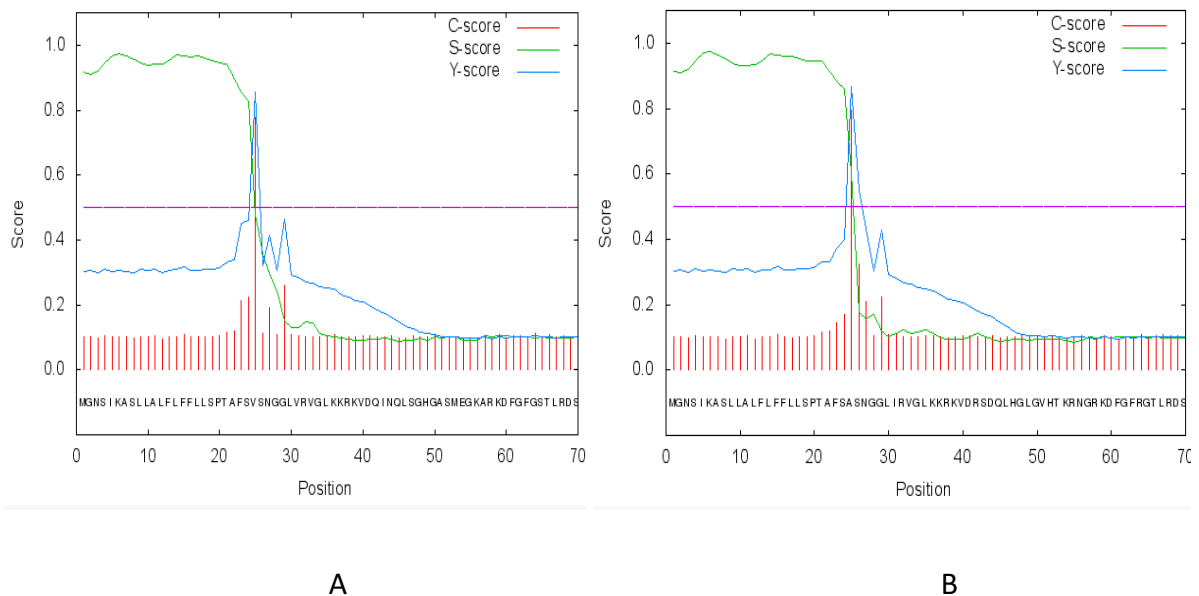


Figura 5. Predicción de sitios de escisión del péptido señal mediante SignalP 4.1. C-score (*score* del sitio de corte en bruto), S-score (*score* del péptido señal) e Y-score (*score* del sitio de corte combinado). **A.** preprosilpepsina 1 y **B.** preprosilpepsina 2

5.3.1. Búsqueda de motivos conservados

Un total de 30 secuencias de PAs de 10 familias diferentes de plantas incluyendo Xanthorrhoeaceae, Poaceae, Polygonaceae, Nepenthaceae, Euphorbiaceae, Solanaceae, Asteraceae, Fabaceae, Myrtaceae y Brassicaceae fueron obtenidas del Centro Nacional de Información Biotecnológica (NCBI, <http://www.ncbi.nlm.nih.gov>) y fueron utilizadas en estos análisis. Sus características incluyendo nombre científico, clase de planta (monocotiledónea o eudicotiledónea), familia,

abreviatura, número de acceso, y longitud de secuencia se enumeran en la Tabla 2.

Un motivo es un patrón, una característica estructural al que se le podría asociar una función, es decir que tiene una significancia biológica. Es común encontrar varios patrones asociados a un motivo, uno que es el más probable y los otros que son las posibles variaciones sobre este. Las familias de proteínas comparten una función o estructura en común, que potencialmente poseen un ancestro común, siendo los motivos o dominios comunes a una familia.

Con el fin de encontrar patrones de los motivos conservados de las dos PAs clonadas se utilizó el programa MEME (Bailey & Elkan, 1994). El método utilizado se basa en localizar estos pequeños motivos conservados y en crear matrices que reflejen la composición de cada una de las posiciones de estos motivos a lo largo de todas las secuencias analizadas. Este análisis permitió encontrar diez motivos conservados (Fig. 6). El diagrama muestra la ubicación de cada motivo; donde cada bloque indica la posición y la fuerza de un motivo, es decir nos da una idea de cuales posiciones son más conservadas en el motivo. Las representaciones gráficas de la conservación de secuencia de cada motivo (*Sequence Logo*) se muestran en la Fig. 1 del Anexo 1.

| Nombre científico | Tipo | Familia | Abreviatura | Accession number | Longitud (aa) |
|---------------------------------------|-----------------|------------------|-------------------|------------------|---------------|
| <i>1- Hemerocallishybrid cultivar</i> | Monocotiledónea | Xanthorrhoeaceae | AP- <i>Hbc</i> | AAC34854.1 | 517 |
| <i>2 - Oryza sativa JaponicaGroup</i> | Monocotiledónea | Poaceae | AP- <i>Os</i> | NP_001043786.1 | 522 |
| <i>3- Zea mays</i> | Monocotiledónea | Poaceae | AP- <i>Zm</i> | NP_001150729.1 | 518 |
| <i>4- Brachypodiumdistachyon</i> | Monocotiledónea | Poaceae | AP- <i>Bd</i> | XP_003567869.1 | 505 |
| <i>5- Triticum aestivum</i> | Monocotiledónea | Poaceae | AP- <i>Ta</i> | BAE20413.1 | 508 |
| <i>6- Hordeumvulgare</i> | Monocotiledónea | Poaceae | AP- <i>Hv</i> | CAA39602 | 508 |
| <i>7- Fagopyrumesculentum</i> | Monocotiledónea | Polygonaceae | AP- <i>Fe</i> | AAV84085.2 | 506 |
| <i>8- Nepenthesalata</i> | Eudicotiledónea | Nepenthaceae | AP- <i>Na1</i> | BAB20969.1 | 514 |
| <i>9- Nepenthes alata</i> | Eudicotiledónea | Nepenthaceae | AP- <i>Na2</i> | BAB20970.1 | 514 |
| <i>10- Ricinus communis</i> | Eudicotiledónea | Euphorbiaceae | AP- <i>Rc</i> | XP_002524837.1 | 456 |
| <i>11- Nicotiana tabacum</i> | Eudicotiledónea | Solanaceae | AP- <i>Nt</i> | ABG37021.1 | 508 |
| <i>12- Helianthus annuus</i> | Eudicotiledónea | Asteraceae | AP- <i>Ha</i> | BAA76870.1 | 509 |
| <i>13- Cynara cardunculus</i> | Eudicotiledónea | Asteraceae | AP- <i>Cc.A</i> | Q9XFX3.1 | 504 |
| <i>14- Cynara cardunculus</i> | Eudicotiledónea | Asteraceae | AP- <i>Cc.B</i> | Q9XFX4.1 | 506 |
| <i>15- Centaurea calcitrapa</i> | Eudicotiledónea | Asteraceae | AP- <i>Cc</i> | CAA70340.1 | 509 |
| <i>16- Cynara cardunculus</i> | Eudicotiledónea | Asteraceae | AP- <i>Cc.cyp</i> | CAA57510.1 | 509 |
| <i>17- Arctium minus</i> | Eudicotiledónea | Asteraceae | AP- <i>Am</i> | AJA72485.1 | 509 |
| <i>18- Castanea mollissima</i> | Monocotiledónea | Fagaceae | AP- <i>Cm</i> | ACX55830.1 | 513 |
| <i>19- Vitis vinifera</i> | Eudicotiledónea | Asteraceae | AP- <i>Vv</i> | XP_002279049.1 | 514 |
| <i>20- Silybum marianum</i> | Eudicotiledónea | Asteraceae | AP- <i>Sm1</i> | AGE15494.1 | 506 |
| <i>21- Silybum marianum</i> | Eudicotiledónea | Asteraceae | AP- <i>Sm2</i> | AGE15495.1 | 509 |
| <i>22- Cirsium vulgare</i> | Eudicotiledónea | Asteraceae | AP- <i>Cv</i> | AFB73927.2 | 509 |
| <i>23- Vigna angularis</i> | Monocotiledónea | Fabaceae | AP- <i>Va</i> | BAT80779.1 | 554 |
| <i>24- Solanum lycopersicum</i> | Monocotiledónea | Solanaceae | AP- <i>Slnep1</i> | Q766C3.1 | 437 |
| <i>25- Solanum lycopersicum</i> | Monocotiledónea | Solanaceae | AP- <i>Slnep2</i> | Q766C2.1 | 438 |
| <i>26- Eucalyptus grandis</i> | Monocotiledónea | Myrtaceae | AP- <i>Eg</i> | XP_010058713.1 | 479 |
| <i>27- Phaseolus vulgaris</i> | Eudicotiledónea | Fagaceae | AP- <i>Pv</i> | AEM05966.1 | 437 |
| <i>28- Arabidopsis thaliana</i> | Eudicotiledónea | Brassicaceae | AP- <i>AtA1</i> | AEE82391.1 | 504 |
| <i>29- Arabidopsis thaliana</i> | Eudicotiledónea | Brassicaceae | AP- <i>AtA2</i> | AEE33947.1 | 513 |
| <i>30- Arabidopsis thaliana</i> | Eudicotiledónea | Brassicaceae | AP- <i>AtA3</i> | AEE82391.1 | 504 |

Tabla 2. Lista de 30 especies diferentes de PAs de plantas

Como se muestra en la Tabla 3 y en la Fig. 6, entre los 10 motivos conservados de las 30 especies de PAs de plantas analizadas, los motivos 1 (50 aa, $E\text{-value} = 1,7 \times e^{-1019}$), 2 (50 aa, $E\text{-value} = 6,1 \times e^{-936}$), 3 (38 aa, $E\text{-value} = 5,3 \times e^{-774}$), 6 (50 aa, $E\text{-value} = 1,6 \times e^{-909}$), 7 (50 aa, $E\text{-value} = 3,1 \times e^{-873}$), y 10 (21 aa, $E\text{-value} = 1,7 \times e^{-245}$) son comunes a casi todas las PAs estudiadas. Mientras que los motivos 5 (50 aa, $E\text{-value} = 6,6 \times e^{-937}$) y 9 (29 aa, $E\text{-value} = 2,0 \times e^{-314}$) sólo se encuentra en 24 de las especies y el motivo 4 (50 aa, $E\text{-value} = 1,5 \times e^{-909}$) sólo en 22 especies de las PAs vegetales estudiadas. Tanto silpepsina 1 (*AP-Sm1*) como silpepsina 2 (*AP-Sm2*) muestran los 10 motivos conservados. Las secuencias catalíticas, DTG y DSG, comunes a todas las PAs de plantas de la familia A1 (Simões I, 2004) se hallan en los motivos 1 y 6 (Fig. 1 del Anexo 1). La PA de *R. communis* no presenta el motivo 4 y cardosina A los motivos 4 y 5 debido a que en ambas peptidasas existen mayores divergencias en estas regiones de la secuencia respecto al patrón con el $p\text{-value}$ empleado como punto de corte. Por otro lado, en las especies *Eucalyptus grandis*, *Hemerocallis hybrid cultivar*, *Solanum lycopersicum*, *Vitis vinífera* y *Phaseolus vulgaris* si bien los motivos están presentes, estos aparecen en la Fig. 6 de un color más claros que el resto; son los denominados “motivos escaneados” que presentan $E\text{-values}$ muy altos, y por lo tanto dichos alineamientos no resultan significativos.

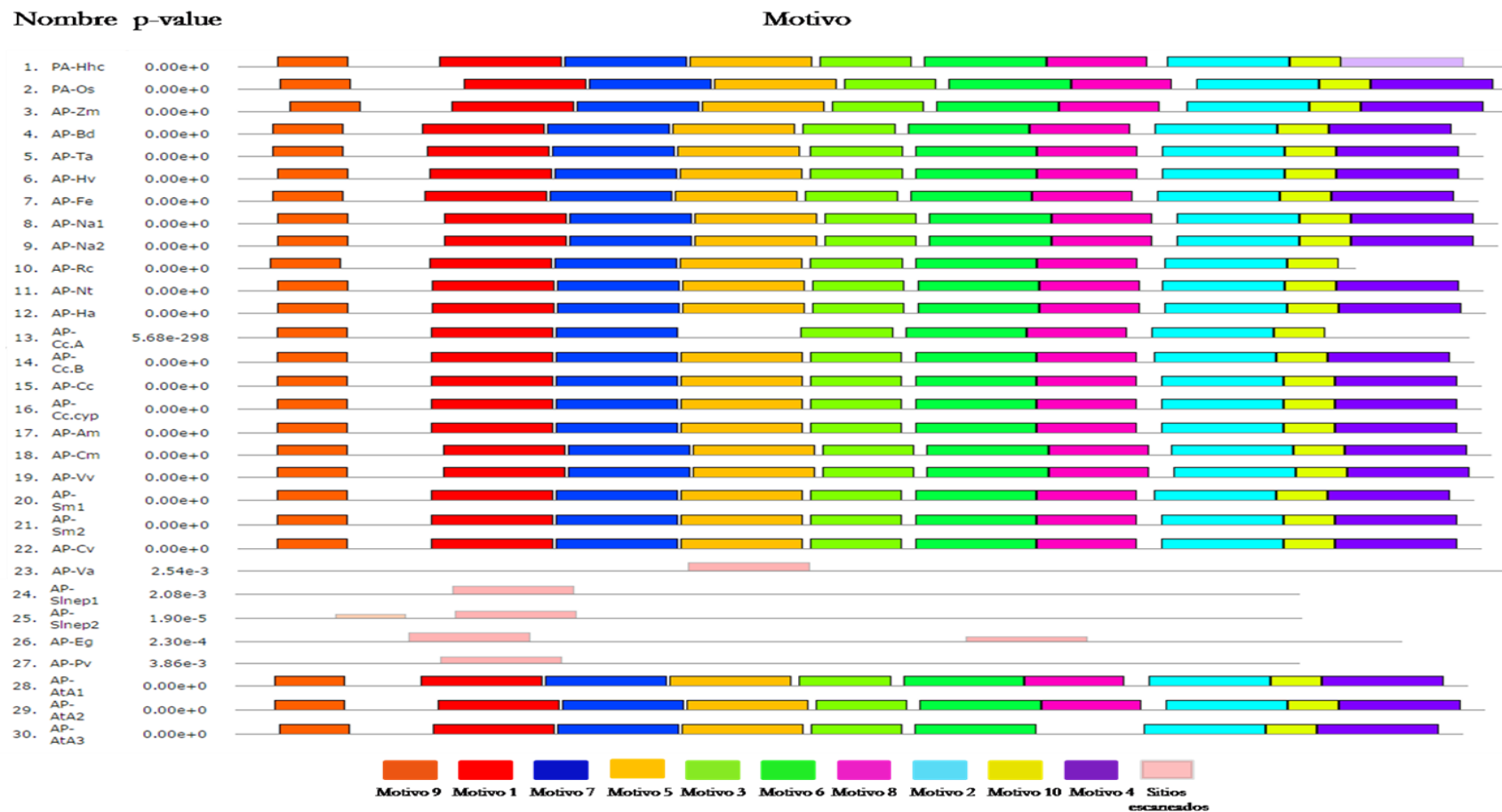


Figura 6. Motivos conservados para las PAs utilizando la herramienta MEME. Los motivos se muestran como cuadros de diferentes colores en las posiciones relativas en las que se encuentran ubicados en cada proteína. Las secuencias y el *Sequence Logo* correspondientes a cada uno de los motivos se muestran en la Fig. 1 del Anexo 1.

| Motivo | E-value | Ancho | sitios | Secuencia |
|--------|-----------------------|-------|--------|--|
| 1 | 1.7e ⁻¹⁰¹⁹ | 25 | 50 | Y[ML][DN][AT]QY[FY]GEI[GA][VI]G[TS]P[PA]Q[KT]FTVIFDTGSS[ND]LWVPS[AS]KCY[FL]S[IV]AC[YL][F L]H[SP][RK]YK |
| 2 | 6.1e ⁻⁹³⁶ | 25 | 50 | HD[EA]MC[TS][FA]CEMAVVWMQN[QE][IL][KAR][RQN][EQ]T[EQ][DE][NR]I[IL][ND]Y[VI]N[EQ]LC[DE] RLPS[PS][MS]GES[AS]VDC |
| 3 | 5.3e ⁻⁷⁷⁴ | 25 | 38 | EGGE[IL]VFGG[VM]DP[NK]H[YF][KR]G[KE]HTYVPVT[QR]KGYWQF[DN]MGDVL |
| 4 | 1.5e ⁻⁹⁰⁹ | 22 | 50 | PE[QE]Y[IV]L[KQ][IV]GEG[EP][AV][AT]QCISGFTA[ML]D[VI][PA]PPRGPLWILGD[VI]FMG[RQ]YHTVFD[YF]G |
| 5 | 6.6e ⁻⁹³⁷ | 24 | 50 | [IV]TF[LV][AV]AKFDGILGLGF[QK]EISVG[KN][SA][VT]P[VL]WYNM[VI][NK]QGL[VI][QK][ED][PA]VFSF W[FL]NR[NH][AV] |
| 6 | 1.6e ⁻⁹⁰⁹ | 25 | 50 | TGFC[SA][DG]GC[AS]AIADSGTSLLAGPT[AT][IV][IV][TA][QE]IN[HE][AK]IGA[SA]GV[MV][SN]Q[QE]CK[TA][LV]V[SD]QY |
| 7 | 3.1e ⁻⁸⁷³ | 25 | 50 | [SG][HQ]SSTYK[KE]NG[KT][SP]A[AS]I[QH]YGTG[SA]I[SA]G[FY][FV]S[QE]D[SH]V[KT][VL]GD[LY][VA]V[KE][EN]Q[DE]FIE[AT]AT[KR]E[PA] |
| 8 | 1.2e ⁻⁶⁴⁰ | 24 | 41 | G[QK][QT]I[LI][DE][LM]L[LY][AS][EQK][TA]QP[KPQ]KICS[QH][VMI][GK]LCTFDG[TAK][RH][GD]VS[AS M][GI][ER]SVVD |
| 9 | 2.0e ⁻³¹⁴ | 24 | 29 | L[LV][SF][PF][TL][AV][FL][SA][VA][SA][NA][GDE]GL[VL]R[VI][GA]LKK[RL][KP][VL]D[QR][IN][ND][QR] |
| 10 | 1.7e ⁻²⁴⁵ | 25 | 21 | [NG]STL[SG][ST]MP[ND][IV][SA]FTIGGK[VK]F[DE]L[TK] |

Tabla 3. Importancia estadística de cada motivo. MEME generalmente encuentra el primer motivo estadísticamente significativo (valor E más bajo). El valor E de un motivo se basa en su relación de verosimilitud, ancho, sitio, frecuencias de las letras de fondo, y el tamaño del conjunto establecido. Entre corchetes se muestran los residuos más probables para una dada posición.

5.3.2. Búsqueda de dominios conservados

La presencia de distintos dominios en combinaciones variables en diferentes proteínas da lugar al diverso repertorio de proteínas que se encuentran en la naturaleza. La identificación de los dominios presentes en una proteína puede ayudar a comprender la función de esa proteína.

Las proteínas o dominios de proteínas se pueden clasificar en familias de secuencias homólogas que derivan de un antecesor común. Existen distintas bases de datos, tales como Pfam y ProDom (Protein Domain Database), que tienen por objeto el estudio de la arquitectura de los dominios de proteínas.

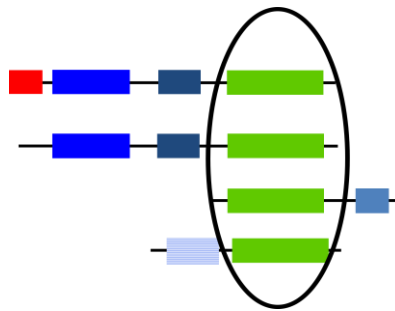


Figura 7. Esquemática de familias de dominios homólogos en distintas proteínas. Cada recuadro coloreado representa un dominio funcional dentro de cada proteína. Los mismos dominios pueden encontrarse en proteínas con distinta función.

ProDom es una base de datos de familias de dominios de proteínas que provienen de la duplicación, pérdida o translocación de dichos dominios (Fig. 7). Por medio de este servidor se determinaron los posibles dominios funcionales que tienen similitudes con las peptidasas estudiadas en este trabajo de Tesis. Un conjunto de familias de dominios de proteínas, se generaron automáticamente desde la base de datos UniProt (*Universal Resource Protein*), el análisis permitió encontrar siete dominios (con mayor score dentro de los dominios analizados) para preprosilpepsina 2, cinco de los cuales son comunes con los presentes en preprosilpepsina 1. La organización de estos dominios en las dos PAs de *S.*

marianum se muestran en la Fig. 8. La posición y validez estadística de los datos se detallan en la Tabla 4.

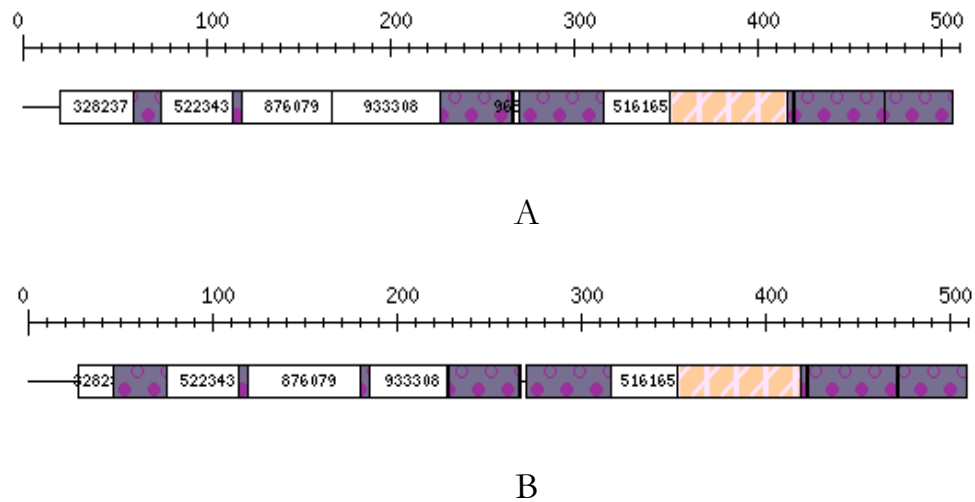


Figura 8. Resultados gráficos de los dominios funcionales mediante el servidor ProDom. A. preprosilpepsina1. B. preprosilpepsina2

| PEPTIDASA | <i>Preprosilpepsina 1 (PA-Sm1)</i> | <i>Preprosilpepsina 1 (PA-Sm2)</i> |
|------------------------|------------------------------------|------------------------------------|
| Dominio ProDom | | |
| <i>PD933308</i> | | |
| Posición | 168-226 | 185-226 |
| Score | 215 | 204 |
| E-value | 2×e ⁻¹⁸ | 5×e ⁻¹⁷ |
| <i>PDA1D2U8</i> | | |
| Posición | 270-315 | 270-315 |
| Score | 182 | 183 |
| E-value | 5×e ⁻¹⁴ | 4×e ⁻¹⁴ |
| <i>PD516165</i> | | |
| Posición | 316-351 | 316-351 |
| Score | 193 | 169 |
| E-value | 2×e ⁻¹⁵ | 3×e ⁻¹² |

| <i>PD001732</i> | | |
|------------------------|--------------------|--------------------|
| Posición | 352-415 | 352-418 |
| Score | 312 | 304 |
| E-value | $7 \times e^{-32}$ | $7 \times e^{-31}$ |
| <i>PDA9Z9E1</i> | | |
| Posición | 420-468 | 423-470 |
| Score | 250 | 231 |
| E-value | $3 \times e^{-23}$ | $1 \times e^{-20}$ |
| <i>PD522343</i> | | |
| Posición | | 75-113 |
| Score | | 176 |
| E-value | | $3 \times e^{-13}$ |
| <i>PDA1D277</i> | | |
| Posición | | 472-508 |
| Score | | 217 |
| E-value | | $9 \times e^{-19}$ |

Tabla 4. Dominios funcionales de las PAs de *S. marianum* por ProDom

Los dominios encontrados están involucrados en dos actividades mayoritarias (mayor probabilidad del término GO, *Gene Ontology*): *PD933308*, *PDA1D2U8*, *PD001732*, *PD522343*, *PDA1D277* y *PDA9Z9E1* presentan actividad peptidasa, en particular endopeptidasa de tipo aspártico, mientras que los dominios *PD001732* y *PD516165* participan en procesos metabólico lipídicos, además el último participa en el metabolismo de esfingolípidos funcionando como activador enzimático.

Por su parte, la base de datos Pfam (Protein Families Database) es una gran colección de familias de dominio de proteína. Cada familia está representada por

múltiples alineamientos de secuencias basados en el modelo oculto de Márkov o HMM (por sus siglas del inglés, Hidden Markov Model). El Pfam (Finn et al., 2014) de los dominios comunes de las peptidasas en estudio fue consistente con los resultados de ProDom, las dos proteínas pertenecen a la familia de aspartilproteasas eucarióticas (PF00026) miembros del clan AA(CL0129), a excepción nuevamente de los dominio PD516165 y PD001732 que pertenecen a la familia saposina B región 2 y saposina B región 1, respectivamente. En la Tabla 5 se indican la posición estimada para estos dominios dentro de las secuencias. El resultado global obtenido se esquematiza en la Fig. 9.

| PEPTIDASA | Dominio | Inicio | Fin | score | | E-value | |
|--------------------|---------|--------|-----|-----------|---------|-----------------------|-----------------------|
| | | | | Secuencia | dominio | Secuencia | dominio |
| Preprosilpepsina 1 | Asp | 84 | 505 | 456 | 455,60 | $2,1 \times e^{-133}$ | $2,8 \times e^{-133}$ |
| | SapB-2 | 318 | 351 | 52,30 | 51,0 | $1,2 \times e^{-10}$ | $2,8 \times 10^{-10}$ |
| | SapB-1 | 378 | 415 | 52,60 | 52,60 | $9,2 \times e^{-11}$ | $9,2 \times e^{-11}$ |
| Preprosilpepsina2 | Asp | 84 | 508 | 432,20 | 431,80 | $3,6 \times e^{-08}$ | $4,6 \times e^{-126}$ |
| | SapB-2 | 318 | 351 | 46,60 | 45,40 | $7,6 \times e^{-09}$ | $1,8 \times e^{-08}$ |
| | SapB-1 | 381 | 418 | 44,30 | 44,30 | $3,7 \times e^{-08}$ | $3,7 \times e^{-08}$ |

Tabla 5. La tabla muestra los límites de los dominios para cada uno de los dominios encontrados en preprosilpepsina 1 y preprosilpepsina 2. Los *E-values* corresponden a la búsqueda de Pfam-A en la base de datos UniProtKB 2015_08 mediante búsqueda HMM.



Figura 9. La imagen muestra la disposición de los dominios Pfam encontrados en preprosilpepsina 1 y preprosilpepsina 2. Los colores en el dibujo corresponden a los colores de los dominios en la tabla.

5.3.3. Búsqueda de las peptidasas homólogas con mayor grado de similitud con las preprosilpepsinas

Para poder comparar los precursores de las preprosilpepsinas con las PAs de plantas homólogas con mayor similitud de secuencia se realizó la búsqueda en el

servidor BLAST. La secuencia aminoacídica de preprosilpepsina 1 mostró un elevado grado de similitud con los precursores de cenprosina (CAA70340) proveniente de *Centaurea calcitrapa* (identidad: 95%, positivos: 97%, E-value: 0), arctiumsina (AJA72485) de *Arctium minus* (identidad: 95%, positivos: 98%, E-value: 0), cirsina (AFB73927) de *Cirsium vulgare* (identidad: 95%, positivos: 97%, E-value: 0), cyprosina B (CAA57510) de *C. cardunculus* (identidad: 94%, positivos: 97%, E-value: 0), Ha-AP de *Helianthus annuus* (BAA76870, identidad: 82%, positivos: 91%, E-value: 0), silpepsina 2 (AGE15495) de *S. marianum* (identidad: 79%, positivos: 86%, E-value: 0), perteneciendo todas estas peptidasas a la familia A1 del clan AA de la base de datos de peptidasas MEROPS.

Por otro lado, preprosilpepsina 2 mostró un elevado grado de similitud con los mismos precursores de la familia A1 pero con un porcentaje menor de identidad y distinto orden. Los resultados en orden decreciente de identidad fueron: arctiumsina (AJA72485) de *A. minus* (identidad: 81%, positivos: 87%, E-value: 0), cardosina B (Q9XFX4) de *C. cardunculus* (identidad: 78%, positivos: 87%, E-value: 0), cenprosina (CAA70340) de *C. calcitrapa* (identidad: 80%, positivos: 87%, E-value: 0), cyprosina B (CAA57510) de *C. cardunculus* (identidad: 79%, positivos: 87%, E-value: 0), cardosina A (Q9XFX3) de *C. cardunculus* (identidad: 76%, positivos: 84%, E-value: 0), cirsina (AFB73927) de *C. vulgare* (identidad: 79%, positivos: 86%, E-value: 0), silpepsina 1 (AGE15494) de *S. marianum* (identidad: 79%, positivos: 86%, E-value: 0), Ha-AP (BAA76870) de *H. annuus* (identidad: 72%, positivos: 84%, E-value: 0).

En base a estas búsquedas de similitud secuencial se alinearon preprosilpepsina 1, preprosilpepsina 2 y las proteínas con mayor grado de similitud con ambas a través del programa Clustal Omega. Este alineamiento fue importado al programa GeneDoc y el resultado se muestra en la Fig. 10. Existe una clara conservación de secuencia, sin embargo puede apreciarse una mayor divergencia

en la secuencia de los propéptidos de los precursores de cardosina A, cardosina B, silpepsina 1 y HA-AP.

5.3.3.1. Comparación de los precursores de peptidasas clonadas con sus homólogos más cercanas

Cuando se compararon el precursor de silpepsina 1 con el de cenprosina, su homólogo más cercano, se pudo observar que las sustituciones se encuentran distribuidas a lo largo de toda la secuencia y que todas ellas son conservativas y semiconservativas (Fig. 2 del Anexo 1). La diferencia más significativa entre ambas enzimas es la presencia de un gap entre los aminoácidos 377-380. Mientras que si se alinean las secuencias aminoacídicas del precursor de silpepsina 2 con su homólogo más cercano, el precursor de arctiumsina, este alineamiento permite observar que existe un gran número de sustituciones, siendo las mismas de carácter conservativo y semiconservativo (Fig. 3 del Anexo 1). Si bien silpepsina 2 y artiumsina están estrechamente relacionadas estas peptidasas presentan mayor divergencia en sus secuencias que silpepsina1 y cenprosina.

5.3.4. Organización de la estructura primaria

Del análisis de los resultados obtenidos mediante las herramientas bioinformáticas empleadas surge que los precursores enzimáticos de *S. marianum* comparten la estructura primaria característica de las peptidasas aspárticas típicas de plantas. Contienen un péptido señal hidrofóbico de 24 aminoácidos, un prosegmento de 44 residuos en el extremo N-terminal y la secuencia polipeptídica de 337 aminoácidos que incluye las dos cadenas de la peptidasa madura, interrumpida por 101 residuos correspondientes al dominio PSI de preprosilpepsina1 y 104 residuos en el caso de preprosilpepsina2. La presencia de este inserto que separa los dominios N-y C-terminales es una característica

distintivas de todas las PAs típicas de plantas y en la mayoría de los casos no está presente en la enzima madura (Vairo Cavalli, Lufrano, Colombo, & Priolo, 2013). Los PSIs han sido llamados swaposins ya que comprenden la porción carboxi-terminal del dominio de una saposina seguido de una secuencia de unión y la porción amino-terminal de dominio de la subsiguiente saposina considerando la disposición en la prosaposina de mamífero (M. H. M. M. S. A. K. S. P. S. and B. P. Brodelius, 2005).

En ambas enzimas se encuentran presentes dos triadas catalíticas altamente conservadas Asp-Thr-Gly (DTG 103-105) y Asp-Ser-Gly (DSG 290-292) y un residuo tirosina en la posición 147 (los cuales se corresponden a los motivos DTG 32-34, DTG 215-217 y el residuo Y 75, respectivamente, usando la numeración de pepsina como referencia).

Asimismo, como se indicó en la Tabla 1 se predijeron para los dos zimógenos, tres sitios de N-glicosilación conservados en varias PAs típicas de plantas (Domingos et al., 2000; Vieira et al., 2001). Para prosilpepsina 1 en los residuos N139 (motivo NGT), N397 (motivo NQT) y N433 (motivo NIS), mientras que en prosilpepsina 2 dos sitios putativos en los residuos N97 (motivo NFT), N139 (motivo NGT) y un tercer sitio presente en el dominio PSI N252 (motivo NHT).

| | * | 20 | * | 40 | * | 60 | * | 80 | * | 100 | * | | | | | | | | | | | | | | |
|-------------|---|-----------|----------------|---------|----------|---------|-----------|-----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|--------|---------|-------|---------|---------|-----|-----|
| cardosinaA | : | MGTSIKANV | LALFLFYLLSPTV | FVSVDGL | IRIGLKKR | VDRIDOL | RGRALMEGN | ARKDFGFRG | IVDS | GSVAVALT | NDRDTSY | FGEIGIGT | PPQKFT | VIFDTGSS | VLWVPSSK | : | 115 | | | | | | | | |
| cardosinaB | : | MGTPIKAS | LLALFLFLLSPTAF | SVSNG | LRLVGLK | RRKVDRL | DOLRAH | CVHMLN | ARKDFGFR | ILSDS | SGSIVALT | NDRDTAY | YGEIGIGT | PPQKFA | VIFDTGSS | DLWVPSTK | : | 115 | | | | | | | |
| silpepsina2 | : | MGNSIKAS | LLALFLFLLSPTAF | SASNG | LRLVGLK | RRKVDRS | DOLHGL | GVHTKR | NGRKDFG | FRGIL | SDS | DGTGII | VLTNER | DTTYF | GEIGIGT | PLOKFT | VIFDTGSS | DLWVPSSK | : | 115 | | | | | |
| Ha-AP | : | MSTTIKTS | LLALFLFLLSPTAF | SSTK | GLLRVGL | KKRKT | NQFN | RVSEH | GLSMEG | TDRN | FGFYD | ILN | SEGD | VI | VLN | YMDA | QYFGE | IGIGT | PPQKFT | VVFD | TGSAN | LWVPSSK | : | 115 | |
| silpepsina1 | : | MGNSIKAS | LLALFLFLLSPTAF | SVSNG | LRLVGLK | RRKVDQ | INOL | SGH | CASMEG | KARKD | FGFS | IL | SDS | SDI | I | ALKN | YMDA | QYYGE | IGIGEP | ACKFT | VIFDT | GSSN | LWVPSAK | : | 115 |
| cyprosinaB | : | MGTAIKAS | VLALFLFLLSPTAF | SVSNG | LRLVGLK | RRKVDQ | INOL | SGH | CVSME | AKARKD | FGG | AL | SDS | SDI | I | ALKN | YMDA | QYYGE | IGIGS | PPQKFT | VIFDT | GSSN | LWVPSAK | : | 115 |
| cenprocina | : | MGTAIKAS | LLALFLFLLSPTAF | SASNG | LRLVGLK | RRKVDQ | INOL | RN | H | CASMEG | KARKD | FGG | SDS | SDI | I | ALKN | YMDA | QYYGE | IGIGS | PAKFT | VIFDT | GSSN | LWVPSAK | : | 115 |
| arctiumsina | : | MGNSIKAS | LLALFLFLLSPTAF | SVSNG | LRLVGLK | RRKVDQ | INOL | R | H | CASMEG | KARKD | FGG | SDS | SDI | I | ALKN | YMDA | QYYGE | IGIG | APPOKFT | VIFDT | GSSN | LWVPSAK | : | 115 |
| cirsina | : | MGTSIKAS | LLALFLFLLSPTA | ISVSD | GLIRVGL | KKRVDQ | INOL | SGH | CASMEG | KARKD | FGG | IL | SDS | SDI | I | ALKN | YMDA | QYYGE | IGIG | APPOKFT | VIFDT | GSSN | LWVPSAK | : | 115 |

| | | 120 | * | 140 | * | 160 | * | 180 | * | 200 | * | 220 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------------|---|-----------|----------|----------|--------|-------|------|------|------|-------|-------|-------|-------|----|------|-------|-------|-------|------|--------|-----|-----|-----|-------|------|---|---|---|---|---|---|---|---|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|---|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| cardosinaA | : | CINSKACRA | HSMEYESS | DSSTYK | ENGTFG | AI | YGTG | SI | IGF | SQDSV | TIGDL | VVKE | QDFIE | AT | DEAD | NVFL | HRLFD | GILGL | SFQ | TIS--- | VPV | WYN | NML | NQGLV | KERR | F | S | F | W | L | N | R | : | 226 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| cardosinaB | : | CDTSLACV | IHPRYD | SGDSSTYK | NGNT | PA | SI | OYGT | AI | VGF | YSQDS | VEV | GD | LV | VEH | QDFIE | T | TEED | DTV | FLK | SEF | D | G | L | G | L | G | F | Q | E | I | S | A | G | K | A | V | P | V | W | Y | N | M | V | N | Q | G | L | V | E | E | A | F | S | F | W | L | N | R | : | 230 | | | | | | | | | | | | | | | | | | | | | | |
| silpepsina2 | : | CYSSVSC | WVHSRY | KSSDS | SSTYK | NGT | SG | AI | OYGT | SI | IGF | YSQDS | VEV | GD | LV | VEH | QDFIE | T | KEAD | NVFL | RAH | F | D | G | L | G | L | G | F | Q | E | I | S | V | R | K | A | V | P | V | W | Y | N | M | V | D | Q | G | L | V | Q | E | P | V | F | S | F | W | L | N | R | : | 230 | | | | | | | | | | | | | | | | | | | | |
| Ha-AP | : | CFLSVAC | LHQK | YKAS | RSSTYK | KNGT | AA | AI | OYGT | AI | SGV | FSR | DS | V | K | L | G | D | L | V | V | K | E | Q | D | F | I | E | A | T | R | E | P | G | I | T | F | L | A | A | K | F | D | G | L | G | L | G | Y | Q | D | I | S | V | G | K | A | V | P | V | W | Y | N | M | V | N | Q | G | L | V | Q | E | P | V | F | S | F | W | F | N | R | : | 230 |
| silpepsina1 | : | CYFSVAC | LH | SKYK | SSHS | SSTYK | KNGT | SAA | AI | OYGT | SI | IGF | YSQDS | V | K | L | G | D | L | V | V | K | E | Q | D | F | I | E | A | T | K | E | P | G | I | T | F | L | A | A | K | F | D | G | L | G | L | G | F | Q | E | I | S | V | G | K | S | V | P | V | W | Y | N | M | V | N | Q | G | L | V | Q | E | P | V | F | S | F | W | F | N | R | : | 230 |
| cyprosinaB | : | CYFSVAC | LH | SKYK | SSHS | SSTYK | KNGT | SAA | AI | OYGT | SI | IGF | YSQDS | V | K | L | G | D | L | V | V | K | E | Q | D | F | I | E | A | T | K | E | P | G | I | T | F | L | A | A | K | F | D | G | L | G | L | G | F | Q | E | I | S | V | G | K | S | V | P | L | W | Y | N | M | V | N | Q | G | L | V | Q | E | P | V | F | S | F | W | F | N | R | : | 230 |
| cenprocina | : | CYFSVAC | LH | SKYK | SSHS | SSTYK | KNGT | SAA | AI | OYGT | SI | IGF | YSQDS | V | K | L | G | D | L | V | V | K | E | Q | D | F | I | E | A | T | K | E | P | G | V | T | F | L | A | A | K | F | D | G | L | G | L | G | F | Q | E | I | S | V | G | K | S | V | P | V | W | Y | N | M | V | N | Q | G | L | V | Q | E | P | V | F | S | F | W | F | N | R | : | 230 |
| arctiumsina | : | CYFSVAC | LH | SKYK | SSHS | SSTYK | KNGT | SAA | AI | OYGT | SI | IGF | YSQDS | V | K | L | G | D | L | V | V | K | E | Q | D | F | I | E | A | T | K | E | P | G | V | T | F | L | A | A | K | F | D | G | L | G | L | G | F | Q | E | I | S | V | G | K | S | V | P | V | W | Y | N | M | V | N | Q | G | L | V | Q | E | P | V | F | S | F | W | F | N | R | : | 230 |
| cirsina | : | CYFSVAC | LH | SKYK | SSHS | SSTYK | KNGT | SAA | AI | OYGT | SI | IGF | YSQDS | V | K | L | G | D | L | V | V | K | E | Q | D | F | I | E | A | T | K | E | P | G | I | T | F | L | A | A | K | F | D | G | L | G | L | G | F | Q | E | I | S | V | G | K | S | V | P | V | W | Y | N | M | V | N | Q | G | L | V | Q | E | P | V | F | S | F | W | F | N | R | : | 230 |

| | | 240 | * | 260 | * | 280 | * | 300 | * | 320 | * | 340 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------------|---|-----------|---------|---------|---|-------|-------|------|------|-------|-------|------|------|------|-------|-------|-------|----|------|------|------|-----|-----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|---|-----|-----|-----|
| cardosinaA | : | VDEEEGGEL | VFGGLDP | NHFRGD | H | TYVPV | YQYY | WQFG | I | GDVLI | GDKS | TGFC | APGC | QAF | ADSGT | SL | SGPT | AI | VTQ | IN | AI | G | A | N | G | V | M | N | Q | O | C | K | T | V | V | S | R | Y | G | R | D | I | I | E | M | L | S | K | I | Q | P | D | K | I | C | : | 341 | | | | |
| cardosinaB | : | VDEEEGGEL | VFGGVD | PNHFRGN | H | TYVPV | YQYY | WQFG | F | MGDVL | I | GDKS | SGFC | AGGC | AA | ADSGT | S | FF | AGPT | AI | ITQ | IN | AI | G | A | K | G | V | L | N | Q | O | C | K | T | L | V | G | O | Y | G | K | N | M | I | Q | M | L | S | E | V | Q | P | D | K | I | C | : | 345 | | |
| silpepsina2 | : | VNEEEGGEL | VFGGVD | PNHFRGN | H | TYVPV | YQYY | WQFG | F | MGDVL | I | GDKS | TGFC | ANGC | AA | ADSGT | T | L | T | GPT | AI | VTQ | IN | AI | G | A | N | R | V | M | N | Q | O | C | K | T | L | V | S | R | Y | G | K | N | I | M | Q | M | L | S | E | A | Q | P | D | K | I | C | : | 345 | |
| Ha-AP | : | TGEEEGGEL | VFGGVD | PNHFKG | K | H | TYVPV | YQYY | WQFG | F | MGDVL | I | GDKT | TGFC | SGGC | AA | ADSGT | S | L | L | AGPT | AI | ITQ | IN | AI | G | A | A | G | V | M | S | Q | O | C | K | T | L | V | D | O | Y | G | K | T | I | I | E | M | L | S | E | A | Q | P | D | K | I | C | : | 345 |
| silpepsina1 | : | ADEEEGGEL | VFGGVD | PNHFKR | H | TYVPV | YQYY | WQFG | F | MGDVL | I | EDKT | TGFC | SDGC | AA | ADSGT | S | L | L | AGPT | AI | ITQ | IN | AI | G | A | K | G | V | M | S | Q | O | C | K | T | L | V | D | O | Y | G | K | S | I | I | E | M | L | S | E | A | Q | P | D | K | I | C | : | 345 | |
| cyprosinaB | : | ADEEEGGEL | VFGGVD | PNHFKG | K | H | TYVPV | YQYY | WQFG | F | MGDVL | I | EDKT | TGFC | SDGC | AA | ADSGT | S | L | L | AGPT | AI | ITE | IN | AI | G | A | K | G | V | M | S | Q | O | C | K | T | L | V | D | O | Y | G | K | T | I | I | E | M | L | S | E | A | Q | P | D | K | I | C | : | 345 |
| cenprocina | : | ADEEEGGEL | VFGGVD | PNHFKG | K | H | TYVPV | YQYY | WQFG | F | MGDVL | I | EDKT | TGFC | ADGC | AA | ADSGT | S | L | L | AGPT | AI | ITQ | IN | AI | G | A | K | G | V | M | S | Q | O | C | K | T | L | V | D | O | Y | G | K | T | I | I | E | M | L | S | E | A | Q | P | D | K | I | C | : | 345 |
| arctiumsina | : | ADEEEGGEL | VFGGVD | PNHFKG | K | H | TYVPV | YQYY | WQFG | F | MGDVL | I | EDKT | TGFC | SDGC | AA | ADSGT | S | L | L | AGPT | AI | VTQ | IN | AI | G | A | K | G | V | M | S | Q | O | C | K | T | L | V | D | O | Y | G | K | T | I | I | E | M | L | S | E | A | Q | P | D | K | I | C | : | 345 |
| cirsina | : | ANEEEGGEL | VFGGVD | PNHFKG | K | H | TYVPV | YQYY | WQFG | F | MGDVL | I | EDKT | TGFC | SDGC | AA | ADSGT | S | L | L | AGPT | AI | ITE | IN | AI | G | A | K | G | V | M | S | Q | O | C | K | T | L | V | D | O | Y | G | K | S | I | I | E | M | L | S | E | A | Q | P | D | K | I | C | : | 345 |

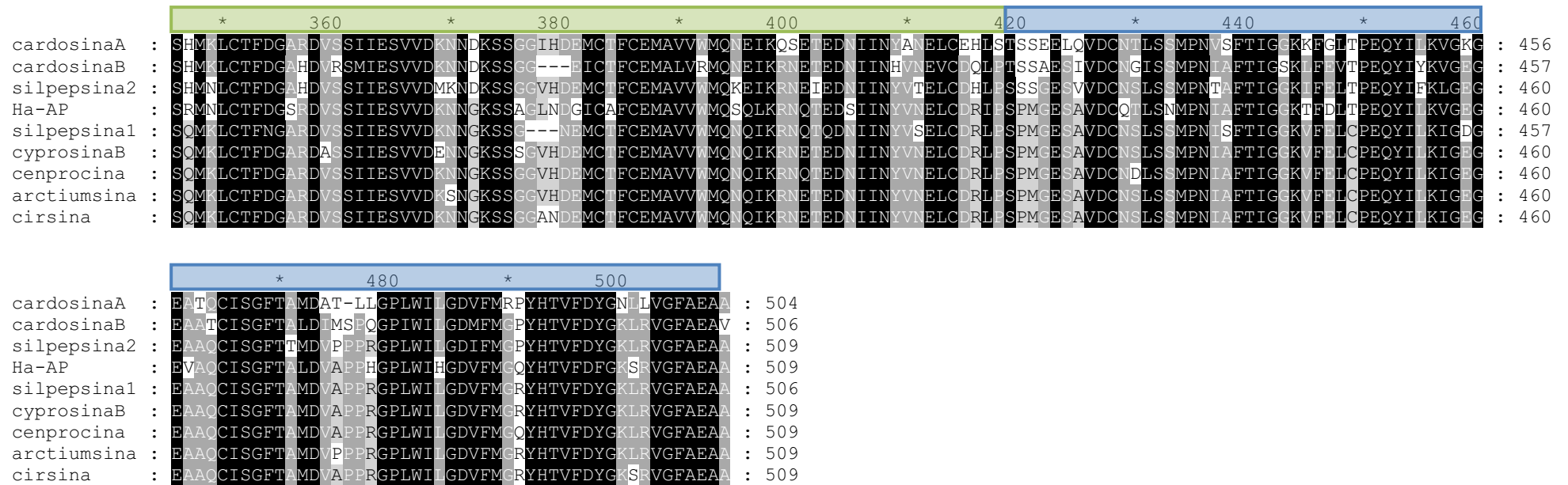


Figura10. Alineamiento múltiple entre los precursores de silpepsina 1 (AGE15494), silpepsina 2 (AGE15495), cardosina A (Q9XFX3.1), cardosina B (CAL07969), cenprosinA (CAA70340), cirsinA (AFB73927.2) arctiunsina (AJA72485.1), ciprosina B (CAA57510) y Ha-AP (BAA76870). Las secuencias fueron alineadas a través del programa Clustal omega e importadas al programa GeneDoc. Se utilizaron cuatro niveles de conservación: 100% de conservación con fondo negro; 80% de conservación con fondo gris y letra blanca; 60% de conservación fondo gris con letra negra y menor al 60% de conservación con fondo blanco. En recuadro de color naranja se muestra el péptido señal, en rojo el prosegmento, en verde el dominio PSI que separa las dos cadenas de la forma

Los esquemas de la Fig. 11 resumen las características de los dos precursores de PAs clonados.

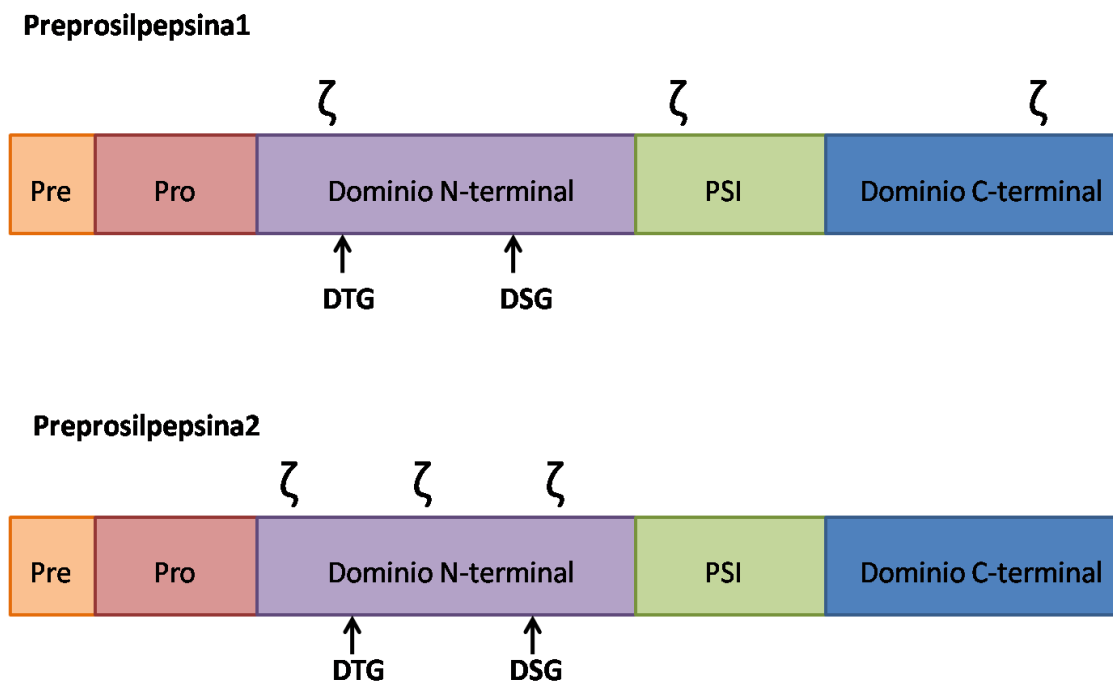


Figura 11. Representación esquemática de la organización de los dominios de los precursores de silpepsina 1 y silpepsina2. Ambas estructuras incluyen un péptido señal (Pre), un prosegmento (Pro) y un dominio PSI que separa las dos cadenas de la forma madura (dominios N-y C-terminal). Se representan las diferencias en las posiciones de los sitios de N-glicosilación entre las dos PAs.

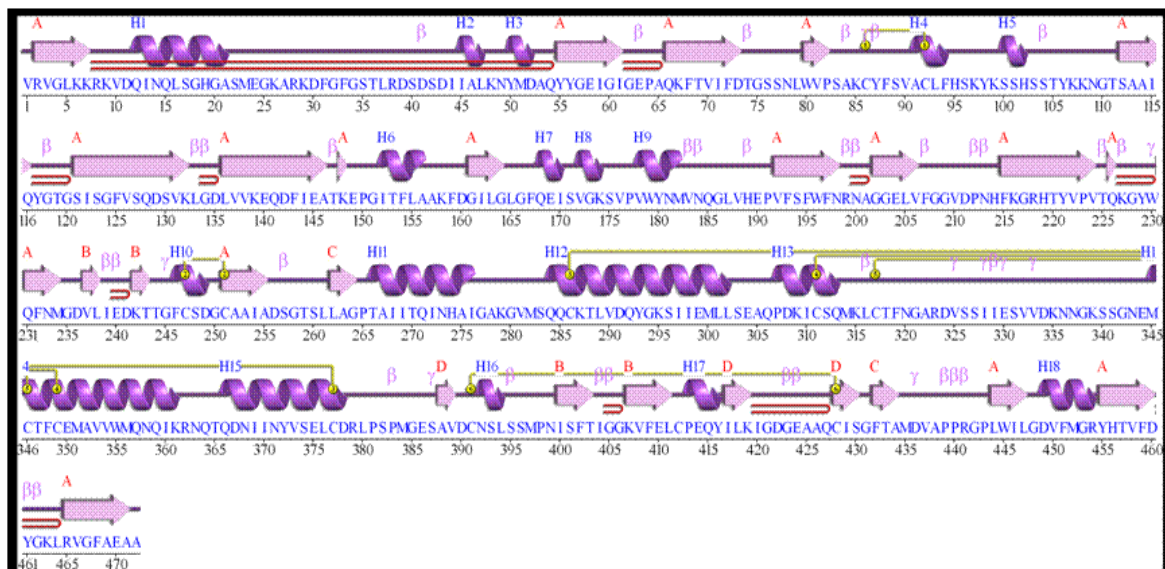
5.3.4.1. Comparación de los precursores de las peptidasas clonadas entre sí

Al ser comparadas entre sí preprosilpepsina 1 y preprosilpepsina 2 se observan algunas diferencias (Fig. 4 del Anexo 1). Estas son tres aminoácidos extra con que cuenta el PSI del segundo precursor y los sitios de N-glicosilación distribuidos en el dominio N-terminal, a diferencia de preprosilpepsina 1 que presenta los tres sitios distribuidos en distintos dominios de la molécula (el dominio N-terminal, el PSI y el dominio C-terminal). Las sustituciones son en su

mayoría conservativas y los tres aminoácidos extra que tiene pro-silpepsina 2 están consecutivos en el PSI.

5.3.5. Predicción de la estructura secundaria

Por medio del programa ProFunc se predijeron los patrones de plegamiento de la estructura secundaria de los zimógenos. Los arreglos regulares de la estructura secundaria (Fig. 12 A y B) son hojas β , harpins β , ψ loops, protuberancias β , hebras β , α -hélices, interacciones α -hélice- α -hélice, giros β , giros γ y puentes disulfuros (Fig. 14 y 15 C). Tal como fuera inferido por otros métodos bioinformáticos los resultados de esta predicción mostraron que las proenzimas presentan seis puentes disulfuros, tres de los cuales se encuentran en el PSI. Se puede observar en las Fig. 14 y 15 .que de los 476 residuos de pro-silpepsina 1 y de los 479 para pro-silpepsina 2, en su mayoría forman hebras β y α -hélices.



A

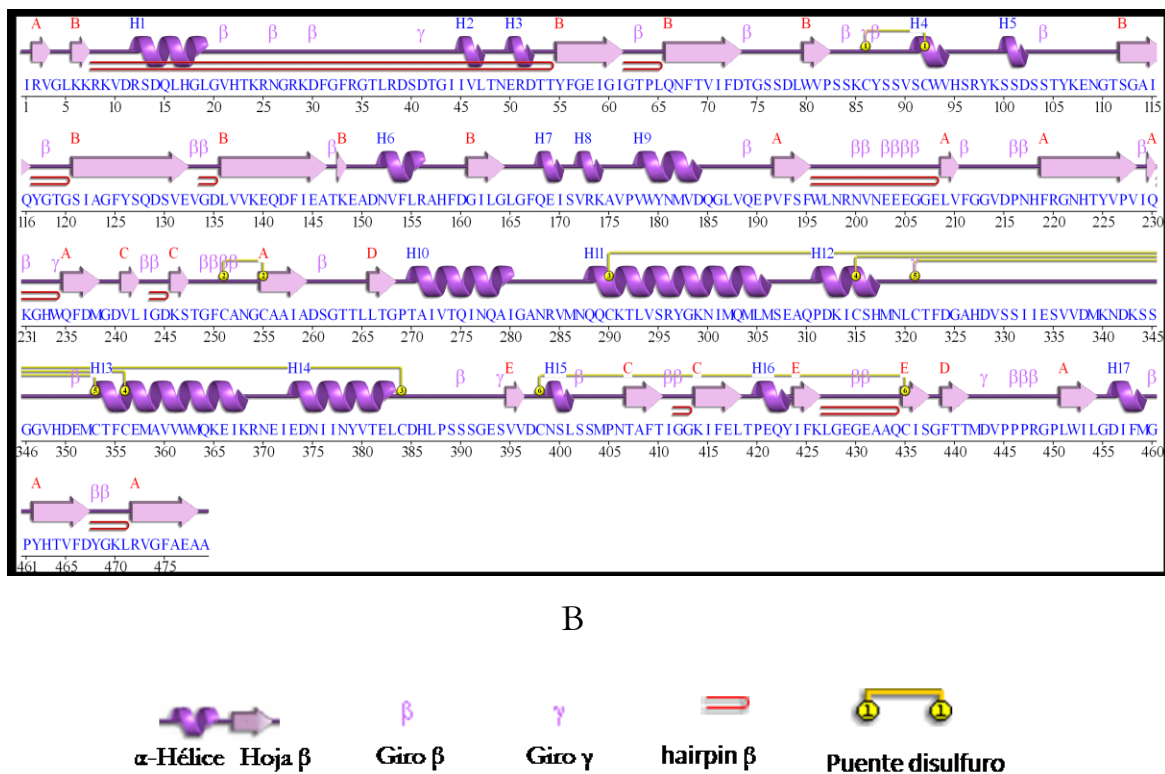


Figura 12. Estructura secundaria: A proinsulina 1. B proinsulina 2. Las hélices se enumeran como H1, H2, etc. y las hojas como A, B, etc.

5.3.6. Predicción de estructura terciaria, evaluación y validación del modelo

Cuando una secuencia de estructura desconocida tiene un homólogo claro de estructura conocida se puede construir un modelo basándose en esta estructura. El modelado comparativo de proteínas utiliza estructuras resueltas previamente como punto de partida o plantillas basado en la suposición de que dos proteínas homólogas comparten estructuras muy similares. Puesto que un plegamiento proteico está evolutivamente más conservado que su secuencia de aminoácidos, una secuencia objetivo puede ser modelada con una precisión razonable sobre una plantilla relacionada muy distante, siempre que la relación entre objetivo y plantilla sea perceptible en el alineamiento de sus secuencias.

Con el objetivo de predecir la estructura tridimensional de prosilpepsina 1 y prosilpepsina 2, se construyeron modelos estructurales mediante modelado por homología.

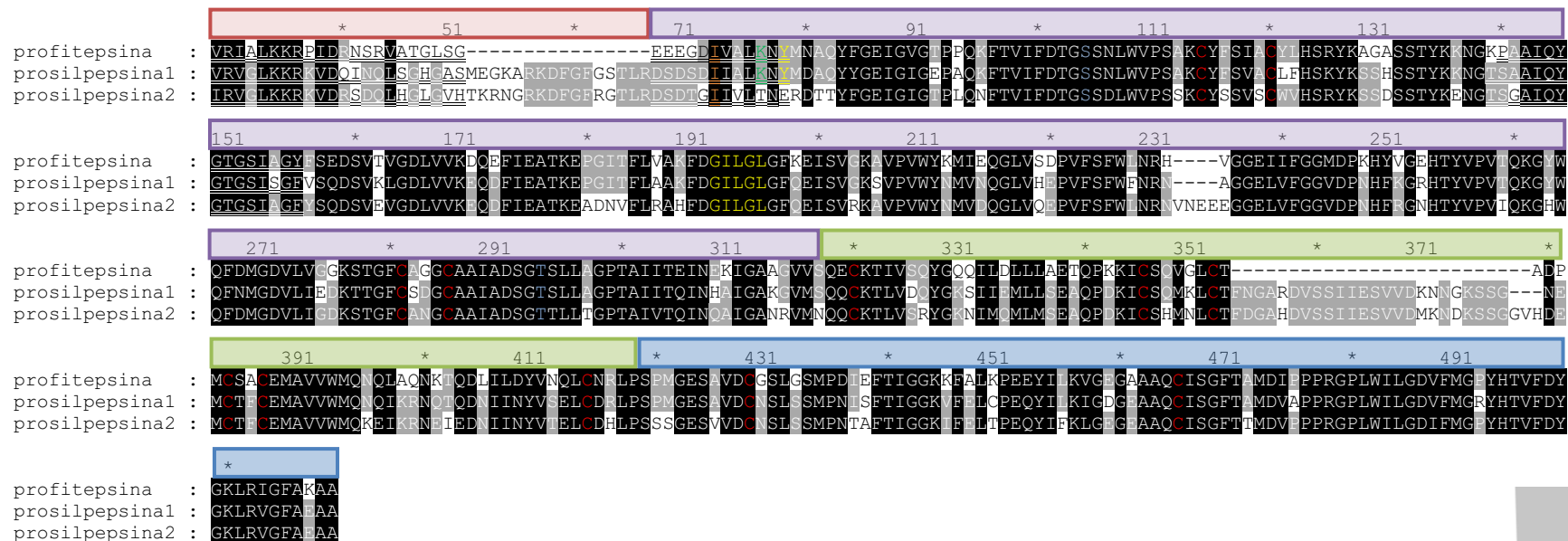


Figura 13. Comparación de la secuencias profitepsina, prosilpepsina 1 y prosilpepsina 2 utilizadas en la construcción de los modelos tridimensionales. En la secuencia de profitepsina (PDB: 1QDM) se indican con guiones las posiciones de los residuos que debido a su muy baja densidad electrónica no fueron resueltos en la cristalografía y por lo tanto tampoco se consideraron en los modelos estructurales. Las secuencias fueron alineadas a través del programa Clustal omega e importadas al programa GeneDoc. Se utilizaron cuatro niveles de conservación: 100% de conservación con fondo negro; 80% de conservación con fondo gris y letra blanca; 60% de conservación fondo gris con letra negra y menor al 60% de conservación con fondo blanco. Se indican en color naranja la Ile74, en verde la Lys78, en amarillo la Tyr80 y en celeste la Ser 106 y Thr 293 (numeración de la preprosilpepsinas) que se corresponden con Lys11, Tyr13, Ile7, Ser39 y Thr 226 numeración de la profitepsina. En rojo se muestran las cisteínas involucradas en los puentes disulfuros y en amarillo la secuencia GILGL (ψ loop). El péptido y el N-terminal de la proteína madura se conectan en la hendidura del sitio activo por interacción iónica entre la Lys78 y los aspárticos catalíticos 103 y 290. El grupo hidroxilo de la Tyr80 forma un enlace de hidrógeno con el Asp103. La Ile74 se coloca en el bolsillo de unión al sustrato S1'. Los grupos hidroxilo de los residuos

La estructura de profitepsina (PDB: 1QDM) se tomó como referencia, por tratarse de la proteína homóloga cristalizada más próxima a ambas proenzimas según la búsqueda de similitud de secuencias realizada. El modelo estructural de profitepsina tomado como plantilla incluye los residuos 6p a 26p del propéptido (de los 41 de su secuencia), 2 al 164, 168 al 37s y 64s–338 (residuos que forman parte del PSI: 1s–104s). El resto de los aminoácidos del propéptido, el primer residuo de la proteína madura y 27 residuos del PSI (38s-64s, según la numeración de profitepsina) no fueron incluidos (Fig. 13) debido a que la densidad electrónica era muy baja y no pudieron tenerse en cuenta en la cristalografía (Kervinen et al., 1999). Dado que estos aminoácidos no se tuvieron en cuenta al resolver la estructura de profitepsina, en el modelado de prosilpepsina 1 y prosilpepsina 2 tampoco pudieron incluirse

Utilizando el programa Modeller se construyeron 100 modelos para ambas proenzimas. Para cada modelo construido se evaluó el valor de la función objetivo (*objective function*) a fin de elegir el mejor de ellos; siendo este el que minimiza el valor de dicho parámetro y se corresponde con el modelo que viola el menor número de restricciones espaciales.

Las Fig. 14 y 15 muestran los modelos moleculares elegidos para prosilpepsina 1 y prosilpepsina 2, respectivamente.

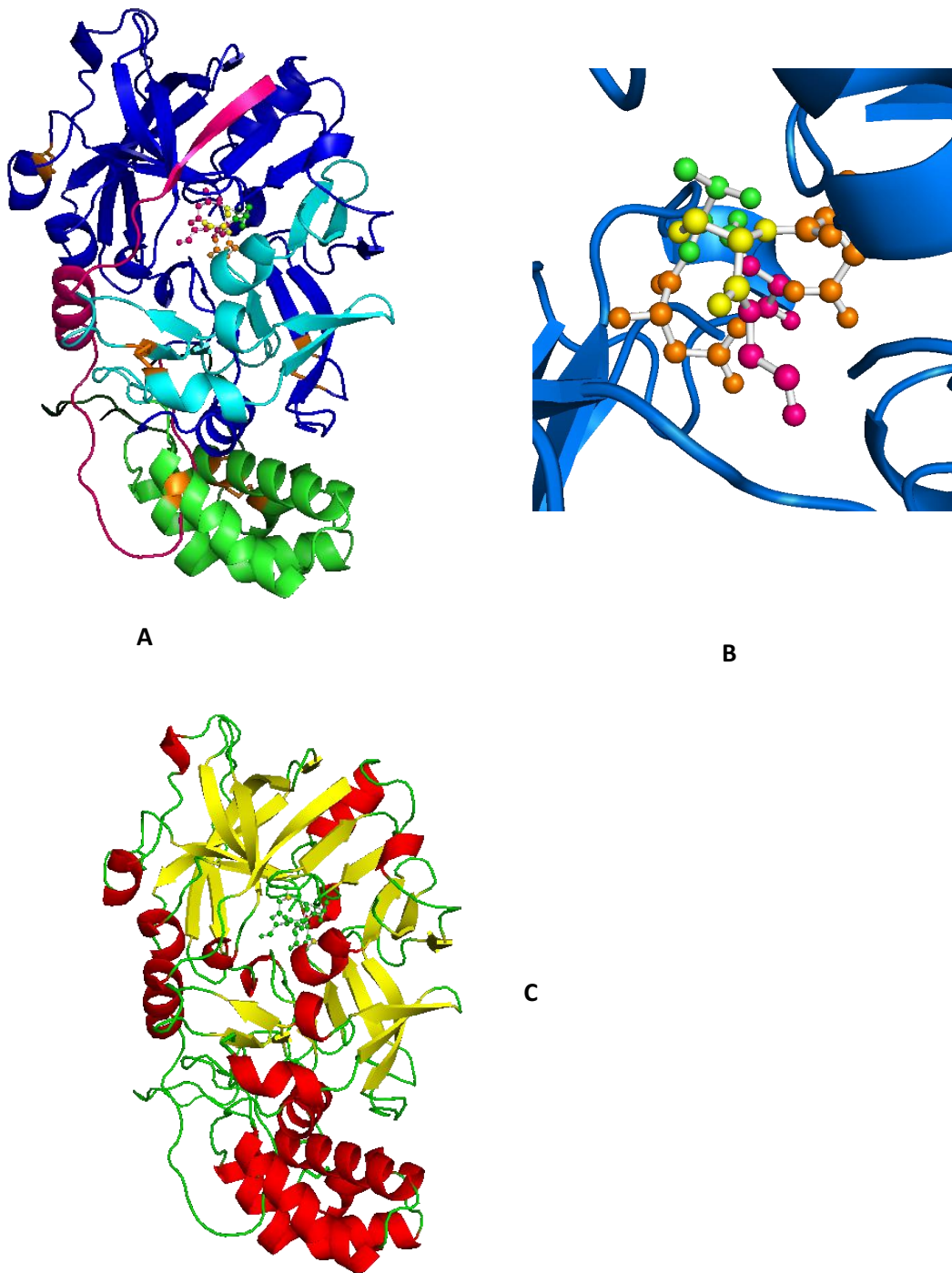


Figura 14. Modelo estructural de proinsulina 1. La secuencia de proinsulina1 se modeló por técnicas de homología con el programa MODELLER usando la estructura cristalográfica conocida de proinsulina. **A.** Modelo de proinsulina 1: en azul se observa la subunidad mayor, en celeste la subunidad menor, en verde el PSI y en fucsia la secuencia propéptido. En naranja se observan los puentes disulfuro. **B.** Sitio activo. Los residuos de las tríadas catalíticas se muestran en la representación de barras y esferas; Asp103 y Asp290 en color naranja, Thr104 en verde, Gly105 y Gly292 en rosa y Ser291 en amarillo (numeración de proinsulina 1). **C.** Modelo mostrando el ordenamiento de hojas beta y alfa hélices. En amarillo se muestran las hojas beta y en rojo las alfa hélices, mientras que en verde se muestran los *loops* y el sitio activo

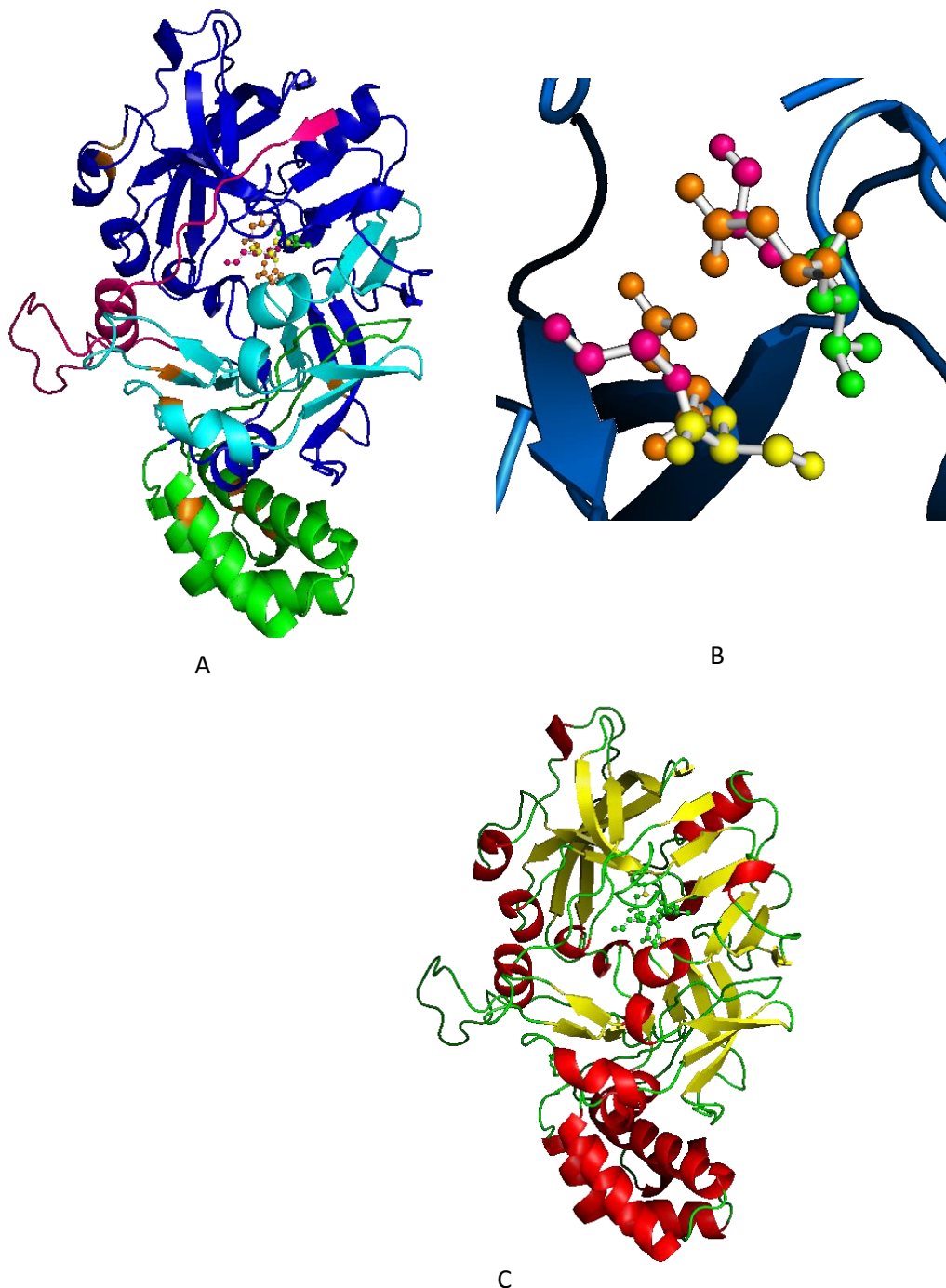


Figura 15. Modelo estructural de prosilpepsina 2. La secuencia de prosilpepsina 2 se modeló por técnicas de homología con el programa MODELLER usando la estructura cristalográfica conocida de profitepsina. **A.** Modelo de prosilpepsina 2: en azul se observa la subunidad mayor, en celeste la subunidad menor, en verde el PSI y en fucsia la secuencia propéptido. En naranja se muestran los enlaces disulfuro. **B.** Sitio activo. Los residuos de las tríadas catalíticas se representan con barras y esferas; Asp103 y Asp290 en color naranja, Thr104 en verde, Gly105 y Gly292 en rosa y Ser291 en amarillo (numeración de preprosilpepsina 2). **C.** Modelo mostrando el ordenamiento de hojas beta y alfa hélices, en amarillo se muestran las hojas beta y en rojo las alfa hélices, mientras que en verde se muestran los *loops* y el sitio activo

El plegamiento de los zimógeno se divide en tres elementos principales: el propéptido, los dos dominios que conjuntamente forman la peptidasa madura y el PSI que aparece como un dominio independiente de la enzima. Las silpepsinas están formadas por dos cadenas polipeptídicas y su plegamiento es típico de otras PAs. Dicho plegamiento consiste en dos dominios topológicamente similares tipo barril β (Figs. 14 C y 15 C) con los residuos de ácido aspártico catalíticos localizados en una hendidura entre los dominios. Seis puentes disulfuro contribuyen al plegamiento y estabilizan la estructura, tres de ellos se encuentran en el PSI (Cys 116-122, 281-285, 319-411, 345-380, 351-383, 425-462 para preprosilpepsina 1 y Cys 116-122, 281-285, 320-414, 345-386, 351-383, 428-465 para preprosilpepsina 2; en color rojo en las Fig. 13 o naranja en las Fig. 14 A y 15 A). Los puentes disulfuros que se infirieron tras la comparación con otras secuencias de PAs coinciden con los que predicen los modelos.

En los modelos aquí presentados el propéptido, junto con la cadena N-terminal, envuelven la hendidura interdominio donde se ubica el sitio activo. El propéptido al cubrir parcialmente esta hendidura conduciría a la inactivación del sitio activo en la proenzima (Fig. 14 A y 15 A, se muestra el propéptido de color fucsia y los dominios de la enzima madura en azul y celeste). En las PAs hay una zona flexible conocida como flap que se proyecta a lo largo de la hendidura del sitio activo encerrando a sustratos e inhibidores. En las prosilpepsinas, al igual que en profitepsina, el propéptido, los 13 residuos del extremo N-terminal de la enzima madura y la horquilla- β que contiene los residuos 75-88 de la enzima madura, que compone el flap, cubren completamente el sitio activo, quedando de este modo bloqueado (Figs. 13 y 14 A). Además el flap contiene residuos esenciales para la especificidad hidrolítica que pueden interactuar con sustratos como parte del subsitio S1. Durante el ciclo catalítico el flap debe abrirse para permitir la entrada de los sustratos en la hendidura así como para liberar los productos de hidrolisis (Vairo Cavalli et al., 2013).

En profitepsina la estructura del zimógeno se estabiliza además de con los puentes disulfuro, por uniones iónicas y varios puentes de hidrógeno entre el propéptido de carga positiva y las regiones de la enzima madura cargadas negativamente, incluyendo las interacciones entre Lys78 (numeración de preprosilpepsina) y los aspárticos del sitio activo (Kervinen et al., 1999). Tanto en la estructura de pepsina A (PDB 4PEP) como en la de otras PAs de la familia A1, Andreeva & Rumsh (2001) demostraron que las interacciones a través de puentes de hidrógeno entre los aminoácidos adyacentes al sitio activo están conservadas. Estos puentes de hidrógeno se establecen entre los grupos hidroxilo de los residuos Ser 106 y Thr 293 (numeración de preprosilpesinas) con los carboxilos de los Asp 103 y Asp 290, respectivamente (Fig. 13).

En prosilpepsina 1, como sucede con profitepsina, el propéptido y el N-terminal de la proteína madura se conectan en la hendidura del sitio activo por interacción iónica entre la Lys78 y los aspárticos catalíticos 103 y 290. Además, el grupo hidroxilo de la Tyr80 forma un enlace de hidrógeno con el Asp103, mientras que su anillo aromático se localiza en el bolsillo hidrofóbico S1 de la enzima, la cadena principal de 77-81 forma los puentes de hidrógeno del flap, y la Ile74 se coloca en el bolsillo de unión al sustrato S1' (de acuerdo a la numeración de fitepsina: Lys11, Tyr13, Ile 7 y Asp36 y 223). La posición de la Tyr80 (numeración de preprosilpepsinas) que se ubica en la región flap y la ubicación de la secuencia GILGL presente en el ψ loop (Figs. 12 y 13) que forma la base del sitio activo es característico de las PAs típicas de plantas (Faro & Gal, 2005). En resumen las posiciones y conformación de los primeros 54 residuos de la molécula tomada como modelo, profitepsina, en relación al resto de la proteína, están ancladas por dos interacciones iónicas, puentes de hidrogeno y contactos hidrofóbicos, incluyendo interacciones en el bolsillo de unión al sustrato S1 y S1'.

Por otro lado, en prosilpepsina 2, como se puede observar en la Fig. 13, entre los 13 residuos del extremo N-terminal de la enzima madura no están presentes ni la Lys78, ni la Tyr80, en su lugar se encuentran Thr78 y Glu80, una sustitución de positivo a neutro polar y de aromático a negativo, respectivamente. Sin embargo, si comparamos dicha secuencia con la de cardosina A (código PDB 1B5F, Frazão et al., 1999) podremos observar que en estas posiciones encontramos Val74, Thr78 y Asp80. Estos residuos de cardosina A, Val y Asp, constituyen sustituciones del mismo tipo que las observadas en prosilpepsina 2 (Fig. 16). De este análisis puede inferirse que silpepsina 2 podría acomodar en el bolsillo S1 sustratos con el mismo tipo de residuos en P1 que los que permite la hendidura del sitio activo de cardosina A.

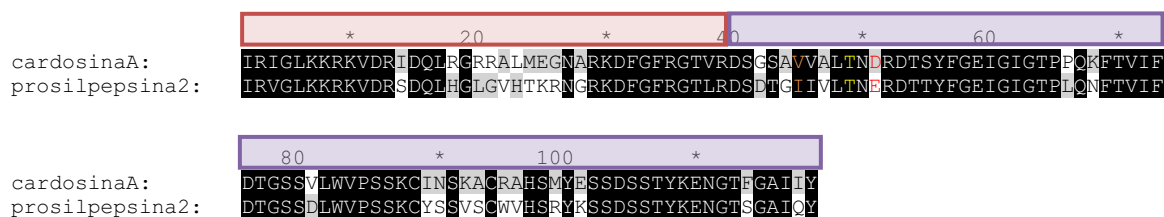


Figura 16. Comparación de la secuencias utilizadas en el modelado prosilpepsina 2 y cardosina A: En naranja se indica la Ile74 y Val 74, en amarillo la Thr78 y en rojo la Glu80 y Asp 80 (numeración de la preprosilpepsinas) que se corresponde con la Lys11, Tyr13 y Ile7 numeración de la profitepsina.

El PSI comprende cinco α -hélices plegadas en un dominio globular que están conectadas entre sí por puentes disulfuro. Este plegamiento es similar a la estructura descrita para NK-lisina y proteínas tipo saposinas (Figs. 14 A y 15 A en color verde, (Egas et al., 2000)). Si se compara la secuencia consenso de 13 PSIs de planta deducidos de las correspondientes secuencias primarias de las PAs con el de las prosilpepsinas, NK-lisina y las saposinas de mamífero se puede observar una notable similitud de estructura (Fig. 17), incluyendo la longitud, la posición de las cinco hélices, los tres puentes disulfuro y el pliegue entre las hélices 3 y 4.

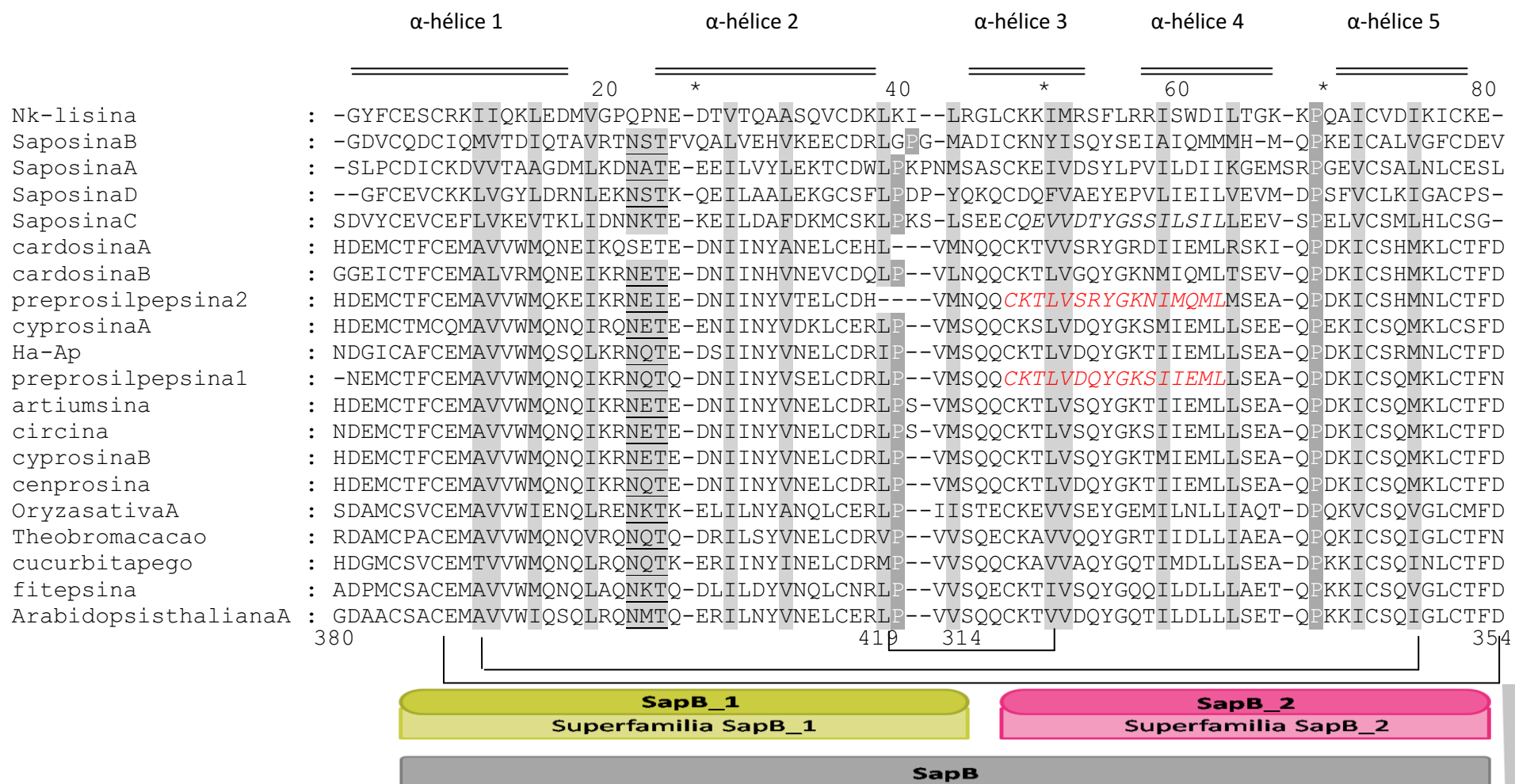


Figura 17. Alineamiento de varias proteínas que contienen el plegamiento típico de saposina. (página anterior) Los sitios de glicosilación conservados están subrayados y marcados en color gris claro. Los aminoácidos hidrófobicos conservados están en recuadros sombreados. Los residuos prolina (P) y aromáticos (Y/F) se muestran en letra blanca y sombreados en gris oscuro, respectivamente. La región de saposina C responsable de la unión a

Entre los aminoácidos 47-61 de saposina C se ha localizado la secuencia responsable de la unión y activación de glucosilceramidasa. La secuencia consenso basada en cinco secuencias de saposina C ha sido definida como CQEVbXTaGXSbbXbhh; donde b= aminoácidos de cadena ramificada; h= aminoácido hidrófobico; a= aminoácido aromático; X=cualquier aminoácido). Brodelius y colaboradores (2005) han sugerido que aunque la secuencia equivalente en el PSI de cipsosina, CKTLVSQYGKTMIEMLL, se corresponde con la secuencia consenso de saposina C sólo en una parte relativamente limitada, para la activación de la enzima glucosilceramidasa la estructura secundaria y/o terciaria de la región del PSI puede ser de mayor importancia que la estructura primaria. La misma secuencia está presente en los PSIs de ambas prosilpepsinas, se muestra en rojo y cursiva en la Fig. 17.

5.3.7. Evaluación del modelo tridimensional obtenido

La calidad y la precisión estereoquímica del modelo se evaluó con PROCHECK por análisis gráfico de Ramachandran (Laskowski et al. 1993) y con el servidor ProSA. Como se muestra en la Fig. 18 en prosilpepsina 1 el 89,9 % de los residuos se encuentran en regiones favorecidas, 8,6 % en regiones permitidas, y un 0,7 % de los residuos se encuentran en regiones generosamente permitidas al igual que en regiones no permitidas. Para prosilpepsina 2 el diagrama de Ramachandran muestra un 85,9 % de los residuos en regiones favorecidas, un 10,9 % en regiones permitidas, un 2,2 % de los residuos se encuentran en regiones generosamente

permitidas, mientras que un 1,0 % se encuentra en regiones no permitidas. Al comparar los diagramas obtenidos con el de profitepsina (Fig. 19), que se utilizó como referencia para el modelado, podremos observar que presenta un pequeño número de aminoácidos que se encuentran en regiones no permitidas como ocurre con los zimógenos de las peptidasas en estudio.

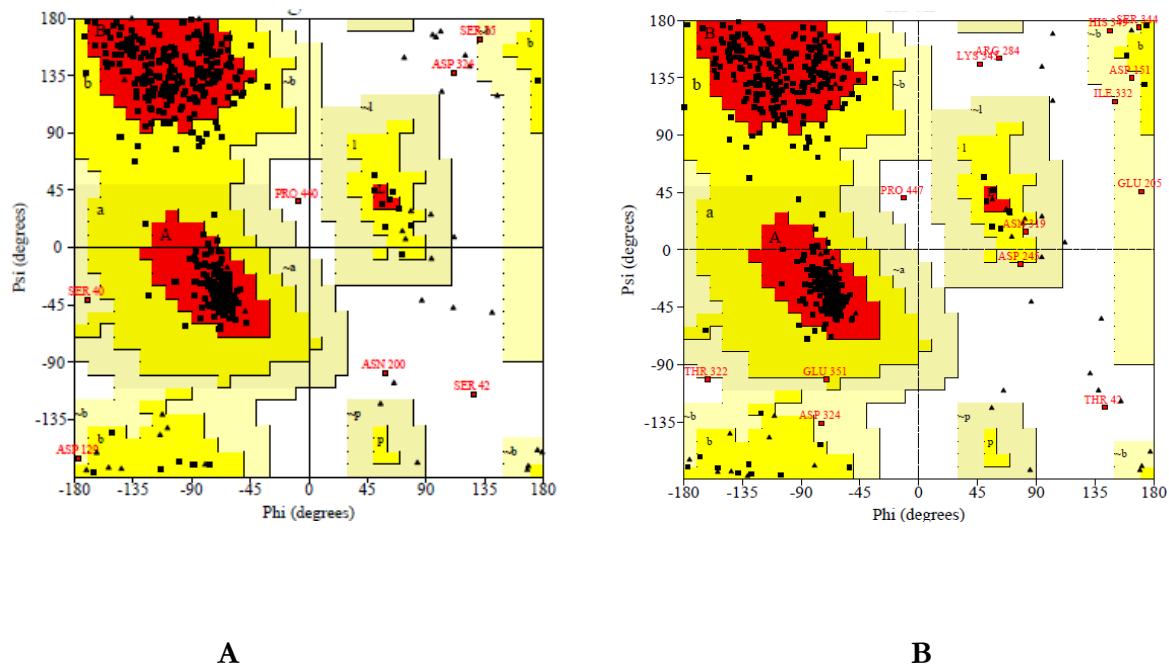


Figura 18. Evaluación estructural del modelo: diagrama de Ramachandran. Diagramas obtenidos a través del programa PROCHEK. **A.** protilpepsina1 **B.** protilpepsina 2. En puntos negros se muestran los residuos marcadores en las regiones favorecidas, mientras que en puntos rojos se muestran los marcadores en las regiones desfavorecidas. Las zonas correspondientes a las regiones más favorecidas se indican en rojo, permitidas en amarillo, generosamente permitidas en beige y no permitidas en color blanco.

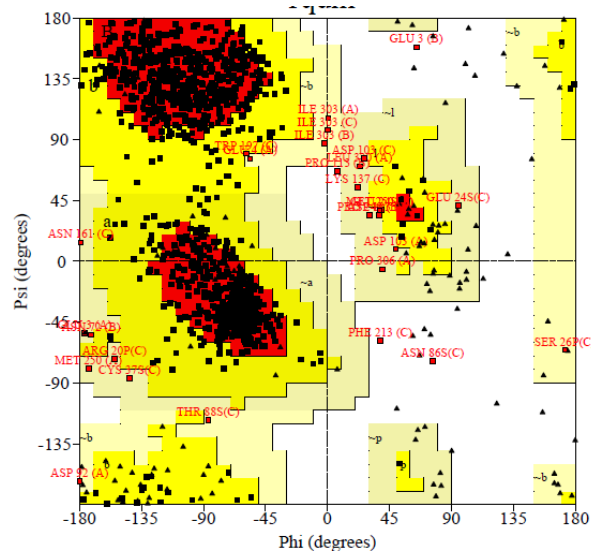
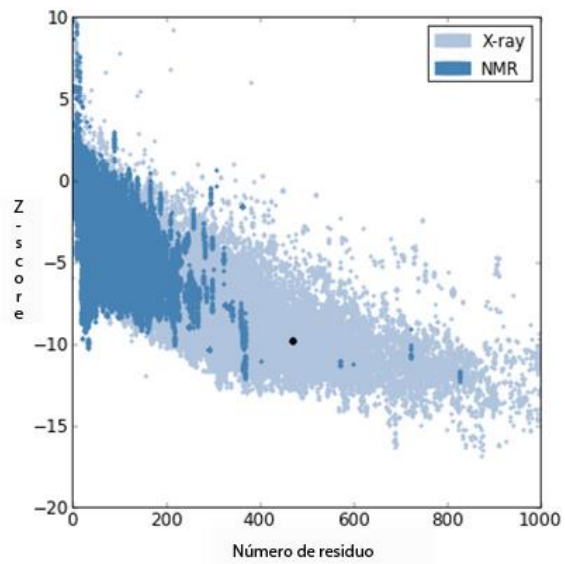


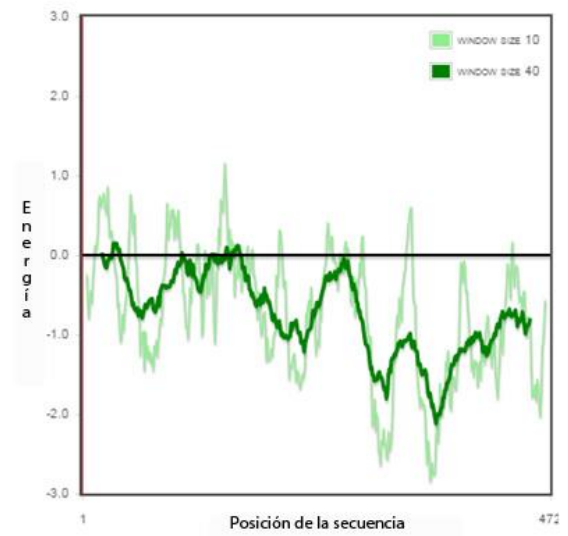
Figura 19. Evaluación estructural del modelo: diagrama de Ramachandran de las cadenas A, B y C de profitepsina. (Página anterior) En puntos negros se muestran los marcadores en regiones favorecidas, mientras que en puntos rojos se muestran los marcadores en regiones desfavorecidas (Kervinen et al.,

Los errores en la predicción de estructura terciaria fueron analizados en el servidor ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>). Se calculó y utilizó el parámetro *Z-score* para el reconocimiento de los errores en la estructura tridimensional que indica la calidad del modelo y el total de desviación de energía de la estructura con respecto a la distribución de la energía derivada de conformaciones aleatorias (Wiederstein & Sippl, 2007). Un *Z-score* positivo muestra que la estructura no es estable, mientras que un valor de cero o negativo indica una estructura ideal. Los gráficos muestran el *Z-score* de todas las cadenas de proteínas cristalizadas y depositadas en la base de datos PDB (*Protein Data Bank*) determinadas por cristalografía de rayos X (azul claro) o espectroscopia RMN (azul oscuro) con respecto a su longitud. El análisis energético global del servidor ProSa-Web dio como resultado un *Z-score* de -9.84 para prosilpepsina 1 y un *Z-score* de -6.96 para prosilpepsina 2, que se destacan como un punto grande en la Fig. 20 A y C,

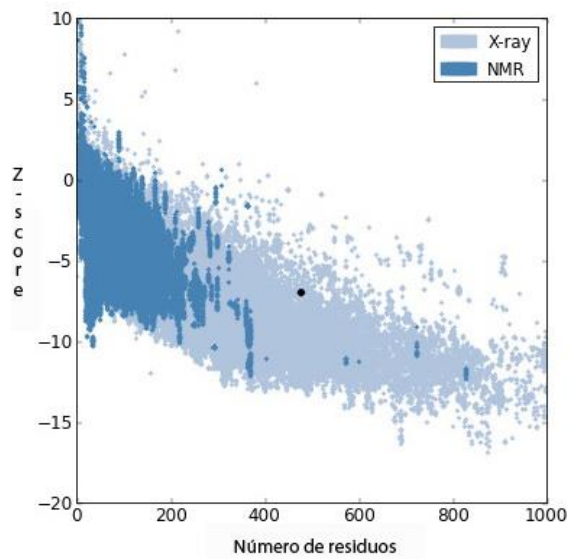
respectivamente, lo que pone en manifiesto la buena calidad del modelo cuando se compara con la estructura cristalina de profitepsina, con *Z-score* de -8.52 (Fig. 20 E). El análisis energético local realizado con el mismo servidor (Fig. 20 B y D) mostró que el modelo no presenta en general partes problemáticas. Este gráfico muestra la calidad del modelo local por el trazado de las energías en función de la posición de la secuencia de aminoácidos. Los valores de energía positivos corresponden a partes erróneas o problemáticas de la estructura de entrada. Un gráfico de energías individuales de residuos por lo general contiene grandes fluctuaciones y tiene un valor limitado para la evaluación del modelo. Por lo tanto el diagrama se afina mediante el cálculo de la energía media de los residuos, tomando una ventana de 40 aminoácidos (curva verde oscura) o de 10 aminoácidos (curva verde claro). Como se observa, prosilpepsina 1 y prosilpepsina 2 en su mayoría poseen valores por debajo de cero en el eje x. Estos resultados sugieren que los modelos obtenidos para prosilpepsina 1 y prosilpepsina 2 son de buena calidad en términos energéticos. El cálculo del potencial DOPE realizado con el programa Modeller, corroboró los resultados obtenidos.

Z-Score: **-9.84**

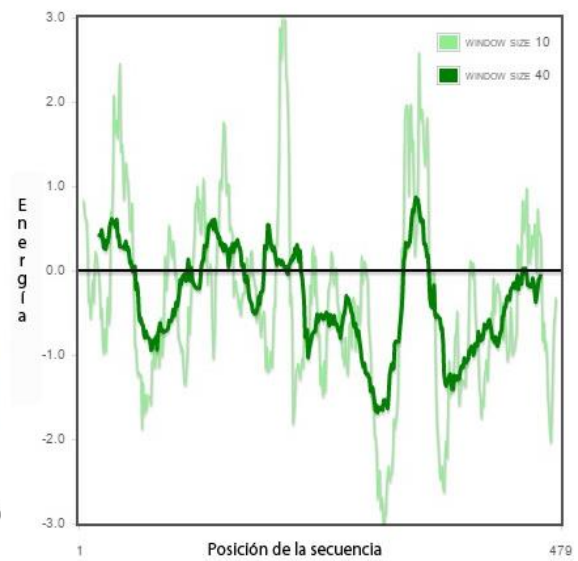
A



B

Z-Score: **-6.96**

C



D

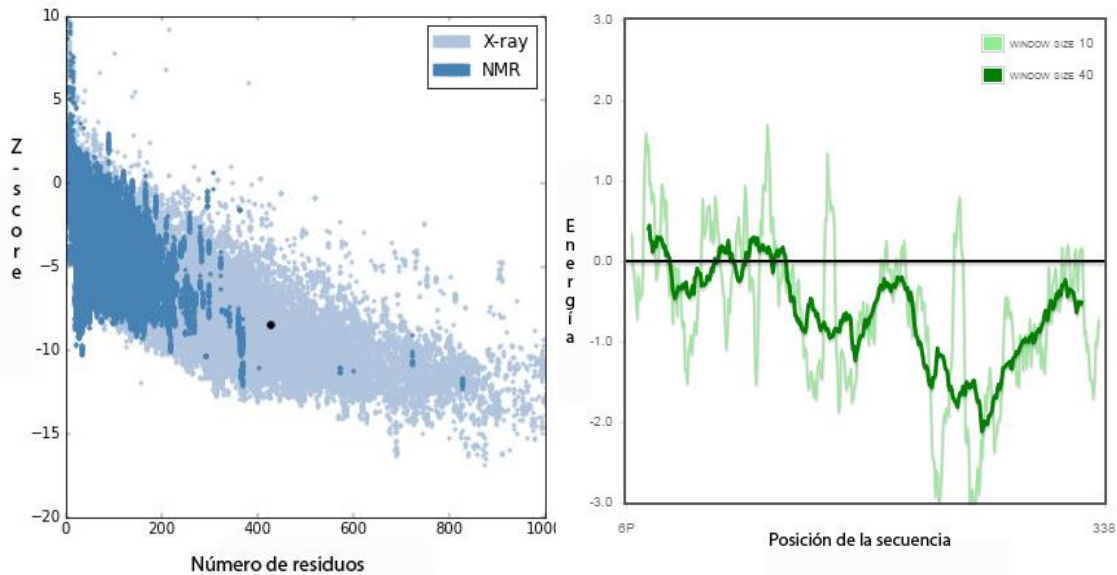
Z-Score: **-8.52****E****F**

Figura 20. Evaluación energética del modelo de prosilpepsina 1, prosilpepsina 2 y profitepsina. Realizada con el servidor ProSA-web. Evaluación global del modelo. A y B en la página anterior. A: *Z-score* -9.84 para prosilpepsina 1. C: *Z-score* -6.96 para prosilpepsina 2. E: *Z-score* -8,52 para profitepsina. **B, D y F:** Evaluación local del modelo. Muestra la energía en función de la posición de cada residuo. El cálculo de la energía se realizó con una ventana de 10 aminoácidos (curva verde claro) y con una ventana de 40 aminoácidos (curva verde oscuro).

5.3.8. Predicción funcional

Con el fin de buscar motivos funcionales en las preprosilpepsinas se analizaron las secuencias de las mismas con el servidor ELM (*The Eukaryotic Linear Motif resource for Functional Sites in Proteins*, Holger et al., 2014). ELM es una base de datos de motivos lineales cortos (SLiMs). Estos motivos son sitios compactos de interacción con otras proteínas compuestos de tramos cortos de aminoácidos adyacentes (Davey et al., 2012, Van Roey et al., 2014). Mediante este análisis se encontró un motivo funcional putativo LIR en preprosilpepsina 1 (accession: ELME000368): ligando de la familia de proteínas Atg8 (LIG_LIR_Gen_1). Para mantener la homeostasis celular, así como en el control de las condiciones de estrés intra y extracelular, la

célula utiliza motivos de interacción de la familia Atg8 (Rogov, Dötsch, Johansen, & Kirkin, 2014). Las proteínas de la familia Atg8 son proteínas relacionadas con la macroautofagia (AuTophaGy, ATG) que se expresan en diversos tejidos (Shpilka, Weidberg, Pietrokovski, & Elazar, 2011). Los motivos LIR se unen a los miembros de esta familia de proteínas para mediar procesos que intervienen en la macroautofagia. Las proteínas que contienen LIR incluyen receptores de carga, miembros del aparato de la autofagia basal, proteínas asociadas a vesículas y del transporte vesicular, proteínas activadoras de GTPasa Rab y proteínas de señalización específicas que se degradan por autofagia selectiva. Este tipo de motivos no han sido previamente reportados en la bibliografía en PAs de plantas (Birgidottir, Lamark, & Johansen, 2013).

```

atg32      : sgswwqaiqp--
CHR        : tprdwiflir--
NBR1       : -sedyiilip--
p62        : -dddwthlss--
CRT        : -eddwdfpp-
Nix        : -nsswveipm--
atg4B      : -tltydtlrf--
atg19      : -altweel----
atg3       : -vgdwedlqg-
preprosilpepsina1: -elvfvggvd

```

Figura 21. Alineamiento de motivos LIR. El alineamiento incluye la secuencia conservada de la región *core* **W/Y---L/I/V** de preprosilpepsina 1. Donde Atg32 es la proteína 32 relacionada con la autofagia de *Saccharomyces cerevisiae* (P40458), CHR es la cadena pesada clatrina de *Homo sapiens* (NP_004850), NBR1 es BRCA1 gen 1 de *Homo sapiens* (Q14596), p62 también conocida como Sequestosome-1 (Q13501), CRT es calreticulina de *Homo sapiens* (P27797), NIX es BCL2/proteína de adenovirus E1B de 19kDa, proteína de interacción 3 de *Homo sapiens*, (EAW63576), Atg4B es la proteína 4 homólogo B relacionada con la autofagia, *Homo sapiens* (QPY4P1), Atg 19 es la proteína 19 relacionada con la autofagia de *Saccharomyces cerevisiae* (P35193), Atg 3 es la proteína 3 relacionada con la autofagia de *Homo sapiens* (Q9NT62). Tomado de Noda et al. 2010.

La Fig. 21 muestra el alineamiento del motivo LIR encontrado en prosilpepsina 1 con otros motivos LIR conocidos (Noda et al., 2010). Los motivos LIRs en la proteína que los contiene están incluidos en una hoja beta. Al producirse la interacción con una hoja beta de la proteína tipo Atg8 ambas hojas beta forman una lámina beta intermolecular paralela (Rogov et al., 2014). En el modelo estructural de prosilpepsina 1 se comprobó que el motivo encontrado está formando una hoja beta como puede observarse en la Fig. 22 (color fucsia).

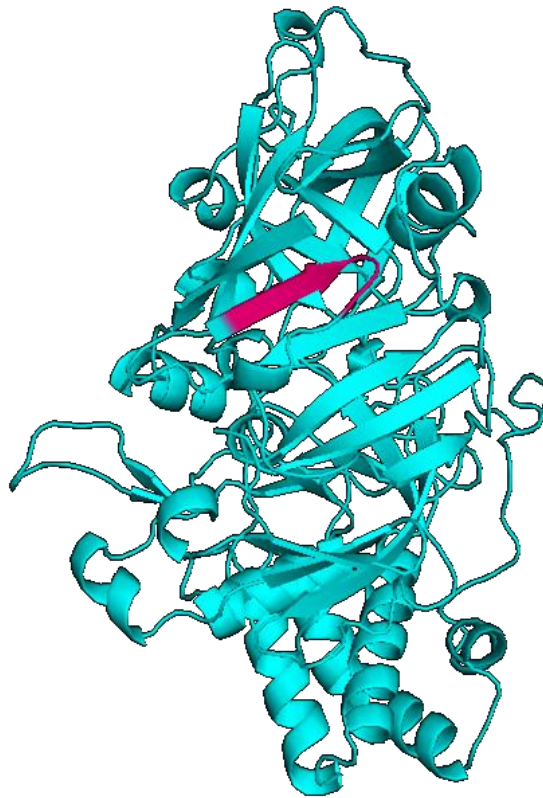


Figura 22. Motivo LIR putativo en prosilpepsina 1. Diagrama de cintas de prosilpepsina 1 indicando en fucsia el motivo LIR putativo: ELVFGGVDP ($X_3 X_2 X_1$ Phe $X_1 X_2$ Val).

La secuencia consenso de los motivos LIR se define como: $X_3-X_2-X_1$ -[Trp/Phe/Tyr]-X1-X2-[Leu/Ile/Val], en el que los aminoácidos alternativos están entre corchetes. El motivo *core* está generalmente precedido por un número variable de residuos ácidos o por residuos serina o treonina que pueden fosforilarse para incorporar una carga negativa. Se ha demostrado que la carga negativa de estos residuos ácidos o fosforilados fortalecen la interacción LIR:Atg8 o proteínas tipo Atg8 (Rogov et al., 2013; Wild, McEwan, & Dikic, 2014).

5.4. ANÁLISIS FILOGENÉTICO

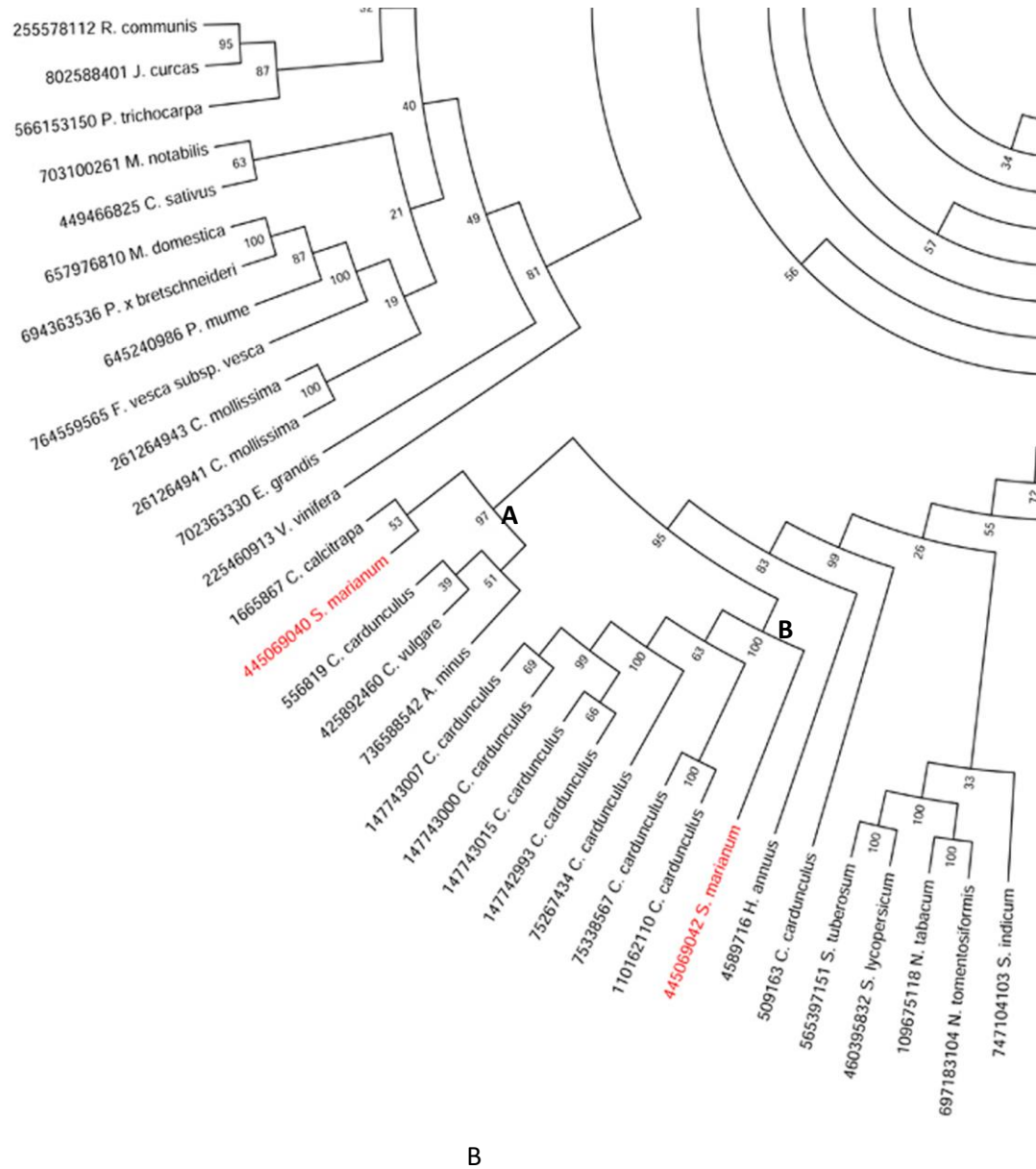
Cuando hablamos de evolución nos referimos al proceso de descendencia con modificación a partir de un ancestro común. El hecho de que los organismos compartan relaciones ancestro-descendiente nos permite utilizar arboles filogenéticos para representar gráficamente sus relaciones de parentesco. Un árbol consta de ramas y nodos que conectan estas ramas entre sí. Las ramas terminales (hojas) corresponden a las secuencias actuales (observadas), mientras que las ramas internas representan sus ancestros hipotéticos. Su longitud refleja la cantidad de cambio acumulado. Los nodos relacionan estas ramas entre sí según su relación ancestro-descendiente. La rama más interna, el ancestro común, representa la raíz del árbol y es esencial para dotar al árbol de un sentido histórico.

En este trabajo de tesis se utilizó el método de máxima verosimilitud para la reconstrucción de la historia evolutiva de las proteasas de distintas familias del reino Plantae. Se realizaron búsquedas de similitud secuencial a través del servidor BLAST–Basic Local Alignment Search Tool– con el objetivo de identificar las proteínas homólogas más cercanas a preprosilpepsina 1 y preprosilpepsina 2. Las secuencias así obtenidas (con un *E-Value* menor a 1×10^{-150}) fueron alineadas a través del programa de alineamientos múltiples ClustalX. Previo a la realización de la

inferencia filogenética se determinó a través del programa HyPhy (Kosakovsky Pond, Frost, & Muse, 2005) el modelo que mejor se ajustaba a los datos. La estimación filogenética se llevó a cabo empleando el programa PhyML (Guindon et al., 2010). Este algoritmo permite estimar grandes filogenias en forma precisa, rápida y sencilla haciendo uso del método de máxima verosimilitud. Dicha estimación se realizó empleando un modelo de evolución de proteínas JTT, con una distribución *gamma* discreta.

La robustez de los resultados fue determinada a partir de un *bootstrap* no paramétrico, que nos informa acerca de la estabilidad de las relaciones filogenética del árbol obtenido. El *bootstrapping* no paramétrico es una técnica estadística de remuestreo con reemplazamiento. En este análisis el soporte de *bootstrap* para cada una de las ramas del árbol fue obtenido usando 500 réplicas, de este modo se supone que una rama es significativa si el valor del *bootstrap* del nodo es mayor al 50% (Fig. 23 A y B, Holmes, 2003).

Las secuencias aminoacídicas de las PAs de *S. marianum* se compararon con 135 secuencias de PAs vegetales (*Viridiplantae*). El análisis filogenético (Fig. 23) mostró que las peptidasas estudiadas se ubican en dos grupos o clúster diferentes. Preprosilpepsina 1 comparte el clúster A con cenprocina (CAA70340) de *C. calcitrapa* su homologo más cercano con un 95% de similitud de secuencia tal como lo indicó el análisis de alineamiento (apartado 5.3.3. de esta sección). En este mismo grupo se encuentran cyprosina B (*C. cardunculus*, CAA57510), cirsina (*C. vulgare*, AFB73927) y artiumsina (*A. minus*, AJA72485). Por su parte preprosilpepsina 2 comparte un clúster más grande (Fig. 23 B) con PAs de especies miembros de la tribu *Cardueae* (cardosinas G, F, E, H, A, B de *C. cardunculus*).



30

Figura 23. Análisis filogenético de proteínas homólogas cercanas a preprosilpepsinas. El árbol filogenético se obtuvo por cálculo de probabilidad máxima usando el programa PhyML tal y como se detalla en *Materiales y Métodos*. Las diferentes proteínas en el árbol están indicadas con el correspondiente número de identificación del NCBI (*National Center for Biotechnology Information*), seguido por la inicial del género y el nombre de la especie. Las preprosilpepsinas se indican en color rojo. El cálculo utilizó 500 réplicas y los valores de *bootstrap* se muestran en cada nodo para valores mayores o iguales al 50% (estadísticamente significativos).

El árbol de la Fig. 23 muestra los valores de *bootstrap* de cada nodo; así para la preprosilpepsina 1, el valor de 97 significa que la especie *C. calcitrapa*, *C. cardunculus*, *C. vulgare* y *A. minus* están emparentados en un 97 %. Por su parte el clúster al que pertenece preprosilpepsina 2 con las especies de la tribu *Cardueae* presenta un *bootstrap* del 100%.

Ambos clústeres, A y B presentan *bootstrap* del 95 %, lo que significa que las secuencias de las especies de la tribu *Carduae*, *C. calcitrapa*, *C. cardunculus*, *C. vulgare*, *A. minus* y *S. marianum* están agrupadas en un grupo monofilético (del griego: de una rama), es decir que todos los organismos incluidos en dicho grupo han evolucionado a partir de una población ancestral común, y todos los descendientes de ese ancestro están incluidos en el grupo. Estos clúster incluyen las PAs de la familia *Asteraceae*, que en su mayoría se han encontrado en flores y por lo tanto se han propuesto como participes de eventos específicos de la flor (Cordeiro et al., 1998; Faro et al., 1999; Vieira et al., 2001). Se cree que las PAs de la familia A1 han evolucionado a través de la duplicación de genes (Revuelta, van Kan, Kay, & ten Have, 2014). La topología sugiere que, en algún momento de la evolución de *C. cardunculus* y *S. marianum*, un gen ancestral de PA se ha duplicado y ha dado lugar a las ramas que comprenden las PAs tipo cyprosinas y las PAs tipo cardosinas. Duplicaciones posteriores dentro del clúster deberían haber ocurrido originando la configuración real del grupo (Pimentel, Van der Straeten, Pires, Faro, & Rodrigues-pousada, 2007; Revuelta et al., 2014).

5.5. LOCALIZACIÓN DE UNA PEPTIDASA DE *S. marianum* EN *Nicotiana benthamiana*

Estudios de localización *in situ* han demostrado que las PAs no están distribuidas al azar dentro de las plantas. Por el contrario, se encuentran en diferentes tejidos específicos, dependiendo de la especie (Simões & Faro, 2004). Por otra parte, en cada parte de la célula, la actividad proteolítica es llevada a cabo por una

peptidasa diferente dirigida específicamente a determinados compartimentos subcelulares (Pesquet, 2012). Dicha especificidad anatómica celular sugiere una enorme diversidad de estas enzimas y explica el hecho de que los genes que codifican a Pas en flores de cardos, cyprosinas y cardosinas, se organizan como una familia multigénica que se expresa durante las primeras etapas en el desarrollo de las flores de *C. cardunculus* (Pimentel, Van Der Straeten, Pires, Faro, & Rodrigues-Pousada, 2007).

Los precursores de silpepsinas como se analizó anteriormente poseen péptidos señal que envía los precursores a la vía secretoria. Al comparar estos precursores con otros precursores de flores de cardo (Fig. 24) se observa que son altamente conservados. Difieren en un único aminoácido con respecto a cardosina B y 5 con respecto a cardosina A, dos de los cuales corresponden a sustituciones conservativas y tres semiconservativas.

| | |
|------------------|--------------------------|
| Silpepsina 1 y 2 | MGNSIKASLLALFLFLLSPTAFS |
| cardosina A | MGTSIKANVLALFLFYLLSPTVFS |
| cardosina B | MGTPIKASLLALFLFLLSPTAFS |

Figura 24. Comparación de péptidos señal de los precursores de silpepsinas. Se compararon los péptidos señal de los precursores de silpepsina 1 y 2 con los de cardosina A y B, dos proteínas que siguen la vía secretoria. Los residuos azules representan sustituciones conservativas entre ambas secuencias y los rojos, aquellas sustituciones semi-conservativas.

Para determinar la localización subcelular de una PA de *S. marianum* se utilizó el programa Plant-mPLOC (Predicting subcellular localization of plant proteins including those with multiple sites, <http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>) que predice la localización subcelular en células de planta. Este programa incluye 1.055 secuencias de proteínas vegetales (978 proteínas diferentes), clasificados en 12 localizaciones subcelulares de planta. Ninguna de las proteínas tiene más de 25% de identidad de secuencia con cualquier otra proteína en el

mismo subconjunto (localización subcelular). Tras analizar la secuencia de prosilpepsina 2 se predice una localización vacuolar.

Está demostrado para cardosina A que presenta dos VSDs, el CtVSD y su PSI que también actúa como VSD que podrían especificar dos vías de tráfico diferentes (Pereira, Pereira, Satiat-Jeunemaitre, & Pissarra, 2013). Se propone que prosilpepsina 2, al igual que cardosina A, posee dos VSDs. Cuando se comparan la secuencia C-terminal de las silpepsinas (VGFAEAA) con la misma secuencia de otras PAs y con lectina de cebada, cuatro aminoácidos se conservaban entre ellos, FAEA (Fig. 25). Si bien cardosina A y cardosina B difieren en sólo un aminoácido en el péptido C-terminal, se ha demostrado que cardosina A se acumula en las vacuolas de las papilas estigmáticas (Ramalho-Santos et al., 1997) mientras que cardosina B se acumula extracelularmente en el tejido de transmisión de *C. cardunculus* (Vieira et al., 2001).

Pereira y colaboradores demostraron que este péptido era suficiente para dirigir a cardosina A a la vacuola, sin embargo cuando se colocaba el péptido C-terminal (GFAEAA) o el péptido FAEA entre el SP y mCherry (SP-Cter-mCherry y SP-FAEA-mCherry) no era capaz de direccionar el reportero a la vacuola, lo que confirmó que el péptido C-terminal o su versión más corta actúa como verdadera ctVSD y pertenecen al grupo de ctVSDs descrito para varias proteínas vacuolares. Además la eliminación de una de las dos señales de direccionamiento no deteriora la localización vacuolar de cardosina A (Pereira et al., 2013).

Se estudió la hipótesis de que una única señal de direccionamiento, presente en el precursor del zimógeno, puede ser suficiente para dirigir la proteína en estudio a su destino final en tejidos transgénicos.

| | |
|-------------|---|
| silpepsina1 | KLRVGF AEAA |
| silpepsina2 | -LRVGF AEAA |
| PA Vigna | --VGF AEAA - |
| catepsina D | --VGF AEAA RL |
| AtAP1 | -QVGF AEAA |
| PA Tomate | --QVGF AEAA |
| cardosinaA | --LVGF AEAA |
| cardosinaB | --RVGF AEAV |
| lectina | -DGV-F AE IA |
| Ha-AP | --RVGF AEAA |
| | * * * * * |

Figura 25. Alineamiento de las secuencias de aminoácidos de los péptidos C-tVSD de varias PAs vegetales, de una PA de mamífero (catepsina D) y de la lectina *Hordeum vulgare* (cebada). *Vigna unguiculata* (AAQ14346.1); catepsina D humana (AAA51922.1); *Arabidopsis thaliana* (AAC49730.1); *Solanum lycopersicum* (tomate) (XM_004243437.1); cardosina A (AJ132884.1); cardosina B (AJ237674.1.); lectina (BAJ87539.1). Los aminoácidos conservados en el extremo C-terminal se representan en color gris claro. Adaptado de Pereira et al., 2013.

5.5.1. Obtención de las construcciones para la expresión de preprosilpepsina 2

Para determinar la localización subcelular de una PA de *S. marianum* se empleó la secuencia completa de la proteína. Se realizaron dos construcciones para sobreexpresar preprosilpepsina 2 bajo el control del promotor 35S virus mosaico de la coliflor (CaMV) fusionando el extremo C-terminal a las proteínas eGFP o mRFP.

A partir de la secuencia de pGEM-pPreproSm2-wt se realizó un subclonado utilizando los cebadores F4R y F4R, los cuales contienen los sitios “attB”. Los oligonucleótidos utilizados fueron diseñados a partir del extremo 5’ y 3’ del marco de lectura abierto de la PA en estudio, incluyendo la secuencia GGGG-attB1, Shine Dalgarno y Kozac, para el cebador *forward* y la secuencia GGGG-attB1 para el cebador *reverse*.

La amplificación se logró con 20 ciclos de PCR ($T_a=52\text{ °C } \Delta 30^\circ$). El producto de amplificación de 1,5 kb (Fig. 26), tamaño esperado, se utilizó para recombinar con el vector pDONR222 que posee los sitios “attP” mediante la

reacción con la mezcla enzimática BP *clonase*. Se siguió trabajando con la amplificación correspondiente a la temperatura 48,3 °C.

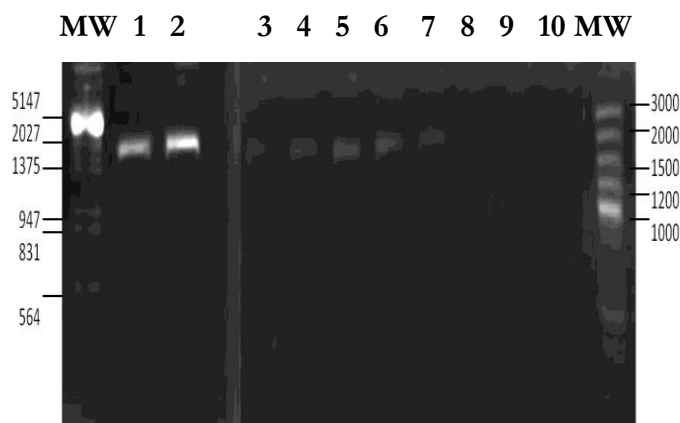


Figura 26 Amplificación y clonado de la secuencia correspondiente al marco de lectura abierto de preprosilpepsina 2. Análisis electroforético en gel de agarosa (1%) del producto amplificado mediante PCR. *Calle 1–9:* PCR preprosilpepsina 2 con cebadores de clonado para pDONR222 en gradiente de temperatura: 46,3; 48,3; 50,8; 52,7; 56,7; 58,7; 61,8; 62,7, 64,5. *Calle 10:* control negativo. *MW:* marcador de masas moleculares de GeneRuler ADN de λ DNA EcoRI+HindIII100pb plus (*ThermoScientific*).

El producto de la recombinación se utilizó para transformar la cepa *E coli* GC10. A partir de las colonias positivas obtenidas se realizó el aislamiento del ADN plasmídico utilizando el método de lisis alcalina (*Materiales y Métodos 3.10.15.1*). Las construcciones fueron analizadas mediante PCR utilizando los cebadores 523 y 524 diseñados por fuera de los sitios att (Fig. 27).

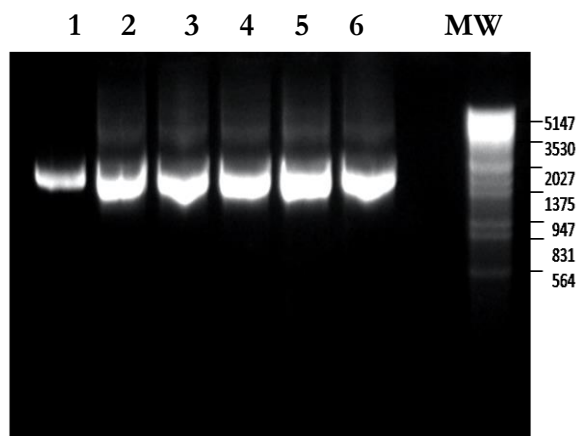


Figura 27. Chequeo de la recombinación para la obtención de las construcciones pDONR222-pPreproSm2-wt. Análisis electroforético en gel de agarosa (1%) del producto amplificado mediante PCR de distintos clones. *Calle 1–6:* pDONR222-pPreproSm2-wt. *MW:* marcador de masas moleculares de ADN: λ DNA EcoRI + HindIII 100 pb plus (*ThermoScientific*).

Los productos de amplificación del tamaño esperado, fueron secuenciados y de este modo se confirmó que las secuencias no presentaban mutaciones. Cuatro muestras correspondientes a la construcción pDONR222-pPreproSm2-wt (calles 2, 3, 4 y 5 de Fig.27) fueron analizadas. Confirmadas las secuencias, se continuó trabajando con único clon. Se realizó el subclonado en los vectores binarios pB7FWG2 y pB7RWG2 mediante la reacción de recombinación LR (*Materiales y Métodos 3.11.4*). La secuencia codificante fue clonada en marco con la región de codificación para la fusión de eGFP y mRFP bajo control del promotor CaMV 35S de expresión ubicua. Los clones recombinantes fueron seleccionados con el antibiótico Sp (resistencia del plásmido) y con la correspondiente resistencia de la célula. Se obtuvieron dos construcciones denominadas: pB7FWG2-pPreproSm2-wt-eGFP, pB7RWG2-pPreproSm2-wt-mRFP. Se realizó el aislamiento del ADN plasmídico utilizando el método de lisis alcalina y/o utilizando un Kit comercial. Los clones fueron analizados mediante PCR para las construcciones con eGFP y mediante digestión con la enzima BsrGI y SalI para las construcciones con mRFP. Para la reacción de PCR se emplearon los cebadores que se utilizaron para amplificar la preprosilpepsina (F4F y F4R) con 20 ciclos de PCR. La amplificación se logró a una temperatura de hibridación de 48,3 °C. La reacción de amplificación dio como resultado una banda intensa entre 1375 y 2027 pb (Fig. 28), que se corresponde con el tamaño esperado de 1594 pb para la preprosilpepsina 2 con los sitios attB.

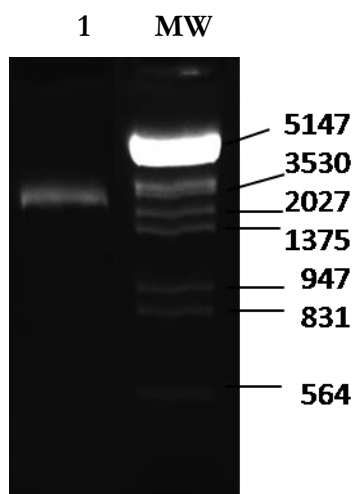


Figura 28. Chequeo de la construcción pB7FWG2-pPreproSm2-wt-eGFP. Análisis electroforético en gel de agarosa (1%) del producto amplificado mediante PCR. *Calle1*: pB7FWG2-pPreproSm2-wt-eGFP. *MW*: marcador de masas moleculares de ADN λ DNA EcoRI + HindIII 100 pb plus (*Thermo Scientific*).

Para corroborar la presencia del inserto en las construcciones con mRFP se realizaron dos análisis de restricción con las enzimas *Bsr*GI y *Sa*I, respectivamente. Los mapas de restricción del vector pB7RWG2 y de la construcción pB7RWG2-pPreproSm2-wt-mRFP se muestran en el Anexo II. De acuerdo a la digestión teórica realizada con el servidor RestrictionMapper, el vector pB7RWG2 tendría un perfil de digestión con *Bsr*GI de tres fragmentos de 9929, 1283 y 402 pb, mientras que la construcción pB7RWG2-pPreproSm2-wt-mRFP dos fragmentos de 9929 y 1575 pb. La construcción posee un fragmento menos que el vector vacío ya que al producirse la recombinación, se pierde el *cassette* que contiene al gen *ccdB* que incluye uno de los sitios de restricción de *Bsr*GI. Al realizar la digestión teórica del vector vacío y de la construcción con la enzima *Sa*I se obtienen diferentes perfiles. Para el vector vacío la digestión tendría cuatro fragmentos de 7293, 1689, 1606 y 1026 pb, mientras que la construcción pB7RWG2-pPreproSm2-wt-mRFP tendría un perfil de restricción de tres fragmentos de 7293, 3185 y 1026 pb.

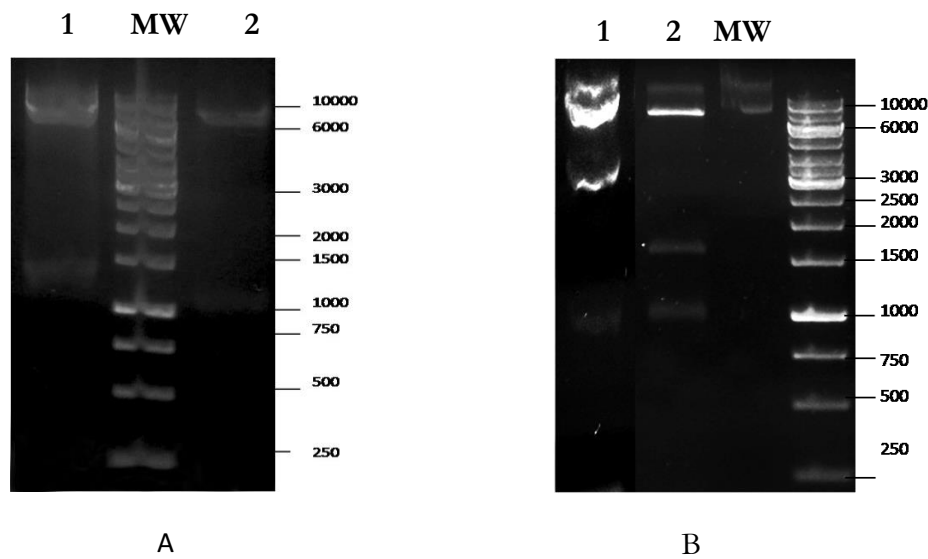


Figura 29. Chequeo de la construcción pB7RWG2-pPreproSm2-mRFP. Análisis de restricción con la enzima *Sall* y *BsrGI* en gel de agarosa (1%) de las minipreparaciones. **A:** *Calle1:* pB7RWG2-pPreproSm2-Wt-mRFP digerido con *BsrGI*, *Calle2:* vector pB7RWG2 vacío digerido con *BsrGI*, *MW:* marcador de masas moleculares de ADN Gene Ruler 1 kb (*Thermo Scientific*). **B:** *Calle 1:* pB7RWG2-pPreproSm2-Wt-mRFP digerido con *Sall*, *Calle2:* vector pB7RWG2 vacío digerido con *Sall*, *Calle3:* vector pB7RWG2 vacío sin digerir, *MW:* marcador de masas moleculares de ADN Gene Ruler 1 kb (*Thermo Scientific*).

Luego de la digestión con la enzima de restricción *BsrGI* durante 3 hs a 37°C se pudieron observar las bandas esperadas para la construcción pB7RWG2-pPreproSm2-mRFP (Fig. 29 A), sin embargo para el vector vacío por el bajo contraste no se observa la banda de 402 pb. En el caso de la digestión con *Sall* se observan tres bandas (Fig. 29B) para el vector vacío en lugar de cuatro dado que son 83 pares de bases de diferencia entre dos de los productos de digestión y no se logra resolver en el sistema utilizado. En el caso de la construcción pB7RWG2-pPreproSm2-mRFP se observaron las tres bandas esperadas, dos de las cuales coinciden en peso con las esperadas para el vector vacío (7293pb y 1026 pb).

5.4.1.2. Mutantes de sitio activo

Como ya se determinó silpepsina 2 es PAs típica de planta y como tal se presenta dos residuos de ácido aspártico organizados en dos motivos conservados, DTG/SG, responsables de la actividad catalítica. A fin de evaluar si el sitio activo tiene alguna influencia en el destino final de las peptidasas, se generaron mutantes del sitio activo en el residuo aspártico D103. El residuo catalítico fue sustituido mediante una mutación puntual por un residuo de alanina. La estrategia seguida se describe en la Fig. 6 del apartado de *Métodos*.

La construcción conteniendo la secuencia mutante de sitio activo se obtuvo utilizando como molde las construcciones pGEM-pPreproSm2-wt. Se diseñaron dos nuevos oligonucleótidos: 618 y 619. La amplificación se logró mediante dos fases, la primera se llevó a cabo con 20 ciclos de PCR y en gradiente de temperatura ($57\text{ }^{\circ}\text{C } \Delta 26^{\circ}$). Los productos obtenidos en las etapas de la mutagénesis para la construcción de pB7FWG2-pPreproSm2-D103A-eGFP se muestran en las Fig. 30 y 31.

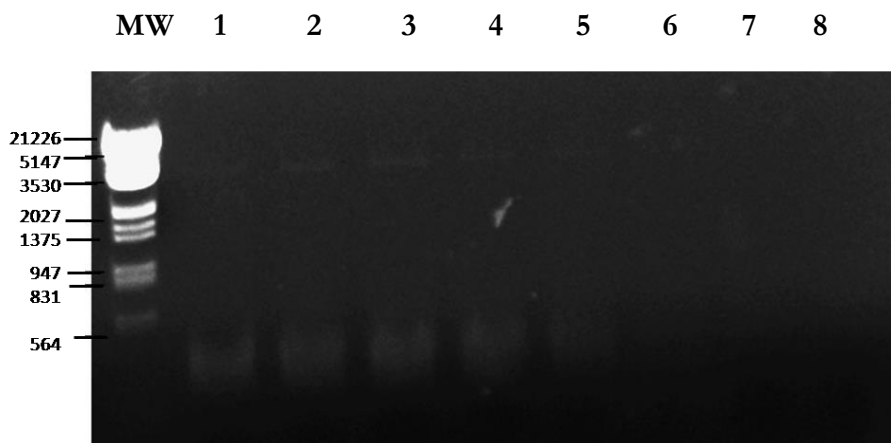


Figura 30. Generación de los casetes mutantes para pDONR222-pPreproSm2-D103A. Calle 1–8: PCR, molde preprosilpepsina 2 con cebadores F4F-618 en gradiente de temperatura: 53,8; 55,8; 55,9; 57,6; 61,6; 62,6; 65,5; 67,8. MW: marcador de masas moleculares de ADN λ DNA EcoRI + HindIII 100 pb plus (*Thermo Scientific*)

El resultado de la reacción de PCR muestra una banda tenue de un peso menor a 564 pb utilizando el juego de cebadores F4F-616, mientras que con el juego de cebadores 618-F4R se obtuvo una banda intensa entre 947 pb y 1375 pb. Estas bandas son del tamaño esperado para el mutante de preprosilpepsina 2 con los sitios attB siendo de 375 pb para el juego F4F-616 y de 1267pb para el juego 617-F4R.

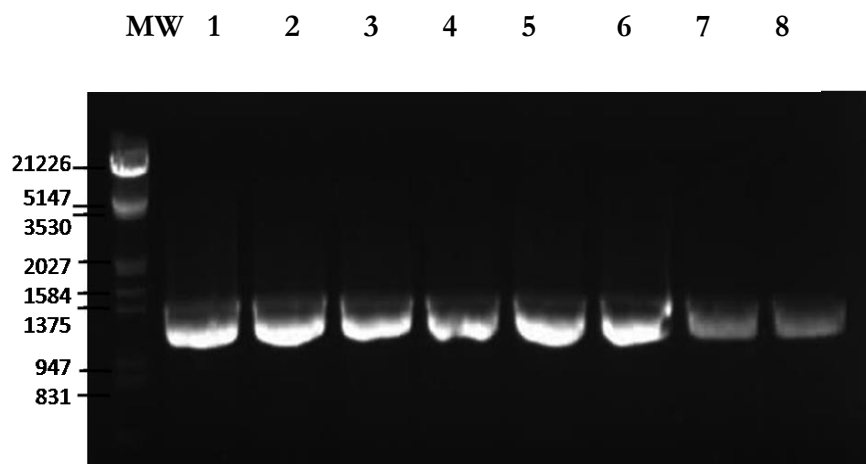


Figura 31. Generación de los casetes mutantes para pDONR222-pPreproSm2-D103A: Calles 1–8: PCR, molde preprosilpepsina 2 con cebadores 619-F4R en gradiente de temperatura: 53,8; 55,8; 55,9; 57,6; 61,6; 62,6; 65,5; 67,8. MW: marcador de masas moleculares de ADN λ DNA EcoRI + HindIII 100 pb plus (*Thermo Scientific*)

Los productos correspondientes a la temperatura de 55,9°C (*Calle3*, Fig.30 y 31) se utilizaron posteriormente como moldes para realizar el segundo ciclo a la misma temperatura. La Fig.32 muestra el resultado obtenidos para la PCR realizada. Se observó el resultado esperado, una banda de entre 1375 y 1584 pb.

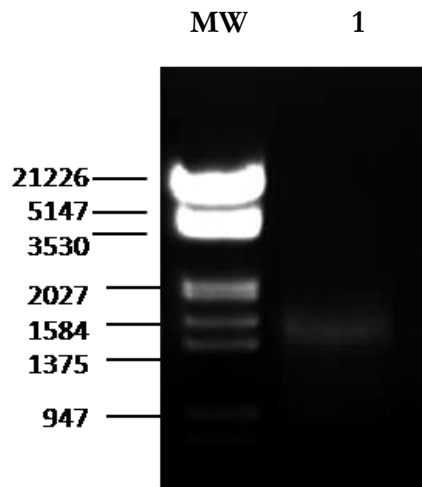


Figura 32: Generación de mutante de pDONR222-pPreproSm2-D103A: *Calle 1* amplificación correspondiente al mutante de preprosilpepsina 2. *MW*: marcador de masas moleculares de ADN λ DNA EcoRI + HindIII 100 pb plus (*Thermos Cientific*)

El producto de amplificación se utilizó posteriormente para llevar a cabo la reacción de recombinación *BP clonase*, para obtener así la construcción: pDONR222-pPreproSm2-D103A. El recombinante se utilizó para transformar *E coli* GC10. A partir de las colonias positivas obtenidas se realizó el aislamiento del ADN plasmídico por el método de lisis alcalina. Luego para corroborar la presencia del inserto las colonias fueron analizadas mediante PCR utilizando los cebadores específicos: 523 y 524. (Fig. 33)

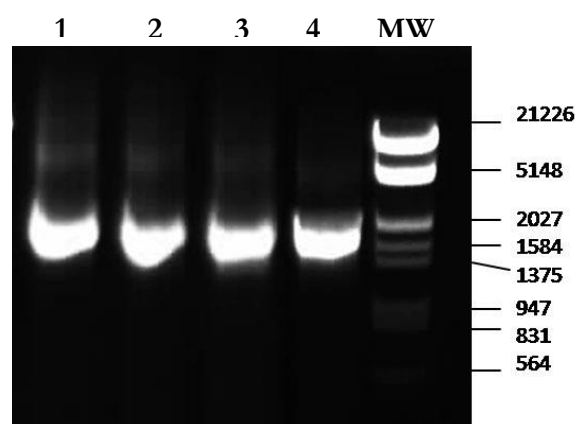


Figura 33. Chequeo de la recombinación del producto de PCR de las proteína mutante. Análisis electroforético en gel de agarosa (1%) del producto amplificado mediante PCR. *Calle 1–4*: pDONR222-pPreproSm2-D103A *MW*: marcador de masas moleculares de ADN λ DNA EcoRI + HindIII 100pb plus (*Thermo Scientific*).

A partir de los resultados obtenidos se mandaron a secuenciar los cuatro clones de pDONR222-pPreproSm2-D103A. El posterior análisis de la secuencia confirmó la presencia del inserto con la mutación en el sitio catalítico, en marco de lectura y sin mutaciones. A continuación, la construcción obtenida correspondiente al clon de la *calle 4* de la Fig. 33 se empleó para realizar la reacción de recombinación LR con el vector binario pB7FWG2. Tal como se realizó con las construcción *wt*, para la fusión la secuencia de la proteína mutante fue clonada en marco con la región de codificante de eGFP bajo el control del promotor CaMV 35S de expresión ubicua. El producto de la recombinación, fue utilizado para transformar *E. coli* GC10. Los clones recombinantes denominados pB7FWG2-pPreproSm2-D103A-eGFP fueron seleccionados con Sp. Se realizó el aislamiento del ADN plasmídico y los clones fueron analizados mediante PCR utilizando los cebadores F4F y F4R y como temperatura de hibridación 48,3°C (Fig. 34).

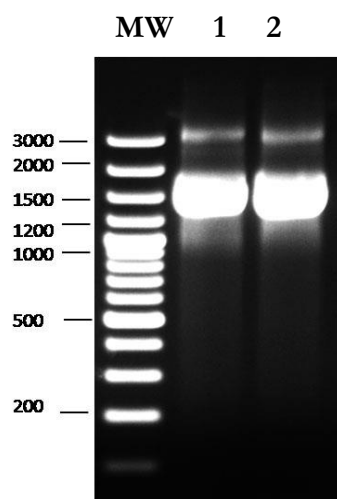


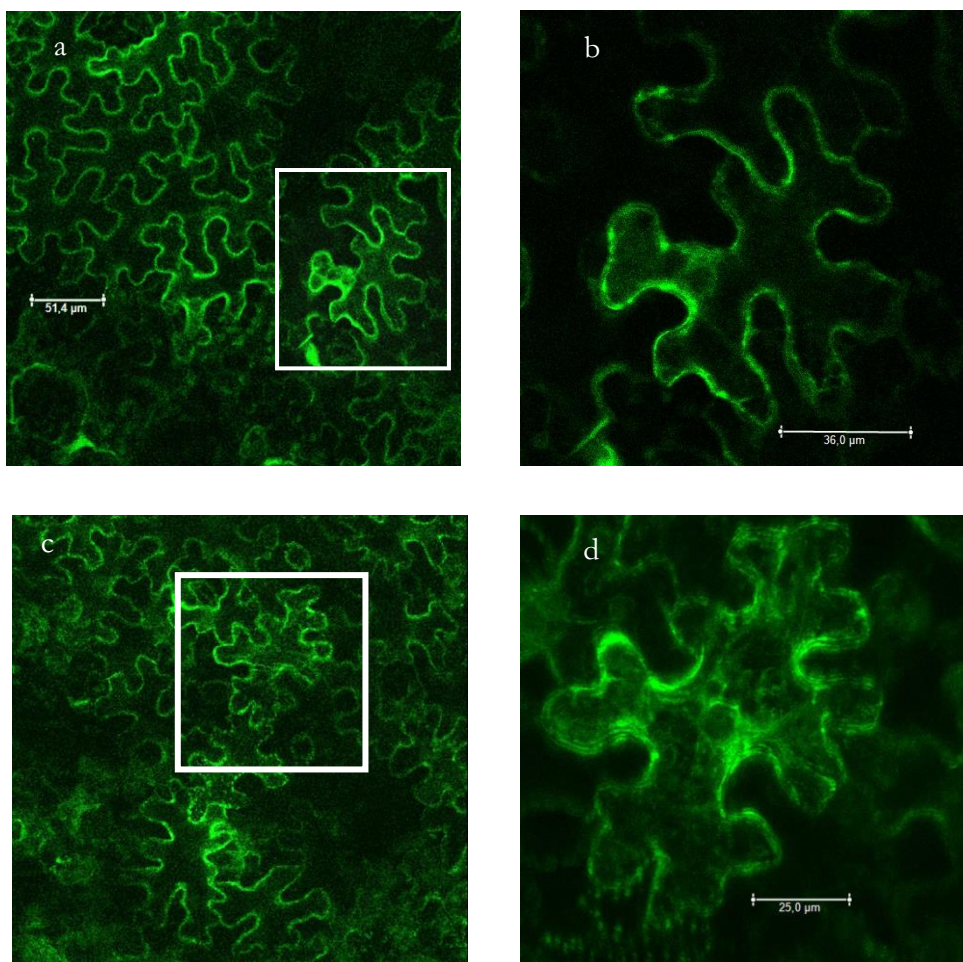
Figura 34. Chequeo de la construcción pB7FWG2-pPreproSm2-D103A-eGFP. Análisis electroforético en gel de agarosa (1%) del producto amplificado mediante PCR. *Calles 1 y 2:* pB7FWG2-pPreproSm2-Wt-D103A. *MW:* marcador de masas moleculares de GeneRuler 100 pb plus (*Thermo Scientific*).

La reacción de amplificación dio como resultado una banda intensa del tamaño esperado de 1594 pb para preprosilpepsina 2 con los sitios attB. Se continuó

con el clon correspondiente a la calle 2 del gel de la Fig. 34 ya que no se observaron diferencias entre ambos clones.

5.5.2. Agroinfiltración de hojas de *N. benthamiana*, expresión transitoria y microscopía confocal

Se realizaron tres construcciones, dos fusionadas a la proteína reportera eGFP (proteína fluorescente verde mejorada): pB7FWG2-pPreproSm2-wt-eGFP, pB7FWG2-pPreproSm2-D103A-eGFP y otra al gen reportero mRFP (proteína fluorescente roja monomérica): pB7RWG2-pPreproSm2-wt-mRFP a fin de determinar la localización celular de la peptidasa y su mutante.



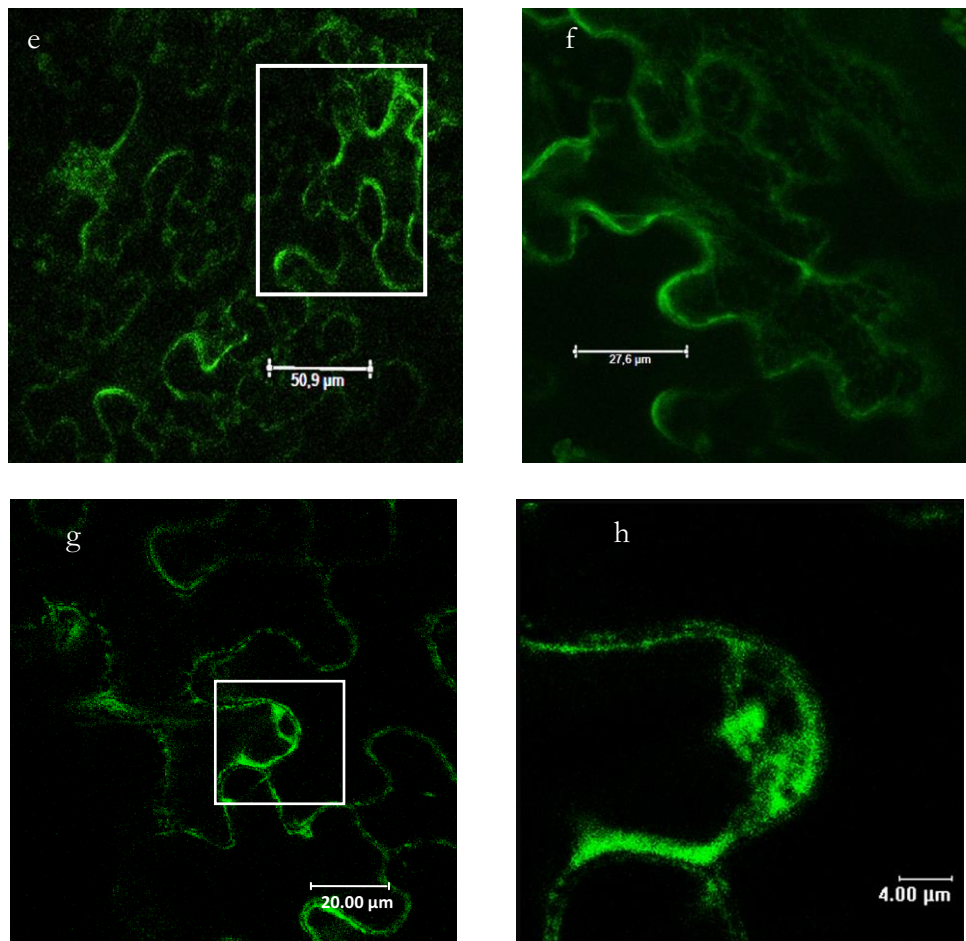


Figura 35. Localización subcelular de la peptidasa *wild-type* o mutante fusionada a eGFP. Expresión de la peptidasa fusionada eGFP luego de 48 hs (a-b) y 72 hs (c-d) post infiltración (página anterior). Expresión de la peptidasa mutante fusionada a eGFP luego de 48 hs (e-f) y 72 hs (g-h). En cada imagen la barra indica la referencia de tamaño en μm . Los cuadros blancos en las figuras de la izquierda muestran la ampliación de la figura de la derecha.

La cara abaxial de células epidérmicas de hojas de *N. benthamiana* se transformaron utilizando la técnica de infiltración con agrobacterium descrita en *Métodos* 3.7.7. Las plantas agroinfiltradas se mantuvieron en las mismas condiciones previas, cultivándose inicialmente en macetas con fotoperíodo normal de 16/8hs luz. Para la optimización de la expresión el nivel final de inoculo se ajustó a DO_{600} entre 0,1 y 1,0. Mediante microscopía confocal se realizaron observaciones a diferentes tiempos después de la agroinfiltración y se evaluó la expresión de cada proteína de fusión por la emisión de la fluorescencia

detectada. Se realizaron tres experimentos independientes para cada condición ensayada.

Las células transfectadas ($DO_{600} = 1$) con las construcciones reporteras eGFP fusionadas a preprosilpepsina 2, fueron detectadas en RE/AG/ compartimiento prevacuolar tras 48 y 72 hs post-agroinfiltración. Las células epidérmicas mostraron una distribución perinuclear, formando una característica red reticular, que es típico del patrón de RE (Fig. 35a-d). Por otra parte, no se pudieron evidenciar diferencias apreciables en el perfil de localización respecto a las construcciones *wild-type* o mutante (Fig. 35 e-h). Estos resultados indicarían que peptidasa *wild-type* o mutante no llegaría a su destino final en este sistema. No pudo evidenciarse ninguna señal de fluorescencia cuando la DO_{600} del inóculo fue de 0,1.

La proteína verde fluorescente (eGFP) permite la visualización directa de la localización subcelular de proteínas de fusión en células vivas y en tiempo real. Así, la localización de distintas proteínas de fusión a eGFP en la vía secretoria ha sido estudiada en diferentes órganos de plantas superiores. Sin embargo, la fluorescencia del reportero no ha podido ser observada en el lumen de vacuolas, especialmente ácidas, de algunos órganos de plantas (Flückiger et al., 2003; Samalova, Fricker, & Moore, 2006; Tamura et al., 2003). La vacuolas mantienen su condición ácida que es generada por la acción de H^+ -ATPasas vacuolares y H^+ -fosfatasa usando la energía derivada de la hidrólisis de sustratos- ATP y pirofosfato, respectivamente- (Heven, Xuhang, & Palmgrenb, 1999; Maeshima, 2000). La inestabilidad de la proteína reportera se atribuye principalmente en un entorno ácido que podría alterar el equilibrio de protonación de la proteína GFP (Tamura et al., 2003) o, al aumento de la actividad de ciertas peptidasas pH-dependientes (equivalentemente a las peptidasas fotodependientes), concordando con lo publicado por Flückiger (2003) donde solo observa fluorescencia verde al utilizar inhibidores de peptidasa. No obstante, al ser

incubadas bajo condiciones de oscuridad en plantas transgénicas de *N. benthamiana* y *A. thaliana*, la fluorescencia de GFP dirigida a vacuolas, ha sido observada de manera estable en vacuolas, siendo esta degradada con menos de 1 hora de iluminación, especialmente si la misma era azul (Tamura et al., 2003).

Teniendo en cuenta lo antedicho, para la observación de fluorescencia debida a GFP las condiciones iniciales de cultivo con fotoperíodo de 16/8hs luz se cambiaron luego de la agroinfiltración, incubandose las plantas en cámara oscura al menos 48hs previo a la observación microscópica.

Como puede observarse en la Fig. 36, las células de las hojas que expresan la peptidasa presentaron una población de vesículas móviles altamente fluorescentes situadas en los filamentos citoplasmáticos, también llamado hebras transvacuolares de las células epidérmicas. Estos filamentos citoplasmáticos son rutas esenciales de transporte para la distribución de los orgánulos y metabolitos. El mecanismo de este transporte genera el movimiento de las proteínas. Los mismos resultados se obtuvieron para la proteína mutante (datos no mostrados). La condición de oscuridad no fue suficiente para que las peptidasas alcanzaran su destino final. Sin embargo, al comprar estos resultados con los obtenidos en condiciones de luz, puede notarse que la intensidad de la fluorescencia es mayor en condiciones de oscuridad, pudiendo observarse además el movimiento de las vesículas en las hebras transvacuolares (Brandizzi, Irons, Johansen, Kotzer, & Neumann, 2004; Esseling-Ozdoba, Houtman, Van Lammeren, Eiser, & Emons, 2008)

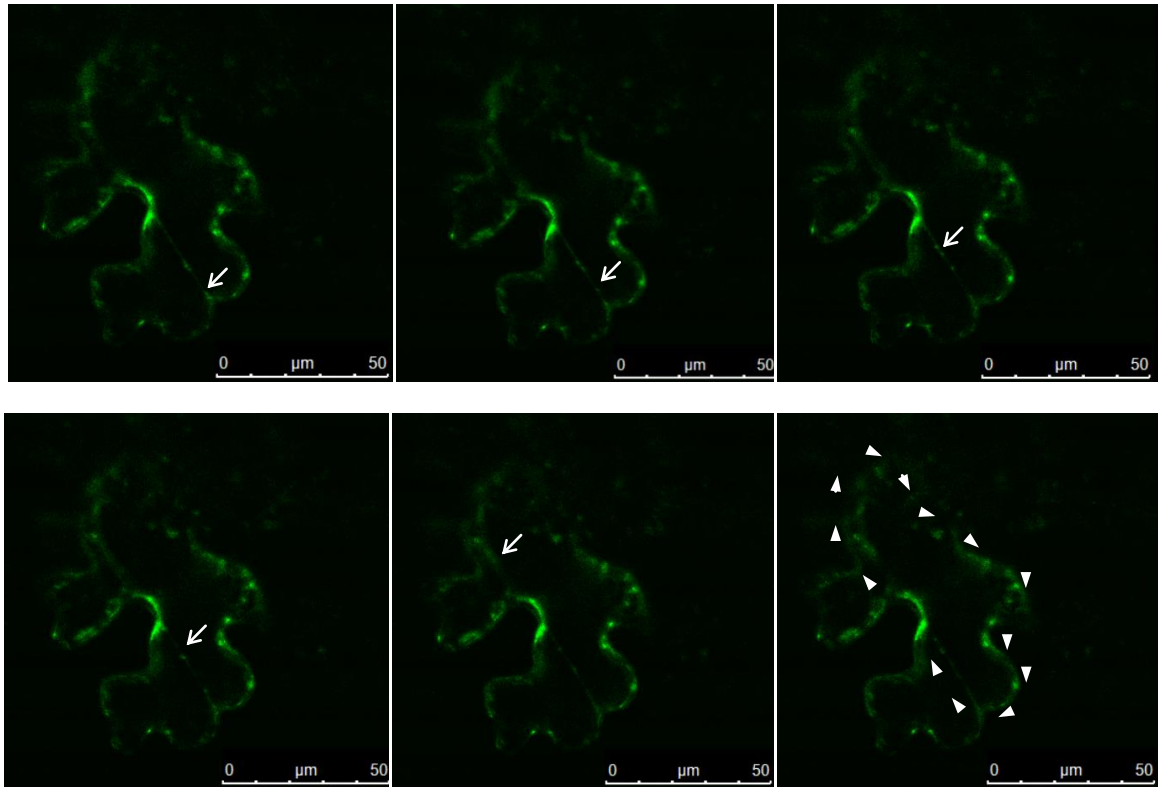


Figura 36. Localización subcelular de la peptidasa *wild-type* fusionada a eGFP en condiciones de oscuridad. Expresión de la peptidasa fusionada a eGFP luego de 72 hs post infiltración en oscuridad. Imagen obtenida por capturas de trama secuenciales que muestra el movimiento de las vesículas en las hebras transvacuolares conteniendo la proteína fluorescente (como múltiples imágenes) a lo largo de una sola hebra del RE. La dirección de movimiento se indica mediante las flechas. La barra indica 50 μm

Dado que los resultados obtenidos en la expresión de las proteínas marcadas con fluorescencia verde no eran suficientes para confirmar el destino final, se coexpresaron las proteínas quimeras con un nanoanticuerpo Anti-GFP de fluorescencia roja (nanobody anti-GFP), este es un dominio de anticuerpo simple cadena VHH con actividad de unión específica a GFP. Las células epidérmicas de una hoja infiltrada pueden ser transfectadas con múltiples proteínas, lo que provee un medio eficiente para evaluar la localización diferencial de proteínas marcadas (Goodin, Dietzgen, Schichnes, Ruzin, & Jackson, 2002). Para aumentar la estabilidad del GFP en entornos ácidos como el apoplasto o la vacuola se coinfiltraron plantas de *N. benthamiana* con las proteínas fusionadas a GFP y el nanobody (Anti-GFP) y se evaluó la expresión

tras 24 hs, 48 hs, 72 hs y 8 días post-infiltración (Fig. 37 a 39). De esta manera durante la síntesis proteica se forma el complejo GFP: nanobody (AntiGFP) estabilizado por interacciones hidrofóbicas y polares en la interfaz de unión de las dos proteínas, esto contribuyen fuertemente a la afinidad y estabilización del complejo, lo que evitaría la degradación del GFP en entornos ácidos.

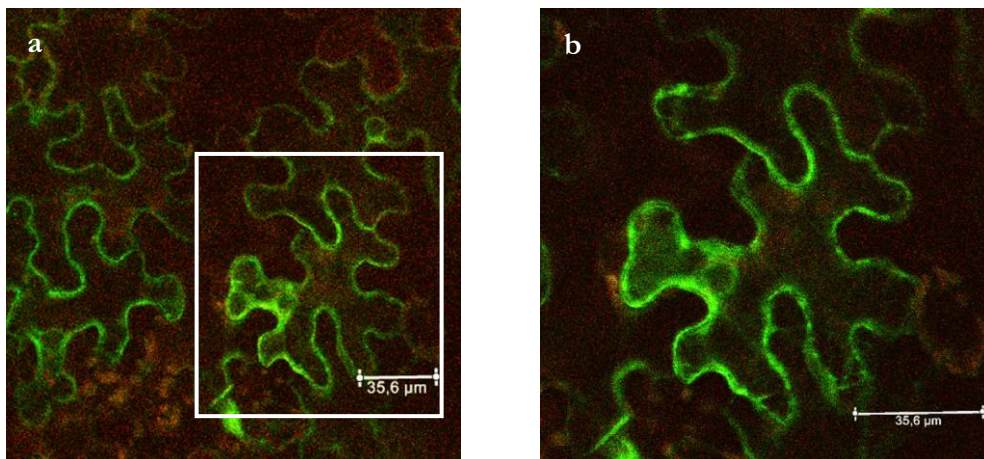


Figura 37. Localización subcelular de la peptidasa *wild-type* fusionada con eGFP coexpresada con nanobody anti-GFP luego de 48 hs post infiltración (a-b). Se muestra fluorescencia verde (eGFP) en co-localización con la fluorescencia roja (nanobody-AntiGFP). En cada imagen la barra indica la referencia de tamaño en μm .

Las células de *N. benthamiana* fueron co-infiltradas con las construcciones reporteras que expresa la peptidasa *wt* o mutante con el nanobody (Anti-GFP). Luego de 24 hs post-transfección no se observó ninguna señal de fluorescencia (datos no mostrados). En la epidermis de las hojas infiltradas, luego de 48 hs, pudo observarse la expresión de la peptidasa, pero la expresión del nanobody (Anti-GFP) no pudo evidenciarse (Fig 37). Se puede notar el patrón reticulado característico y la envoltura nuclear propia del RE, resultados similares a los obtenidos en la infiltración sin el nanobody para los mismos tiempos post-infiltración (Fig. 35 a-b).

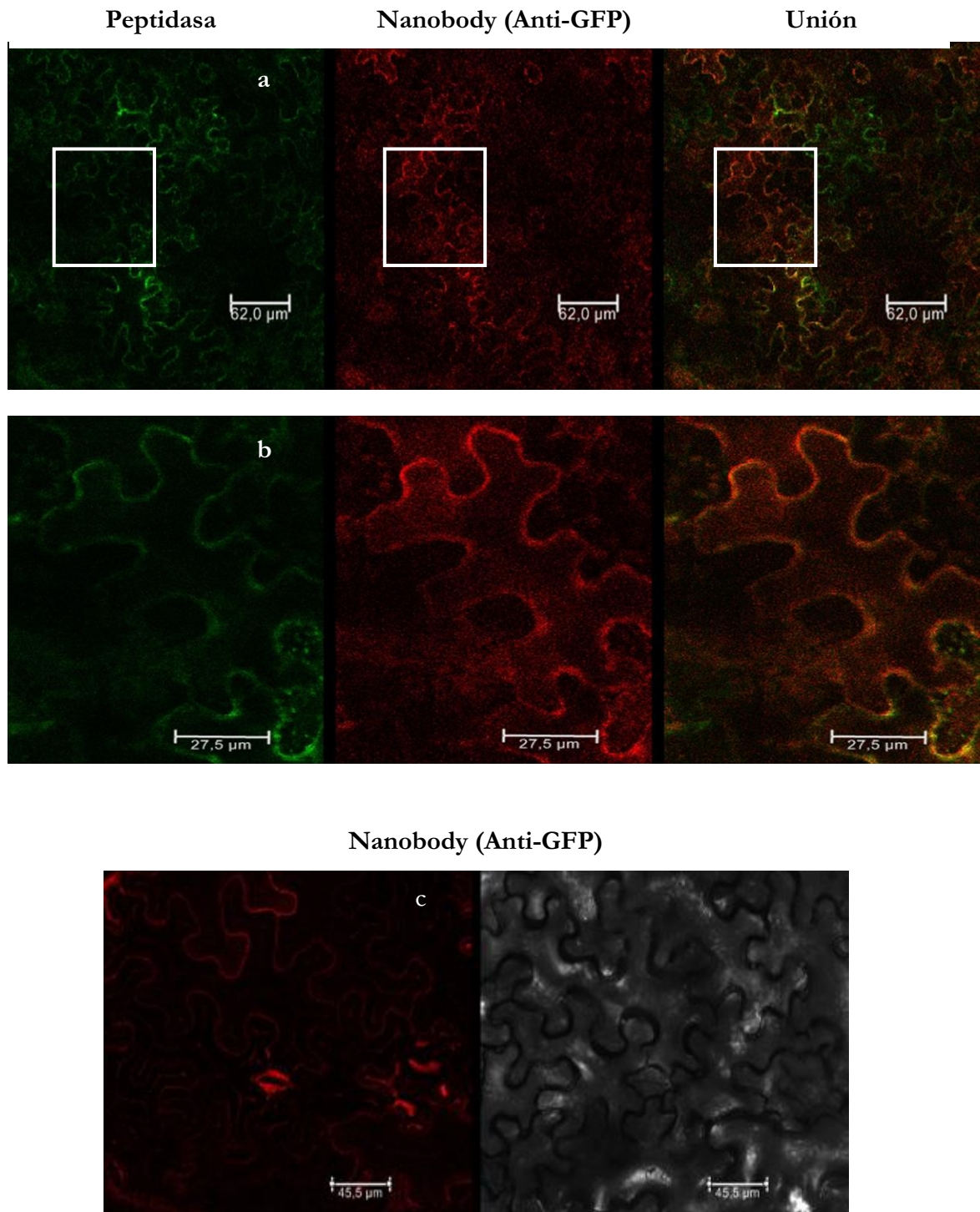


Figura 38. Localización subcelular de la peptidasa *wild-type* fusionada con eGFP coexpresada con nanobody anti-GFP luego de 72 hs post infiltración (a-b). A la izquierda se muestra la fluorescencia verde (eGFP), al centro la fluorescencia roja (nanobody-antiGFP), y la imagen de la derecha muestra la superposición de ambas. (c) control: nanobody-(AntiGFP) (izquierda), imagen de campo claro (derecha). En cada imagen la barra indica la referencia de tamaño en μm

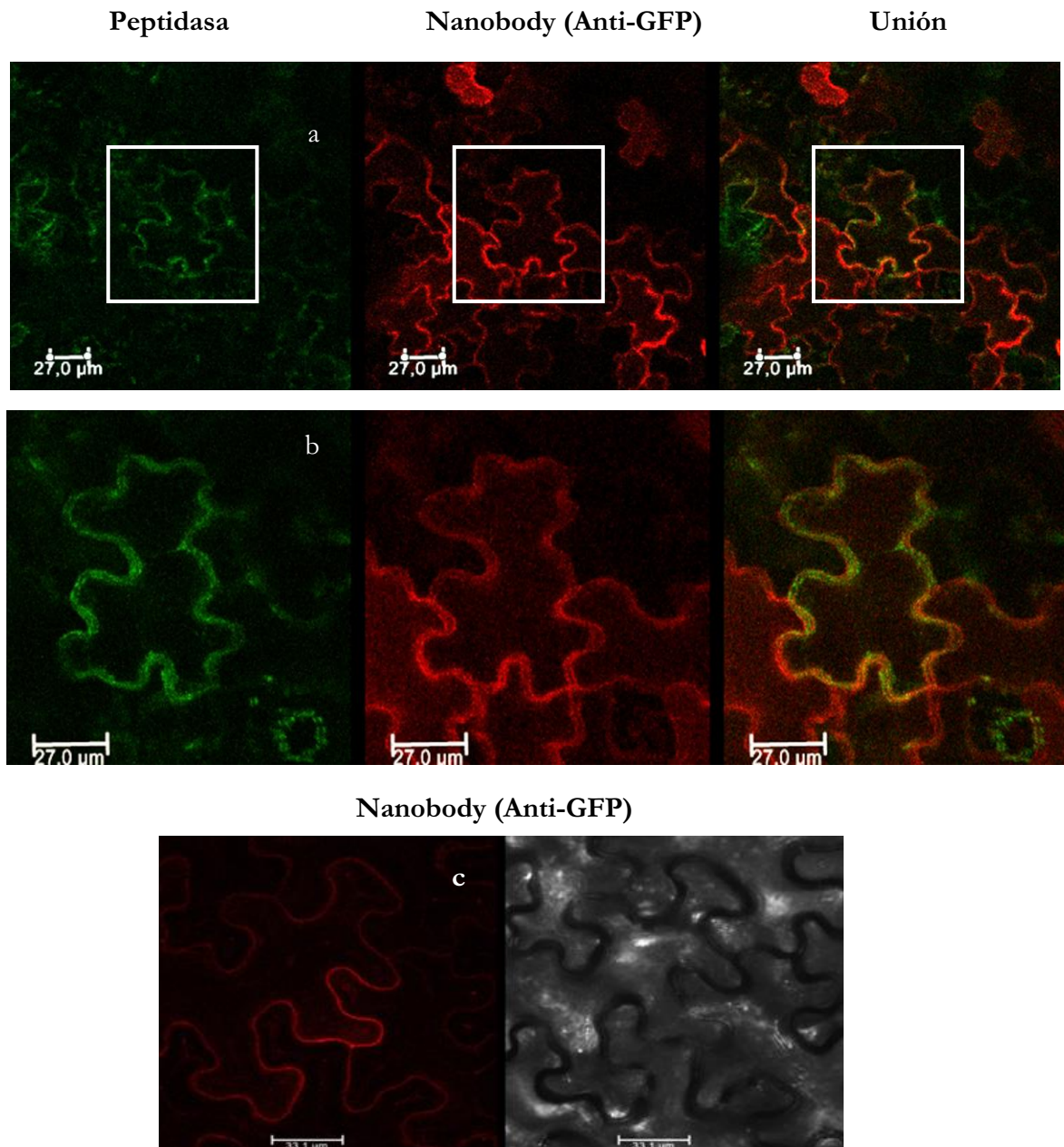


Figura 39. Localización subcelular de la peptidasa *wild-type* fusionada con eGFP coexpresada con nanobody anti-GFP luego de 8 días post infiltración (a-b). A la izquierda se muestra la fluorescencia verde (eGFP), al centro se muestra la fluorescencia roja (nanobody-antiGFP), y la imagen de la derecha muestra la superposición de ambas. (c) control: nanobody-(AntiGFP) (izquierda), imagen de campo claro (derecha). En cada imagen la barra indica la referencia de tamaño en μm

Las Fig. 38 y 39 muestran los resultados de colocalizaron de la peptidasa y el nanobody (Anti-GFP) obtenidos tras 72 hs y 8 días post-infiltración, respectivamente. El control agroinfiltrado únicamente con la construcción del

nanobody anti-GFP (Fig 38 y 38 c) mostró un patrón de distribución apoplástica compatible con el hecho de no portar señales de direccionamiento (Fig 38 y 39, centro). Sin embargo, la peptidasa, continuó mostrando el mismo patrón de localización observado a las 48 hs más una muy débil fluorescencia en el lumen vacuolar. Al superponer los canales verdes y rojo de estas imágenes, pudo observarse colocalización parcial entre la peptidasa y el nanobody (Anti-GFP), que se visualiza en tonos de amarillo/naranja lo que surge de la combinación de fluorescencia verde y roja. No obstante, dado que este patrón no fue observado en todas las células en los campos analizados, los resultados obtenidos no pueden considerarse concluyentes. Probablemente en este sistema la proteína no es completamente estabilizada por el nanobody por lo que se observa en tránsito a través de la vía secretoria, no se pliega correctamente y es degradada. Por otra parte, no se pudieron evidenciar diferencias apreciables en el perfil de localización respecto a las construcciones *wild-type* o mutante (datos no mostrados).

Debido a que los resultados con las construcciones con el reportero GFP en todas las condiciones ensayadas no fueron concluyentes, se realizaron ensayos de expresión transiente de la peptidasa fusionada a mRFP en los tejidos de hojas de *N. benthamiana* luego de agroinfiltración. La epidermis abaxial de las hojas fue infiltrada con suspensiones de *Agrobacterium* portando la secuencia codificante de preprosilpepsina 2 fusionada en su extremo 5' con la secuencia codificante de la proteína fluorescente roja (mRFP), integradas en el vector de expresión pB7RWG2. El reportero mRFP fue elegido debido a su capacidad para emitir fluorescencia aún al pH bajo de las vacuolas (Samalova et al., 2006; Shaner, Steinbach, & Tsien, 2005). Como se explicó anteriormente, un entorno ácido podría alterar el equilibrio de protonación de la proteína GFP volviéndola inestable, marcando una disminución en la fluorescencia, no así para la proteína roja que es estable en el lumen vacuolar (Samalova et al., 2006).

Las células transfectadas (DO_{600nm} del inóculo=1) con las construcciones reporteras mRFP fusionadas a la peptidasa tras 48hs, 72hs y 8 días post-infiltración se muestran en la Fig. 40, 41 y 42 respectivamente. La fluorescencia del RFP fue bien visualizada en las células epidérmicas de tejidos frescos por microscopía confocal. La peptidasa se encuentra principalmente en el lumen vacuolar, la fluorescencia roja se dispersa definitivamente en el interior celular, no así en los límites de las mismas propio de una localización en vacuolas de tipo líticas o vacuola central (Brandizzi et al., 2004; Di Sansebastiano, Paris, Marc-Martin, & Neuhaus, 2001). Este resultado evidencia el tráfico de la peptidasa a través del sistema endomembranas para llegar al destino final, la vacuola. Puede observarse que no existen diferencias apreciables en la localización de la peptidasa en los sucesivos días post-infiltración, salvo luego 24 hs post-transfección que no presentó ninguna señal de fluorescencia.

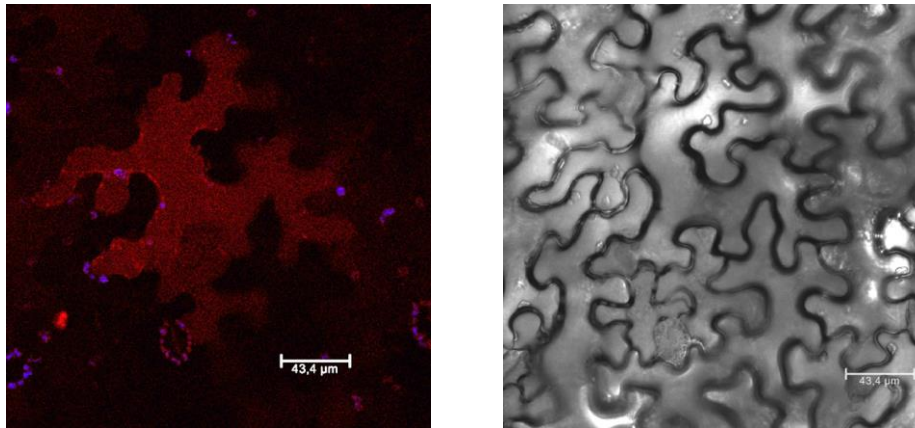


Figura 40. Localización subcelular de la peptidasa *wild-type* fusionada a mRFP A la izquierda se observa la fluorescencia roja (mRFP) luego de 48 hs post infiltración. En azul se observa la clorofila. A la derecha se observa la imagen digital con luz visible del mismo campo observado con fluorescencia. La barra indica tamaño en μm .

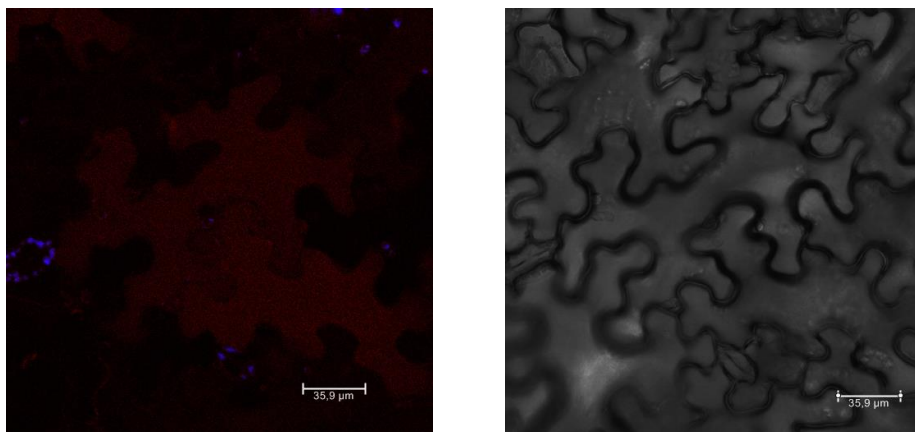


Figura 41. Localización subcelular de la peptidasa *wild-type* fusionada a mRFP A la izquierda se observa la fluorescencia roja (mRFP) luego de 72 hs post infiltración. En azul se observa la clorofila. A la derecha se observa la imagen digital con luz visible del mismo campo observado con fluorescencia. La barra indica tamaño en μm .

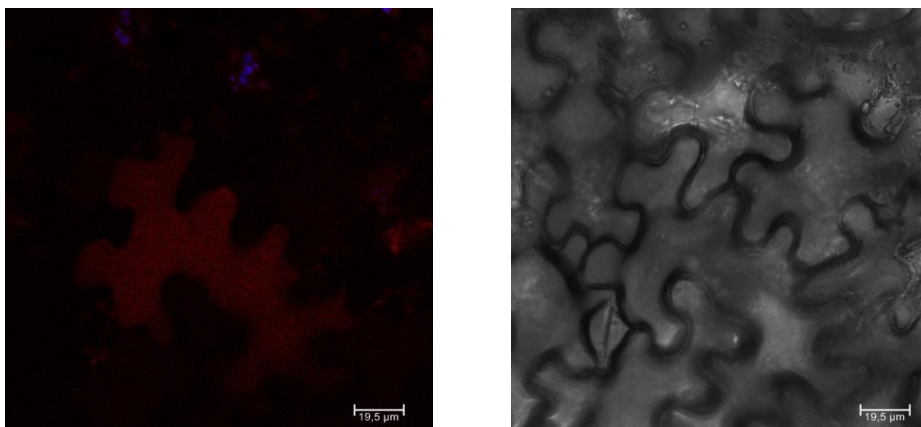
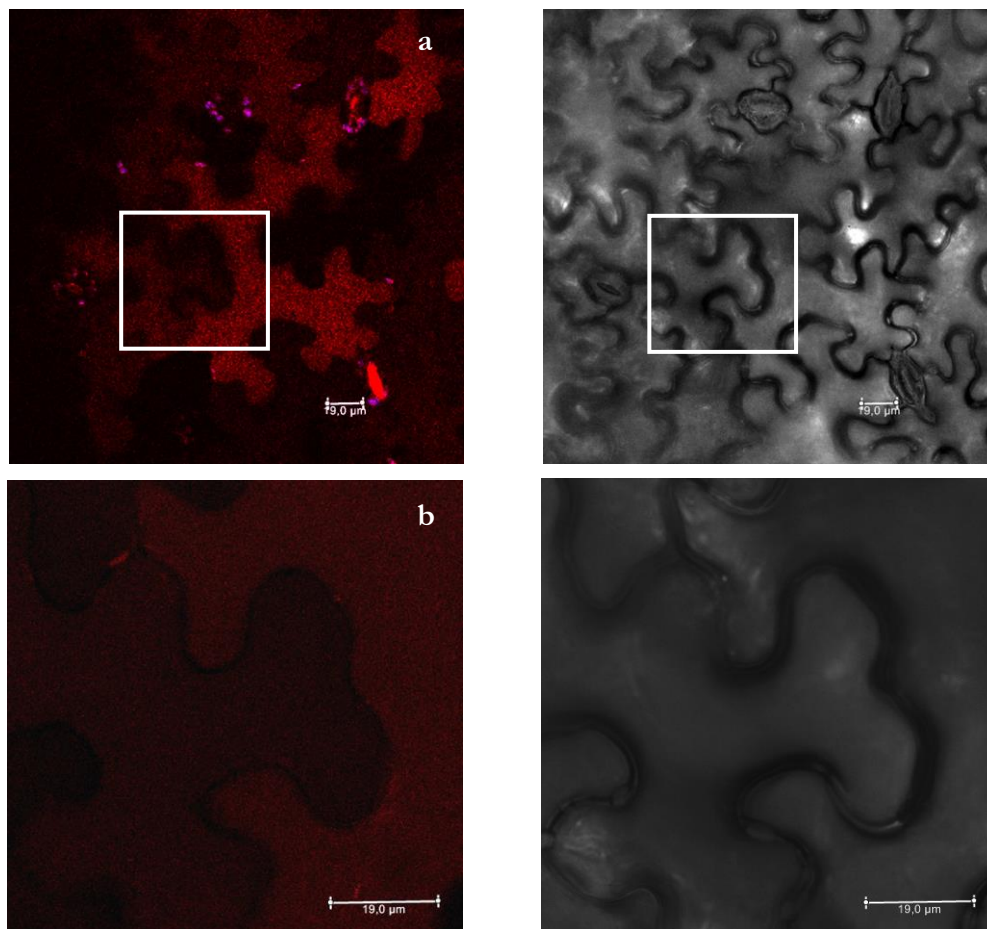


Figura 42. Localización subcelular de la peptidasa *wild-type* fusionada a mRFP A la izquierda se observa la fluorescencia roja (mRFP) luego de 8 días post infiltración. En azul se observa la clorofila. A la derecha se observa la imagen digital con luz visible del mismo campo observado con fluorescencia. La barra indica tamaño en μm .

En base a estos resultados se decidió realizar los posteriores ensayos con observaciones a 48 hs y 72 hs post-infiltración. Se probaron diferentes condiciones de DO_{600nm} del inóculo (0,1; 0,3; 0,6 y 1,0) para confirmar el resultado y evitar el problema de la sobrecarga de los sistemas de transporte de la expresión transiente (Scabone, Frigerio, & Petruccelli, 2011).

Cuando las plantas fueron agroinfiltradas con una DO_{600} de 0,1 no se observó señal de fluorescencia. En todas las otras condiciones ensayadas (DO_{600} 0,3; 0,6 y 1,0) luego de agroinfiltrar hojas enteras en tres experimentos independientes, ya sea a 48 como a 72 hs post-infiltración, casi todas las células epidérmicas mostraron la misma distribución de la proteína de fusión en el lumen vacuolar (Fig. 43 a 45).



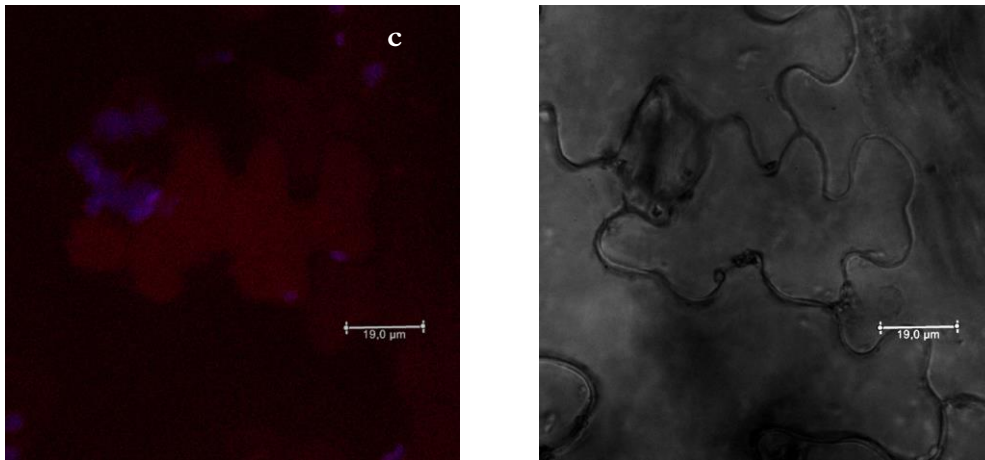
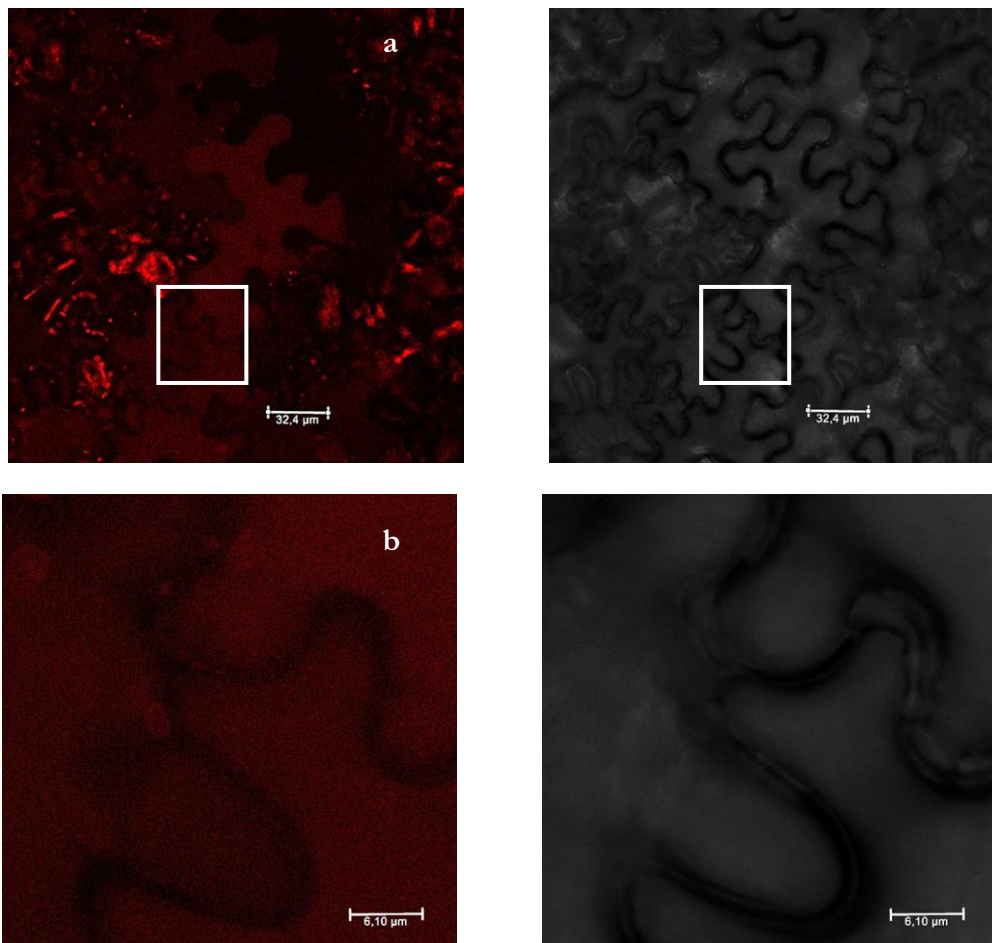


Figura 43. Localización subcelular de la peptidasa *wild-type* fusionada a mRFP, epidermis abaxial de hojas de *N. benthamiana* infiltrada con suspensiones de *Agrobacterium* de $DO_{600nm} = 0,3$. A la izquierda se observa la fluorescencia roja (mRFP) luego de 48 hs (a-b, página anterior) y 72 hs (c) post infiltración. En azul se observa la clorofila. A la derecha se observa la imagen digital con luz visible del mismo campo. La barra indica tamaño en



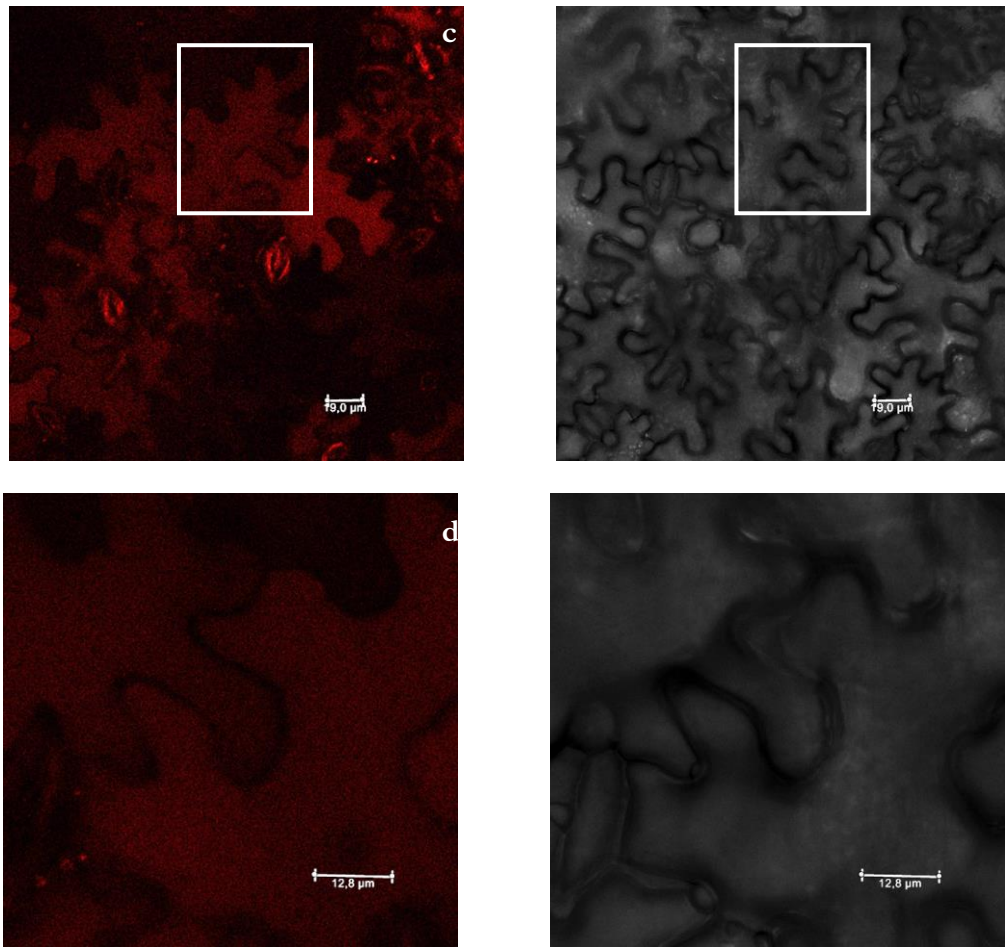


Figura 44. Localización subcelular de la peptidasa *wild-type* fusionada a mRFP, epidermis abaxial de hojas de *N. benthamiana* infiltrada con suspensiones de *Agrobacterium* de $DO_{600nm} = 0,6$. A la izquierda se observa la fluorescencia roja (mRFP) luego de 48 hs (a-b, página anterior) y 72 hs (c-d) post infiltración. A la derecha se observa la imagen de campo claro de lo que se observa en (Izquierda). La barra indica tamaño en μm .

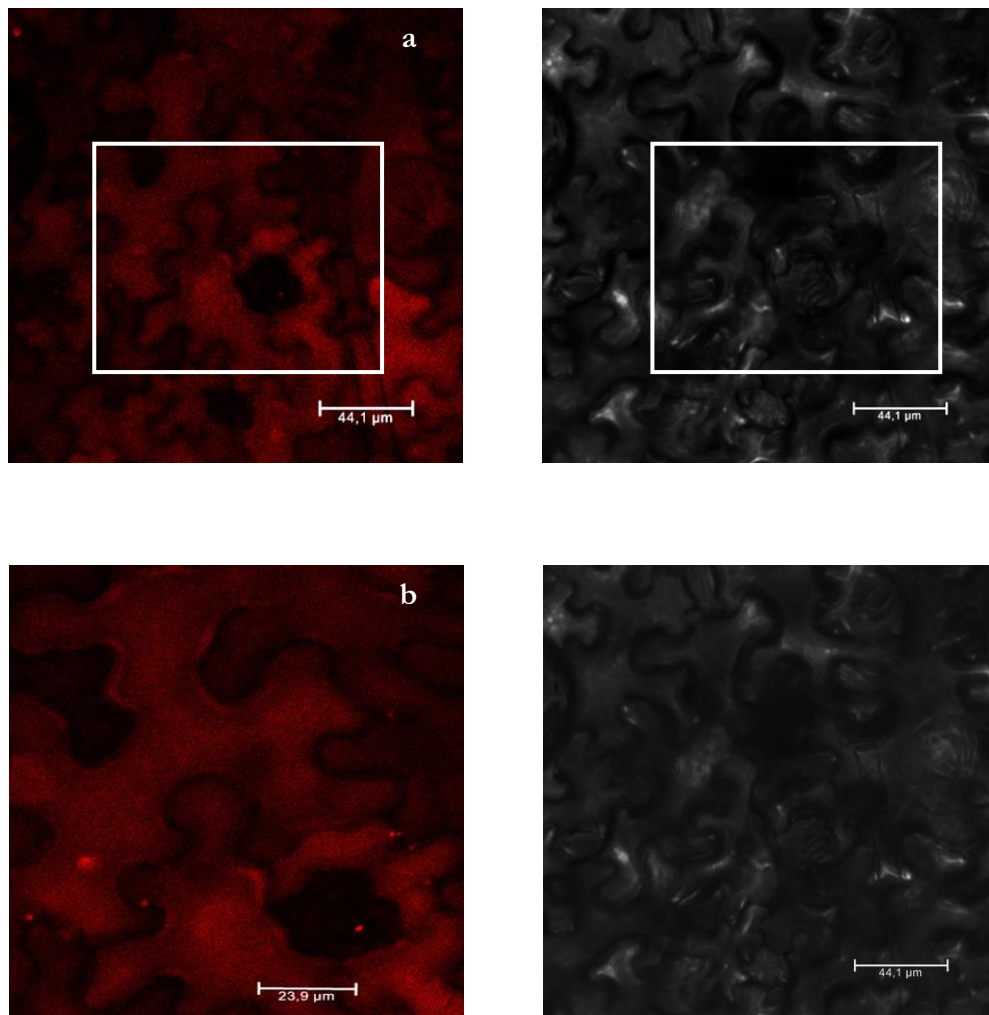


Figura 45. Localización subcelular de la peptidasa *wild-type* fusionada a mRFP, epidermis abaxial de hojas de *N. benthamiana* infiltrada con suspensiones de *Agrobacterium* de $DO_{600nm} = 1$. A la izquierda se observa la fluorescencia roja (mRFP) luego de 48 hs (**a-b**) post infiltración. A la derecha se observa la imagen digital con luz visible del mismo campo. La barra indica tamaño en μm .

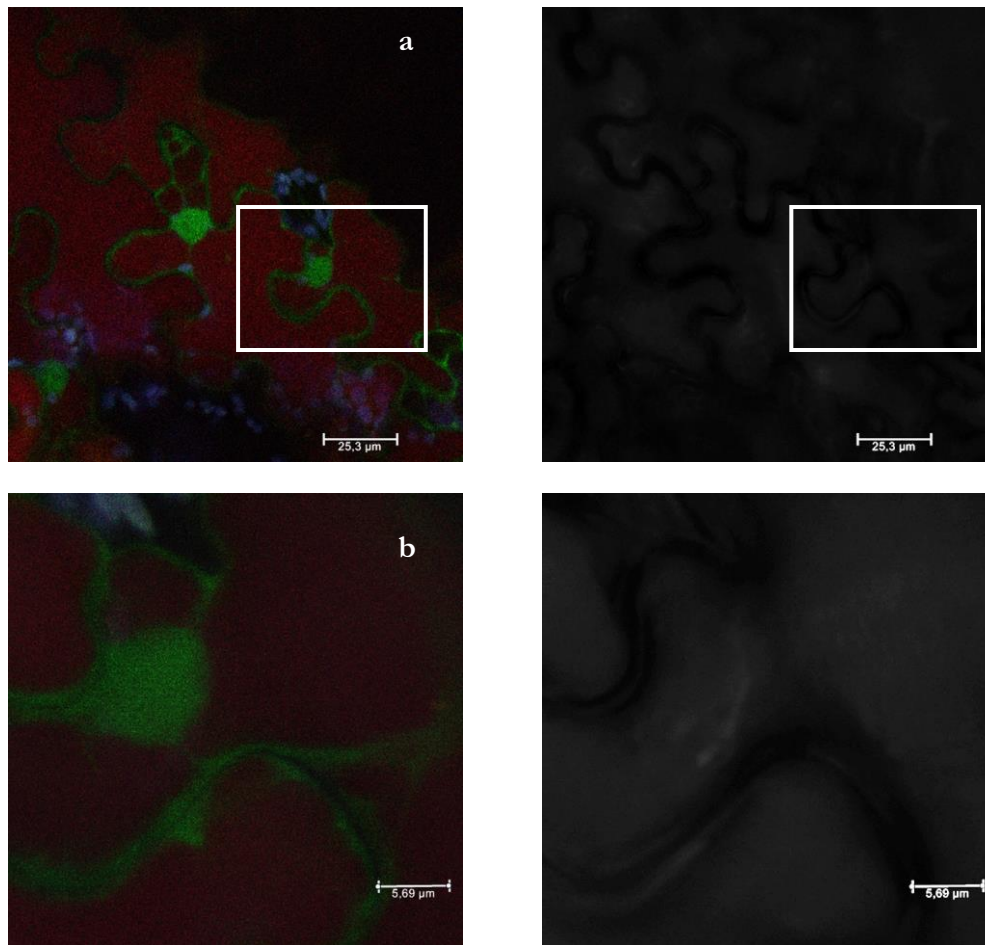
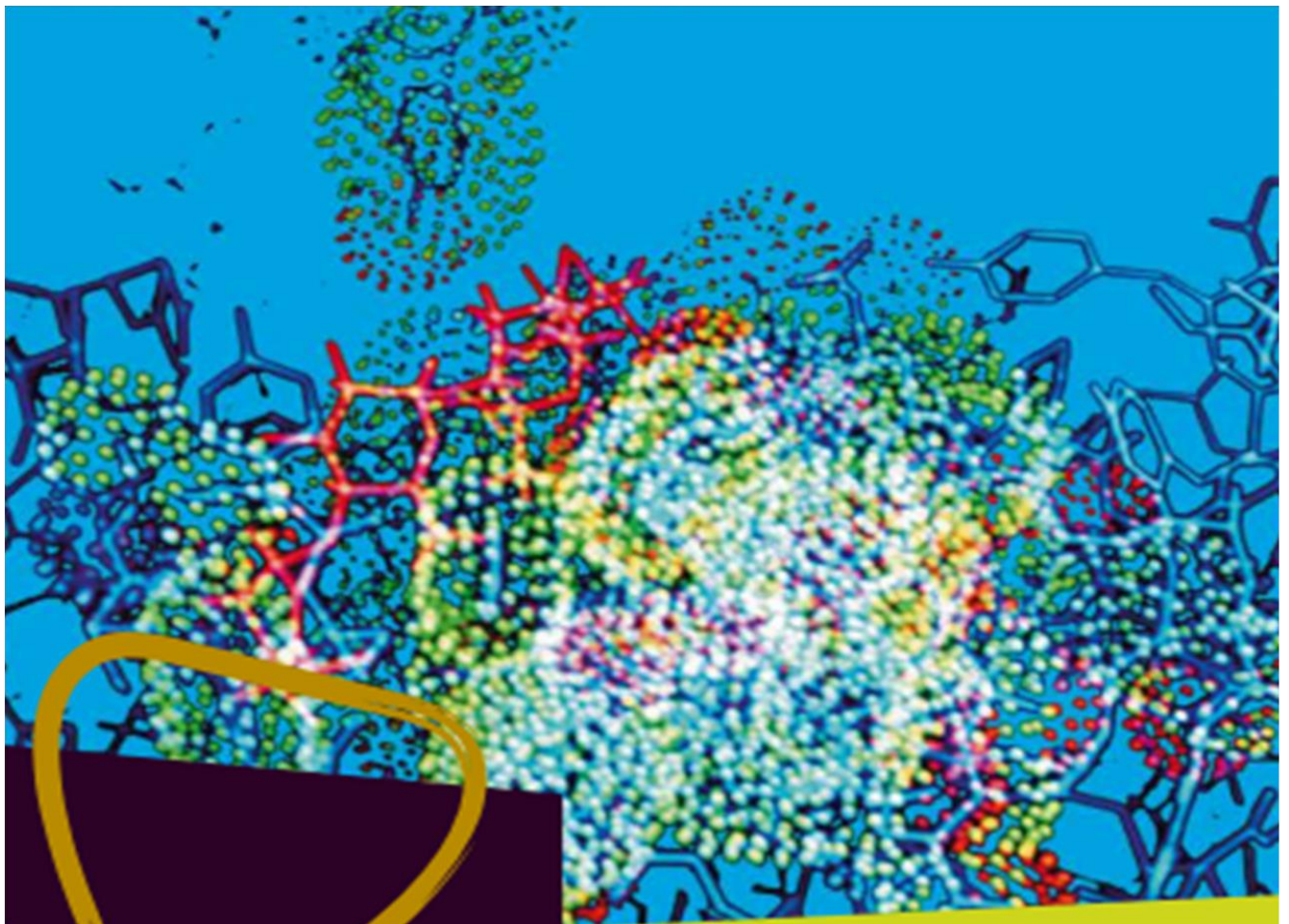


Figura 46. Localización subcelular de la peptidasa *wild-type* fusionada a mRFP coexpresada con pGDG (página anterior) (a-b). A la izquierda se observa la fluorescencia roja (mRFP) y verde (eGFP) luego de 72hs (a-b) post infiltración. En azul se observa la clorofila. A la derecha se observa la imagen digital con luz visible del mismo campo. La barra

Cuando las hojas de *N. benthamiana* fueron co-agroinfiltradas con la construcción de peptidasa fusionada a RFP y el plásmido pGDR que expresa un marcador de núcleo y citoplasma fusionado a GFP, pudo observarse la fluorescencia roja de la proteína quimera en el lumen vacuolar, mientras que la proteína verde se localizó en citoplasma y núcleo (Fig 46). En la imagen ampliada puede distinguirse también el espacio correspondiente a la pared celular y apoplasto.

Considerando las limitaciones de los experimentos de expresión transiente en el estudio de secuencias señal de direccionamiento por la posibilidad de sobrecarga

en la maquinaria de transporte (Scabone et al., 2011), nuestros resultados muestran que preprosilpepsina 2 es capaz direccionar el reportero RFP a la vacuola en las células epidérmicas de *N. benthamiana*, aun cuando la señal CtVSD no interviene como señal de direccionamiento. Estos resultados muestran que la secuencia de preprosilpepsina 2 presenta una señal adicional de direccionamiento vacuolar.



*Discusión
y Conclusiones*

6. DISCUSIÓN Y CONCLUSIONES

La expresión de PAs no tiene una distribución aleatoria en las plantas sino más bien tejido específica (Guo et al., 2013; Mutlu & Gal, 1999). Más aún, algunas especies de plantas tienen múltiples genes de PAs que se expresan en distintos tejidos, lo que sugiere cierta especialización funcional (Duarte, Pissarra, & Moore, 2008; Guo et al., 2013; Simões & Faro, 2004). Incluso cuando la actividad aspartilendopeptidasa en las plantas se encuentra en niveles relativamente bajos, las flores de varias especies de la tribu *Cardueae* poseen PAs con una alta actividad específica en ciertos tejidos. Las PAs de *S. marianum*, han sido aisladas previamente a partir de flores, aunque, tal como sucede con las PAs de *Arabidopsis* (Mazorra-Manzano & Yada, 2008) debido a la cantidad limitada obtenida y a su co-purificación su caracterización completa no pudo llevarse a cabo.

En este trabajo de Tesis Doctoral se reportó la clonación de dos aspartilpeptidasas a partir de ARNm obtenido de flores. La expresión del gen de una de ellas, preprosilpepsina 1, solo se encontró en flores maduras mientras que la segunda, preprosilpepsina 2, se encontró en ambos tipos de flores. Cordeiro y colaboradores (1994) demostraron que las cyprosinas se acumula en flores maduras, aunque la síntesis de su ARNm comienza durante las primeras etapas de desarrollo floral y finaliza en la maduración de la flor. Por otra parte, la expresión de cardosina A ocurre en inflorescencias jóvenes, desde el inicio de desarrollo de la flor, pero no cuando están completamente abiertas, lo que indica que la expresión del gen está regulada por el desarrollo floral (Frazão et al., 1999; Pereira et al., 2008).

La secuencia aminoacídica derivada del ADNc de ambas preprosilpepsinas, muestra la usual organización de los precursores de PAs típicas de plantas. La similitud global entre la secuencia de aminoácidos de estas enzimas con las de

otras PAs típicas de plantas, junto con la presencia de los motivos característicos de PAs, clasifican a las silpepsinas como nuevos miembros de este grupo de enzimas incluido en la familia A1 de MEROPS. En general, las PAs de plantas presentan una gran similitud de secuencia en sus dominios N- y C-terminales, con una identidad mayor del 60% y aproximadamente 45% de identidad con catepsina D, la PA más cercana de origen no vegetal (Simões & Faro, 2004). El análisis filogenético mostró que preprosilpepsina 1 y preprosilpepsina 2 se ubican en dos grupos o clúster diferentes. En ambos clústeres, se encuentran PAs de especies de la tribu *Carduae*, *C. calcitrapa*, *C. cardunculus*, *C. vulgare*, *A. minus* y *S. marianum* y forman un grupo monofilético. Los resultados obtenidos sugieren que, en algún momento de la evolución de *C. cardunculus* y *S. marianum*, un gen ancestral de PA se ha duplicado y ha dado lugar a las ramas que comprenden las PAs tipo cyprosinas y tipo cardosinas. Del análisis de 30 secuencias de PAs de 10 familias diferentes de plantas incluyendo Xanthorrhoeaceae, Poaceae, Polygonaceae, Nepenthaceae, Euphorbiaceae, Solanaceae, Asteraceae, Fabaceae, Myrtaceae y Brassicaceae se encontraron 10 motivos conservados presentes en casi todas las PAs estudiadas. Cada uno de estos motivos presenta un arreglo de residuos de aminoácidos altamente conservado en el grupo estudiado. Los avances biotecnológicos, en particular, de ingeniería de proteínas permitirán en el futuro cercano el diseño de nuevas peptidasas, tomando como plantillas los motivos altamente conservados de peptidasas de planta, con más y mejoradas propiedades para su uso industrial (Feijoo-Siota & Villa, 2011).

Las enzimas se utilizan en una variedad de procesos industriales. Una propiedad crítica en muchas aplicaciones es la alta estabilidad térmica necesaria en detergentes o el tratamiento del cuero (Güracar, 2011). El aumento de las temperaturas puede romper las interacciones no covalentes dentro de una proteína plegada, lo que lleva a la desnaturalización de la proteína. Durante el

proceso de desnaturalización, la exposición de los residuos hidrofóbicos puede potencialmente inducir la agregación de las proteínas. Además, las proteínas también se pueden inactivar a través de algunos mecanismos covalentes, tales como la desamidación de los residuos de Asn y Gln, hidrólisis de enlaces peptídicos, beta-eliminación de los enlaces disulfuro, y la oxidación de residuos de Cys y Met (Qing, Yi, Marek, & Iverson, 2013). Un alto valor de índice alifático indica que la proteína tiene un alto contenido de cadenas alifáticas que le proveen termorresistencia y por lo tanto estabilidad en diferentes condiciones (Sivakumar, Balaji, & Gangaradhakrishnan, 2007). Tanto silpepsina 1 como silpepsina 2 presentaron valores de índice alifático intermedios. Los índices de alifaticidad calculados para las enzimas clonadas junto con la especificidad acotada de las PAs sugieren que estas enzimas no serían buenas candidatas para este tipo de aplicaciones. Para las PAs de flores de cardos se han desarrollado diversas aplicaciones biotecnológicas y biomédicas que incluyen la manufactura de quesos – un proceso que data del periodo en que los Romanos ocupaban la Península Ibérica – así como otros usos absolutamente contemporáneos (Vairo Cavalli, Lufrano, Colombo, & Priolo, 2013). Es por esta razón que en las últimas dos décadas, las PAs presentes en flores de cardo han sido objeto de renovado interés. Las PAs de *S. marianum* presentan actividad proteolítica sobre hemoglobina con un pH óptimo de 3,8 como se observa para otras PAs como quimosinas de ternero, cerdo y cordero. Además por su capacidad coagulante de leche han sido propuestas como sustitutos de rennet (Vairo Cavalli, Claver, Priolo, & Natalucci, 2005; Vairo Cavalli, Silva, Cimino, Malcata, & Priolo, 2008). La estructura primaria de PAs típicas de plantas se caracteriza por la presencia de la secuencia señal que dirige el zimógeno al retículo endoplásmico, seguido por el prosegmento (Pro) que participa ya sea en la inactivación o en el correcto plegamiento, la estabilidad y direccionamiento intracelular de varios zimógenos. Además presentan un dominio N-terminal, un dominio C-terminal y el PSI, que

es una inserción única identificada en las PAs típicas de planta, altamente similar a las saposinas y proteínas tipo saposina, cuya función biológica no ha sido completamente esclarecida (Brodelius et al., 2005; Bryksa et al., 2011; Muñoz et al., 2010; Pereira et al., 2008; Simões & Faro, 2004). En cuanto a la estructura secundaria y terciaria las prosilpepsinas están formadas por dos cadenas polipeptídicas y su plegamiento es típico de otras PAs. Dicho plegamiento consiste en dos dominios topológicamente similares tipo barril β y el PSI que aparece como un dominio independiente de la enzima.

Los Pro de las PAs poseen típicamente alrededor de 50 aminoácidos, con la excepción de las plasmepsina del parásito *Plasmodium falciparum* que posee 120 residuos de aminoácidos (Bernstein & James, 1999). En muchas PAs, el Pro además de estar involucrado en estabilización e inhibición de la forma nativa, contribuye a la activación dependiente de pH y al direccionamiento intracelular del zimógeno. La función inhibitoria del Pro ha sido demostrada no solamente en los precursores de PAs sino también para precursores de peptidasas de otras clases, incluyendo serin-peptidasas como la peptidasa α -lítica (α LP), subtilisina y procarboxipeptidasa Y, la metalopeptidasa termolisina, papaína y otras peptidasas cisteínicas (Derek R. Dee, Filonowicz, Horimoto, & Yada, 2009).

Para fitepsina, una PA vacuolar de *H. vulgare*, se ha descrito que la presencia del Pro es esencial para el control de la proteólisis no deseada. El Pro y el N-terminal de la proteína están anclados en la hendidura del sitio activo, unidos a través de la interacción iónica entre la Lys11 y los residuos de ácido aspárticos catalíticos 36 y 223, haciendo que la enzima se encuentre inactiva (Kervinen et al., 1999). Esta interacción se desestabiliza al pH bajo del compartimiento vacuolar, lo que conduce a la activación de la enzima. En los modelos estructurales de ambas prosilpepsinas el propéptido, junto con la cadena N-terminal, envuelven la hendidura interdominio donde se ubica el sitio activo. El propéptido al cubrir parcialmente esta hendidura podría estar implicado en la

inactivación del sitio activo en ambas proenzimas. Castanheira y colaboradores (2005) propusieron que la eliminación del Pro en cardosina A, ocurre cuando llega a la vacuola, sugiriendo que esta región puede ser parte en una estrategia de inactivación en cardosina A. De hecho, se ha descrito que para que cardosina A recombinante sea activa necesita la eliminación parcial del prosegmento.

El requerimiento del Pro para el correcto plegamiento también se ha demostrado entre las distintas clases de peptidasas. Específicamente, el papel del Pro en la catálisis del plegamiento de proteínas ha sido estudiado para las serin-peptidasas subtilisina, α LP y la peptidasa B de *Streptomyces griseus*. La pepsina, una PA arquetípica, se desnaturaliza irreversiblemente cuando se somete a pH casi neutro, desnaturalizantes químicos o calor, mientras que su forma zimogénica, el pepsinógeno, se despliega reversiblemente en las mismas condiciones. Esto indica que el Pro es necesario para el plegamiento de la pepsina (Derek R. Dee et al., 2009; Wang, Horimoto, Dee, & Yada, 2014). Recientemente, mediante el empleo de cuatro péptidos correspondientes a secuencias truncadas del Pro de pepsinógeno, (Pro1–16, Pro1–29, Pro17–44 y Pro30–44, numeración del pepsinógeno) han sido atribuidas tres funcionalidades a regiones discretas altamente conservadas del Pro. Los primeros 16 aminoácidos del Pro proveería los contactos para la unión a los dominios de la peptidasa, el segmento Pro 17–29 inhibiría el sitio activo, mientras que la unión y/o catalización del estado de transición del plegamiento estaría a cargo del segmento Pro1-29 del pepsinógeno (Dee, Myers, & Yada, 2015). La alta conservación de la secuencia de Pro1-29 entre las PAs tipo pepsina y su papel en la catálisis del plegamiento de pepsina sugieren que existe un mecanismo de plegamiento catalizado por dicho segmento Pro. No se ha demostrado aún un papel evidente en la catálisis del plegamiento mediada por Pro para el segmento Pro30-44. Dee et al. (2015) han sugerido que esta región por ser la más variable puede servir como un enlazador entre Pro1-29 y el dominio maduro. Si bien, como puede observarse en la Fig. 10

de la Sección Resultados, existe una mayor divergencia en la secuencia de los propéptidos de los precursores de cardosina A, cardosina B, silpepsina 1 y HA-AP entre los residuos 47-57 (equivalente a Pro25-35, numeración de pepsinógeno), la secuencia RKDFGF (ubicada en el segmento correspondiente a Pro30-44 numeración de pepsinógeno, 58-63 numeración de preprosilpepsinas) está conservada en PAs de *Asteraceae*. Esta conservación de secuencia implicaría alguna función específica de esta región del Pro aun no demostrada.

En ambas prosilpepsinas se predijeron dominios funcionales involucrados en la actividad aspartilpeptidasa y en actividades implicadas en procesos metabólicos lipídicos. El dominio peptidasa aspártica eucariota Asp: PF00026 encontrado por Pfam para los dos precursores de silpepsinas está presente en pepsinas, catepsinas, y reninas. Los dos dominios aspárticos, subunidad mayor y menor, probablemente surgen de una duplicación ancestral. Esta familia no incluye proteasas retrovirales ni retrotransposones (PF00077) ya que son mucho más pequeños y parecen ser homólogos a cada uno de los dominios de las peptidasas aspárticas de eucariotas (Tang, 2004).

Los dos motivos de secuencia catalíticos DTG/DSG se encuentran en todas las PAs de plantas pertenecientes a la familia A1 con la excepción de chlapsina, la PA de *Chlamydomonas reinhardtii* que contiene DTG/DTG al igual que las peptidasas homólogas animales y microbianas (Almeida et al., 2012). En algunas PAs de hongos y de protozoos, los residuos de ácido aspártico catalíticos también se encuentran dentro de los motivos DTG/DSG. El significado evolutivo o biológico de esta variación observada en las PAs de diferentes reinos aún no se ha establecido (Simões & Faro, 2004). Además de la estructura de dominios, las tríadas catalíticas y la Tyr del ψ loop, otra característica compartida con las PAs típicas es la distribución de cisteínas a lo largo de toda la secuencia (Faro & Gal, 2005). Existe una sorprendente conservación de los enlaces disulfuro de las peptidasas secretadas de la familia A1 en mamíferos y sus

homólogas de plantas, con tres enlaces disulfuro, el primero y tercero de las cuales también está presentes en muchas PAs homólogas de hongos. Los enlaces disulfuro son diferentes en los diferentes lóbulos, lo que implica que los enlaces se introdujeron en algún momento de la evolución después de la duplicación de genes que dio origen a los dos dominios (N. D. Rawlings & Salvesen, 2013).

De la comparación de las secuencias y el análisis de las estructuras puede inferirse que silpepsina 1 y silpepsina 2 tendrán especificidades diferentes ya que sus bolsillos S1 en la hendidura del sitio activo presentan distintos residuos de aminoácidos, siendo la primera más parecida a fitepsina y la segunda a cardosina A. El grupo hidroxilo de la Tyr80 de prosilpepsina 1, como en profitepsina, formaría un enlace de hidrógeno con el Asp103, mientras que su anillo aromático se localizaría en el bolsillo hidrofóbico S1 de la enzima, la cadena principal de 77-81 formaría los puentes de hidrógeno del flap, y la Ile74 se colocaría en el bolsillo de unión al sustrato S1' (de acuerdo a la numeración de fitepsina: Lys11, Tyr13, Ile 7 y Asp36 y 223). Mientras que otros residuos están presentes en esas posiciones en prosilpepsina 2 (Ile74, Thr78 y Glu80), y las sustituciones son conservativas con respecto a los residuos presentes en cardosina A (Val74, Thr78 y Asp80). Castanheira y colaboradores (2005) demostraron que el bolsillo hidrofóbico S1 de cardosina A, obtenida tanto de su fuente natural como en forma recombinante, sólo es capaz de acomodar Phe y en menor medida Leu; mientras que el subsitio S1' presenta una especificidad más amplia, con preferencia por Tye y Phe. Fitepsina de cebada tiene una especificidad primaria más amplia que cardosina A; cliva enlaces entre residuos con cadenas hidrofóbicas (Leu, Ile, Val, Phe) o cerca de residuos de este tipo (Kervinen & Wlodawer, 2013).

Al igual que cardosina B las secuencias de preprosilpepsina 1 y preprosilpepsina 2 no contienen el motivo Arg-Gly-Asp (RGD) ni el motivo Lys-Gly-Glu (KGE) presentes en cardosina A, C, E, F, G, y H (Sarmiento et al., 2009). La presencia

de esta secuencia funcional RGD en la porción C-terminal de la cadena de 31 kDa es una característica distintiva de estas cardosinas entre las peptidasas de planta. RGD es una secuencia conocida de unión a integrinas en tejidos de mamíferos en donde facilita muchas funciones de reconocimiento celular tales como adhesión, migración, señalización, diferenciación y crecimiento. Esta secuencia se localiza en un giro entre dos cadenas, sobre la superficie molecular y opuesta al surco del sitio activo en cardosina A (Frazão et al., 1999). El motivo KGE mimetiza al RGD tanto en su carga como por su posición protuyendo del *core* proteico, expuesto en la superficie molecular (Simões et al., 2005; Vairo Cavalli et al., 2013). En silpepsina 1, silpepsina 2, cardosina B y cyprosina la ausencia de estos motivos RGD/KGE, implicaría que estas enzimas desempeñan un rol bien diferenciado en las flores del rol desempeñado por las cardosinas que si contienen estos motivos.

En prosilpepsina 1 se encontró un motivo: ligando de la familia de proteínas Atg8 (LIR). Los motivos LIRs en la proteína que los contiene están incluidos en una hoja beta y le permiten a dicha proteína interactuar con proteínas tipo Atg8 de la vía de la macroautofagia. Al producirse la interacción con una hoja beta de la proteína tipo Atg8 ambas hojas beta forman una lámina beta intermolecular paralela (Noda, Ohsumi, & Inagaki, 2010; Rogov, Dötsch, Johansen, & Kirkin, 2014). En el modelo estructural de prosilpepsina 1 se comprobó que el motivo encontrado está formando una hoja beta. La caracterización de proteínas que contienen motivos LIR y la elucidación de sus roles en la macroautofagia se encuentra en sus inicios (Birgidottir, Lamark, & Johansen, 2013). La presencia de este motivo se podría relacionar con una posible función de silpepsina 1 en el mecanismo de macroautofagia en las flores. La macroautofagia es una de las vías principales de degradación en eucariotas. Estudios recientes han revelado la importancia de la macroautofagia en variados aspectos de la vida de la planta, incluyendo el establecimiento de las plántulas,

desarrollo, resistencia al estrés, metabolismo y reproducción. Esto se manifiesta por la doble capacidad de la macroautofagia para ejecutar la degradación generalizada en condiciones ambientales severas, mientras que al mismo tiempo puede ser altamente selectiva en la orientación de compartimentos específicos y complejos de proteínas para regular procesos celulares clave, incluso durante condiciones de crecimiento favorables. Esta vía podría tener un rol importante en el desarrollo y la senescencia floral (Michaeli, Galili, Genschik, Fernie, & Avin-Wittenberg, 2015).

Los dominios tipo saposina (SapB 2 y SapB 1: PF03489 y PF05184) que surgen del análisis del Pfam para prosilpepsina 1 y 2 están incluidos en el PSI (Bryksa et al., 2011). Las saposinas en los mamíferos son pequeñas proteínas lisosomales que sirven como activadores de varias enzimas que degradan lípidos lisosomales. Probablemente actúan aislando el sustrato lipídico de los alrededores de la membrana lo que hace más accesible este sustrato a las enzimas de degradación (Alattia, Shaw, Yip, & Privé, 2007). En los mamíferos se han identificado cuatro clases de saposina (A, B, C y D), su especificidad y modo de activación de las hidrolasas de esfingolípidos difiere entre las saposinas individuales. Estructuralmente poseen seis cisteínas que forman tres enlaces disulfuro, un sitio de glicosilación conservado y la presencia de prolina y tirosina conservadas (Munford, Sheppard, & O'Hara, 1995). Los datos estructurales y homología de secuencia sugieren que las saposina, proteínas tipo saposinas y dominios tipo saposinas en otras proteínas comparten la misma estructura compacta unida por puentes disulfuros con una topología de cinco α -hélices. Estas características estructurales también se predijeron para los PSIs de ambas prosilpepsinas. Aunque las regiones Sap2 y Sap1 están invertidas en el PSI con respecto a las saposinas animales formando la *swaposina* (Brodelius et al., 2005). A pesar de esta diferencia en la localización de los extremos, la estructura global tipo saposina no se encuentra alterada en los PSIs. De hecho, estudios cristalográficos del

dominio *swaposina* presente en una PA de papa (StAP-PSI), revelaron para el mismo una configuración global en forma de V similar a la estructura abierta de la saposina C humana (Bryksa et al., 2011).

Brodelius y colaboradores (2005) postularon que el PSI puede activar la galactosilceramidasa en las células vegetales y de ese modo estar involucrado en la hidrólisis de los esfingolípidos. El dominio *PD516165* se ubica entre el aminoácido 316 y 351 (MSQQCKTLVDQYGKSIIEML[ML]SEAQPDKICSQMKLC) de las prosilpepsinas (equivalente a PF03489 de Pfam). Dentro de este dominio se encuentra la secuencia consenso del péptido responsable de la unión y activación de la glucosilceramidasa en saposina C. La glucosilceramidasa es una enzima del metabolismo de los esfingolípidos que cataliza en mamíferos la hidrólisis de glucosilceramidas en ceramida y glucosa. La interacción entre saposina C y las membranas que contienen fosfolípidos ácidos, es un paso crucial para la activación de la enzima, saposina C induce la restructuración y desestabilización de la membrana (Morimoto et al., 1990; You, Qi, Grabowski, & Yu, 2003). La glucosilceramida es el principal esfingolípidos de las membranas vegetales. Aunque la actividad de glucosilceramidasa ha sido detectada, es poco lo que se conoce sobre la hidrólisis de la glucosilceramida en los tejidos vegetales (Dunn, Lynch, Michaelson, & Napier, 2004).

Más allá del posible rol biológico sugerido para los homólogos de planta de las saposinas – activación de enzimas involucradas en el metabolismo lipídico – también se ha estudiado su función en defensa (Brodelius et al., 2005). Egas y colaboradores (2000) demostraron que el PSI de cardosina A es un potente inductor de permeabilización vesicular y propusieron la idea de que los precursores de PA son moléculas bifuncionales, que contienen un dominio desestabilizador de membrana en adición al dominio peptidasa. Además, la presencia del PSI en PAs de plantas que en su forma madura son monoméricas ha sido asociada con este rol bifuncional. El dominio PSI de la peptidasa

monomérica StAsp, sobreexpresado en *E. coli*, exhibe actividad citotóxica contra un amplio espectro de patógenos, incluidos patógenos de papa (Muñoz et al., 2010); cabe destacar que el PSI de cirsinina (*C. vulgare*), aunque es una PA heterodimérica en su forma madura, al ser producido en *Kluyveromyces lactis*, también fue activo contra fitopatógenos fúngicos (Curto et al., 2014).

Las PAs pueden acumularse en compartimentos diferentes, y la misma proteína puede ser secretada al apoplasto o ser dirigida a la vacuola dependiendo del tipo de célula, y estado de desarrollo, sugiriendo un mecanismo estricto en la regulación del tránsito (Pereira, Pereira, Satiat-Jeunemaitre, & Pissarra, 2013). Los datos generales sobre localización de cardosinas sugieren una expresión regulada en el desarrollo de tejidos específicos para ambas cardosinas. Cardosina A se acumula masivamente en vacuolas de almacenamiento de proteínas de las papilas estigmáticas de la flor de cardo (Ramalho-Santos et al., 1997); mientras que, cardosina B se encuentra en la matriz extracelular del tejido de transmisión en flores maduras, en el lado ventral de la pared del ovario y micrópilo (Figueiredo, Duarte, Pereira, & Pissarra, 2006; Vieira et al., 2001). Además de las funciones antes descritas, la presencia del dominio PSI ha sido vinculado al tránsito intracelular de las PAs heterodiméricas. La delección del PSI de fitepsina – PA que normalmente se localiza en vacuolas– produce su direccionamiento al apoplasto (Törmäkangas et al., 2001); sin embargo, al estudiar dos PAs vacuolares de soja, (Terauchi et al., 2006) encontraron que la remoción de los PSIs, provocaba que uno de los mutantes siguiera siendo enviado a la vacuola mientras que el otro no. Se ha propuesto que el PSI podría actuar como señal vacuolar sólo en condiciones específicas o en determinadas etapas de desarrollo y dependiendo del tipo de PA y de PSI (Terauchi et al., 2006).

El transporte de proteínas solubles a la vacuola es un proceso saturable (Frigerio, Prada, Faoro, & Vitale, 1998) que depende de la presencia de VSD en su estructura (Vitale & Raikhel, 1999). Estos VSDs están situados en el extremo N

o C-terminal, o dentro de las secuencias internas de las proteínas vacuolares. Los primeros son los ssVSDs y consisten en motivos conservados NPIRL o similares (Nakamura & Matsuoka, 1993). Los CtVSDs no tienen similitud en la composición de aminoácidos y en la longitud, pero en general son ricos en residuos hidrófobos y deben estar accesible en el extremo C-terminal (Matsuoka & Neuhaus, 1999; Vitale & Pedrazzini, 2005). Las señales de direccionamiento internas están menos definidas y se ha sugerido que podrían actuar como determinantes físicos (psVSD) provocando la agregación que resulta luego en el direccionamiento (Holkeri & Vitale, 2001; Petruccelli, Molina, Lareu, & A, 2007). Pereira y colaboradores, 2013 propusieron que las PAs pueden seguir tres rutas principales de direccionamiento vacuolar. Una de estas rutas implicaría el direccionamiento por el dominio PSI, este funcionaría como VSD de una manera similar a la participación de saposinas en el cotransporte de procatepsina D al lisosoma (da Costa, Pereira, Moore, & Pissarra, 2010; Domingos et al., 2000; Simões & Faro, 2004). El PS de preprosilpesina 1 y 2 difiere del PS del precursor de cardosina B en un único residuo (sustitución semiconservativa); Pereira y colaboradores (2013) demostraron que este último es capaz de dirigir proteínas reporteras a la vía secretoria. Asimismo, demostraron que secuencia C-terminal de cardosina A: VGFAEAA es un CtVSD. Esta secuencia, capaz de dirigir proteínas reporteras a la vacuola, también está presente en ambas silpepsinas. Por otro lado los mismos autores fusionaron el dominio PSI de cardosina A a la proteína fluorescente mCherry y por expresión transiente detectaron ambas proteínas quiméricas en vacuolas de células de *N. tabacum*. Esto indica que ambos dominios, PSI y CtVSD de cardosina A actúan como dominios de direccionamiento vacuolar en este sistema. El dominio PSI, sin embargo, no cumple ninguno de los criterios habituales que definen un VSD. No puede ser considerado como un ssVSD en el sentido estricto de la palabra la secuencia de aminoácidos del PSI no se

conserva; la definición psVSD tampoco se aplica, ya que sólo una única región de la proteína está implicada (Jolliffe, Ceriotti, Frigerio, & Roberts, 2003). Aunque no todos los dominios de PSI de PAs desempeñan un papel esencial en el direccionamiento vacuolar, el dominio PSI definiría una nueva clase de determinante de direccionamiento vacuolar (Pereira, Pereira, Satiat-Jeunemaitre, & Pissarra, 2013).

Cuando se compararon los dominios PSI de procardosina A y procardosina B y los de prosilpepsinas 1 y 2 se observó que tanto procardosina B como prosilpepsina 1 tienen un sitio de glicosilación putativo en el dominio PSI, mientras que el mismo está ausente en procardosina A y prosilpepsina 2. La presencia o la ausencia de un sitio de glicosilación en el dominio PSI han sido propuestas como una estructura clave en la determinación de la ruta a ser seguida hasta la localización impulsada por el PSI (Pereira et al., 2013). En este trabajo de tesis, la secuencia codificante de preprosilpepsina 2 y un mutante de sitio activo se fusionaron en C-terminal a genes reporteros y se expresaron en forma transiente en plantas de *N. benthamiana*. La fusión C-terminal enmascara la función del dominio CtVSD presente en el precursor silpepsina 2, permitiendo evaluar la existencia de otra señal de direccionamiento en la secuencia proteica. La proteína reportera GFP es inestable en entornos ácidos por alteración del equilibrio de protonación y activación de peptidasas pH dependientes (Flückiger et al., 2003; Tamura et al., 2003), se abordaron dos diferentes estrategias de estabilización pero no se logró observar la peptidasa ni en vacuola, ni en apoplasto. Sin embargo, la fusión de preprosilpepsina 2 a RFP, que es capaz de emitir fluorescencia aún a pH bajo en la vacuola, permitió observar en este sistema a la proteína quimera en vacuola. Lo que demuestra que prosilpepsina 2 posee otra señal de direccionamiento vacuolar además de CtVSD. Estos resultados son coincidentes con los observados por Pereira y colaboradores (2013), quienes postularon que la presencia de un solo dominio, ya sea el PSI o

el péptido C-terminal, puede ser necesaria y suficiente para la acumulación vacuolar de la proteína reportera. La aparición de dos determinantes de direccionamiento vacuolar en la misma proteína ha sido previamente informado, además de cardosina A, para la proteína de almacenamiento de *Glicina max* (Nishizawa et al., 2006). La presencia de dos VSDs en una proteína podría asegurar una capacidad de *poli-sorting*, dando a las PAs múltiples funciones según su localización final, y finalmente permitiendo modular el tráfico vacuolar según las limitaciones de las células o necesidades a través de una vía de señalización específica del PSI, que puede convivir con otra vía vacuolar (Pereira et al., 2013).

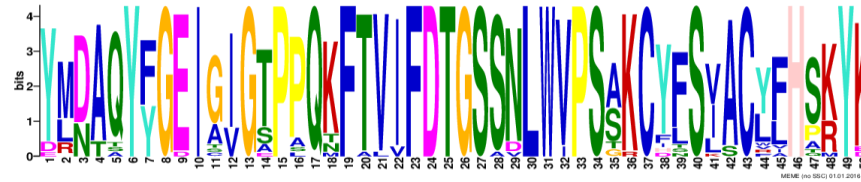


Anexos

7.1. ANEXO I

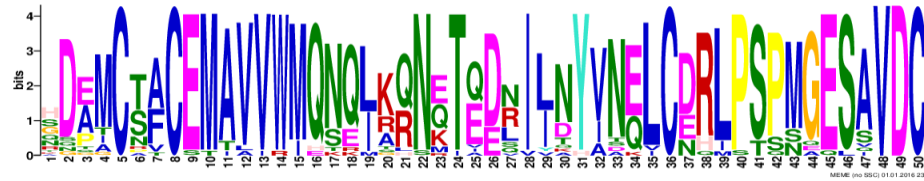
Con el fin de encontrar patrones de los motivos conservados de las dos PAs clonadas se utilizó el programa MEME. El diagrama de la Fig. 6 de la sección de Resultados muestra la ubicación de cada motivo; donde cada bloque indica la posición y la fuerza de un motivo, es decir nos da una idea de cuales posiciones son más conservadas en el motivo. Aquí se muestran las representaciones gráficas de la conservación de secuencia de cada motivo. La altura de un bloque indica la importancia del sitio, siendo los bloques más altos los más significativos. La altura es proporcional al logaritmo negativo del *p-value* del sitio, siendo el máximo de altura un *p-value* de $1 \times e^{-10}$, es decir es directamente proporcional a su nivel de conservación. El *p-value* se define como la probabilidad de que una secuencia aleatoria (con la misma longitud y teniendo en cuenta el fondo) tenga valores de *p-value* en dicha posición, tal que el producto sea menor o igual que el valor calculado para la secuencia en estudio.

MOTIVO 1



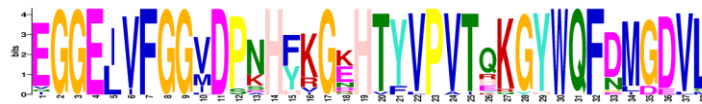
| Nombre | comiezo | p-value | Motivos |
|----------|---------|----------|--|
| Cm-AP | 85 | 9.46e-63 | EDIDIVSLKN YMDAQYFGEIGVGTTPPQKFTVIFD TDGSS NLWVPSSKCYFSVACYF#SKYK SSSSSTYKKN |
| Bd-AP | 76 | 1.91e-62 | DEGDIVALQN YMNAQYFGEIGVGTTPPQKFTVIFD TDGSS NLWVPSAKCYFSIACYF#SRYK AGQSSTYKKN |
| Vv-AP | 85 | 6.30e-62 | KHTYVVALHN YMDAQYFGEISIGTTPPQKFTVIFD TDGSS NLWVPSSKCYFSVACYF#SKYK SSQSSTYKKN |
| Cccyp-AP | 80 | 6.30e-62 | SGSDIIALKN YMDAQYFGEIGIGSPPQKFTVIFD TDGSS NLWVPSAKCYFSVACLF#SKYK SSHSSTYKKN |
| Cv-AP | 80 | 1.40e-61 | SDSDIIALKN YMDAQYFGEIGIGAPPQKFTVIFD TDGSS NLWVPSAKCYFSVACLF#SKYK SSHSSTYKKN |
| Am-AP | 80 | 1.40e-61 | SDSDIIALKN YMDAQYFGEIGIGAPPQKFTVIFD TDGSS NLWVPSAKCYFSVACLF#SKYK SSHSSTYKKN |
| Na1-AP | 85 | 1.40e-61 | DDADIISLKN YMNAQYFGEIGIGTTPPQKFTLIFD TDGSS NLWVPSAKCYFSIACYF#SKYK SSLSSSYTKN |
| Hv-AP | 49 | 1.92e-61 | EEGDIVALKN YMNAQYFGEIGVGTTPPQKFTVIFD TDGSS NLWVPSAKCYFSIACYL#SRYK AGASSTYKKN |
| Na2-AP | 85 | 4.15e-61 | DDADIISLKN YMDAQYFGEIGIGSPPQKFTVIFD TDGSS NLWVPSAKCYFSIACYL#PKYK SFKSSTYAKN |
| Ta-AP | 78 | 2.42e-60 | EEGDIVSLKN YMNAQYFGEIGVGTTPPQKFTVIFD TDGSS NLWVPSAKCYFSIACYL#ARYK AGASSTYKKN |
| Cc-AP | 80 | 4.26e-60 | SDSDIIELKN YMDAQYFGEIGIGSPAQKFTVIFD TDGSS NLWVPSAKCYFSVACLF#SKYK SSHSSTYKKN |
| Hhc-AP | 83 | 9.76e-60 | DSTDIISLKN YMNAQYFGEIGVGTTPPQKFTVIFD TDGSS NLWVPSAKCYFSIACLL#TKYK SGRSSTYHKN |
| Sm1-AP | 80 | 1.47e-59 | SDSDIIALKN YMDAQYFGEIGIG EP AQKFTVIFD TDGSS NLWVPSAKCYFSVACLF#SKYK SSHSSTYKKN |
| Zm-AP | 88 | 2.20e-59 | DDSDVIALNX YXNAQYFGEIGVGTTPPQKFTVIFD TDGSS NLWVPSSKCYFSIACYF#SRYK SGQSSTYKKN |
| Rc-AP | 79 | 6.34e-59 | DDVDIVELKN YLDAQYFGEIAIGTTPPQ TF VIFD TDGSS NLWIPSSKCYFSVACYF#SKYK ASESSTYQKN |
| AtA1-AP | 77 | 7.95e-58 | GDADVVLKN YLDAQYFGEIAIGTTPPQKFTVV FD TDGSSNLWVPSSKCYFSLACLL#PKYK SRSSTYEKN |
| Nt-AP | 80 | 2.68e-57 | SDTDIIALKN YLDAQYFGEICIGSPPQKFTVIFD TDGSS NLWVP SARC YFSLACYL#PKYK SSHSSTYKKN |
| Os-AP | 93 | 6.93e-57 | AEGDIVALKN YLNAQYFGEIAIGTTPPQ MF VIFD TDGSS NLWVPSSK CL SIACYF#SRYK AGQSSTYKKN |
| AtA2-AP | 84 | 8.76e-57 | GDADIVPLKN YLDAQYFGEIAIGTTPPQKFTVIFD TDGSS NLWVP SGKCF SLSCYF#AKYK SRSSTYKKS |
| Ha-AP | 80 | 5.54e-56 | SEGDVIVLKN YMDAQYFGEIGIGTTPPQKFTVV FD TDGANLWVPSSK CF LSVACLF#QKYK ASRSSTYKKN |
| AtA3-AP | 82 | 8.71e-56 | ENADMVPLKN YLDAQYFGEITIGTTPPQKFTVIFD TDGSS NLWIP STKCY LSVACYF#SKYK ASQSSSYRKN |
| Fe-AP | 77 | 1.90e-55 | DDTGIISLKN YMDAQYFGEIAIGT PSQT FVIFD TDGSS NLWVP SGKCY LSIACFF#SKYK SSKSSTYVKN |
| Sm2-AP | 80 | 3.65e-48 | SDTGIIVLTN ERD TTYFGEIGIGT PLQ NFTVIFD TDGSS DLWVPSSK CV SSVSCWV#SRYK SSDSSTYKEN |
| CcA-AP | 80 | 1.32e-46 | SGSAVVALTN DRD TSYFGEIGIGTTPPQKFTVIFD TDGSS VLWVPSSK CIN SKACRA#SMYK SSDSSTYKEN |
| CcB-AP | 80 | 3.31e-45 | SGSGIVALTN DRD TAYYFGEIGIGTTPPQ N FAVIFD TDGSS DLWVP STKCD TSLACVI#PRYK SGDSSTYKGN |

MOTIVO 2



| Nombre | comiezo | p-value | Motivos |
|-----------|---------|----------|---|
| Am-AP | 379 | 2.56e-65 | KSNNGKSSGGV HDEMCTFC EMAVVWMQ NI KRN ETED NIINYVNE LC DRL PS PMGESAVDC NSLSSMPNIA |
| Cc.cyp-AP | 379 | 2.56e-65 | ENNGKSSSGV HDEMCTFC EMAVVWMQ NI KRN ETED NIINYVNE LC DRL PS PMGESAVDC NSLSSMPNIA |
| Cc-AP | 379 | 3.36e-65 | KNNGKSSGGV HDEMCTFC EMAVVWMQ NI KRN QTED NIINYVNE LC DRL PS PMGESAVDC NDLSSMPNIA |
| Cv-AP | 379 | 9.45e-64 | KNNGKSSGGA NDEMCTFC EMAVVWMQ NI KRN ETED NIINYVNE LC DRL PS PMGESAVDC NSLSSMPNIA |
| Na1-AP | 384 | 1.06e-58 | MNVDGSSSGL QDATCTAC EMTVVWMQ NL KQ NTQ ERILNYVNE LC NRL PS PMGESAVDC SSLSSMPGVS |
| Cm-AP | 383 | 1.39e-58 | DNTHKASNGL RDAMCSTC EMTVVWMQ NL KQ NTQ DRILTYVNE LC DRL PS PMGESAVDC GSLSSLPNVS |
| Hhc-AP | 380 | 1.39e-58 | VNGNVDKSVA SDAMCSAC EMAVVWMQ NI K NKTQ LILNYIN Q LCER LP SPMGESAVDC SVLSTMP SIS |
| Na2-AP | 384 | 1.59e-58 | KNPGNSSDGL QDAMCTAC EMAVVWMQ NL RQ NRTE EQILNYVNE LC NRL PS PMGESAVDC GSLSSMPNVS |
| Sm1-AP | 376 | 3.57e-58 | VVDKNNKSS GNEMCTFC EMAVVWMQ NI KRN QTD NIINYV SE LC DRL PS PMGES AVDC NSLSSMPNIS |
| Ta-AP | 378 | 1.03e-57 | DEPVKSNGLH TDPMCSAC EMAVVWMQ NLA Q NKTQ DLILDYV NQ LCNRL PS PMGESAVDC ASLGSMPIE |
| Hv-AP | 348 | 1.17e-57 | DEPVKSNGLR ADPMCSAC EMAVVWMQ NLA Q NKTQ DLILDYV NQ LCNRL PS PMGESAVDC GSLGSMPIE |
| Fe-AP | 376 | 1.52e-57 | KNVDKSSGNL KDATCSAC EMAVVWMQ NL KQ NTQ EDRILDY ANQ LCER LP SPMGESAVDC GSLSTLPTVS |
| Zm-AP | 388 | 2.45e-56 | KAGKSNGLK SDPMCNAC EMAVVWMQ NLA Q NKTQ ELILTYIN Q LCER LP SPMGESAVDC ASLGSMPIA |
| Bd-AP | 375 | 8.28e-56 | DEAEKSNGLH SDPMCNAC EMAVVWMQ NLS Q NKTQ DVILNYIN Q LC DRL PS PMGES SVDC GSLASMPEIE |
| AtA1-AP | 376 | 7.02e-55 | KENAKLSNGV GDAACSAC EMAVV WI Q S QLRQ NMTQ ERILNYVNE LC ER LP SPMGESAVDC AQLSTMP TVS |
| Ha-AP | 379 | 4.44e-54 | KNNGKSSAGL NDGICAF CEMAVVWMQ S QL KRNQ TE DS IINYVNE LC DRL PS PMGESAVDC QTLSNMPNIA |
| Nt-AP | 378 | 1.72e-53 | DKHNGASNGL GDEMCRV CEMAV IWM Q N MRR NETA DSIYDYV NQ LC DRL PS PMGES AVDC SSLASMPNVS |
| Vv-AP | 384 | 6.30e-51 | ENNSKASDGL HDTMCSAC SMAVV WI Q N KL GQ NET ID RILKYVNE LC DRL PS PMGESAVDC GSLSSMPNVS |
| Sm2-AP | 379 | 2.16e-50 | MKNDKSSGGV HDEMCTFC EMAVVWMQ KE IKR NEI EDNIINYV TE LC DRL PS SSGES VVDC NSLSSMPNTA |
| Rc-AP | 379 | 1.21e-49 | SKE ^T ASDGLQ QDSAC^T FC EM IVVWMQ NRL R LN ET VD QILNYV NK LC DRL PS PNGES AVDC SSLSSMPIVS |
| AtA2-AP | 383 | 2.96e-49 | KENTRSSSGL RDAGCFAC EMAVV WI Q S QLRQ NMTQ ERIVNYINE IC ER MP SPNGESAVDC SQLSKMPTVS |
| Cc.A-AP | 375 | 1.75e-47 | KNNDKSSGGI HDEMCTFC EMAVVWMQ NE IK Q SE TE DNIIINYAN EL CE FL ST SSE ELQVDC NTLSSMPNVS |
| Os-AP | 392 | 5.05e-46 | DEVGKSSGPF SSAMCNAC ETAVV WM T Q LAQ NTQ DLVL QY ID Q LC DRL PS PMGES SVDC SSLASMPDIA |
| Cc.B-AP | 376 | 6.74e-43 | VVDKNNDKSS GGEIC^T FC EM ALVR M Q NE IKR NETED NIINHVN EV C D QL PT SS AE SVDC NGISSMPNIA |
| AtA3-AP | 374 | 1.71e-40 | VVDGTSGLL NQAMCSAC EMAAV W ME SE L T Q N Q TQ ERILAY AE LC DRL PS PMGES AVDC GRVSSMPIVT |

MOTIVO3



| Nombre | comiezo | p-value | Motivos |
|-----------|---------|----------|---|
| Cm-AP | 240 | 8.54e-49 | FWFNRTDEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL IDGQTTGFCV |
| AtA1-AP | 232 | 3.31e-48 | FWLNRNADEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL IGGAPTGFCE |
| Ha-AP | 235 | 3.31e-48 | FWFNRTGEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL IGDKTTGFCS |
| Nt-AP | 235 | 1.35e-47 | FWFNRAQEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL VGGETTGFCS |
| Cc-AP | 235 | 4.31e-47 | FWFNNADEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL IEDKTTGFCA |
| Am-AP | 235 | 5.83e-47 | FWFNNADEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL IEDKTTGFCS |
| Cc.cyp-AP | 235 | 5.83e-47 | FWFNNADEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL IEDKTTGFCS |
| Na2-AP | 240 | 7.95e-47 | FWLNRKSEE EGGELVFGGVDNHFYKGR...TYVPVTRKGYWQFDMDDVL VGGETTGYCS |
| Sm1-AP | 235 | 2.34e-46 | FWFNNADEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL IEDKTTGFCS |
| Cv-AP | 235 | 5.70e-46 | FWFNANEE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL IEDKTTGFCS |
| Rc-AP | 234 | 7.48e-46 | FWLNRNVQAE EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGEVL IGNEITGLCA |
| Na1-AP | 240 | 8.57e-46 | FWLNRNATEE EGGELVFGGVDNHFYKGR...TFVPVTQKGYWQFDMDDVL VGGETTGYCS |
| Bd-AP | 231 | 1.11e-45 | FWFNRHAGEG EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL VGGKSTGFCA |
| Hv-AP | 204 | 7.29e-45 | FWLNRHVDEG EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL VGGKSTGFCA |
| Hhc-AP | 238 | 8.24e-45 | FWLNRKSEDG EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL VGGQSTGFCE |
| Cc.B-AP | 235 | 1.18e-44 | FWLNRNVDEE EGGELVFGGVDNHFYKGR...TYVPVTRKGYWQFEMGDVL IGDKSSGFCA |
| Sm2-AP | 235 | 5.46e-44 | FWLNRNVNEE EGGELVFGGVDNHFYKGR...TYVPVIQKGYWQFDMGDVL IGDKSTGFCA |
| Zm-AP | 243 | 1.21e-43 | FWFNRHAGEG EGGELVFGGVDNHFYKGR...TYVPVTQKGYWQFDMGDVL VDGKSTGFCA |
| Ta-AP | 234 | 1.69e-43 | WLNRRHADDEG EGGELVFGGVDNHFYKGR...TYVPATQKGYWQFDMGDVL VGGKSTGFCA |
| Os-AP | 248 | 2.95e-43 | FWFNRADEE QGGELVFGGVDNHFYKGR...TYVPVTRKGYWQFDMGDVL IGGNSTGFCA |
| AtA2-AP | 239 | 4.11e-43 | FWLNRDPKSE EGGELVFGGVDNHFYKGR...TFVPVTQKGYWQFDMGEVL IAGESTGYCG |
| AtA3-AP | 237 | 5.71e-43 | FWLNRNPKDP EGGELVFGGVDNHFYKGR...TFVPVTQKGYWQFDMGDVL IAGKPTGYCA |
| Fe-AP | 232 | 2.51e-39 | FWLNRNADEE IGGELVFGGVDNHFYKGR...TYLPVTQKGYWQFDLDDVL VGGESTGFCS |
| Vv-AP | 240 | 1.94e-38 | FWFNRSNEE VGGELVFGGVDNHFYKGR...TYVPVTKKGYWQFDLGDVM IGGKTTGFCA |
| Cc.A-AP | 231 | 1.85e-37 | FWLNRNVDEE EGGELVFGGVDNHFYKGR...TYVPVTYQYWQFQIGDVL IGDKSTGFCA |

MOTIVO4



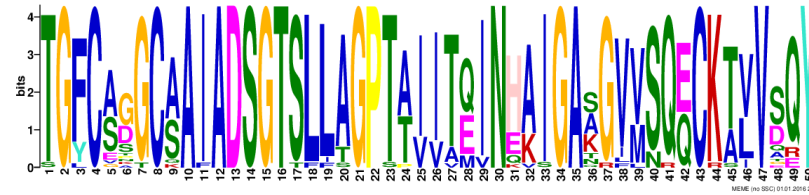
| Nombre | comiezo | p-value | Motivos |
|-----------|---------|----------|---|
| Am-AP | 450 | 5.46e-64 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDVPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Cv-AP | 450 | 4.45e-63 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDVAPPRGPLWILGDVFMGRYHTVFDYG KSRVGF AEAA |
| Cc.cyp-AP | 450 | 4.45e-63 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDVAPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Cc-AP | 450 | 1.28e-62 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDVAPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Bd-AP | 446 | 2.69e-62 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDVAPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Bd-AP | 446 | 2.69e-62 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDIPPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AKAA |
| Cm-AP | 454 | 3.23e-62 | TIGGRVFDLS PEQYVLKIGEGEAAQCISGFTALDVPPRGPLWILGDVFMGRYHTVFDYG NQRVGF AEAA |
| Hv-AP | 419 | 2.03e-61 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDIPPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Ta-AP | 449 | 2.03e-61 | TISGKVFELC PEQYILKIGEGEAAQCISGFTAMDIPPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Sm1-AP | 447 | 3.38e-60 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDVAPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Os-AP | 463 | 1.43e-59 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDIPPPRGPLWILGDVFMGRYHTVFDYG NLKVGF AEAA |
| Sm2-AP | 450 | 1.08e-58 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDVAPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Zm-AP | 459 | 1.76e-58 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTAMDIPPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAA |
| Nt-AP | 449 | 2.85e-58 | TVGNQTFGLT PQQYVLQVGEQAAQCISGFTALDVPPRGPLWILGDVFMGRYHTVFDYG NSRVGF AEAA |
| AtA2-AP | 454 | 9.30e-58 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTALDVPPRGPLWILGDVFMGRYHTVFDYG NEQVGF AEAV |
| Vv-AP | 455 | 1.48e-57 | TIGGKVFELC PKQYILKIGEGEAAQCISGFTALDVPPRGPLWILGDVFMGRYHTVFDYG NMKVGF AEAA |
| Fe-AP | 447 | 1.48e-57 | TLGGKVFELC PEQYILKIGEGEAAQCISGFTALDVPPRGPLWILGDVFMGRYHTVFDYG NMQVGF AEAA |
| Ha-AP | 450 | 2.63e-57 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTALDVAPPRGPLWILGDVFMGRYHTVFDYG KSRVGF AEAA |
| Na2-AP | 455 | 3.30e-57 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTALDVAPPRGPLWILGDVFMGRYHTVFDYG NLSVGF AEAA |
| AtA1-AP | 447 | 1.14e-56 | TIGGKVFELC PEQYILKIGEGEAAQCISGFTALDVAPPRGPLWILGDVFMGRYHTVFDYG NEQVGF AEAA |
| Na1-AP | 455 | 3.81e-56 | TVGNQTFGLT PQQYVLQVGEQAAQCISGFTALDVPPRGPLWILGDVFMGRYHTVFDYG NMRVGF AEAV |
| AtA3-AP | 445 | 2.84e-53 | SIGGRSFDLT PQQYILKIGEGEAAQCISGFTALDVAPPRGPLWILGDVFMGRYHTVFDYG KGRVGF AKAA |
| Cc.B-AP | 447 | 1.36e-52 | TIGSKLFEVT PEQYILKIGEGEAAQCISGFTALDVAPPRGPLWILGDVFMGRYHTVFDYG KLRVGF AEAV |

MOTIVO 5



| Nombre | comiezo | p-value | Motivos |
|-----------|---------|----------|---|
| Cv-AP | 182 | 1.35e-61 | DFIEATKEPG ITFLAAKFDGILGLGFQEISVGKSPVWYNMVNQGLVQEPVFSFWFNRNA NEEEGGELVF |
| Sm1-AP | 182 | 4.40e-61 | DFIEATKEPG ITFLAAKFDGILGLGFQEISVGKSPVWYNMVNQGLVQEPVFSFWFNRNA DEEEGGELVF |
| Cc-AP | 182 | 3.02e-60 | DFIEATKEPG VTFLAAKFDGILGLGFQEISVGKSPVWYNMVNQGLVQEPVFSFWFNRNA DEEEGGELVF |
| Nt-AP | 182 | 1.40e-59 | DFIEATREPG ITFLAAKFDGILGLGFQEISVGKSPVWYNMVNQGLVKKPVFSFWFNRNA QEEEGGELVF |
| Na1-AP | 187 | 1.40e-59 | DFIEATREPS ITFVAAKFDGILGLGFQEISVGNAPVWYNMVKQGLVNEPVFSFWLNRNA TEEEGGEIVF |
| Cc.cyp-AP | 182 | 1.92e-59 | DFIEATKEPG ITFLAAKFDGILGLGFQEISVGKSPVWYNMVNQGLVQEPVFSFWFNRNA DEEEGGELVF |
| Am-AP | 182 | 3.08e-59 | DFIEATKEPG VTFLAAKFDGILGLGFQEISVGKSPVWYNMVNQGLVPEPVFSFWFNRNA DEEEGGELVF |
| Cm-AP | 187 | 4.19e-59 | EFIEATREPS ITFLVAKFDGILGLGFKEISVGNAPVWYNMVKQGLVKEPVFSFWFNRT DEEEGGEIVF |
| Fe-AP | 179 | 5.67e-59 | EFIEATREPS LTFVAAKFDGILGLGFQEISVGKAVPVWYNMVNQGLVNEPVFSFWLNRNA DEEIGGEIVF |
| Ta-AP | 180 | 2.15e-57 | EFIEATKEPG VTFLVAKFDGILGLGFKEISVGKAVPVWYNMVEQGLISDPVFSFWLNRA DDEGEGGEII |
| AtA1-AP | 179 | 1.02e-56 | EFIEATKEPG ITFVVAKFDGILGLGFQEISVGKAAVPWYNMLKQGLIKEPVFSFWLNRNA DEEEGGELVF |
| Ha-AP | 182 | 1.68e-56 | DFIEATREPG ITFLAAKFDGILGLGYQDISVGKAVPVWYNMVNQGLVQEPVFSFWFNRET GEEEGGELVF |
| Zm-AP | 190 | 5.10e-56 | EFIEATKEPG LTFMVAKFDGILGLGFQEISVGNATPVWYNMVKQGLISDPVFSFWFNRA GEGEGGEIVF |
| Os-AP | 195 | 8.27e-56 | DFIEATREPS ITFMVAKFDGILGLGFKEISVGNAPVWYNMVRQGLVVDPEVFSFWFNRA DEGQGGGEIVF |
| Hv-AP | 151 | 1.50e-55 | EFIEATKEPG ITFLVAKFDGILGLGFKEISVGKAVPVWYKMIQGLVSDPEVFSFWLNRAV DEGEGGEIIF |
| Bd-AP | 178 | 1.69e-55 | EFIEATKEPG VTFMVAKFDGILGLGFQEISVGKAVPVWYKMIQGLISDPVFSFWFNRA GEGEGGEIVF |
| Rc-AP | 181 | 5.45e-55 | DFIEATKEPG VTFLAAKFDGILGLGFQEISVGKAVPVWYNMVEGLVKEQVFSFWLNRAV QAEEGGEIVF |
| Vv-AP | 187 | 3.00e-54 | EFIEATKEPS ITFALAKFDGILGLGFQEISVGNAPVWYNMINQELIKEPIFSFWFNRS NEEVGGEIVF |
| Na2-AP | 187 | 9.08e-54 | DFIEATKEPS ITFVAAKFDGILGLGFQEISVGDAVPWYNMIDQGLVNEPVFSFWLNRRKS EEEEGGEIVF |
| AtA3-AP | 184 | 1.57e-53 | EFIEATSEPG ITFLLAKFDGILGLGFKEISVGNSTPVWYNMVEKGLVKEPIFSFWLNRRNP KDPEGGEIVF |
| Sm2-AP | 182 | 8.69e-53 | DFIEATKEAD NVFLRA FDGILGLGFQEISVRKAVPVWYNMVDQGLVQEPVFSFWLNRAV NEEEGGELVF |
| Hhc-AP | 185 | 1.07e-52 | EFIEATKEPG VTFLVAKFDGILGLGFKEISVGGAVPLWYNMVEQGLVKEAVFSFWLNRRKS EDGEGGEIVF |
| AtA2-AP | 186 | 7.02e-51 | EFIEATSEPG LTFVAKFDGILGLGFQEIAVGNATPVWYNMLKQGLIKRPFVFSFWLNRRDP KSEEGGEIVF |
| Cc.B-AP | 182 | 1.42e-49 | DFIETTEEDD TVFLKSEFDGILGLGFQEISAGKAVPVWYNMVNQGLVBEAVFSFWLNRAV DEEEGGELVF |

MOTIVO 6



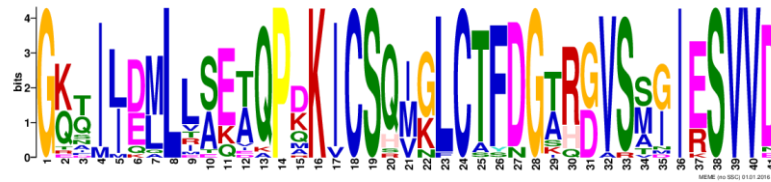
| Nombre | comiezo | p-value | Motivos |
|----------|---------|----------|---|
| Cc-AP | 278 | 4.48e-60 | MGDVLIEDKT TGFCADGCAA IADSGT SLLAGPTAIITQIN HAI GAKGVMSQQCKTLVDQY GKTIIEMLLS |
| Sm1-AP | 278 | 3.50e-59 | MGDVLIEDKT TGFCSDGCAA IADSGT SLLAGPTAIITQIN HAI GAKGVMSQQCKTLVDQY GKSIIEMLLS |
| Cccyp-AP | 278 | 7.13e-59 | MGDVLIEDKT TGFCSDGCAA IADSGT SLLAGPTAIITQIN HAI GAKGVMSQQCKTLVDSQY GKTMIEMLLS |
| Ha-AP | 278 | 1.10e-58 | MGDVLIGDKT TGFCSDGCAA IADSGT SLLAGPTAIITQIN HAI GAGVMSQQCKTLVDQY GKTIIEMLLS |
| Am-AP | 278 | 5.14e-58 | MGDVLIEDKT TGFCSDGCAA IADSGT SLLAGPTAVITQIN HAI GAKGVMSQQCKTLVDSQY GKTIIEMLLS |
| Bd-AP | 274 | 6.80e-57 | MGDVLVGGKS TGFCAGGCAA IADSGT SLLAGPTAIITQIN EA GAGVVSQECKTVVDSQY GQQILDLLLA |
| Zm-AP | 286 | 6.80e-57 | MGDVLVDGKS TGFCAGGCAA IADSGT SLLAGPTAIITQIN EA NEKIGAGVVSQECKTVVDSQY GQQILDLLLA |
| Cv-AP | 278 | 3.25e-56 | MGDVLIEDKT TGFCSDGCAA IADSGT SLLAGPTAIITQIN HAI ASGAKGVMSQQCKTLVDSQY GKSIIEMLLS |
| AtA1-AP | 275 | 4.89e-56 | MGDVLIGGAP TGFCSESGCA IAIDSGT SLLAGPTTIITMIN HAI GAGVVSQQCKTVVDQY GQTILDLLLS |
| Hv-AP | 247 | 6.39e-56 | MGDVLVGGKS TGFCAGGCAA IADSGT SLLAGPTAIITQIN EA NEKIGAGVVSQECKTIVDSQY GQQILDLLLA |
| Ta-AP | 277 | 6.39e-56 | MGDVLVGGKS TGFCAGGCAA IADSGT SLLAGPTAIITQIN EA NEKIGAGVVSQECKTIVDSQY GQQILDLLLA |
| Fe-AP | 275 | 7.80e-54 | LDDVLVGGES TGFCSDGCAA IAIDSGT SLLAGPTVVAQIN HAI GAGVVSQECKTVVDSQY GKQILDLLVS |
| Os-AP | 291 | 7.80e-54 | MGDVLIGGNS TGFCAGGCAA IAIDSGT SLLAGPTAIITQIN EA NEKIGAGVVSQECKAVVDSQY GQQILDQLRA |
| Hhc-AP | 281 | 1.10e-53 | MGDVLVGGQS TGFCSEGGCAA IAIDSGT SLLAGPTTVITEIN EA NEKIGAGVVSQECKAVVQY GQQILDMLIA |
| Na1-AP | 283 | 1.55e-53 | MDDVLVGGET TGYCSGGCSA IAIDSGT SLLAGPTTIVAQIN HAI GAGVVSQECKAVVAQY GTAILDMLIS |
| Na2-AP | 283 | 8.19e-53 | MDDVLVGGET TGYCSGGCSA IAIDSGT SLLAGPTTIVQIN HAI GAGVVSQECKAVVDSQY GKAILDALVA |
| Rc-AP | 277 | 1.02e-52 | MGEVLIGNEI TGLCADGCKA IAIDSGT SLLAGPTTVITQIN HAI GAGVVSQECKTVVEQY GKFILEMLTA |
| Vv-AP | 283 | 5.04e-52 | LGDVMIGGKT TGFCASGCSA IAIDSGT SLLAGPTTITEVNV HAI GAGVVSQECKRAVVQY GQIIDMLLT |
| Nt-AP | 278 | 9.43e-52 | MGDVLVGGET TGFCSDGCAA IAIDSGT SLLAGPTTITQIN HAI GAGVVSQECKSLVTEY GKTILDLLS |
| CcA-AP | 274 | 1.31e-50 | IGDVLIGDKS TGFCAPGCA IAIDSGT SLLAGPTAIITQIN HAI GAGVVSQECKTVVSR GRDIIEMLR |
| AtA2-AP | 282 | 3.16e-50 | MGEVLIAGES TGYCSGGCSA IAIDSGT SLLAGPTAVVAMINKA IA GAGVVSQQCKTVVDQY GQTILDLLA |
| Sm2-AP | 278 | 3.84e-50 | MGDVLIGDKS TGFCANGCAA IAIDSGT LLTGTPTAIITQIN QA IGANRVMNQCKTLVSR GKNIMQMLMS |
| Cm-AP | 283 | 6.86e-50 | MGDVLIDGQT TGFCVTTCSA IAIDSGT SLLAGPTTITEVNV HAI GAGVVSQECKAVVAEY GETIIRKMLE |
| CcB-AP | 278 | 2.36e-49 | MGDVLIGDKS SGFCAGGCAA IAIDSGT SLLAGPTAIITQIN QA IGAGVVSQECKTVLQY GKNMIQMLTS |
| AtA3-AP | 280 | 3.44e-49 | MGDLQIAGKP TGYCAKCSA IAIDSGT SLLAGPTVITMIN HAI GAGVVSQECKAVVDQY GKTMLNSLLA |

MOTIVO 7



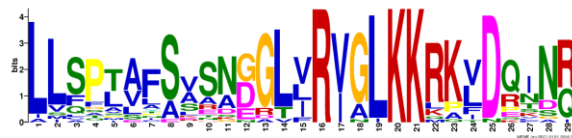
| Nombre | comiezo | p-value | Motivos |
|-----------|---------|----------|--|
| Cv-AP | 131 | 4.80e-57 | ACLFHSHKYKS SHSSTYKKN GTSAAIQYGTGSISGFVSQDSVKLGDLVVKEQDFIEATKEP GITFLAAKFD |
| Sm1-AP | 131 | 4.80e-57 | ACLFHSHKYKS SHSSTYKKN GTSAAIQYGTGSISGFVSQDSVKLGDLVVKEQDFIEATKEP GITFLAAKFD |
| Am-AP | 131 | 4.80e-57 | ACLFHSHKYKS SHSSTYKKN GTSAAIQYGTGSISGFVSQDSVKLGDLVVKEQDFIEATKEP GVTFLAAKFD |
| Cc.cyp-AP | 131 | 4.80e-57 | ACLFHSHKYKS SHSSTYKKN GTSAAIQYGTGSISGFVSQDSVKLGDLVVKEQDFIEATKEP GITFLAAKFD |
| Cc-AP | 131 | 4.80e-57 | ACLFHSHKYKS SHSSTYKKN GTSAAIQYGTGSISGFVSQDSVKLGDLVVKEQDFIEATKEP GVTFLAAKFD |
| Cm-AP | 136 | 4.04e-56 | ACYFHSHKYKS SSSSTYKKN GKPAIHYGTGAISGYFSQDHSVKGDLVVKNEQFIEATREP SITFLVAKFD |
| Bd-AP | 127 | 2.40e-55 | ACYFHSHRYKA QSSSTYKKN GKPAAIQYGTGSISGYFSQDHSVTVGDLVVKDQEFIEATKEP GVTFMVAKFD |
| Hv-AP | 100 | 1.11e-54 | ACYLHSHRYKA GASSTYKKN GKPAAIQYGTGSISGYFSQDHSVTVGDLVVKDQEFIEATKEP GITFLVAKFD |
| Ta-AP | 129 | 1.11e-54 | ACYLHARYKA GASSTYKKN GKPAAIQYGTGSISGYFSQDHSVTVGDLVVKDQEFIEATKEP GVTFLVAKFD |
| Nt-AP | 131 | 1.26e-54 | ACYLHPKYKS SHSSTYKKN GTSAAIRYGTGSISGYFSNDNVKVGDLVVKDQDFIEATREP GITFLAAKFD |
| Ha-AP | 131 | 1.22e-53 | ACLFHQKYKA SRSSTYKKN GTAAAIQYGTGAISGVFSRDSVKLGDLVVKEQDFIEATREP GITFLAAKFD |
| Os-AP | 144 | 4.27e-53 | ACYFHSHRYKA QSSSTYKKN GKPAIHYGTGAISGYFSQDHSVKGDLVAVKNEQFIEATREP SITFMVAKFD |
| Fe-AP | 128 | 1.15e-52 | ACFFHSHKYKS SKSSTYVKN GKSAIHYGTGAISGYFSQDNDVKVGDLVVENQEFIEATREP SLTFVAAKFD |
| Na2-AP | 136 | 1.60e-52 | ACYLHPKYKS FKSSTYAKN GKSAIHYGTGAISGFFSQDHSVKGDLVVENQDFIEATKEP SITFVAAKFD |
| AtA1-AP | 128 | 1.98e-52 | ACLLHPKYKS SRSSTYEKN GKAAAIHYGTGAIAGFFSNDDAVTVGDLVVKDQEFIEATKEP GITFVAKFD |
| Zm-AP | 139 | 2.46e-52 | ACYFHSHRYKS QSSSTYKKN GKPAAIQYGTGAIAGFFSQDHSVKGDLVNDQEFIEATKEP GLTFMVAKFD |
| Rc-AP | 130 | 1.19e-51 | ACYFHSHRYKA SESSTYQKN GTSAAIRYGTGSISGFFSQDNDVKVGDLVIRNDFIEATKEP GVTFLAAKFD |
| Vv-AP | 136 | 1.46e-51 | ACYFHSHRYKS QSSSTYKKN GTSAIHYGTGAISGFFSKDDVKVGDLVINQEFIEATKEP SITFALAKFD |
| Hhc-AP | 134 | 2.20e-51 | ACLLHTKYKS GRSSTYKKN GKPAIHYGTGAIAGYFSQDHSVKGDLVVKQEFIEATKEP GVTFLVAKFD |
| Na1-AP | 136 | 2.70e-51 | ACYFHSHRYKS SLSSSYTKN GKSAIHYGTGAISGFFSQDHSVKGDLVVENQDFIEATREP SITFVAAKFD |
| Sm2-AP | 131 | 2.64e-50 | SCVWHSRYKS SDSSTYKEN GTSGAIQYGTGSISGYFSQDSVEVGDVVKEQDFIEATKEA DNVFLRAHFD |
| AtA3-AP | 133 | 1.78e-49 | ACYFHSHRYKA QSSSYRKN GKPAIHYGTGAISGYFSNDNVKVGDLVVKQEFIEATSEP GITFLLAKFD |
| AtA2-AP | 135 | 2.13e-48 | SCYFHAKYKS SRSSTYKKS GKRAAIHYSGSISGFFSYDAVTVGDLVVKDQEFIEATSEP GLTFVAKFD |
| Cc.A-AP | 131 | 2.31e-46 | ACRAHSMYES SDSSTYKEN GTFGAIYGTGSITGFFSQDSVTIGDLVVKQDFIEATDEA DNVFLHRLFD |
| Cc.B-AP | 131 | 7.82e-45 | ACVIHPRYDS GDSSTYKCN GTTASIQYGTGAIVGFYSQDSVEVGDVVKEQDFIEATTEED DTVFLKSEFD |

MOTIVO 8



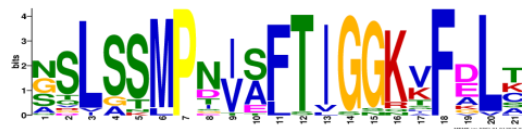
| Nombre | comiezo | p-value | Motivos |
|-----------|---------|----------|---|
| Am-AP | 328 | 2.22e-47 | QQCKTLVSQY GKTIEMLLSEAOQPKDKICSQMKLCTFDGARDVSSIIESVVD KSNKGSSGGV |
| Cc-AP | 328 | 2.22e-47 | QQCKTLVDQY GKTIEMLLSEAOQPKDKICSQMKLCTFDGARDVSSIIESVVD KNNKGSSGGV |
| AtA1-AP | 325 | 8.33e-47 | QQCKTVVDQY GQTILDLLLEETQPKKICSQIGLCTFDGTRGVSMGIESVVD KENAKLSNGV |
| Cv-AP | 328 | 5.36e-46 | QQCKTLVSQY GKSIIEMLLSEAOQPKDKICSQMKLCTFDGARDVSSIIESVVD KNNKGSSGGA |
| Hv-AP | 297 | 1.54e-44 | QECKTIVSQY GQQILDLLLEETQPKKICSQVGLCTFDGTRGVSAIRSVVD DEPVKSNGLR |
| Bd-AP | 324 | 1.54e-44 | QECKTVVSQY GQQILDLLLEETQPKKICSQVGLCTFDGTRGVSAIRSVVD DEAEKSNGLH |
| Sm1-AP | 328 | 3.00e-44 | QQCKTLVDQY GKSIIEMLLSEAOQPKDKICSQMKLCTFNGARDVSSIIESVVD KNNKGSSGNE |
| Cc.cyp-AP | 328 | 3.00e-44 | QQCKTLVSQY GKTIEMLLSEAOQPKDKICSQMKLCTFDGARDASSIIESVVD ENNGKSSSGV |
| Ha-AP | 328 | 9.57e-44 | QQCKTLVDQY GKTIEMLLSEAOQPKDKICSRMNLCTFDGSRDVSSIIESVVD KNNKGSSAGL |
| Fe-AP | 325 | 1.23e-43 | QECKTVVSQY GKQILDLLVSQTQPKKICSQIGLCTFDGTRGVSMGIESVVD KNVDKSSGNL |
| Ta-AP | 327 | 3.30e-43 | QECKTIVSQY GQQILDLLLEETQPKKVCQVGLCTFDGTRGVSAIRSVVD DEPVKSNGLH |
| Hhc-AP | 331 | 4.74e-43 | QECKAVVQY GQQILDMLIAQTQPMKICSQIGLCTFDGTRGVSMGIESVVD GNVDKSVASD |
| Zm-AP | 336 | 6.02e-43 | QECKTVVSQY GQQILDLLLEETQPAKICSQVGLCTFDGTRGVSTGIRSVVD DKAGKSNGGL |
| AtA2-AP | 332 | 1.09e-42 | QQCKTVVDQY GQTILDLLLEETQPKKICSQIGLCAYDGTGVMGIESVVD KENTRSSSGL |
| Sm2-AP | 328 | 9.42e-42 | QQCKTLVSRY GKNIMQMLMSEAOQPKDKICS MNLCTFDGARDVSSIIESVVD MKNDKSSGGV |
| Cc.A-AP | 324 | 3.87e-41 | QQCKTVVSRY GRDIIEMLRSKIQPKDKICS MKLCTFDGARDVSSIIESVVD KNNDKSSGGI |
| Na2-AP | 333 | 2.82e-40 | QECKAVVSQY GKAILDALVAEQPKKICSQIGLCTFDGKRGVSMGIESVVE KNPGNSSDGL |
| Na1-AP | 333 | 4.68e-40 | QECKAVVAQY GTAILDMLISETQPKKICSQIGLCTFDGKRGVSVGIKSVVD MNVDGSSSGL |
| Cm-AP | 333 | 2.54e-39 | QECKAVVAEY GETIIKMLLEKQPMKICSQIGLCTFDGTQGVSMIDIESVVD NTHKASNGLR |
| Cc.B-AP | 328 | 2.80e-39 | QQCKTLVGQY GKNMIQMLTSEVQPKDKICS MKLCTFDGARDVSRMIESVVD KNNDKSSGGE |
| Rc-AP | 327 | 5.52e-39 | QECKTVVEQY GKFILEMLTAQAPQPKKICSQIGFCTFDGTQGVSTNIESVVD KSKETASDGL |
| Vv-AP | 333 | 6.08e-39 | QECKAVVQY GQIIDMLLTKEQPKKICSQIGLCAFNGIRGVSMGIESVVD ENNSKASDGL |
| Nt-AP | 328 | 2.74e-37 | QECKSLVTEY GKTILDLLLESKAAPQKIKCSQIGLCSSDGSRDVSMIIESVVD KHNGASNGLG |
| Os-AP | 341 | 7.98e-37 | QECKAVVSQY GQQILDQLRAETKPAKVCSSVGLCTFDGTGVSAGIRSVVD DEVGKSSGPF |

MOTIVO 9



| Nombre | comiezo | p-value | Motivos |
|-----------|---------|----------|---|
| Sm1-AP | 17 | 8.76e-34 | ASLLALFLFF LLSPTAFSVSNGLRVGLKKRKVDQINQ LSGHGASMEG |
| Am-AP | 17 | 1.70e-33 | ASLLALFLFF LLSPTAFSVSNGLRVGLKKRKVDQINQ LRGHGASMEG |
| Cc.cyp-AP | 17 | 3.24e-33 | ASVLALFLFF LLSPTAFSVSNGLRVGLKKRKVDQINQ LSGHGVSMEA |
| Cc-AP | 17 | 3.98e-33 | ASLLALFLFV LLSPTAFSASNGGLRVGLKKRKVDQINQ LRNHGASMEG |
| Cv-AP | 17 | 5.76e-31 | ASLLALFLFF LLSPTAISVSNGLRVGLKKRKVDQINQ LSGHGASMEG |
| Sm2-AP | 17 | 7.24e-30 | ASLLALFLFF LLSPTAFSASNGGLRVGLKKRKVDRSDQ LHGLGVHTKR |
| Cc.B-AP | 17 | 1.23e-29 | ASLLALFLFF LLSPTAFSVSNGLRVGLKKRKVDRLDQ LRAHGVHMLG |
| Nt-AP | 17 | 1.59e-29 | ACLSALCLLL LLSPMVFSVSNGLRVGIKKRKLQINQ AFGGIDSNGA |
| Cc.A-AP | 17 | 4.99e-29 | ANVLALFLFY LLSPTVFSVSDGLRIRIGLKKRKVDRIDQ LRGRRALMEG |
| Na1-AP | 17 | 2.40e-28 | SVLASLFLLL LLSPLVNSSSNDRLRVGLKKRKLQINR FSSLYGCKGK |
| Cm-AP | 17 | 1.51e-27 | TVVATFFLCF LLFPLVFSASNGGLRVIRIGLKKMKLDKNNR VAAQLESKDG |
| Na2-AP | 17 | 3.01e-26 | TALTTLFLLL LLSPLVTSLSSDRLRVGLKKRKLQINR LSSHYGCKGK |
| Ha-AP | 17 | 7.83e-25 | TSLALFLFF ILSPTAFSSTKGGLRVGLKKRKTNQFNR VSEHGLSMEG |
| AtA2-AP | 17 | 1.35e-23 | AVAFSVFVSF LLFFTAYSKRNDGTFRVGLKKLKLDENNR LATRFGSKQE |
| Vv-AP | 17 | 1.48e-23 | TVAVALFLSI LMFSPFESASDGLRVIRIGLKKRAFQDQNR LAARIESKQG |
| Fe-AP | 15 | 2.11e-23 | SRISLLLLVL LVSPIISLVANNDLRVGLKKRKLDPQNR PASRFGCKKH |
| Rc-AP | 14 | 1.30e-22 | KSSALWISFV LLLPVVFSLHNDALRVGLKKKFDQVNI PAGTVDFKEG |
| Bd-AP | 15 | 2.15e-22 | RVVLVLFQAV LLQALLPASEAEGLVRIALKKRPIQDQNR VATRLSGEEG |
| AtA1-AP | 17 | 4.95e-22 | TVAVSLIVSF LLCFSAFARNNDGTFRVGLKKLKLDKSKNR LAARVESKQE |
| Ta-AP | 15 | 8.87e-21 | GLAVLLAAV LLQQLLLASEAEGLVRIALKKRAIDRNSR VAKSLSDREE |
| Os-AP | 18 | 1.53e-20 | LLLLLLAAV LLQQLLAAAAEGVRIALKKRQVDETGK VGGHLAGEDA |
| AtA3-AP | 19 | 2.42e-20 | LLVFLSCLI LISTASCERNGDGTFIRIGLKKRKLDRSNR LASQLFLKNR |
| Hhc-AP | 17 | 4.79e-20 | GILPALILSM LVFQLALSASEGLVRINLKKKPFDEKSR VSSRLSADED |
| Zm-AP | 22 | 4.21e-17 | AXALLQALAP ASAXAAAAAXAAGGLVRVALKKQFVDQNR VAARLSAEER |

MOTIVO 10



| Nombre | comiezo | p | Motivos |
|-----------|---------|----------|---|
| Cv-AP | 429 | 6.91e-26 | SPMGESAVDC NSLSSMPNIAFTIGGKVFELC PEQYILKIGE |
| Sm1-AP | 426 | 6.91e-26 | SPMGESAVDC NSLSSMPNISFTIGGKVFELC PEQYILKIGD |
| Am-AP | 429 | 6.91e-26 | SPMGESAVDC NSLSSMPNIAFTIGGKVFELC PEQYILKIGE |
| Cc.cyp-AP | 429 | 6.91e-26 | SPMGESAVDC NSLSSMPNIAFTIGGKVFELC PEQYILKIGE |
| Cc-AP | 429 | 9.55e-23 | SPMGESAVDC NDLSSMPNIAFTIGGKVFELC PEQYILKIGE |
| Sm2-AP | 429 | 5.34e-22 | SSSGESVVDC NSLSSMPNTAFTIGGKIFELT PEQYIFKLGE |
| Cc.A-AP | 425 | 5.34e-22 | TSSEELQVDC NTLSSMPNVSFTIGGKKFGLT PEQYILKV GK |
| Vv-AP | 434 | 2.27e-21 | SPMGESAVDC GSLSSMPNVSLTIGGKVF DLS PKQYILKVGE |
| Na2-AP | 434 | 2.27e-21 | SPMGESSVDC GSLSSMPNVSLTIGGKVF DLS PEKYVLKVGE |
| Ha-AP | 429 | 2.03e-20 | SPMGESAVDC QTLSNMPNIAFTIGGKT FDLT PEQYILKVGE |
| Rc-AP | 429 | 2.03e-20 | SPNGESAVDC SSLSSMPIVSFTIGGKAF KLT ADQVKGC |
| Hv-AP | 398 | 2.03e-20 | SPMGESAVDC GSLGSMPIEFTIGGKK FALK PEEYILKVGE |
| Zm-AP | 438 | 2.26e-20 | SPMGESAVDC ASLGSMPDIAFTIGGKK FKLK PEQYILKVGE |
| Bd-AP | 425 | 4.17e-20 | SPMGESSVDC GSLASMPDIEFTIGGKK FALK PEEYILKVGE |
| Na1-AP | 434 | 2.22e-19 | SPMGESAVDC SSLSSMPGVSFTVGGK VF D LL PEQYILQVGE |
| AtA2-AP | 433 | 3.90e-19 | SPNGESAVDC SQLSKMPTVSFTIGGK VF D LA PEEYVLKIGE |
| Hhc-AP | 430 | 8.93e-19 | SPMGESAVDC SVLSTMP SISFTIGGKQ FDLT AEQYVLKVGE |
| AtA1-AP | 426 | 9.77e-19 | SPMGESAVDC AQLSTMP TVSLTIGGK VF D LA PEEYVLKVGE |
| Ta-AP | 428 | 9.77e-19 | SPMGESAVDC ASLGSMPDIEFTISGKK FALK PEEYILKVGE |
| Os-AP | 442 | 1.67e-18 | SPMGESSVDC SSLASMPDIAFTIGGNK FV LK PEQYILKVGE |
| Cm-AP | 433 | 3.10e-18 | SPMGESAVDC GSLSSLPNVSLTIGGR VF D LS PEQYVLKVGE |
| Cc.B-AP | 426 | 4.16e-17 | TSSAESIVDC NGISSMPNIAFTIGSK LF E V T PEQYIYKVGE |
| Nt-AP | 428 | 5.29e-17 | SPMGESAVDC SSLASMPNVSFTVGNQ TFGLT PQQYVLQVGE |
| Fe-AP | 426 | 5.29e-17 | SPMGESAVDC GSLSTLPTVSFTLGGK TF FALA PEQYILQVGE |
| AtA3-AP | 424 | 1.42e-15 | TQNQQSAVDC GRVSSMPIVTF S I GGRS FDLT PQDYIFKIGE |

Figura 1: Motivos conservados 1 a 10 para PAs de plantas (página anterior). Los motivos 1 y 6 muestran las secuencias catalíticas (recuadros rojos) y en recuadros verdes se indican la silpepsina 1 y silpepsina 2. El logo indica el motivo de la proteína. Los logotipos son una herramienta de visualización de motivos. La altura de una letra indica su frecuencia relativa en la posición dada (Eje x) en el motivo

Cada motivo tiene un cierto número de columnas alineadas, indicadas por letras de color en la figura. Las columnas alineadas pueden presentar deleciones. Los residuos grises no son alineados entre sí, dado a que su identidad no es importante y se omiten del logo.

Los colores corresponden al tipo de residuos con mayor prevalencia en dicha posición: los residuos hidrófobos se muestran en azul (ACFILVM), los residuos polares no cargados en verdes (NQST), los residuos ácidos en color magenta (DE) y los residuos con carga positiva en rojo (KR), mientras que en rosa se muestra la histidina, en naranja la glicina, en amarillo la prolina y en turquesa el triptófano.


```

preprosilpepsin1|JX535387| <-----> <----->
MGNSIKASLLALFLFFLLSPTAFSVSNGLLVRVGLKRRKVDQINQLSGHGASMEGKARKD
preprosilpepsin2|JX535358| MGNSIKASLLALFLFFLLSPTAFSA SNGGLIRVGLKRRKVD RSDQLHGLGVHTKRNGRKD
*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****
----->
preprosilpepsin1|JX535387| FGFGSTLRDSDSDIIALKNYMDAQYIGEIGEPAQKFTVIFDTGSSNLWVPSAKCYFSV
preprosilpepsin2|JX535358| FGFRGTLRSDSDTGIIVLTNERDTTYFGEIGIGTPLQNFVTFIDTGS SLDLWVPS SKCYSSV
***.*****:***.*.* * :*:***** * *:*****:*****:*****:*** **

preprosilpepsin1|JX535387| ACLFHSKYKSSHSTYKNGTSAAIQYGTGSI SGFVSQDSVKLGDLVVKEQDFIEATKEP
preprosilpepsin2|JX535358| SCWVHSRYKSSDSSTYKENGTSAGAIQYGTGSIAGFYSDSVEVGDLVVKEQDFIEATKEA
:* .*:***.*****:***.*****:*** *****:*****:*****:*****

preprosilpepsin1|JX535387| GITFLAAKFDGILGLGFQEISVSKVSPVWYNMVDQGLVHEPVFSFWFNRADEEEGGELV
preprosilpepsin2|JX535358| DNVFLRAHFDGILGLGFQEISVRKAVPVWYNMVDQGLVQEPVFSFWLNRNVNEEEGGELV
.* * :***** ***** *:*****:*****:*****:*****:*****

preprosilpepsin1|JX535387| FGGVDPNHFGRKGRHTYVPVTQKGYWQFNMGDVLIEDKTTGFCSDGCAAIADSGT SLLAGPT
preprosilpepsin2|JX535358| FGGVDPNHFGRNHTYVPVIQKGHWQFDMGDVLIQDKSTGFCANGCAAIA DSGT TLLTGPT
*****:* .***** ***:***:***** ***:*****:*****:*****:*****

preprosilpepsin1|JX535387| AIITQINHAIGAKGVM SQQCKTLVDQYKGS IIEMLLSEAQDPKIC SQMKLCTFNGARDVS
preprosilpepsin2|JX535358| AIVTQINQAIGANRVMNQCKTLVSRYGKNI MQMLMSEAQDPKICSHMNLCTFDGAHDVS
**:* ***:***: ** .***** .:***.*:***:*****:***:***:***

preprosilpepsin1|JX535387| SIIESVVDKNNKSSG---NEMCTFCEMAVVWMQNIKRNQTQDNIINYVELCDRLPSP
preprosilpepsin2|JX535358| SIIESVVDKNDKSSGGVHDEMCTFCEMAVVWMQKEIKRNEIEDNIINYVELCDHLPSS
***** :* **** :*****:*****:*****:*****:*****:*****

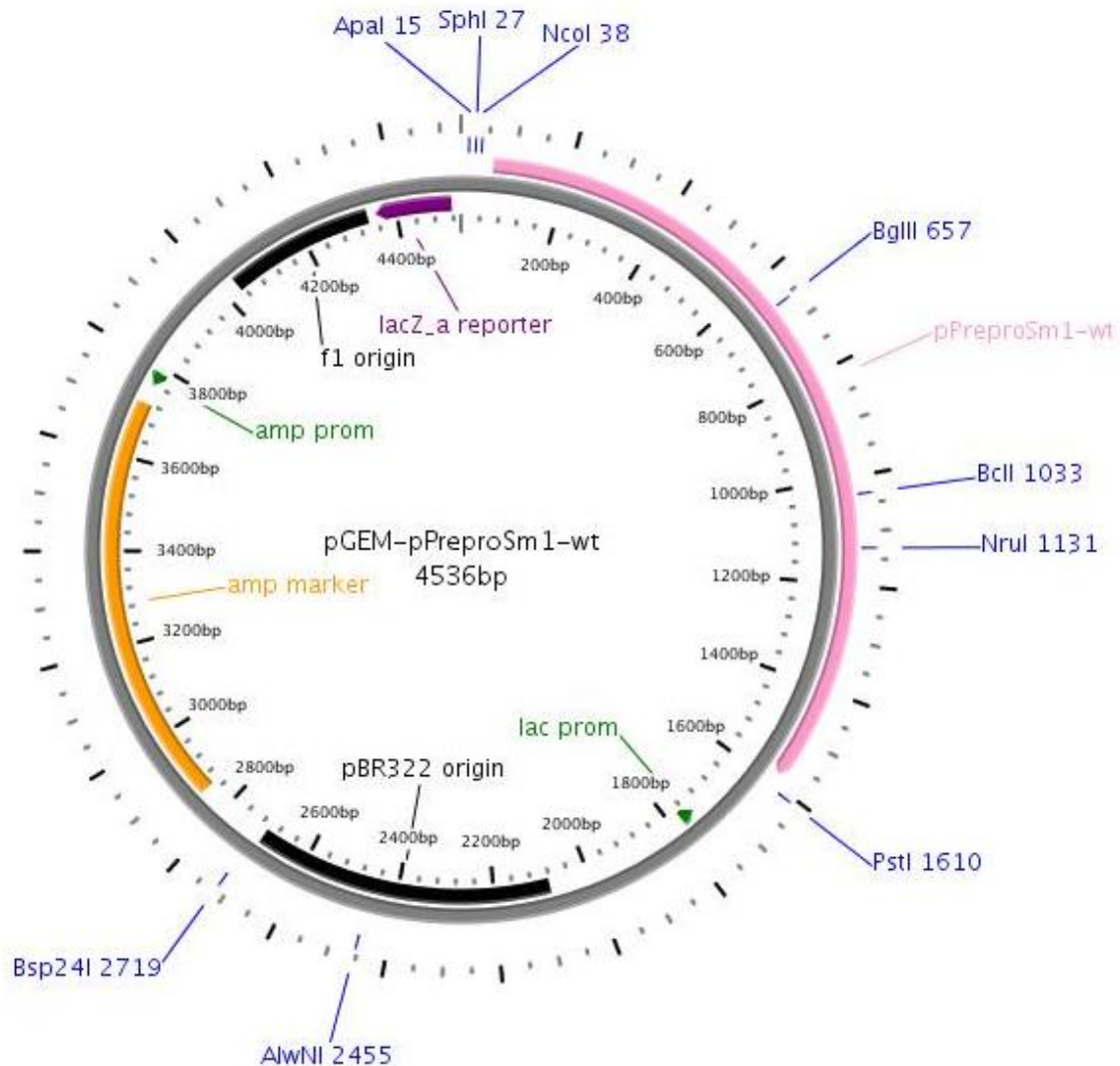
preprosilpepsin1|JX535387| MGESAVDCNSLSSMPNISFTIGGKVFELCPEQYILKIGDGEAAQCISGFTAMDVAPPRGP
preprosilpepsin2|JX535358| SGESVVDKNSLSSMPNTAFTIGGKIFELTPEQYIFKLGEAAQCISGFTTMDVPPPRGP
***.***** :*****:*** *****:***:*****:*****:*****

preprosilpepsin1|JX535387| LWILGDVFMGRYHTVFDYGKLRVGF A E A A
preprosilpepsin2|JX535358| LWILGDI FMGPYHTVFDYGKLRVGF A E A A
*****:*** *****:*****

```

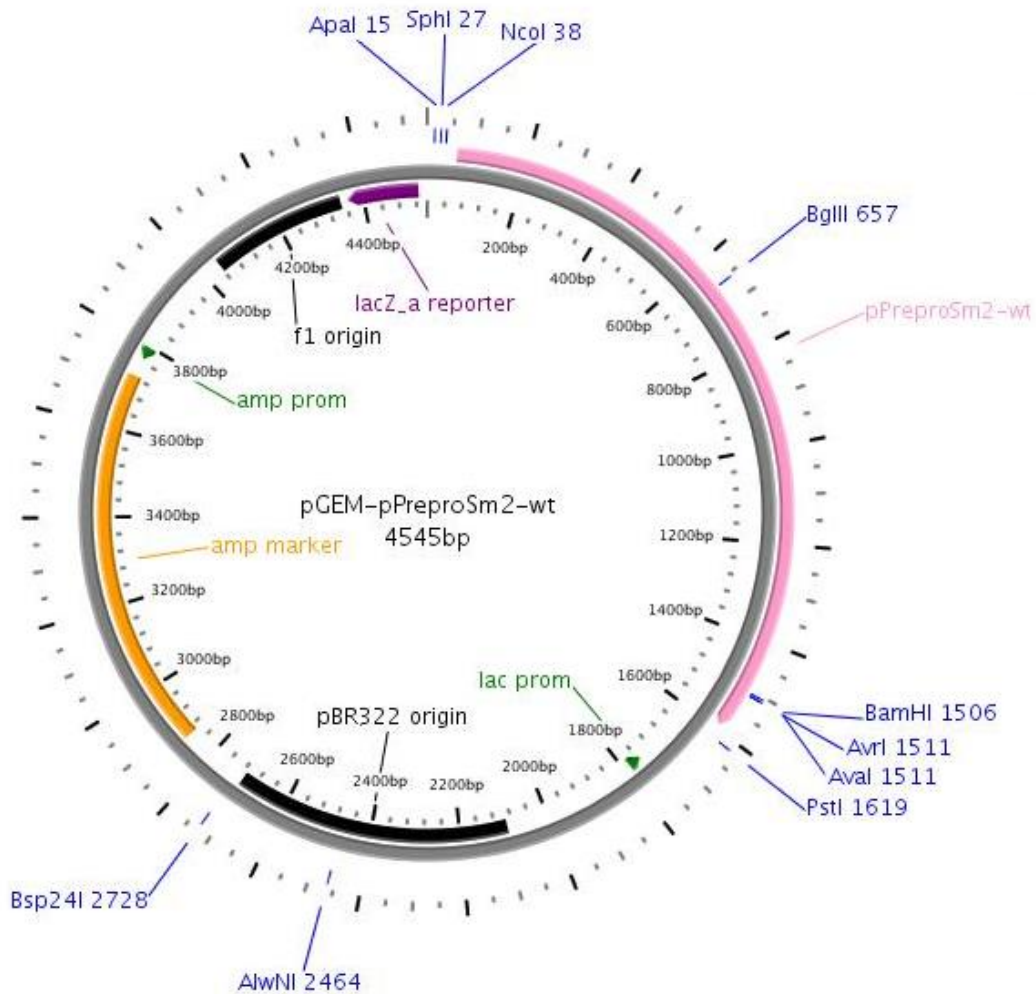
Figura 4. Alineamiento de las secuencias aminoacídicas de los precursores desilpepsina 1 y silpepsina 2. Las regiones correspondientes al péptido señal del N-terminal (-----), el prosegmento (---) y el inserto específico de planta (—) están indicadas por flechas sobre las respectivas secuencias. Las triadas catalíticas se hallan recuadradas. Los residuos azules representan sustituciones conservativas entre ambas secuencias y los rojos, aquellas sustituciones semi-conservativas. En negrita se indican los sitios de N-glicosilación. El alineamiento se realizó con el servidor Clustal Omega.

7.2. ANEXO II



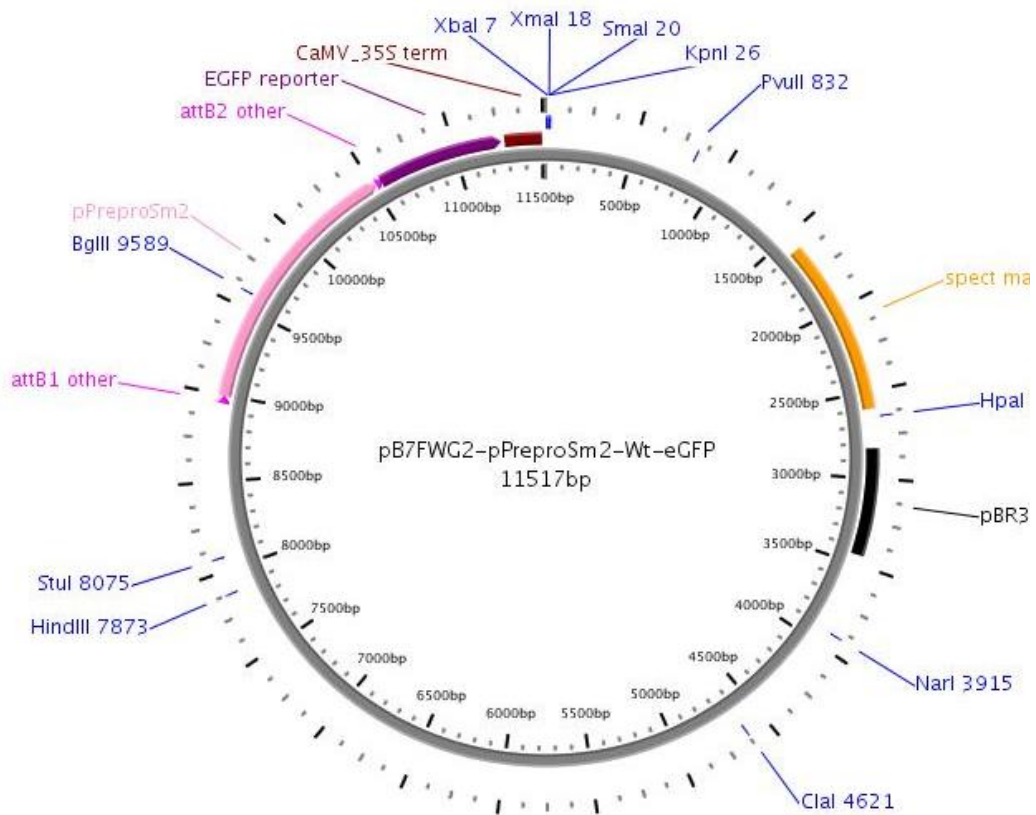
- Gen reportero: LacZ
- Origen f1
- Origen de replicación
- Gen de resistencia ampicilina
- Promotor ampicilina
- Promotor Lac
- pPreproSm1-wt

Mapa del vector pGEM-pPreproSm1-wt. Se indican varios sitios de restricción. La secuencia de los orígenes de replicación, gen reportero: LacZ, origen de replicación, el gen de resistencia a Am, promotor Am y LacZ y la secuencia de la preprosilpepsina 1 (pPreproSm1-wt), se muestra en color rosa. Este mapa fue realizado con el servidor *PlasMapper 2.0*.



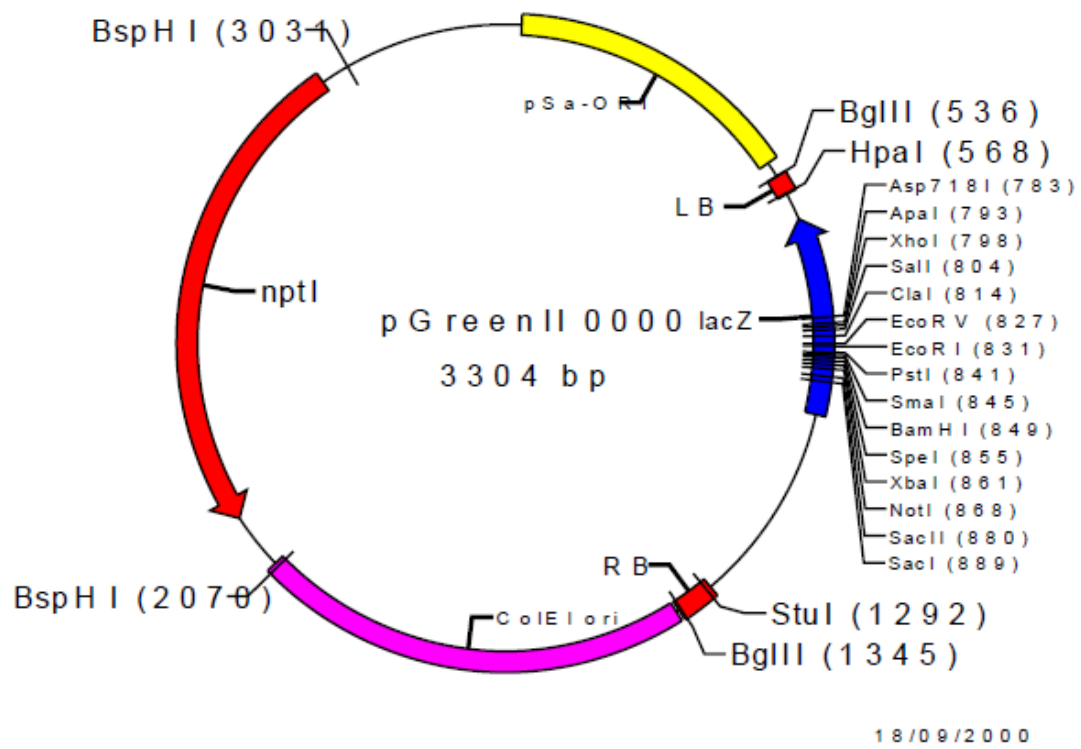
- Gen reportero: LacZ
- Origen f1
- Origen de replicación
- Gen de resistencia ampicilina
- Promotor ampicilina
- Promotor Lac
- pPreproSm2-wt

Mapa del vector pGEM-pPreproSm2-wt. Se indican varios sitios de restricción. La secuencia de los orígenes de replicación, gen reportero: LacZ, origen de replicación, el gen de resistencia a Am, promotor Am y LacZ y la secuencia de la preprosilpepsina 2 (pPreproSm2-wt), se muestra en color rosa. Este mapa fue realizado con el servidor *PlasMapper 2.0*.

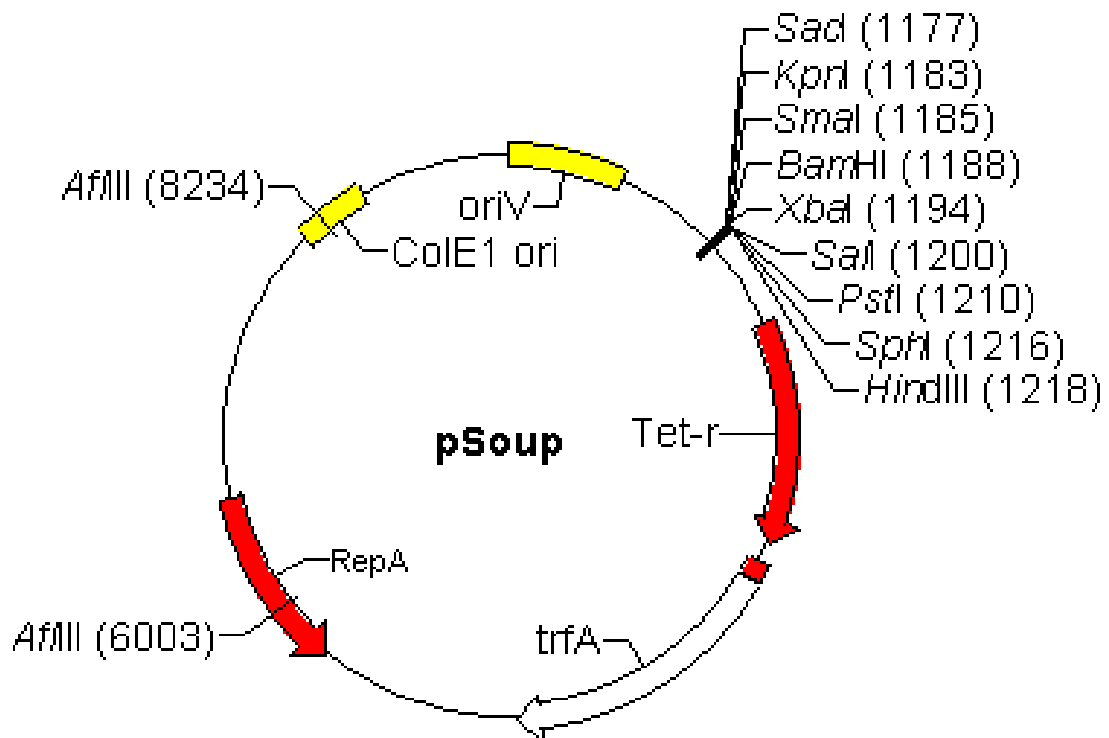


- pPreproSm2
- Gen reportero: eGFP
- Origen de replicación
- Gen de resistencia espectinomicina
- Terminador CaMV-35S
- attB1-attB2

Mapa de la construcción pB7RWG2-pPreproSm2-wt-eGFP. Se indican varios sitios de restricción. La secuencia de la preprosilpepsina 2 (pPreproSm2), se muestra en color rosa, mientras que los sitios *attB* se indican con fucsia. También se muestran el origen de replicación, el gen de resistencia a Sp, promotor y terminador. El mapa de la construcción mutante es equivalente. Este mapa fue realizado con el servidor *PlasMapper 2.0*.



Mapa génico del vector pGreenII.



Mapa génico del vector Soup



Abreviaturas

8. ABREVIATURAS

ADNc: Acido Desoxiribonucleico complementario

AG: Aparato de Golgi (Golgi apparatus)

ARN: Acido Ribonucleico

Atg: proteínas relacionadas con la macroautofagia (AuTophagy related proteins)

CtVSD: determinante C-terminal de direccionamiento vacuolar (C-terminal vacuolar sorting determinant)

Dnasa: Desoxiribonucleasa

DTT: Ditioneitol

EDTA: Acido Etilen Diamin Tetraacetico

GFP: Proteína Verde Fluorescente (green fluorescent protein)

LIR: Región de interacción LC3 (LC3-interacting región)

LV: Vacuola Lítica (lytic vacuoles)

ON: Overnight (durante toda la noche)

PA: Proteasa Aspartica (aspartic proteinase)

PCR: Reaccion en Cadena de la Polimerasa (Polymerase Chain Reaction)

Pro: propéptido

PS: Péptido Señal (signal peptide)

PSI: Inserto Especifico de Plantas (Plant Specific Insert)

PSV: Vacuola de Almacenamiento de Proteínas (protein storage vacuole)

PsVSD: determinante de direccionamiento vacuolar de estructura fisica (physical structure vacuolar sorting determinant)

RE: Retículo Endoplásmico (endoplasmic reticulum)

RFP: Proteína Roja Fluorescente (red fluorescent protein)

SAPLIPs: Saposin-Like Protein Family (Familia de Proteinas Similares a Saposinas)

TGN: Red del *trans*-Golgi (trans-Golgi network)

VSD: determinante de direccionamiento vacuolar (vacuolar sorting determinant)

wt: del ingles *wild-type*, refiere a la proteína cuya secuencia se corresponde con la nativa



Bibliografía



- Alattia, J. R., Shaw, J. E., Yip, C. M., & Privé, G. G. (2007). Molecular imaging of membrane interfaces reveals mode of beta-glucosidase activation by saposin C. *Proceedings of the National Academy of Sciences of the United States of America*, 104(44), 17394–17399.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). *Molecular Biology of the Cell* (5th Edition). New York: Garland Science.
- Almeida, C. M., Pereira, C., Soares da Costa, D., Pereira, S., Pissarra, J., Faro, C., & Simões, I. (2012). Chlpsin, a chloroplastidial aspartic proteinase from the green algae *Chlamydomonas reinhardtii*. *Planta*, 236, 283–296.
- Andreeva, N. S., & Rumsh, L. D. (2001). Analysis of crystal structures of aspartic proteinases: On the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. *Protein Science*, 10, 2439–2450.
- Ariza Espinar, L., & Delucchi, G. (1998). Flora Fanerogámica Argentina. In *Fascículo 60. Asteraceae, parte 11 Tribu XI. Carduae. Programa PROFLORA CONICET* (p. 280).
- Asakura, T., Matsumoto, I., Funaki, J., Arai, S., & Abe, K. (2000). The plant aspartic proteinase-specific polypeptide insert is not directly related to the activity of oryzasin 1. *European Journal of Biochemistry / FEBS*, 267, 5115–22.
- Asakura, T., Watanabe, H., Abe, K., & Arai, S. (1997). Oryzasin as an aspartic proteinase occurring in rice seeds: purification, characterization, and application to milk clotting. *Journal of Agricultural and Food Chemistry*, 45, 1070–1075.
- Asakura, T., Watanabe, H., Abe, K., Arai, S., Asakura, T., Watanabe, H., Arai, S. (1995). Rice aspartic proteinase, oryzasin, expressed during seed ripening and germination, has a gene organization distinct from those of animal and microbial aspartic proteinases. *European Journal of Biochemistry*, 232, 77–83.
- Athauda, S. B., Matsumoto, K., Rajapakshe, S., Kuribayashi, M., Kojima, M., Kubomura, N., ... Takahashi, K. (2004). Enzymic and Structural Characterization of Nepenthesin, a Unique Member of a Novel Subfamily of Aspartic Proteinases. *Biochemical Journal*, 381, 295–306.
- Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, 2, 28–36.
- Balandin, M., Royo, J., Gomez, E., Muniz, L. M., Molina, A., & Hueros, G. (2005). A protective role for the embryo surrounding region of the maize endosperm, as evidenced by the characterisation of ZmESR-6, a defensin gene specifically expressed in this region. *Plant Molecular Biology*, 58, 269–282.
- Barrett, A. J., & Rawlings, N. D. (1991). Types and families of endopeptidases. *Biochemistry Society Transaction*, 19(3), 707–15.
- Barrett, A. J., & Rawlings, N. D. (2007). “Species” of peptidases. *Biology Chemistry*, 388(11), 1151–7.
- Barrett, A. J., Rawlings, N. D., & O’Brien, E. (2001). The MEROPS database as a protease information system. *Journal Structure Biology*, 134(2-3), 95–102.
- Bernstein, N. K., & James, M. N. G. (1999). Nobel ways to prevent proteolysis – prophytepsin and proplasmepsin II. *Journal - Elsevier European Journal of Cell Biology*, 684–689.

- Bertucci, J. I., Liggieri, C. S., Colombo, M. L., Vairo Cavalli, S. E., & Bruno, M. A. (2015). Application of peptidases from *Maclura pomifera* fruit for the production of active biopeptides from whey protein. *LWT - Food Science and Technology*, *64*(1), 157–163.
- Bethke, P., & Jones, R. L. (2000). Vacuoles and prevacuolar compartments. *Current Opinion in Plant Biology*, *3*, 469–475.
- Birgidottir, A. B., Lamark, T., & Johansen, T. (2013). The LIR motif - crucial for selective autophagy. *Journal of Cell Science*, *126*(15), 3537–47.
- Bliven, S., & Prlic, A. (2012). Circular Permutations in Proteins. *PLoS Computational Biology*, *8*(3). <http://doi.org/DOI:10.1371>
- Brandizzi, F., & Hawes, C. (2004). *A long and winding road*. *EMBO Reports*, *5*(3) 245–249.
- Brandizzi, F., Iron, S. I, Johansen, J., Kotzer, A., & Neumann, U. (2004). GFP is the way to glow: Bioimaging of the plant endomembrane system. *Journal of Microscopy*, *214*(2), 138–158.
- Brodelius, M., Hiraiwa, M., Marttila, S., Al Karadaghi, S., Picaud, S., & Brodelius, P. E. (2005). Immunolocalization of the saposin-like insert of plant aspartic proteinases exhibiting saposin C activity. Expression in young flower tissues and in barley seeds. *Plant Biochemistry*, *125*(4), 405–418.
- Bryksa, B. C., Bhaumik, P., Magracheva, E., De Moura, D. C., Kurylowicz, M., Zdanov, A., ... Yada, R. Y. (2011). Structure and Mechanism of the Saposin-like Domain of a Plant Aspartic Protease. *The Journal of Biological Chemistry*, *286*(32), 28265–75.
- Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P., Baird, G. S., Zacharias, D. A., & Tsien, R. Y. (2002). A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7877–82.
- Castanheira, P., Samyn, B., Sergeant, K., Clemente, J. C., Dunn, B. M., Pires, E., Faro, C. (2005). Activation, proteolytic processing, and peptide specificity of recombinant cardosin A. *The Journal of Biological Chemistry*, *280*(13), 13047–54.
- Chen, J., Ouyang, Y., Wang, L., Xie, W., & Zhang, Q. (2009). Aspartic proteases gene family in rice: Gene structure and expression, predicted protein features and phylogenetic relation. *Gene*, *442*, 108–118.
- Cimino, C. V, Colombo, M. L., Liggieri, C. S., Bruno, M. A., & Vairo-Cavalli, S. E. (2015). Partial Molecular Characterization of *Arctium minus* Aspartylendopeptidase and Preparation of Bioactive Peptides by Whey Protein Hydrolysis. *Journal of Medicinal Food*, *18*(8), 856–64.
- Conner, G. E., & Richo, G. (1992). Isolation and characterization of a stable activation intermediate of the lysosomal aspartyl protease cathepsin D. *Biochemistry*, *31*, 1142–1147.
- Cordeiro, M. C., Lowther, T., Dunn, B. M., Guruprasad, K., Blundell, T., Pais, M. S., & Brodelius, P. E. (1998). Substrate specificity and molecular modelling of aspartic proteinases (Cyprosin) from flowers of *Cynara cardunculus* subsp. *flavescens* cv. *cardoon*. *Aspartic Proteinases*, *436*, 473–479.
- Cordeiro, M. C., Pais, M. S., & Brodelius, P. E. (1994). Tissue-specific expression of multiple forms of cyprosin (aspartic proteinase) in flowers of *Cynara carduncubis*. *Physiologia Plantarum*, *92*, 645–653.
- Cordeiro, M. C., Zhong-tian, X., Pietrzak, M., Pais, M. S., & Brodelius, P. E. (1994). Isolation and characterization of a cDNA from flowers of *Cynara cardunculus* encoding cyprosin (an aspartic

- proteinase) and its use to study the organ-specific expression of cyprosin. *Plant Molecular Biology*, 24, 733–741.
- Curto, P., Lufrano, D., Pinto, C., Custódio, V., Gomes, A. C., Trejo, S., & Simões, I. (2014). Establishing the yeast *Kluyveromyces lactis* as an expression host for production of the saposin-like domain (plant-specific insert) from the aspartic protease cirsin., 80, 86–96. *Applied and Environmental Microbiology*, 80, 86–96.
- da Costa, D. S., Pereira, S., Moore, I., & Pissarra, J. (2010). Dissecting cardosin B trafficking pathways in heterologous systems. *Planta*, 232(6), 1517–1530.
- Davey, N., Van Roey, K., Weatheritt, R., Toedt, G., Uyar, B., Altenberg, B., Gibson, T. (2012). Attributes of short linear motifs. *Molecular BioSystems*, 1 268-281.
- Dee, D. R., Filonowicz, S., Horimoto, Y., & Yada, R. Y. (2009). Recombinant prosegment peptide acts as a folding catalyst and inhibitor of native pepsin. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1794(12), 1795–1801.
- Dee, D. R., Myers, B., & Yada, R. Y. (2015). Foldase and inhibitor functionalities of the pepsinogen prosegment are encoded within discrete segments of the 44 residue domain. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1854, 1300–1306.
- di Sansebastiano, G. P., Paris, N., Marc-Martin, S., & Neuhaus, J. M. (2001). Regeneration of a Lytic Central Vacuole and of Neutral Peripheral Vacuoles Can Be Visualized by Green Fluorescent Proteins Targeted to Either Type of Vacuoles. *Plant Physiology*, 126(1), 78–86.
- di Tommaso, P., Moretti, S., Xenarios, I., Orbitg, M., Montanyola, A., Chang, J. M., Notredame, C. (2011). T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, 39, 13–17.
- Domingos, A., Cardoso, P. C., Xue, Z. T., Clemente, A., Brodelius, P. E., & Pais, M. S. (2000). Purification, cloning and autoproteolytic processing of an aspartic proteinase from *Centaurea calcitrapa*. *European Journal of Biochemistry*, 267, 6824–31.
- Dong, X., Stothard, P., Forsythe, I. J., & Wishart, D. S. (2004). PlasMapper: A web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Research*, 32, 660–664.
- Duarte, A. S., Duarte, E. P., Correia, A., Pires, E., & Barros, M. T. (2009). Cardosins improve neuronal regeneration after cell disruption: a comparative expression study. *Cell Biology Toxicology*, 25, 99–108.
- Duarte, A. S., Pereira, A. O., Cabrita, A. M. S., Moir, A. J. G., Pires, E. M. V., & Barros, M. M. T. (2005). The characterisation of the collagenolytic activity of cardosin a demonstrates its potential application for extracellular matrix degradative processes. *Current Drug Discovery Technologies*, (2), 37–44.
- Duarte, A. S., Rosa, N., Duarte, E. P., Pires, E., & Barros, M. T. (2007). Cardosins: a new and efficient plant enzymatic tool to dissociate neuronal cells for the establishment of cell cultures. *Biotechnology and Bioengineering*, 97, 991–996.
- Duarte, P., Pissarra, J., & Moore, I. (2008). Processing and trafficking of a single isoform of the aspartic proteinase cardosin A on the vacuolar pathway. *Planta*, 227(6), 1255–68.
- Dunn, B. M. (2002). Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chemical*

- Reviews*, 102, 4431–58.
- Dunn, B. M., & Hung, S. (2000). The two sides of enzyme ^ substrate speci c city : lessons from the aspartic proteinases. *Engineering*, 1477, 231–240.
- Dunn, T. M., Lynch, D. V, Michaelson, L. V, & Napier, J. A. (2004). A post-genomic approach to understanding sphingolipid metabolism in *Arabidopsis thaliana*. *Annual Botanical*, 93(5), 483–97.
- Dykes, C. W., & Kay, J. (1976). Conversion of pepsinogen into pepsin is not a one-step process. *Biochem. Journal*, 153(1), 141–144.
- Egas, C., Lavoura, N., Resende, R., Brito, R. M., Pires, E., de Lima, M. C., & Faro, C. (2000). The saposin-like domain of the plant aspartic proteinase precursor is a potent inducer of vesicle leakage. *The Journal of Biological Chemistry*, 275(49), 38190–6.
- Esseling-Ozdoba, A., Houtman, D., Van Lammeren, A. A. M., Eiser, E., & Emons, A. M. C. (2008). Hydrodynamic flow in the cytoplasm of plant cells. *Journal of Microscopy*, 231(2), 274–283.
- Faro, C., & Gal, S. (2005). Aspartic proteinase content of the *Arabidopsis* genome. *Current Protein & Peptide Science*, 6(6), 493–500.
- Faro, C., Ramalho-Santos, M., Vieira, M., Mendes, A., Simões, I., Andrade, R., ... Pires, E. (1999). Cloning and characterization of cDNA encoding cardosin A, an RGD-containing plant aspartic proteinase. *The Journal of Biological Chemistry*, 274(40), 28724–9.
- Feijoo-Siota, L., & Villa, T. G. (2011). Native and Biotechnologically Engineered Plant Proteases with Industrial Applications. *Food and Bioprocess Technology*, 4(6), 1066–1088.
- Figueiredo, R., Duarte, P., Pereira, S., & Pissarra, J. (2006). The embryo sac of *Cynara cardunculus* : ultrastructure of the development and localisation of the aspartic proteinase cardosin B. *Sexual Plant Reproduction*, 19, 93–101.
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., ... Punta, M. (2014). Pfam: The protein families database. *Nucleic Acids Research*, 42(1), 222–230.
- Fitchette-Lainé, A C Gomord, V., Cabanes, M., Michalski, J. C., Saint Macary, M., Foucher, B., Cavelier, B., Faye, L. (1997). N-glycans harboring the Lewisia epitope are expressed at the surface of plant cells. *Plant Journal*, 12, 1411–1417.
- Flückiger, R., De Caroli, M., Piro, G., Dalessandro, G., Neuhaus, J. M., & Di Sansebastiano, G. P. (2003). Vacuolar system distribution in *Arabidopsis* tissues, visualized using GFP fusion proteins. *Journal of Experimental Botany*, 54(387), 1577–1584.
- Frazão, C., Bento, I., Costa, J., Soares, C. M., Veríssimo, P., Faro, C., Carrondo, M. (1999). Crystal structure of cardosin A, a glycosylated and Arg-Gly-Asp-containing aspartic proteinase from the flowers of *Cynara cardunculus* L. *The Journal of Biological Chemistry*, 274(39), 27694–27701.
- Frigerio, L., Prada, A., Faoro, F., & Vitale, A. (1998). Sorting of Phaseolin to the Vacuole Is Saturable and Requires a Short C-Terminal Peptide. *The Plant Cell*, 10, 1031–1042.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*. (N. J. M. Walker © Humana Press Inc., Totowa, Ed.) *The Proteomics Protocols Handbook*, Humana Press.

- Glathe, S., Kervinen, J., Nimtz, M., Li, G. H., Tobin, G. J., Copeland, T. D., Costa, J. (1998). Transport and activation of the vacuolar aspartic proteinase phytepsin in barley (*Hordeum vulgare* L.). *The Journal of Biological Chemistry*, 273(47), 31230–6.
- González-Rábade, N., Badillo-Corona, J. A., Aranda-Barradas, J. S., & Oliver-Salvador, M. C. (2011). Production of plant proteases in vivo and in vitro-A review. *Biotechnology Advances*, 29(6), 983–96.
- Goodin, M. M., Dietzgen, R. G., Schichnes, D., Ruzin, S., & Jackson, A. O. (2002). pGD vectors: Versatile tools for the expression of green and red fluorescent protein fusions in agroinfiltrated plant leaves. *Plant Journal*, 31(3), 375–383.
- Guevara, M. G., Oliva, C. R., Machinandiarena, M., & Daleo, G. R. (1999). Purification and properties of an aspartic protease from potato tuber that is inhibited by a basic chitinase. *Physiologia Plantarum*, 106(2), 164–169.
- Guevara, M. G., Veríssimo, P., Pires, E., Faro, C., & Daleo, G. R. (2004). Potato Aspartic Proteases : Induction , Antimicrobial Activity and Substrate Specificity, *Journal Plant Pathol*, 86, 233–238.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximim-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3), 1–37.
- Guo, R., Xu, X., Carole, B., Li, X., Gao, M., Zheng, Y., & Wang, X. (2013). Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape. *BMC Genomics*, 14(1), 554.
- Gupta, D., Yedidi, R. S., Varghese, S., Kovari, L. C., & Woster, P. M. (2010). Mechanism-based inhibitors of the aspartyl protease plasmepsin II as potential antimalarial agents. *Journal Medicine Chemistry*, 53(10), 4234–47.
- Gupta, R., Jung, E., & Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. unpublished.
- Güracar, S. (2011). *Production , Purification and Characterization of Thermostable Protease from Alkaliphilic and Thermophilic Geobacillus sp .Instituto İzmir de Tecnología.*
- Hara-Nishimura, I., Shimada, T., Hatano, K., Takeuchi, Y., & Nishimura, M. (1998). Transport of Storage Proteins to Protein Storage Vacuoles Is Mediated by Large Precursor-Accumulating. *Vesicles Plant Cell*, 10, 825–821.
- Hartley, B. S. (1960). Proteolytic enzymes. *Annual Revie. Biochemistry*, 29, 45–72.
- Hawes, C., & Satiat-Jeunemaitre, B. (2005). The Plant Golgi Apparatus – Going with the Flow. *Biochimica et Biophysica Acta*, 1744(2), 93–107.
- Heikal, A. A., Hess, S. T., Baird, G. S., Tsien, R. Y., & Webb, W. W. (2000). Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: coral red (dsRed) and yellow (Citrine). *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), 11996–2001.
- Heim, R., Cubitt, A. B., & Tsien, R. Y. (1995). Improved green fluorescence. *Nature*, 373, 663–664.
- Herman, E., & Larkins, B. A. (1999). Protein storage bodies and vacuoles. *Plant Cell*, 11, 601–614.

- Heven, S., Xuhang, L., & Palmgren, M. G. (1999). Energization of Plant Cell Membranes by H⁺-Pumping ATPases: *Regulation and Biosynthesis*, 11(4), 677–689.
- Hillmer, S., Movafeghi, A., Robinson, D. G., & Hinz, G. (2001). Vacuolar storage proteins are sorted in the cis-cisternae of the pea cotyledon Golgi apparatus. *Journal Cell Biology*, 152, 41–50.
- Hoh, B., Hinz, G., Jeong, B. K., & Robinson, D. G. (1995). Protein storage vacuoles form de novo during pea cotyledon development. *Journal Cell Science*, 108, 299–310.
- Holger, D., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Gibson, T. (2016). ELM 2016—data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Research*, 44(1), 294–300.
- Holger, D., Van Roey, K., Uyar, B., Weatheritt, R., Dinkel, H., Seiler, M., Davey, N. (2014). Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.*, 114(13), 6733–78.
- Holkeri, H., & Vitale, A. (2001). Vacuolar sorting determinants within a plant storage protein trimer act cumulatively. *Traffic*, 2, 737–741.
- Holmes, S. (2003). Bootstrapping Phylogenetic Trees: Theory and Methods. *Statistical Science*, 18(2), 241–255.
- Hooper, N. M. (2002). Proteases : a primer. In *Proteolysis Research Group, School of Biochemistry and Molecular Biology* (pp. 1–8). University of Leeds, LS2 9JT, Leeds, UK:
- Horimoto, Y., Dee, D. R., & Yada, R. Y. (2009). Multifunctional aspartic peptidase prosegments. *New Biotechnology*, 25(5), 318–324.
- Ikai, A. J. (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem*, 88(6), 1895–1898.
- James, M. N. (2004). Catalytic pathway of aspartic peptidases. In *Handbook of Proteolytic Enzymes* (Elsevier A, pp. 12–19).
- Jolliffe, N. A., Ceriotti, A., Frigerio, L., & Roberts, L. M. (2003). The position of the proricin vacuolar targeting signal is functionally important. *Plant Molecular Biology*, 51, 631–641.
- Jürgens, G. (2004). Membrane Trafficking in Plants. *Annual Reviews in Cell and Developmental Biology*, 20, 481–504.
- Kahn, D., Rezvoy, C., & Vivien, F. (2008). Parallel large scale inference of protein domain families. In *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS* (pp. 72–79).
- Kato, Y., Murakami, S., Yamamoto, Y., Chatani, H., Kondo, Y., Nakano, T., Sato, F. (2004). The DNA-Binding Protease, CND41, and the Degradation of Ribulose- 1,5-biphosphate Carboxylase/Oxygenase in Senescent Leaves of Tobacco. *Planta*, 220(1), 97–104.
- Kervinen, J., Tobin, G. J., Costa, J., Waugh, D. S., Wlodawer, A., & Zdanov, A. (1999). Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. *EMBO Journal*, 18(14), 3947–3955.
- Kervinen, J., & Wlodawer, A. (2013). Phytepsin. In N. Rawlings & G. Salvesen (Eds.), *Handbook of Proteolytic Enzymes* (3rd ed., pp. 118–124). Elsevier Ltd. Oxford: Academic Press.

- Kishimoto, Y., Hiraiwa, M., & O'Brien, S. J. (1992). Saposins: structure, function, distribution, and molecular genetics. *The Journal of Lipid Research*, *33*, 1255–1267.
- Koelsch, G., Mares, M., Metcalf, P., & Fusek, M. (1994). Multiple functions of pro-parts of aspartic proteinase zymogens. *FEBS Letters*, *343*(1), 6–10.
- Kosakovsky Pond, S. L., Frost, S. D. W., & Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*, *21*(5), 676–679.
- Kuwabara, T., & Suzuki, K. (1995). Reversible changes in conformation of the 23-kDa protein of photosystem II and their relationship to the susceptibility of the protein to a proteinase from photosystem II membranes. *Plant Cell Physiology*, *36*, 495–504.
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–132.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*, 2947–2948.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291.
- Lufrano, D., Faro, R., Castanheira, P., Parisi, G., Veríssimo, P., Vairo-Cavalli, S., Faro, C. (2012). Molecular cloning and characterization of procirsin, an active aspartic protease precursor from *Cirsium vulgare* (Asteraceae). *Phytochemistry*, *81*, 7–18.
- Maeshima, M. (2000). Vacuolar H⁺-pyrophosphatase. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1465*(1), 37–51.
- Matsuoka, K., & Neuhaus, J. (1999). Cis-elements of protein transport to the plant vacuoles. *Journal of Experimental Botany: Oxford Journals* *50*, 165–174.
- Mazorra-Manzano, M. A., Perea-Gutiérrez, T. C., Lugo-Sánchez, M. E., Ramirez-Suarez, J. C., Torres-Llanez, M. J., González-Córdova, A. F., & Vallejo-Cordoba, B. (2013). Comparison of the milk-clotting properties of three plant extracts. *Food Chemistry*, *141*(3), 1902–7.
- Mazorra-Manzano, M. A., & Yada, R. Y. (2008). Expression and characterization of the recombinant aspartic proteinase A1 from *Arabidopsis thaliana*. *Phytochemistry*, *69*(13), 2439–48.
- Mendieta, J. R., Pagano, M. R., Muñoz, F. F., Daleo, G. R., & Guevara, M. G. (2006). Antimicrobial activity of potato aspartic proteases (StAPs) involves membrane permeabilization. *Microbiology*, *152*, 2039–47.
- Michaeli, S., Galili, G., Genschik, P., Fernie, A. R., & Avin-Wittenberg, T. (2015). Autophagy in Plants - What's New on the Menu? *Journals Trends in Plant Science*, *21*, 134–144.
- Morimoto, S., Kishimoto, Y., Tomich, J., Weiler, S., Ohashi, T., Barranger, J. A., O'Brien, J. S. (1990). Interaction of saposins, acidic lipids and glucosylceramidase. *The Journal of Biological Chemistry*, *265*, 1933–1937.
- Morré, D. J., & Mollenhauer, H. H. (1974). The endomembrane concept: a functional integration of

- endoplasmic reticulum and Golgi apparatus. *In Dynamics aspects of plant ultrastructure* (Robards, A, pp. 84–137). London, New York.
- Muller- Esterl, W. (2008). *Bioquímica: fundamentos para medicina y ciencias de la vida*. (Reverté S.). Barcelona. Bogotá. Buenos Aires. Caracas. México.
- Munford, R. S., Sheppard, P. O., & O'Hara, P. J. (1995). Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure. *Journal of Lipid Research*, 36(8), 1653–63.
- Muñoz, F. F., Mendieta, J. R., Pagano, M. R., Paggi, R. A., Daleo, G. R., & Guevara, M. G. (2010). The swaposin-like domain of potato aspartic protease (StAsp-PSI) exerts antimicrobial activity on plant and human pathogens. *Peptides*, 31, 777–85.
- Muñoz, F., Palomares-Jerez, M., Daleo, G., Villalaín, J., & Guevara, M. G. (2014). Possible mechanism of structural transformations induced by StAsp-PSI in lipid membranes. *Biochimica et Biophysica Acta*, 1838, 339–347.
- Mutlu, A., & Gal, S. (1999). Plant aspartic proteinases: enzymes on the way to a function. *Physiologia Plantarum*, 105, 569–576.
- Nakamura, K., & Matsuoka, K. (1993). Protein targeting to the vacuole in plant cells. *Plant Physiology*, 101, 1–6.
- Nakano, T., Murakami, S., Shoji, T., Yoshida, S., Yamada, Y., & Sato, F. (1997). A novel protein with DNA binding activity from tobacco chloroplast nucleoids. *Plant Cell*, 9, 1673–1682.
- Neuhaus, J. M., & Rogers, J. C. (1998). Sorting of proteins to vacuoles in plant cells. *Plant Molecular Biology* 38, 127–144.
- Noda, N. N., Ohsumi, Y., & Inagaki, F. (2010). Atg8-family interacting motif crucial for selective autophagy. *FEBS Letters*, 584, 1379–1385.
- Ohyanagi, H., Tanaka, T., Sakai, H., Shigemoto, Y., Yamaguchi, K., Habara, T., Sasaki, T. (2006). The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acid Res*, 34, 741–744.
- Olivares, J. E., Díaz-Camino, C., Estrada-Navarrete, G., Alvarado-Affantranger, X., Rodríguez-Kessler, M., Zamudio, F. Z., Olamendi-Portugal, T., Sánchez, F. (2011). Nodulin 41, a novel late nodulin of common bean with peptidase activity. *BMC Plant Biology*, 11, 134–147.
- Oliveira, A., Pereira, C., da Costa, D. S., Teixeira, J., Fidalgo, F., Pereira, S., & Pissarra, J. (2010). Characterization of aspartic proteinases in *C. cardunculus* L. callus tissue for its prospective transformation. *Plant Science*, 178(2), 140–146.
- Page, M. J., & Di Cera, E. (2008). Evolution of Peptidase Diversity. *The Journal of Biological Chemistry*, 283(44), 30010–30014.
- Paris, N., Stanley, C. M., Jones, R. L., & Rogers, J. C. (1996). Plant cells contain two functionally distinct vacuolar compartments. *Cell*, 85(4), 563 – 572.
- Park, H., Kusakabe, I., Sakakibara, Y., & Kobayashi, H. (2001). Autoproteolytic processing of aspartic proteinase from sunflower seeds. *Bioscience, Biotechnology, and Biochemistry*, 65(3), 702– 705.

- Pereira, A. O., Cartucho, D. J., Duarte, A. S., Gil, M. H., Cabrita, A. M. S., Patricio, J. A., & Barros, M. M. T. (2005). Immobilisation of cardosin A in chitosan sponges as a novel implant for drug delivery. *Current Drug Discovery Technologies*, 2, 231–238.
- Pereira, C. (2012). *Cardosin A Molecular Determinants and Biosynthetic Pathways*. Faculty of Sciences, University of Porto, Portugal.
- Pereira, C., da Costa, D. S., Pereira, S., de Moura Nogueira, F., Albuquerque, P. M., Teixeira, J., Pissarra, J. (2008). Cardosins in postembryonic development of cardoon: towards an elucidation of the biological function of plant aspartic proteinases. *Protoplasma*, 232, 203–213.
- Pereira, C., Pereira, S., & Pissarra, J. (2014). Delivering of Proteins to the Plant Vacuole—An Update. *International Journal of Molecular Sciences*, 15(5), 7611–7623.
- Pereira, C., Pereira, S., Satiat-Jeunemaitre, B., & Pissarra, J. (2013). Cardosin A contains two vacuolar sorting signals using different vacuolar routes in tobacco epidermal cells. *Plant Journal*, 76(1), 87–100.
- Perpetuo, E. A., Juliano, L., & Lebrun, I. (2003). Biochemical and pharmacological aspects of two bradykinin-potentiating peptidases obtained from tryptic hydrolysis of casein. *Journal of Protein Chemistry*, 22, 601–6.
- Pesquet, E. (2012). Plant proteases - from detection to function. *Physiologia Plantarum*, 145(1), 1–4.
- Petersen, T. N. N., Brunak, S., von Gunnar, H., Nielsen, H., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 785–786.
- Petrucelli, S., Molina, I. M., Lareu, J. F., & A, C. (2007). Two short sequences from amaranth 11S globulin are sufficient to target green fluorescent protein and betaglucuronidase to vacuoles in Arabidopsis cells. *Plant Physiology and Biochemistry*, 45, 400–409.
- Pimentel, C., Van Der Straeten, D., Pires, E., Faro, C., & Rodrigues-Pousada, C. (2007). Characterization and expression analysis of the aspartic protease gene family of *Cynara cardunculus* L. *The FEBS Journal*, 274(10), 2523–2539.
- Ponting, C. P., & Russel, R. B. (1995). Swaposins: Circular Permutations Within Genes Encoding Saposin Homologues. *Trends in Biochemical Sciences*, 20, 179–180.
- Puntervoll, P., Linding, R., Gemund, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S., Gibson, T. J. (2003). ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Research*, 31(13), 3625–3630.
- Qing, L., Yi, L., Marek, P., & Iverson, B. L. (2013). Commercial proteases: present and future. *FEBS Letters*, 587(8), 1155–63.
- Ramvalho-Santos, M., Pissarra, J., Veríssimo, P., Pereira, S., Salema, R., Pires, E., & Faro, C. J. (1997). Cardosin A, an abundant aspartic proteinase, accumulates in protein storage vacuoles in the stigmatic papillae of *Cynara cardunculus* L. *Planta*, 203(2), 204–212.
- Ramvalho-Santos, M., Verissimo, P., Cortes, L., Samyn, B., & Faro, C. (1998). Identification and proteolytic processing of procardosin A. *Biochemistry*, 255, 133–138.
- Rawlings, N.D., Barrett, A. (2004). *Introduction: aspartic peptidases and their clans*, in: Barrett, A.J., Rawlings,

- N.D., Woessner, J.F. (Eds.). *Handbook of Proteolytic Enzymes*. Elsevier Academic Press, London, pp.
- Rawlings, N. D., & Barrett, A. J. (1995). Families of aspartic peptidases, and those of unknown catalytic mechanism. *Methods in Enzymology*, 248, 105–120.
- Rawlings, N. D., & Barrett, A. J. (1999). MEROPS: the peptidase database. *Nucleic Acids Research*, 27(1), 325–31.
- Rawlings, N. D., & Barrett, A. J. (2014). Peptidases. In R. Beynon & J. S. Bond (Eds.), *Handbook of Proteolytic Enzymes* (2nd edn, pp. 1–9). Ltd: Chichester: eLS: Encyclopedia of Life Sciences.
- Rawlings, N. D., Barrett, A. J., & Bateman, A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 40, 343–50.
- Rawlings, N. D., & Bateman, A. (2009). Pepsin homologues in bacteria. *BMC Genomics*, 10, 437.
- Rawlings, N. D., & Salvesen, G. (2013). Introduction: Aspartic and Glutamic Peptidases and Their Clans. In A. J. Barrett, N. D. Rawlings, & J. Fred (Eds.), *Handbook of Proteolytic Enzymes* (3er ed., Vol. 1, pp. 3–19). Elsevier Ltd. Oxford: Academic Press.
- Rawlings, N. D., Tolle, D. P., & Barrett, A. J. (2004). Evolutionary families of peptidase inhibitors. *Biochemistry J*, 378, 705–716.
- Revuelta, M., van Kan, J. A. L., Kay, J., & ten Have, A. (2014). Extensive expansion of A1 family aspartic proteinases in fungi revealed by evolutionary analyses of 107 complete eukaryotic proteomes. *Genome Biology and Evolution*, 6(6), 1480–1494.
- Rodrigo, I., Vera, P., & Conejero, V. (1989). Degradation of tomato pathogenesis-related proteins by an endogenous 37 kDa aspartyl endoproteinase. *European Journal of Biochemistry*, 184, 663–669.
- Rodrigo, I., Vera, P., van Loon, L. C., & Conejero, V. (1991). Degradation of tobacco pathogenesis-related proteins: evidence for conserved mechanisms of degradation of pathogenesis-related proteins in plants. *Plant Physiology*, 95, 616–622.
- Rogov, V., Dötsch, V., Johansen, T., & Kirkin, V. (2014). Interactions between Autophagy Receptors and Ubiquitin-like Proteins Form the Molecular Basis for Selective Autophagy. *Molecular Cell*, 53(2), 167–178.
- Rogov, V., Suzuki, H., Fiskin, E., Wild, P., Kniss, A., Rozenknop, A., Dötsch, V. (2013). Structural basis for phosphorylation-triggered autophagic clearance of Salmonella. *Biochemistry Journal*, 45(3), 459–66.
- Salanoubat, M., Lemcke, K., Rieger, M., Ansorge, W., Unseld, M., Fartmann, B., Tabata, S. (2000). Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature*, 408(6814), 820–822.
- Sali, A., & Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Molecular Biology*, 234(3), 779–815.
- Salmia, M. A. (1981). Proteinase activities in resting and germinating seeds of Scots pine, *Pinus sylvestris*. *Physiology Plant.*, 53, 39–47.
- Salvesen, G. S., Hempel, A., & Coll, N. S. (2015). Protease Signaling in Animal and Plant Regulated Cell Death. *FEBS Journal*.

- Samalova, M., Fricker, M., & Moore, I. (2006). Ratiometric fluorescence-imaging assays of plant membrane traffic using polyproteins. *Traffic*, 7, 1701–1723.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Preparation of plasmid DNA by alkaline lysis with SDS: Miniprep. In *Molecular Cloning. A Laboratory Manual*. (pp. 1.32–1.34). New York: Press, Cold Spring Harbor Laboratory.
- Sampaio, P. N., Fortes, A. M., Cabral, J. M. S., Pais, M., & Fonseca, L. P. (2008). Production and characterization of recombinant cyprosin B in *Saccharomyces cerevisiae* (W303-1A) strain. *J Biosci Bioengineering*, 105(4), 305–12.
- Sarkkinen, P., Kalkkinen, N., Tilgmann, C., Siuro, J., Kervinen, J., & Mikola, L. (1992). Aspartic proteinase from barley grains is related to mammalian lysosomal cathepsin D. *Planta*, 186, 317–323.
- Sarmiento, A. C., Lopes, H., Oliveira, C. S., Vitorino, R., Samyn, B., Sergeant, K., Barros, M. T. (2009). Multiplicity of aspartic proteinases from *Cynara cardunculus*. *Planta*, 230(2), 429–39.
- Sarmiento, A. C., Silvestre, L., Barros, M., & Pires, E. (1998). Cardosins A and B, two new enzymes available for peptide synthesis. *Journal of Molecular Catalysis B Enzymatic*, 5, 327–330.
- Sarmiento, C., Oliveira, C., Pires, E., Amado, F., & Barros, M. (2004). Reverse hydrolysis by cardosin A: specificity considerations. *Journal of Molecular Catalysis B: Enzymatic*, 28, 33–37.
- Scabone, C. M., Frigerio, L., & Petruccioli, S. (2011). A fluorescent reporter protein containing AtRMR1 domains is targeted to the storage and central vacuoles in *Arabidopsis thaliana* and tobacco leaf cells. *Plant Cell Reports*, 30(10), 1823–1833.
- Schaller, A. (2004). A cut above the rest: The regulatory function of plant proteases. *Planta*, 220, 183–197.
- Schechter, I., & Berger, A. (1967). On the size of the active site in proteases: Pronase. *Biochemical and Biophysical Research Communications*, 27(2), 157–162.
- Schuetz, C. G., Pierstorff, B., Huettler, S., & Sandhoff, K. (2001). Sphingolipid activator proteins: proteins with complex functions in lipid degradation and skin biogenesis. *Glycobiology*, 11(6), 81–90.
- Shaner, N. C., Steinbach, P. A., & Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nature Methods*, 2(12), 905–909.
- Shmyrova, A. V., Oliveira, C. S., Sarmiento, A. C., Barros, M. T., Zhadan, G. G., Roig, M. G., & Shnyrov, V. L. (2006). Effect of acetonitrile on *Cynara cardunculus* L. cardosin A stability. *International Journal of Biological Macromolecules*, 39, 273–279.
- Shpilka, T., Weidberg, H., Pietrokovski, S., & Elazar, Z. (2011). Atg8: an autophagy-related ubiquitin-like protein family. *Genome Biology*, 12(7), 226.
- Shuang, L., Yang, X., Yang, S., Zhu, M., & Wang, X. (2012). Technology prospecting on enzymes: application, marketing and engineering. *Computational and Structural Biotechnology Journal*, 2(3), 1–11.
- Silhavy, D., Molnar, a, Luciola, a, Szitty, G., Hornyik, C., Tavazza, M., & Burgyan, J. (2002). A viral protein suppresses RNA silencing and binds silencing-generated, 21-to 25-nucleotide double-stranded RNAs. *Embo Journal*, 21(12), 3070–3080.

- Silva, S., & Malcata, F. X. (2005). Partial identification of water soluble peptides released at early stages of proteolysis in sterilized ovine cheese-like systems: influence of type of coagulant and starter. *Journal of Dairy Science*, *88*, 1947–54.
- Simões, I., & Faro, C. (2004). Structure and function of plant aspartic proteinases. *European Journal of Biochemistry / FEBS*, *271*(11), 2067–75.
- Simões, I., Faro, R., Bur, D., Kay, J., & Faro, C. (2011). Shewasin A, an active pepsin homolog from the bacterium *Shewanella amazonensis*. *FEBS J*, *278*, 3177–3186.
- Simões, I., Mueller, E. C., Otto, A., Bur, D., Cheung, A. Y., Faro, C., & Pires, E. (2005). Molecular analysis of the interaction between cardosin A and phospholipase D(alpha). Identification of RGD/KGE sequences as binding motifs for C2 domains. *The FEBS Journal*, *272*, 5786–98.
- Sivakumar, K., Balaji, S., & Gangaradhakrishnan. (2007). In silico characterization of antifreeze proteins using computational tools and servers. *Journal of Chemical Sciences*, *119*(5), 571–579.
- Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(2), 244–248.
- Staelin, L. A., & Driouch, A. (1997). Brefeldin A Effects in Plants (Are Different Golgi Responses Caused by Different Sites of Action?). *Plant Physiology*, *114*, 401–403.
- Staelin, L. A., & Moore, I. (1995). The plant golgi apparatus: Structure, functional organization and trafficking mechanisms. *Plant Physiology*, *46*, 261–288.
- Tamura, K., Shimada, T., Ono, E., Tanaka, Y., Nagatani, A., Higashi, S., Hara-Nishimura, I. (2003). Why green fluorescent fusion proteins Plants., have not been observed in the vacuoles of higher. *Plant Journal*, *35*, 545–555.
- Tang, J. (2004). *Pepsin A*. In: Barrett, A.J., Rawlings, N.D., Woessner, J.F. (Eds.). *Handbook of Proteolytic Enzymes*. Elsevier, London, pp. (Vol. 19–28).
- Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A., & Blundell, T. L. (1978). Structural evidence for gene duplication in the evolution of the acid proteases. *Nature*, *271*(9), 618–621.
- Terauchi, K., Asakura, T., Ueda, H., Tamura, T., Tamura, K., Matsumoto, I., Abe, K. (2005). Plant Specific Insertions in the Soybean Aspartic Proteinases, soyAP1 and soyAP2, Perform Different Functions of Vacuolar Targeting. *Journal of Plant Physiology*, *163*, 856–862.
- Terauchi, K., Asakura, T., Ueda, H., Tamura, T., Tamura, K., Matsumoto, I., Abe, K. (2006). Plant-specific insertions in the soybean aspartic proteinases, soyAP1 and soyAP2, perform different functions of vacuolar targeting. *Journal of Plant Physiology*, *163*(8), 856–862.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, *15*(25), 4876–82.
- Timmer, J. C., & Salvesen, G. S. (2007). Caspase substrates. *Cell Death Differentiation*, *14*(1), 66–72.
- Törmäkangas, K., Hadlington, J. L., Pimpl, P., Hillmer, S., Brandizzi, F., Teeri, T. H., & Denecke, J. (2001). A vacuolar sorting domain may also influence the way in which proteins leave the

- endoplasmic reticulum. *Plant Cell*, 13(9), 2021–32.
- Tsiatsiani, L., Gevaert, K., & Van Breusegem, F. (2011). Natural substrates of plant proteases: How can protease degradomics extend our knowledge? *Physiologia Plantarum*, 145(1), 28–40.
- Vairo Cavalli, S., Claver, S., Priolo, N., & Natalucci, C. (2005). Extraction and partial characterization of a coagulant preparation from *Silybum marianum* flowers. Its action on bovine caseinate. *Journal Dairy Research*, 72(3), 271–275.
- Vairo Cavalli, S., Lufrano, D., Colombo, M. L., & Priolo, N. (2013). Properties and applications of phytepsins from thistle flowers. *Phytochemistry*, 92, 16–32.
- Vairo Cavalli, S., Silva, S. V., Cimino, C., Malcata, F. X., & Priolo, N. (2008). Hydrolysis of caprine and ovine milk proteins, brought about by aspartic peptidases from *Silybum marianum* flowers. *Food Chemistry*, 106(3), 997–1003.
- van der Hoorn, R. A. L. (2008). Plant proteases: from phenotypes to molecular mechanisms. *Annual Review of Plant Biology*, 59, 191–223.
- Van Deurs, B., Holm, P. K., Kayser, L., Sandvig, K., & Hansen, S. H. (1993). Multivesicular bodies in Hep-2 cells are maturing endosomes. *European Journal of Cell Biology* 61, 208–224.
- Verissimo, P., Esteves, C., Faro, C., & Pires, E. (1995). The vegetable rennet of *Cynara cardunculus* L. contains two proteinases with chymosin and pepsin-like specificities. *Biotechnology Letter*, 17(6), 621–626.
- Vieira, M., Pissarr, J., Verissimo, P., Castanheira, P., Costa, Y., Pires, E., & Faro, C. (2001). Molecular cloning and characterization of cDNA encoding cardosin B, an aspartic proteinase accumulating extracellularly in the transmitting tissue of *Cynara cardunculus* L. *Plant Molecular Biology*, 45(5), 529–39.
- Viotti, C. (2014). ER and vacuoles: never been closer. *Frontiers in Plant Science*, 5(February), 20. h
- Vitale, A., & Pedrazzini, E. (2005). Recombinant pharmaceuticals from plants: The plant endomembrane system as bioreactor. *Molecular Interventions*, 5, 216–225.
- Vitale, A., & Raikhel, N. (1999). What do proteins need to reach different vacuoles? *Plant Science*, 4(4), 149–155.
- Voinnet, O., Rivas, S., Mestre, P., & Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *The Plant Journal*, 33, 949–956.
- Wang, J., Li, Y., Lo, S. W., Hillmer, S., Sun, S. S. M., Robinson, D. G., & Jiang, L. (2007). Protein mobilization in germinating mung bean seeds involves vacuolar sorting receptors and multivesicular bodies. *Plant Physiology*, 143, 1628–1639.
- Wang, S., Horimoto, Y., Dee, D. R., & Yada, R. Y. (2014). Understanding the mechanism of prosegment-catalyzed folding by solution NMR spectroscopy. *Journal Biology Chemistry* 289, 697–707.
- White, P. C., Cordeiro, M. C., Arnold, D., Brodelius, P. E., & Kay, J. (1999). Processing, activity, and inhibition of recombinant cyprosin, an aspartic proteinase from cardoon (*Cynara cardunculus*).

Journal Biology Chemistry, 274(24), 16685–93.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(2), 407–410.

Wild, P., McEwan, D. G., & Dikic, I. (2014). The LC3 interactome at a glance. *Jurnal Cell Science*, 127, 3–9.

Xiang, L., Etxeberria, E., & Van Den Ende, W. (2013). Vacuolar protein sorting mechanisms in plants. *FEBS Journal*, 280(4), 979–993.

You, H. X., Qi, X., Grabowski, G. A., & Yu, L. (2003). Phospholipid membrane interactions of saposin C: in situ atomic force microscopic study. *Biophysiology Journal*, 84, 2043–2057.