



---

---

**UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO.**



---

**INSTITUTO DE BIOTECNOLOGÍA.**

**“Co-evolución de residuos:  
Restricción evolutiva de la  
estructura y la función de la  
profilina”**

**TESIS**

**QUE PARA OBTENER EL GRADO DE  
MAESTRO EN CIENCIAS BIOQUÍMICAS  
PRESENTA**

**Lucio Ricardo Montero Valenzuela**

Asesor de tesis

Dr. Federico Sánchez Rodríguez.

Cuernavaca, Morelos

Enero de 2008.



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**Esta tesis fue dirigida por el Comité Tutorial formado por:**

**Dr. Federico Sánchez Rodríguez (Tutor principal).**

**Dr. Enrique Rudiño Piñera.**

**Dra. Carmen Nina Pastor Colón.**

**Agradezco la asesoría y participación de M. en C. Gabriel Guillén Solís para el desarrollo de esta tesis.**

**Esta tesis fue realizada con el apoyo económico de una beca otorgada por el CONACYT.**

**Esta tesis está dedicada a mis padres, pues por ellos soy lo que soy.**

## **Agradecimientos.**

-A Federico Sánchez, mi tutor, que me dio las facilidades para hacer mi tesis de maestría.

-A Gabriel Guillén, por su paciencia y por su asesoría para el trabajo de tesis y para su redacción.

-A Rama Ranganathan, por permitirme usar su software "Statistical Coupling Analysis" dentro de mi trabajo de tesis.

-A mis padres, que siempre me alentaron en mis estudios.

-A Sara, mi hermana, por su invaluable compañía y apoyo.

-A Hugo, por ser un cuñado excepcional.

-A la Universidad Autónoma de Aguascalientes, donde me inicié en el fascinante mundo de la investigación al estudiar la carrera de Ingeniería Bioquímica.

-Al CONACYT, por la beca que me permitió realizar mis estudios de maestría.

## Contenido.

Agradecimientos.....	iii
Contenido.....	iv
Lista de figuras.....	v
Abreviaturas.....	1
Resumen.....	2
Abstract.....	3
Introducción.....	5
Antecedentes.....	10
Hipótesis.....	17
Objetivos generales.....	17
Objetivos específicos.....	17
Metodología.....	18
Alineamiento múltiple de las secuencias de la familia de la profilina.....	18
<i>Validación del alineamiento múltiple de las secuencias de amino ácidos de la familia de proteínas de la profilina.</i> .....	18
Análisis Estadístico de Acoplamiento (SCA) en la familia de proteínas de profilina.....	20
Representación molecular de la profilina.....	21
Anclaje molecular de la profilina con moléculas orgánicas.....	21
Resultados.....	23
Discusión.....	38
Conclusiones.....	45
Perspectivas.....	45
Referencias.....	46

## Lista de figuras.

Figura 1: Fundamento matemático del Análisis Estadístico de Acoplamiento .....	7
Figura 2: Estructura tridimensional de diferentes isoformas de profilina .....	11
Figura 3: Representación grafica del grado de frecuencia de cada una de las posiciones de las secuencias de profilinas eucarióticas.....	15
Figura 4: Validación de la significancia de las energías estadísticas de conservación y de acoplamiento de las profilinas.....	19
Figura 5: Distribución de amino ácidos de las posiciones del alineamiento de amino ácidos de las profilinas.....	24
Figura 6: Energía estadística de conservación en cada una de las posiciones de las profilinas .....	25
Figura 7: Energía estadística de coevolución entre las posiciones de la secuencia de amino ácidos de la familia de profilina .....	26
Figura 8: Comparación de los perfiles de energías estadísticas de coevolución de los sitios perturbados .....	28
Figura 9: Energía estadística de acoplamiento entre posiciones de la secuencia de amino ácidos de la familia de la profilina .....	29
Figura 10: Zona de mayor energía de acoplamiento entre posiciones de la secuencia de amino ácidos de la familia de la profilina .....	30
Figura 11: Localización estructural de los pares de posiciones con altos valores de energía estadística de acoplamiento en la familia de proteínas de las profilinas .....	32
Figura 12: Localización estructural de las zonas de posiciones con altos valores de energía estadística de acoplamiento en la familia de proteínas de las profilinas .....	33
Figura 13: Posible sitio de unión de las profilinas vegetales a una molécula orgánica	42

## Abreviaturas.

kDa	Kilodaltones.
PLP	Poli-L-prolina.
PIP <sub>2</sub>	Fosfatidil inositol 4,5-bifosfato.
SCA	Statistical Coupling Analysis. Análisis Estadístico de Acoplamiento.
PI3K	Fosfatidil inositol 3-cinasa.
PRM	Proline-rich motifs. Motivos ricos en prolina.
ADN	Ácido desoxiribonucleico.
ARN	Ácido ribonucleico.
ADP	Adenosine difosfate. Difosfato de adenosina.
ATP	Adenosine trifosfate. Trifosfato de adenosina.
NCBI	National Center for Biodiversity Information©. Centro Nacional de Información sobre la Biodiversidad (en EUA).
SRS	Sequene Retrieval System©. Sistema de Recuperación de Secuencias.
PDB	Protein Data Bank©. Banco de Datos de Proteínas.



## Resumen.

Uno de los principios fundamentales de la bioquímica es que la secuencia de amino ácidos de una proteína determina su estructura tridimensional y su función bioquímica [1]. Sin embargo, la manera en que la secuencia de aminoácidos de una proteína determina éstos no se conoce completamente. El análisis de los alineamientos múltiples de las secuencias de amino ácidos en las familias de proteínas ha demostrado que la mayoría de los miembros de una familia suelen tener el mismo amino ácido en ciertas posiciones; también se ha visto que otras posiciones de la secuencia tienen cambios conservativos y en otras puede haber sustituciones de amino ácidos de cualquier tipo. Recientemente, se ha visto que la conservación de amino ácidos en ciertas posiciones no es suficiente para mantener la estructura y la función de una proteína en los diferentes organismos; ahora se sabe que existen interacciones importantes entre los amino ácidos adyacentes entre sí en la estructura terciaria, que necesitan permanecer en los distintos miembros de una familia de proteínas para que mantengan su integridad estructural y funcional. Muchas de estas interacciones se reflejan en la coevolución de las posiciones de los amino ácidos participantes en ellas [1]. Actualmente, existen varios métodos que intentan medir la coevolución de las posiciones de amino ácidos en una familia de proteínas; sin embargo, el Análisis de Acoplamiento Estadístico (SCA) se ha usado en diferentes familias de proteínas, y los resultados obtenidos han permitido predecir los amino ácidos importantes para la función y el plegamiento [1-5]. En este estudio se aplicó el Análisis Estadístico de Acoplamiento a la familia de proteínas de la profilina para determinar las posiciones que son importantes para mantener la similitud estructural y sus propiedades bioquímicas. La profilina es una proteína de 14 kDa que modula la formación del citoesqueleto de actina, ya sea secuestrando la actina monomérica o acelerando el intercambio del ADP por ATP; por lo que la profilina puede favorecer la polimerización y la despolimerización de los filamentos de actina. La comparación de las secuencias de amino ácidos de diferentes miembros de la familia de las profilinas ha revelado que las isoformas vegetales tienen un bajo porcentaje de identidad con las isoformas de otros organismos eucariontes. El análisis de las estructuras de la profilina demostró que todas las isoformas presentan una alta similitud estructural. Por lo tanto, parece ser que esa similitud de las profilinas a nivel estructural puede

explicarse por que, además de que unos pocos amino ácidos están conservados, también se conservan las relaciones entre los amino ácidos de ciertas posiciones, esta correlación se puede estudiar mediante el SCA.

El SCA detectó que los grupos de amino ácidos de la profilina que coevolucionan están en las zonas de interacción de la profilina con sus distintos ligandos. Por ejemplo, el SCA de las isoformas vegetales de la profilina detectó grupos de residuos que coevolucionan cercanos a los sitios de unión a PI3K, a PLP y a actina. También se encontró un grupo de residuos coevolucionados cerca de una cavidad que existe sólo en las profilinas vegetales, que podría interactuar con una molécula orgánica [6]. Por otro lado, el SCA de todas las profilinas eucarióticas detectó coevolución entre los residuos de la interfase de la interacción de la profilina con la actina. Esto es congruente con el hecho de que se ha visto que las distintas clases de profilina coevolucionan con las clases de actina con las que interactúan *in vivo* [7]. Esas posiciones no aparecen en el SCA de las profilinas vegetales, lo cual parece deberse a la alta conservación de estas posiciones entre las profilinas vegetales.

## **Abstract.**

A fundamental tenet in biochemistry is that the amino acid sequence of a protein determines its tridimensional structure and its biochemical function [1]. However, the way this occurs is still not well understood. The analysis of the multiple alignments of the amino acid sequences of the protein families has demonstrated that most of the members of a protein family tend to have the same amino acid in certain positions, other positions have conservative changes and others can have any type of amino acid change. Recently, it has been shown that the amino acid conservation at certain positions is not enough to maintain the structure and function of a protein. It's known that there are important interactions between the amino acids that are adjacent in the tertiary structure, which need to be conserved in different members of a protein family, to maintain their structural and functional integrity. Many of these interactions are reflected in the coevolution of those positions participating in the structure and function of a protein [1]. Actually, there are several methods to try to measure the coevolution of the positions in a protein sequence, however, the Statistical Coupling Analysis (SCA) has been successfully used in several protein families to predict those amino acids that are important for proteins to

fold and function [1-5]. In this study, the SCA was applied to the profilin protein family to determine the important positions for maintaining its structural similarity and its biochemical properties. Profilin is a 14 kDa protein that modulates the actin cytoskeleton formation by two ways: it can promote the actin filament depolymerization by sequestering the monomeric actin and it can promote the actin filament polymerization by accelerating the exchange of ADP for ATP. The comparison of the amino acid sequences of the different members of the profilin family has revealed that plant isoforms have a low identity with profilins from other eukaryotes. Profilin structures show that all the isoforms are structurally similar, so it seems that this similarity could be explained by the fact that, in addition to the conservation in few positions, there are conserved relationships among amino acids from some particular positions, correlation that can be studied by the SCA program.

The SCA detected that the amino acid groups that coevolve are in the zones of the profilin that interacts with its ligands. For example, the SCA for the vegetal profilin isoforms detected residue groups that coevolve and are close to the sites of interaction with PI3K, PLP and actin. Also, a coevolving residue group close to a plant-specific pocket, that could represent an interacting site for an unknown plant-specific ligand, was found. On the other hand, the SCA for all the eukaryotic profilin isoforms detected coevolution among the residues of the interface of interaction of the profilin with the actin. This issue is congruent with the fact that different classes of profilin have coevolved with the actin classes they interact with *in vivo* [7]. These set of positions do not appear in the SCA for the plant profilins since a high conservation of these positions, among the plant profilins, exists.

## Introducción.

Las proteínas se forman por la unión secuencial de moléculas de amino ácidos, los cuales son moléculas orgánicas que tienen un extremo amino, un extremo carboxilo y una cadena lateral que es distinta según el tipo de amino ácido. De todos los amino ácidos que existen, solamente 20 pueden ser usados por las células para la síntesis de proteínas, por lo que, para una secuencia de 100 amino ácidos existen  $100^{20}$  (alrededor de  $10^{130}$ ) combinaciones posibles de éstos. Esta es una manera fascinante que tiene la naturaleza para producir moléculas que llevarán a cabo funciones muy diversas entre sí, pues todas las proteínas se sintetizan del mismo modo como una cadena lineal de amino ácidos, pero según su secuencia de amino ácidos se pliegan en una cierta estructura tridimensional y ese plegamiento les permite llevar a cabo una función específica [8].

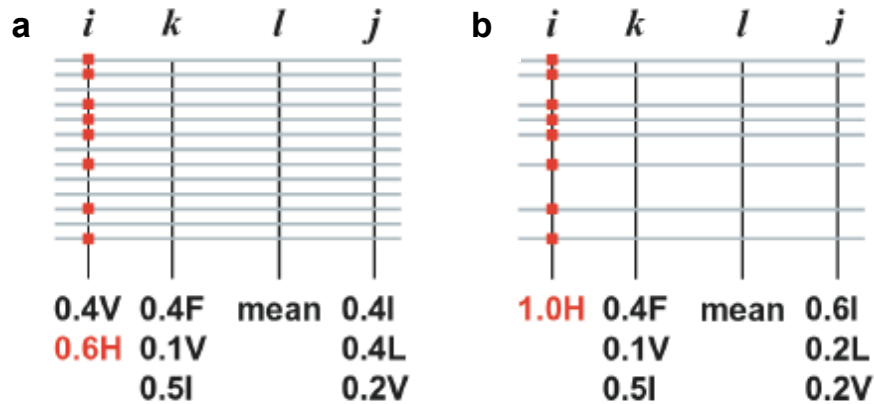
Al conjunto de proteínas con funciones similares distribuidas en diversos organismos se les denomina familias de proteínas ortólogas. Las proteínas de una misma familia no son idénticas pero sí tienen una alta similitud a nivel de secuencias. Estas diferencias en sus secuencias se deben a que han evolucionado de una manera independiente desde un ancestro común. Durante la evolución de las especies han existido mutaciones (cambios) en el ADN que se manifiestan como sustituciones, inserciones y deleciones en ciertos amino ácidos en cada uno de los miembros de la familia. Estas mutaciones son el producto de la selección natural debido a que estos cambios le permiten a la proteína tener una función igual o mejor, en un contexto particular, a la que tenía antes de la mutación, por lo tanto un análisis de similitudes o diferencias entre los miembros de una familia de proteínas puede proporcionar información necesaria que nos permita predecir los amino ácidos que son importantes para la función o para el plegamiento de las proteínas.

Las proteínas están compuestas por uno o más segmentos denominados dominios, que se pliegan con independencia del resto de la proteína, y contribuyen a cierta funcionalidad de la misma, tal como la adhesión a otra proteína o al ADN o la transformación catalítica de un sustrato. Se cree que esos dominios han mantenido el mismo plegamiento que su ancestro común, a pesar de que haya sustituciones, inserciones y deleciones de amino ácidos en algunas posiciones. Por medio de un alineamiento múltiple de secuencias se puede ver dónde

es más probable que se hayan dado estos cambios en las distintas secuencias de una familia de proteínas.

Comparando las secuencias en el alineamiento múltiple de una familia de proteínas, se puede ver que algunas posiciones de amino ácidos están muy conservadas, lo cual sugiere que podrían ser importantes para la estructura y la función de las proteínas. Sorprendentemente, en el caso de algunas familias, la cantidad de amino ácidos que se necesita conservar es muy pequeña: por ejemplo, en el caso de las profilinas, sólo existe una identidad de entre un 5 y un 30% entre las isoformas vegetales y las de los animales, sin embargo sus estructuras son muy parecidas [9].

Recientemente, se ha visto que además de la conservación de los amino ácidos en algunas posiciones de una familia de proteínas, es necesario que se conserven las interacciones entre los residuos de otras posiciones para determinar la estructura y la función de una proteína [1, 10-12]. La interacción entre dos residuos de amino ácidos se puede determinar experimentalmente con la técnica del ciclo termodinámico de la doble mutante, que consiste en mutar uno de estos residuos para que la proteína pierda actividad y luego observar cómo se rescata esta actividad al introducir otra mutación en otro residuo [2]. Sin embargo, este método es costoso y laborioso por lo que sólo se pueden medir unas pocas interacciones [2]. Estas interacciones también se pueden inferir bioinformáticamente analizando el alineamiento múltiple de las secuencias de amino ácidos de una familia de proteínas, dado que cuando la presencia de un amino ácido en una posición influye sobre el amino ácido presente en otra posición (Figura 1) hay coevolución entre ellos y esto sugiere que hay una interacción entre ellos que se ha mantenido a lo largo de la evolución [2].



**Figura 1: Fundamento matemático del Análisis Estadístico de Acoplamiento.** a) Alineamiento múltiple de secuencias de una familia de proteínas, cuyas secuencias están representadas por líneas horizontales, y que tiene las posiciones *i*, *j*, *k* (medianamente conservadas) y *l* (no conservada). En la parte inferior se especifica los valores de frecuencia de los amino ácidos en esas posiciones. Las secuencias que tienen una histidina (H) en la posición *i* están marcadas con un punto rojo. b) Subalineamiento resultante de perturbar *i* (eliminar las secuencias que no tienen la His en esa posición). La frecuencia de amino ácidos tiene una variación en la posición *j* comparado con el alineamiento anterior, sin embargo no se ve ningún cambio en las posiciones *k* y *l*, por lo que, existe una coevolución del amino ácido de la posición *i* con los amino ácidos de la posición *j* pero no con los de la posición *k* y la *l*. Figura tomada de Evolutionarily conserved networks of residues mediate allosteric communication in proteins, Ranganathan et al, Nat. Struct. & Mol. Biol., 2003.

La coevolución entre las posiciones de una familia de proteínas se puede medir mediante el Análisis Estadístico de Acoplamiento (SCA, por sus siglas en inglés). Este método fue desarrollado por el Dr. Rama Ranganathan, de la "University of Texas Southwestern Medical Center at Dallas" [4], y se ha utilizado para analizar distintas familias de proteínas como los dominios PDZ [2], los cuales son segmentos de proteína que reconocen motivos ricos en prolina (PRM) y se pliegan independientemente de la proteína de la que forman parte [2], la familia de las proteínas G [3], los receptores acopladas a proteínas G, las proteasas de serina, las hemoglobinas [4] y los receptores X de retinoides [5]. Para estas familias de proteínas, el SCA detectó redes de residuos que interaccionan entre sí, indistinguibles en la estructura terciaria de las proteínas, importantes en la respuesta de las mismas a la unión con sus ligandos.

Asimismo, el Análisis Estadístico de Acoplamiento se utilizó con la familia de proteínas que contienen los dominios WW; estos dominios, nombrados así por que tienen un motivo conservado Trp-Trp, reconocen regiones ricas en prolina y son segmentos proteicos que se pliegan independientemente del resto de la proteína de la que forman parte. Estos dominios se dividen en cuatro clases en donde la clase I tiene afinidad por regiones ricas en prolinas del

tipo PPxY, la II por las de tipo PPLP, la III se une a las de tipo PPR y la IV reconoce las secuencias de tipo pS/pT-P (donde P es prolina, x significa cualquier amino ácido y p significa que el amino ácido está fosforilado). Los resultados obtenidos por Lockless y colaboradores les permitieron diseñar secuencias de amino ácidos con plegamientos similares a los dominios WW y con una afinidad a las secuencias ricas en prolina [1].

El propósito de este proyecto es aplicar el Análisis Estadístico de Acoplamiento a la familia de proteínas de la profilina. Esta familia está compuesta por proteínas citosólicas de bajo peso molecular ( $\approx 14$  kDa) que se encuentran en todos los organismos eucarióticos [13] y regulan positiva y negativamente la polimerización de los filamentos de actina. Esta proteína está organizada estructuralmente en tres regiones que le permiten interactuar *in vitro* con los monómeros de actina [14], con el fosfatidilinositol 4-5-bifosfato ( $\text{PIP}_2$ ) [15] y con secuencias ricas en prolina, que pueden ser cadenas de poli L-prolina (PLP) [14] o motivos ricos en prolina (PRM, por sus siglas en inglés) de algunas proteínas [14, 16]. La profilina une los tres tipos de ligando antes mencionados con afinidades que difieren entre las profilinas de distintas isoformas, lo que la hace una proteína multifuncional [17, 18].

La profilina parece estar implicada en conectar vías de señalización con la polimerización de la actina. Por ejemplo, la profilina podría conectar la vía de los fosfoinosítidos con la polimerización de la actina [15, 19], ya que las regiones de interacción de la profilina con el  $\text{PIP}_2$  y la actina se sobrelapan parcialmente, lo que puede hacer que la profilina libere a la actina en los filamentos en crecimiento en presencia del  $\text{PIP}_2$ , proceso que se interrumpe cuando se activa la vía del fosfoinosítido y la fosfolipasa C degrada el  $\text{PIP}_2$  [20]. Por otro lado, la interacción de la profilina con motivos ricos en prolina también puede permitir la actividad de la profilina como reguladora de la polimerización de la actina. El módulo de interacción de la profilina con la PLP y los PRM fue descrito inicialmente como un dominio que interacciona con proteínas que contenían al menos 6 prolinas consecutivas en su secuencia de amino ácidos [21]; sin embargo, recientemente Witke y colaboradores evidenciaron que este dominio también puede reconocer secuencias con motivos ricos en prolinas del tipo ZPPX (donde Z puede ser Pro, Gly o Ala; y X cualquier amino ácido hidrofóbico) [22]. Se ha demostrado *in vitro* que la profilina interacciona con proteínas con motivos ricos en prolina involucradas en las adhesiones focales, la endocitosis [16], el tráfico vesicular, el remodelamiento de membrana, la formación

de los axones neuronales y las fibras de estrés [20], y se ha resuelto la estructura terciaria de la profilina en complejo con poli-L-prolina [23], también se han resuelto dos estructuras de complejos de profilina con sendos péptidos ricos en prolina de VASP [16] (vasodilator-stimulated phosphoprotein), la cual es una proteína implicada en la vascularización y la endocitosis. Además, se ha comprobado *in vivo* que la profilina es necesaria para el desarrollo embrionario, y que la bacteria patógena *Listeria monocytogenes* activa la profilina de su hospedero para desplazarse dentro de él. Recientemente se ha visto que la profilina también se localiza en el núcleo celular y tiene funciones reguladoras de la expresión genética. Un ejemplo de esta función es que regula dentro del núcleo al factor transcripcional p42<sup>POP</sup> que se encuentra en los tejidos de ratón [24]. Además se ha visto que la profilina optimiza, en complejo con la actina monomérica, la transcripción del ADN del virus sincitial respiratorio [25]. También se ha visto que los anticuerpos anti-profilina interfieren *in vivo* con el *splicing* del ARN mensajero, y que las profilinas interaccionan con la proteína SMN, que es un factor nuclear regulador del *splicing*, y colocalizan con ésta en el núcleo.

Actualmente la familia de la profilina está compuesta por 257 miembros que se han aislado de diferentes organismos eucarióticos, de las cuales existen 16 estructuras cristalográficas. A partir de esta información se ha visto que, a pesar de que el porcentaje de identidad entre las profilinas animales y las vegetales es del 5 al 30%, sus estructuras terciarias son muy similares entre sí, pues la superposición de la estructura de la profilina de *Arabidopsis thaliana* y la estructura de la bovina da una raíz de la desviación media cuadrática (RMSD) de 1.61 Ångstroms para los carbonos  $\alpha$  [6]. Por ello, se cree que no es suficiente la conservación de algunos amino ácidos para que la profilina tenga su estructura y su función, sino que también necesitan conservarse las relaciones entre algunos otros residuos, las cuales se pueden medir mediante el SCA.

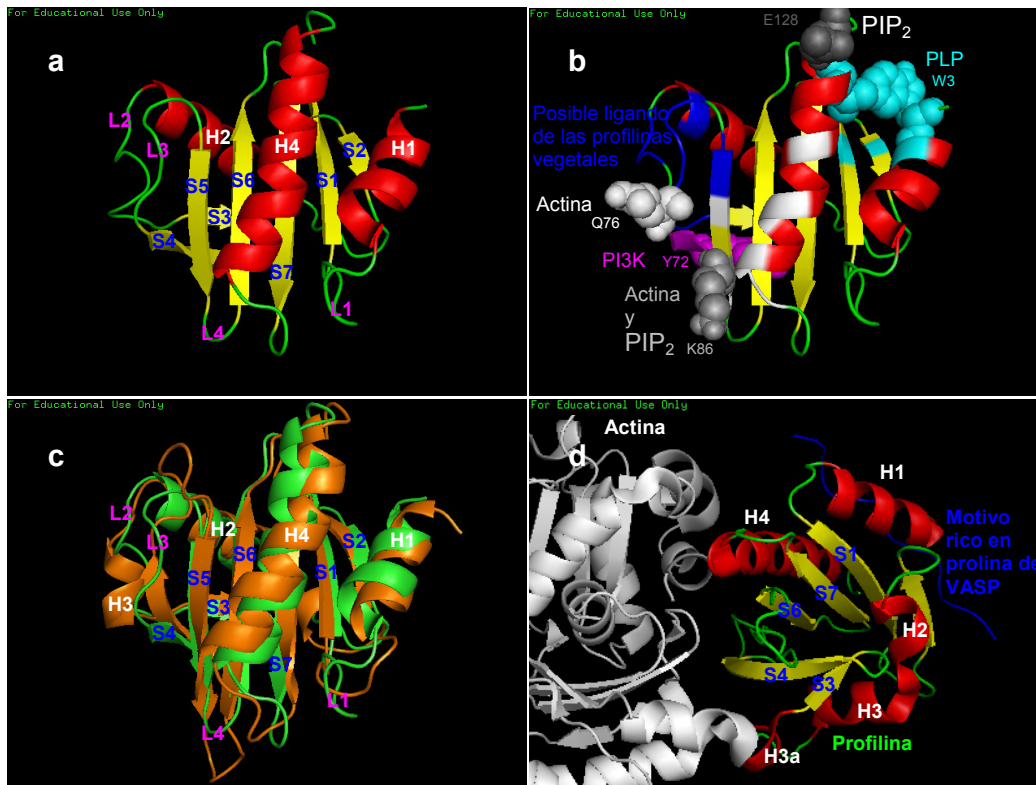


## Antecedentes.

La interacción de la profilina con la actina tiene un efecto dual en la dinámica de los microfilamentos, ya que puede inhibir su polimerización secuestrando los monómeros de actina en una relación 1:1 (actina:profilina) [26] o favorecerla, intercambiando el ADP por ATP en los monómeros de actina [20]. Se ha visto que las profilinas animales aceleran 1,000 veces [27] el intercambio de ATP por ADP en los monómeros de actina [20, 27]; sin embargo, las profilinas de levadura [28] y vegetales tienen [29] poca eficiencia en acelerar ese intercambio, pero sí pueden favorecer la polimerización de la actina, lo que indica que este intercambio no es muy importante para la formación de los filamentos de actina en estos organismos [28, 29].

La estructura de la profilina (Figura 2 a) tiene la forma de un sándwich; tiene dos regiones de  $\alpha$ -hélices separadas por una hoja  $\beta$  principal. De un lado de la hoja  $\beta$  están las  $\alpha$ -hélices amino y carboxilo terminal (H1 y H3, respectivamente), las cuales se ha demostrado que interaccionan con motivos ricos en prolina (PRM) [16] y con poli-L-prolina (PLP) [16, 21]. Los residuos de estas  $\alpha$ -hélices forman un interior hidrofóbico, en el que se alojan la PLP [23] y los PRM [16] al unirse a la profilina. La  $\alpha$ -hélice H1 se empaqueta estrechamente con las hebras  $\beta$  S1 y S2, y la  $\alpha$ -hélice H3 está junto a ésta, empaquetada con las hebras  $\beta$  S5, S6 y S7. Del otro lado de la hoja  $\beta$  están las pequeñas  $\alpha$ -hélices H2 y H3, localizadas entre dos cadenas aleatorias largas, que junto con una horquilla  $\beta$  que contiene las hebras  $\beta$  S3 y S4 forman un interior hidrofóbico ocupado por residuos alifáticos. Este interior hidrofóbico está sellado por arriba por el giro  $\beta$  que hay entre las hebras  $\beta$  S1 y S2 (Figura 2 a).

La interacción de la profilina con la poli-L-prolina [23] y con algunos motivos ricos en prolina de proteínas como VASP (proteína estimulada durante la vasodilatación) [16] se lleva a cabo en un motivo muy conservado en todas las profilinas, el cual forma un canal hidrofóbico donde se alojan las prolinas y se encuentra localizado en el amino y el carboxilo terminal de la profilina [23] (Figura 2 d). En la profilina de *Arabidopsis thaliana*, este motivo incluye los residuos aromáticos Trp3 y Tyr6 en la  $\alpha$ -hélice N-terminal (H1), Ile25 en la hebra  $\beta$  S1, Trp33 al final del anillo que conecta las hebras  $\beta$  S1 y S2, así como Tyr125 y Leu126 en la  $\alpha$ -hélice C-terminal (H3). Además, se ha visto que algunas proteínas con motivos ricos en prolina, como la fosfatidil



**Figura 2: Estructura tridimensional de diferentes isoformas de profilina.** a) Profilina de *Arabidopsis thaliana* (PDB: 3NUL). Las letras H indican las  $\alpha$ -hélices, las S las hebra  $\beta$  y las L las asas que tienen diferente tamaño en las profilinas animales que en las vegetales. b) Sitios de interacción de la profilina de *Arabidopsis thaliana* con PLP (azul claro), PIP<sub>2</sub> (gris), PI3K (lila), actina (blanco) y un posible ligando de las profilinas vegetales (azul marino). c) Alineamiento estructural de la profilina de *Arabidopsis thaliana* con la profilina del complejo profilactina bovino (PDB: 2BTF) d) Profilina humana en complejo con actina (blanco) y con un motivo rico en prolina de VASP (azul) (PDB: 2PAV).

inositol 3-cinasa, interaccionan con la tirosina 72 de la profilina de frijol [30], que se encuentra en un motivo conservado en todas las profilinas vegetales [30] localizado en la estructura de la profilina I de *Arabidopsis thaliana* en una curvatura  $\beta$  que incorpora la hebra  $\beta$  S4 en la hoja  $\beta$  principal de la profilina.

En la profilina II de *Acanthamoeba castellanii* se cree que la interfase de interacción de la profilina con el PIP<sub>2</sub> está formada por 2 módulos de residuos positivos dispersos en toda una cara de la estructura de la profilina. Uno de los módulos está cercano al sitio de interacción con la actina, y se forma por los residuos Arg57, Arg67, Arg72, Lys116, Lys81 y Lys82. Los residuos Lys81 y Lys82 forman parte de la interfase de la interacción de la profilina tanto con el PIP<sub>2</sub> como con la actina, lo que explica la competitividad entre el PIP<sub>2</sub> y la actina en su interacción con la profilina [31]. Este solapamiento entre los sitios de unión al PIP<sub>2</sub> y a la actina

también se ha visto en el residuo correspondiente a la Lys81 de la profilina II de *Acanthamoeba castellanii* en la profilina humana, donde hay una sustitución por una arginina (Arg88) [32]. Estos residuos están separados del otro módulo por unos residuos negativos. Este otro módulo está cercano al sitio de interacción de la profilina con los PRM y tiene a His25, Lys51, Arg76, Lys91 y Lys94 en la profilina II de *Acanthamoeba castellanii*. La profilina I de *Acanthamoeba castellanii* tiene dos residuos positivos menos que la II, pues tiene una leucina en vez de la His25 y una asparagina en vez de la Lys51. Esto puede ser suficiente para determinar que la afinidad de la profilina I por el PIP<sub>2</sub> sea 10-50 veces menor que esta afinidad en la isoforma II. Las profilinas de mamíferos son más afines al PIP<sub>2</sub> que las de *Acanthamoeba castellanii*, lo que se puede explicar por que las primeras tienen una línea de residuos básicos en la  $\alpha$ -hélice amino-terminal que no tienen las profilinas de *Acanthamoeba castellanii* junto al sitio de interacción con los PRM [31].

La estructura de la interfase de la interacción de la profilina con la actina (Figura 2 d), está pobremente conservada entre las isoformas de la profilina de animales y vegetales. Las posiciones de la interfase de la interacción de la profilina con la actina están poco conservadas, a excepción de las correspondientes a la Lys86 y la Gly113 de la profilina de *Arabidopsis thaliana* [6, 28].

Los resultados obtenidos en la determinación de la estructura cristalina del complejo profilina-actina bovino, así como las mutaciones en la profilina de levadura y los estudios bioquímicos en profilina bovina y de *Acanthamoeba* han revelado que los residuos responsables del reconocimiento de los monómeros de actina son algunos amino ácidos de las  $\alpha$ -hélices H3 y H4 además de algunos de las hebras  $\beta$  plegadas S4, S5 y S6 [33-36].

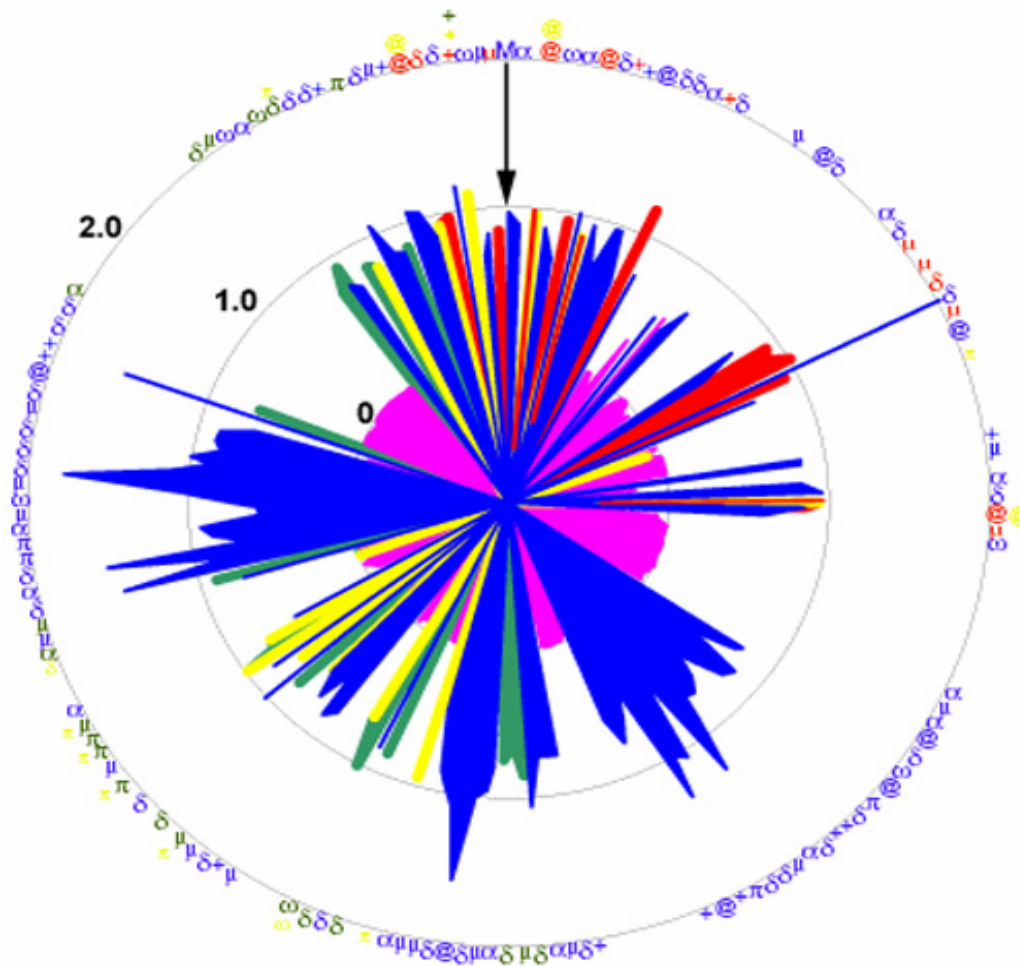
Se ha demostrado por medio de ensayos de interacción *in vitro* de profilinas mutantes que los residuos Phe59, Phe60, Gly121 y Lys126 de la profilina bovina son importantes para la interacción con los monómeros de actina [37]. Para la profilina humana, los ensayos de interacción *in vitro* han demostrado la importancia de las posiciones Trp4, Arg56, Lys70, Arg89, Lys91, Lys126 e His134 para la interacción con la actina [32, 35, 38-41]. En el caso de la profilina de levadura, los estudios de interacción *in vitro* [28] e *in vivo* [34, 40] han demostrado que las posiciones Arg72, Arg76, Arg81 y His82 son importantes para el pegado con la actina [28, 34, 40]. También se ha demostrado que algunos amino ácidos son importantes para la

exclusión del solvente en el sitio de interacción con la actina, por ejemplo, los residuos de las posiciones Glu56, Thr60, Thr82 y Thr122 de la profilina II humana [19].

Por otro lado, se ha demostrado en experimentos de afinidad *in vitro* que la eliminación de los amino ácidos Pro96 y Thr97 de la profilina I humana da como resultado una disminución de la afinidad de por la actina. Esta mutante, al ser expresada en levaduras que tienen suprimido el gen de la profilina, produce células de levadura con un fenotipo anormal en donde las fibras de estrés (filamentos de actina) no se deshacen [39]. Esto puede deberse a que los residuos antes mencionados están localizados en el asa L4 la cual, al igual que L5, conforman el dominio de unión a la actina [41]: Estas dos asas tienen una mayor longitud en las profilinas de mamíferos que en las profilinas vegetales, mientras que las profilinas vegetales presentan el asa L1, localizada entre la  $\alpha$ -hélice H1 y la hebra  $\beta$  S1, más grande que la correspondiente en las profilinas de mamíferos. Estas diferencias que se presentan en la interfase de interacción con la actina podrían explicar las diferencias bioquímicas que se presentan entre las profilinas vegetales y las de mamíferos [6].

Como podemos ver, las mutaciones de ciertos amino ácidos en la secuencia primaria de la profilina tienen un efecto en la afinidad, reflejada en la constante de disociación ( $K_d$ ), de la interacción de la profilina con sus ligandos. Estos resultados podrían explicar en parte la existencia de diferentes isoformas de profilinas en cada uno de los tejidos de los organismos, ya que cada una de estas isoformas tiene diferentes afinidades por la actina, los motivos ricos en prolina y el PIP<sub>2</sub>. Por otro lado, las secuencias de amino ácidos de las distintas isoformas de la profilina de un organismo presentan una identidad del 70% al 90% [42], lo cual sugiere que las variaciones entre las isoformas de la profilina podrían ser el factor importante que determine la afinidad de cada isoforma por cada uno de los ligandos. Esto puede verse al mapear las zonas de variación sobre la estructura de una de las profilinas. Para comprobar esta hipótesis, en un trabajo previo del laboratorio se realizó el análisis filogenético de todas las isoformas de la profilina depositadas en las bases de datos de proteínas [9]. El alineamiento múltiple de las secuencias de amino ácidos de las isoformas de la profilina reveló que las profilinas que pertenecen al mismo reino tienen un porcentaje de identidad entre el 60 y el 90 % y que cuando se comparan con las isoformas de otros reinos esta identidad disminuye al 5-30 %. En otras palabras, esto quiere decir que existen algunos amino ácidos que no han sido cambiados

durante el proceso evolutivo debido a que han sido seleccionados por su función. Por otro lado, cuando se determina el porcentaje de similitud de las secuencias de amino ácidos de todas las profilinas se observa que este porcentaje se incrementa de tres a cuatro veces comparado con el porcentaje de identidad, lo que muestra que la presión selectiva en la familia de las profilinas se da por la función de cada una de las posiciones, no por la secuencia. Existen algunas posiciones que reciben una exigencia funcional que les impide tener mutaciones, hay otras que toleran cambios por aminoácidos similares y otras más toleran cambios de cualquier tipo. Para poder medir el grado de variabilidad, en el trabajo de tesis de Bianca Flores-Bustos [9] se desarrolló un algoritmo el cual permite medir el índice de frecuencia en cada una de las posiciones de las secuencias de amino ácidos de la familia de la profilina. Los resultados del análisis son representados gráficamente en donde el valor de uno para una posición indica que ese residuo no cambia en todas las secuencias analizadas y un valor mayor a uno denota aquellas posiciones que tienen sustituciones por amino ácidos relacionados y finalmente aquellas posiciones que tienen un valor menor a uno implican que esas posiciones pueden tener sustituciones de amino ácidos no conservativas (Figura 3). Los resultados obtenidos con el algoritmo correlacionan con los resultados bioquímicos previamente reportados por otros autores [19, 28, 32-41]. Por ejemplo, los amino ácidos que en el análisis se encontró que tenían un valor de uno o mayor son los mismos que se han demostrado que son esenciales para la interacción de la profilina con el PLP, el PIP<sub>2</sub> y la actina. Sin embargo, en este análisis se encontraron otros amino ácidos que tienen un valor mayor a uno y que no se ha demostrado que sean importantes para la estructura o para la interacción de la profilina con la actina y con otros ligandos.



**Figura 3: Representación grafica del grado de frecuencia de cada una de las posiciones de las secuencias de profilinas eucarióticas.** El alineamiento múltiple de las secuencias de profilina de eucariontes generadas en el programa ClustalX 1.8 se utilizó para determinar el grado de variabilidad de cada posición de la secuencia de profilinas. Los amino ácidos que participan en la interacción con PLP, PIP<sub>2</sub> y actina están indicados en rojo, amarillo y verde respectivamente. Los grupos de amino ácidos se representan de la siguiente manera: δ: hidrofóbicos (Met/Pro/Ile/Leu/Val), @: aromáticos (Phe/Tyr/Trp/His), +: ácidos (Asp/Glu), π: básicos (Lys/Arg), ω: amidas (Asn/Gln), α: nucleofílicos (Ser/Thr/Cys), ψ: amino ácidos pequeños (Ala/Gly). Los círculos representan la escala del valor del factor de variabilidad (G<sub>F</sub>).

En la actualidad se ha visto que también es importante que se conserven las relaciones entre los amino ácidos de algunas posiciones del alineamiento múltiple de secuencias de una familia de proteínas. Estas relaciones conservadas entre aminoácidos se reflejan en la coevolución de las posiciones que les corresponden en el alineamiento. Esta coevolución también se puede dar entre amino ácidos de distintas proteínas, por ejemplo, se ha visto que las distintas clases de profilina han coevolucionado con las clases de actina con las que interactúan *in vivo*, pues la profilina de una clase suprime el fenotipo dado por la sobreexpresión de la actina de la clase

con la que interacciona *in vivo* pero no suprime el fenotipo dado por la actina de otra clase [7]. En el presente trabajo se buscará, mediante el Análisis Estadístico de Acoplamiento, la correlación que existe entre los amino ácidos de las distintas posiciones de la familia de las profilinas, para ver cómo se puede explicar que las profilinas de distintos reinos tengan tan alta similitud estructural y la misma función a pesar de tener tan bajo porcentaje de identidad entre sus secuencias.

## **Hipótesis.**

Por medio del Análisis Estadístico de Acoplamiento de las secuencias de los dominios WW, se ha podido identificar coevoluciones entre amino ácidos que reflejan interacciones químicas, mecánicas o funcionales entre éstos, por lo que se cree que aplicando este análisis a la familia de proteínas de la profilina se podrían predecir aquellas interacciones que son importantes para su plegamiento o para la interacción de ésta con sus ligandos.

## **Objetivos generales.**

Predecir mediante el Análisis de Acoplamiento Estadístico la coevolución entre los amino ácidos de distintas posiciones de la familia de la profilina, para encontrar interacciones que son importantes para el plegamiento o la función de esta familia de proteínas.

## ***Objetivos específicos.***

1. Identificar mediante el SCA las posiciones de amino ácidos de la profilina que han coevolucionado durante la especiación.
2. Identificar mediante el SCA las coevoluciones entre posiciones de amino ácidos que son exclusivas de las isoformas vegetales de la profilina.
3. Analizar la importancia biológica de las relaciones entre amino ácidos predichas por la coevolución entre las posiciones de la profilina.



## **Metodología.**

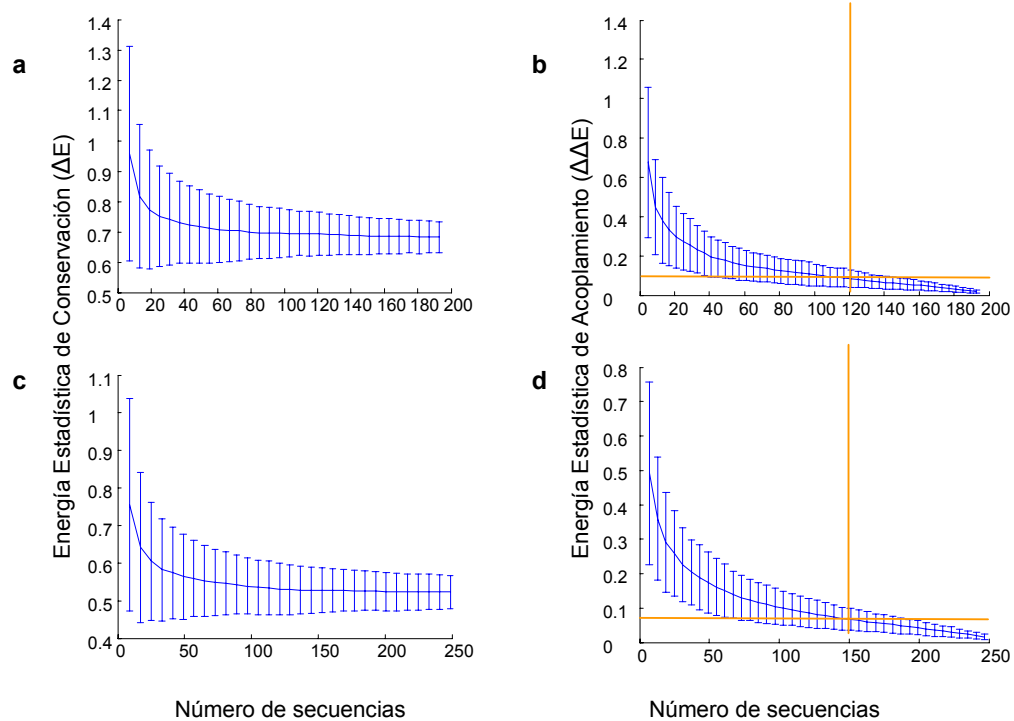
### ***Alineamiento múltiple de las secuencias de la familia de la profilina.***

Las secuencias de amino ácidos de la familia de las profilinas se obtuvieron de las bases de datos de NCBI<sup>©</sup> [43], SwissProt<sup>©</sup>, SRS<sup>©</sup> [44, 45] y UniProt<sup>©</sup> [46]. Las secuencias se filtraron para descartar aquellas iguales, incompletas, o pertenecientes a los virus. Por otro lado, se obtuvieron las coordenadas de las estructuras cristalográficas de las profilinas depositadas en la base de datos de estructuras de proteínas (PDB<sup>©</sup>) (<http://www.rcsb.org/pdb/home/home.do>) [47]. En los casos en que había dos o más estructuras para la misma isoforma de profilina, se seleccionó la que tuviera mayor resolución y estuviera cristalizada y resuelta en condiciones más similares a las usadas para cristalizar y resolver las demás estructuras. En el caso de las estructuras cristalográficas que contuvieron más de una molécula de profilina, se seleccionó la molécula más similar a las demás estructuras cristalográficas de la profilina. Las estructuras escogidas se alinearon por medio del algoritmo "Combinatorial Extension"<sup>©</sup> [48] y las secuencias de las demás profilinas se alinearon a éstas mediante el algoritmo ClustalW [49] en su versión 1.83 usando sus parámetros predeterminados, en ambos casos mediante la interfaz gráfica del programa STRAP<sup>©</sup> [50]. El alineamiento obtenido se refinó usando BioEdit<sup>©</sup> [51].

### ***Validación del alineamiento múltiple de las secuencias de amino ácidos de la familia de proteínas de la profilina.***

El alineamiento múltiple de las secuencias de amino ácidos de la familia de las profilinas se validó de diferentes formas. En primer lugar, se vio si la distribución de frecuencias de amino ácidos de las 5 posiciones más variables del alineamiento cambiaba al perturbar las frecuencias de los amino ácidos en otro sitio (Figura 5). Esta validación nos ayuda a determinar si existe suficiente divergencia evolutiva entre las secuencias de la familia de la profilina para que las coevoluciones detectadas por el SCA sean resultado de las necesidades funcionales de la profilina y no únicamente de la historia de la divergencia de las secuencias, ya que se espera que estas posiciones no tengan un acoplamiento funcional con otras [4]. Posteriormente, se analizaron los cambios en la energía estadística de conservación (valor calculado por el SCA que cuantifica la conservación de una posición) de esas 5 posiciones al

disminuir gradualmente el número de secuencias (permutaciones) en el SCA. Esta validación nos permite determinar si el número de secuencias utilizadas es estadísticamente significativo (ver Figura 4 a, b), ya que cuando la eliminación de secuencias da un cambio pequeño para todas las 5 posiciones más variables, esta energía de conservación se debe a las posiciones de amino ácidos de la familia de proteínas que tienen una función particular y no a una escasa cantidad de secuencias [4]. Finalmente, se determinó el tamaño mínimo que debe tener el alineamiento después de las perturbaciones para que la energía de acoplamiento sea estadísticamente significativa (Figura 4 c, d), usando la consideración de que este subalineamiento debe incluir una cantidad tal de secuencias que la eliminación de unas pocas de ellas resulte en un cambio pequeño en la energía estadística de acoplamiento, por una razón similar a la expuesta en la validación anterior [4].



**Figura 4: Validación de la significancia de las energías estadísticas de conservación y de acoplamiento de las perfilinas.** Se graficó el comportamiento de las energías estadísticas de conservación (a y b) y de acoplamiento (c y d) de las 5 posiciones más variables del alineamiento múltiple de secuencias de las perfilinas vegetales (a y c) y de todas las perfilinas eucarióticas (b y d) al ir eliminando secuencias aleatoriamente. La línea azul es el promedio de estas 5 energías estadísticas y las barras azules son la desviación estándar. Las energías estadísticas son válidas en el intervalo de cantidad de secuencias donde la línea azul tiene una pendiente constante y la barra de error es pequeña.

## ***Análisis Estadístico de Acoplamiento (SCA) en la familia de proteínas de profilina.***

El análisis de acoplamiento estadístico (SCA) se aplicó al alineamiento múltiple de secuencias de la familia de las profilinas vegetales con o sin incluir a las isoformas de profilinas de otros organismos eucarióticos. El análisis se realizó como lo describe Rama Ranganathan y colaboradores [2] para el software escrito en Java (versión 1.5) y adaptado de su versión original en C. Este software fue prestado generosamente por el Dr. Rama Ranganathan de la "University of Texas Southwestern Medical Center at Dallas". Este algoritmo calcula la frecuencia de los amino ácidos en cada una de las posiciones del alineamiento múltiple de las secuencias de amino ácidos de una familia de proteínas, y compara las variaciones que existen en estas frecuencias cuando se modifica la frecuencia de un amino ácido en ciertas posiciones (perturbaciones). [2] Con estos datos de frecuencias el SCA calcula la energía estadística de conservación de cada posición ( $\Delta G$ ) y con la comparación de la energía de conservación entre dos posiciones con y sin perturbación se calcula la energía estadística de acoplamiento ( $\Delta\Delta G$ ) de cada perturbación con cada posición [2]. El algoritmo permite calcular los residuos que tienen frecuencia cero (nunca se presentan) en cierta posición y da puntuaciones más altas a los cambios de la distribución de frecuencia de amino ácidos cuando éstos se presentan en posiciones conservadas [2].

Las energías estadísticas de acoplamiento se almacenaron en una matriz de  $n$  perturbaciones por  $m$  posiciones, y los datos de esta matriz se agruparon para obtener la lista de residuos que muestran altas energías estadísticas de acoplamiento. Este agrupamiento se realizó mediante un método de agrupamiento iterativo que se ha usado para estudiar los patrones de expresión en microarreglos de ADN [52]. La idea general de este método es llevar rondas secuenciales de agrupamiento de la matriz por perturbaciones y por posiciones, identificando cada vez el grupo de posiciones que dan alto acoplamiento con un grupo de perturbaciones, para usar sólo la submatriz compuesta por estas posiciones y estas perturbaciones en la siguiente ronda de agrupamiento [4]. En el caso en que se encontraron dos submatrices de alto acoplamiento, cada una se usó independientemente para una siguiente ronda de agrupamiento. Para realizar los agrupamientos, se calcularon las distancias entre las posiciones y entre las perturbaciones por medio del método "city-block metric" (el cual suma las diferencias de cada elemento de dos

listas de datos del mismo tamaño), y con base en ellas se construyeron sendos dendrogramas por el método del enlace completo, y éstos se usaron para ordenar la matriz tanto por posiciones como por perturbaciones.

### ***Representación molecular de la profilina.***

Para extrapolar las energías estadísticas de conservación de cada posición en la estructura terciaria de la profilina, se representó gráficamente la estructura terciaria de la profilina de *Arabidopsis thaliana* en forma de esferas atómicas en un espectro de colores que representaba estas energías, usando el programa Discovery Studio® Visualizer© de Accelrys en su versión 1.6 ([http://www.accelrys.com/products/downloads/ds\\_visualizer/](http://www.accelrys.com/products/downloads/ds_visualizer/)). La imagen resultante se guardó como un archivo de escena de POV-Ray®. La imagen final de ese modelo fue elaborada por el software POV-Ray® 3.6 (<http://www.povray.org/download/>).

Para representar gráficamente los resultados del SCA, se utilizó la estructura terciaria de la profilina I de *Arabidopsis thaliana* (PDB: 3NUL). Esta estructura se representó, usando la versión educativa del programa PyMol® 1.03 [53], en forma de carbonos  $\alpha$ , y los grupos de residuos altamente acoplados que resultaron del SCA se representaron en forma de esferas atómicas. La posición de la actina se expresó en forma de esqueleto de carbonos  $\alpha$ , en su posición predicha a partir de la superposición del complejo profilactina bovino mediante el mismo programa PyMol®. En el caso en que resultaron del SCA residuos cercanos a unas moléculas de glicerol que aparecieron en la estructura terciaria de la profilina I de *Arabidopsis thaliana* dentro de una cavidad que se encuentra en las profilinas vegetales de estructura conocida, estas moléculas se representaron en forma de esferas atómicas.

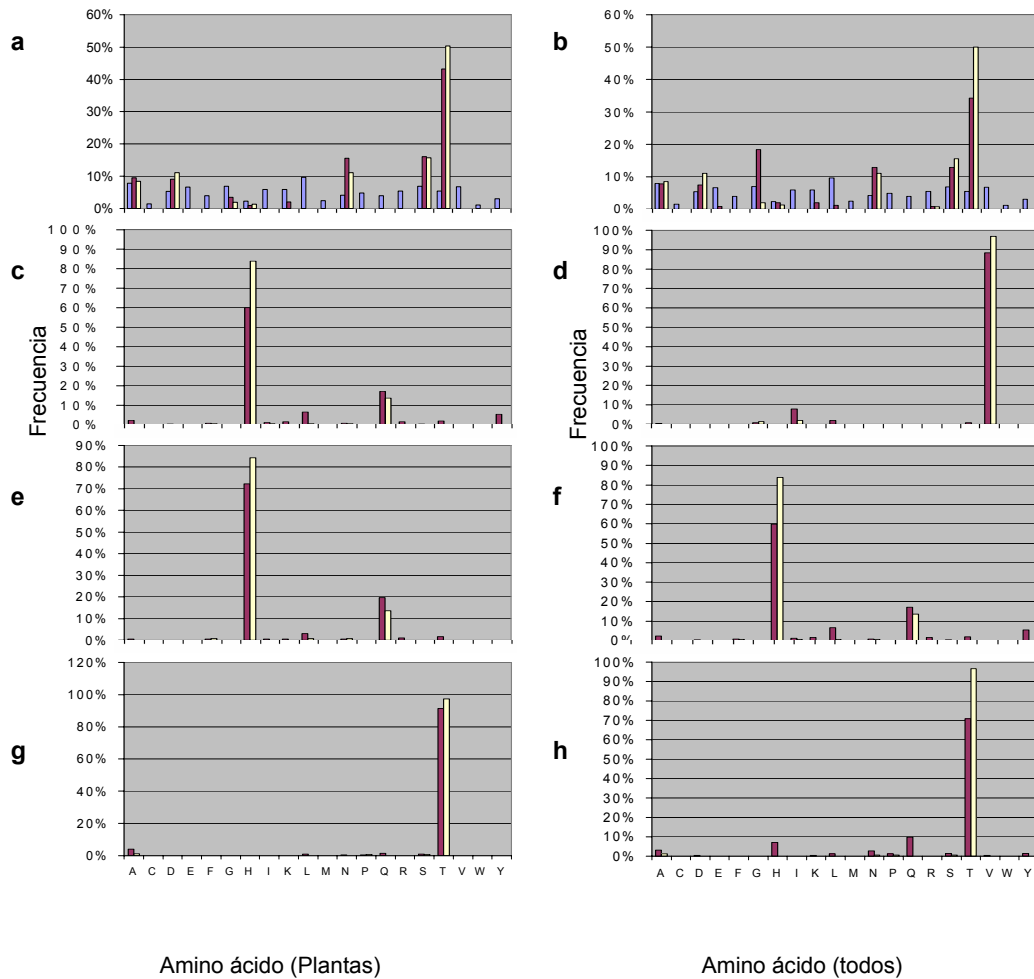
### ***Anclaje molecular de la profilina con moléculas orgánicas.***

Con el fin de predecir si la cavidad presente en las profilinas vegetales, en cuya proximidad se encontraron amino ácidos coevolucionados, podría unir alguna molécula orgánica, se hizo un anclaje molecular de algunas de éstas con la profilina I de *Arabidopsis thaliana*. Este anclaje se realizó usando el programa AutoDock 4© [54] en su versión para Linux. La estructura de la profilina se preparó usando la interfaz gráfica del programa AutoDockTools© [55] en su versión 1.4.5 para Windows, mediante la adición de hidrógenos, el cálculo de cargas atómicas de

Gasteiger [56, 57] y la fusión de los hidrógenos no polares a los carbonos con los que están enlazados. Por su lado, en el caso de la estructura de las moléculas orgánicas se agregaron los dobles enlaces y los hidrógenos usando el programa Discovery Studio® Visualizer [58] de Accelrys en su versión 1.6, y mediante AutoDockTools© se calcularon las cargas atómicas de Gasteiger, se fusionaron los hidrógenos no polares a los carbonos con los que se enlazan y se determinó el enlace rotatable que serviría mejor como raíz del árbol de exploración de torsiones de los enlaces rotables. El anclaje molecular se hizo mediante el algoritmo genético Lamarckiano, haciendo 25'000,000 evaluaciones, usando una posición inicial aleatoria de la molécula orgánica y con todos los demás parámetros predeterminados de AutoDock 4©. El anclaje molecular antes mencionado se hizo para la profilina I de *Arabidopsis thaliana* (PDB: 3NUL) con la zeatina, la cinetina, la adenina, la adenosina y el mono-, di- y trifosfato de adenosina.

## Resultados.

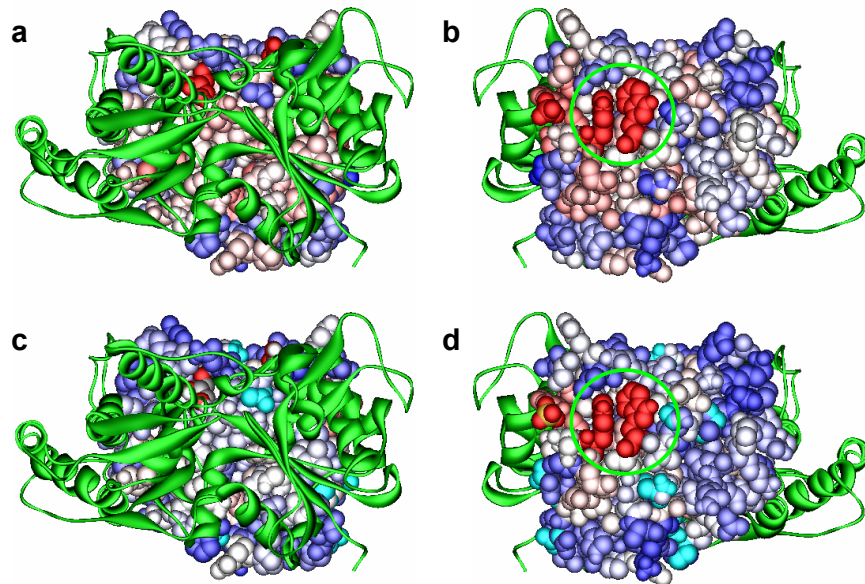
La profilina es una proteína que ha sido identificada en diferentes organismos de todos los reinos eucarióticos. En la actualidad existen 16 estructuras terciarias de profilina conocidas y se ha visto que éstas son bastante similares a pesar de que la similitud de sus estructuras primarias es muy baja. En este trabajo se realizó un alineamiento múltiple de las 257 secuencias de profilinas eucarióticas depositadas en las bases de datos, y se determinó la distribución de amino ácidos en cada una de las posiciones del alineamiento. En la Figura 5 se muestra la frecuencia de los amino ácidos en una de las posiciones menos conservadas del alineamiento de las 199 profilinas vegetales (Figura 5 a) y del alineamiento de todas las secuencias reportadas de profilinas (Figura 5 b). En ambos alineamientos podemos ver que en esta posición generalmente está una treonina (T), haya o no haya una perturbación de frecuencias en otra posición. Por otro lado podemos ver que algunas posiciones conservadas pero sin acoplamiento no varían mucho al hacer una perturbación, tanto en el caso de las profilinas vegetales (Figura 5 c) como en el de todas las eucarióticas (Figura 5 d). Los resultados obtenidos en este análisis, nos indican que la información de coevolución dada por el SCA para otras posiciones, es verdadera y no un artefacto debido a la escasez de información derivada de las secuencias. Por lo tanto, se decidió realizar el Análisis Estadístico de Acoplamiento en los dos alineamientos de isoformas de profilina.



**Figura 5: Distribución de amino ácidos de las posiciones del alineamiento de amino ácidos de las profilinas.** Distribución de frecuencias de la posición 38 (a y b), 7 (c y d), 28 (e y f), y 111 (g y h) del alineamiento de las isoformas vegetales (a, c, e y g) y del alineamiento de todas las isoformas de profilinas que se han reportado (b, d, f y h). La frecuencia de amino ácidos se determinó antes (barras en crema) y después (barras en vino) de la perturbación de la posición 51. En los paneles a y b se graficó la frecuencia media de amino ácidos de las proteínas depositadas en la base de datos de SwissProt (barras en azul).

El análisis de conservación de amino ácidos en cada una de las posiciones del alineamiento múltiple de las secuencias de la familia de las profilinas se efectuó usando el parámetro llamado energía estadística de conservación ( $\Delta G$ ) que se encuentra dentro de la paquetería del SCA. Los resultados obtenidos muestran que muchas posiciones de la profilina tienen valores bajos de energía estadística de conservación. En la Figura 6 se representaron las energías estadísticas de conservación, de cada una de las posiciones, en la estructura terciaria de la profilina de *Arabidopsis thaliana* (PDB: 3NUL). En esta figura se puede observar que las posiciones que presentan los valores más altos de energía estadística de conservación (color

rojo) corresponden a las posiciones de los residuos que se ha demostrado bioquímicamente que participan en la interacción con los PRM (Figura 6 b, d). Además se observan otras posiciones que están medianamente conservadas (colores claros) que se localizan en la interfase de la interacción de la profilina con la actina (Figura 6 a, c). Sin embargo, estas posiciones están más conservadas entre las profilinas vegetales que entre todas las profilinas. Por último, se puede apreciar que las posiciones menos conservadas (azul marino) están en las caras laterales de la profilina, lejos de las interfases de interacción con la actina y los PRM. En la Figura 6, también podemos ver que en general se obtuvieron los mismos resultados de distribución de energía estadística de conservación tanto en el alineamiento de las profilinas vegetales (Figura 6 a, b) como en el alineamiento de todas las profilinas reportadas (Figura 6 c, d). Debido a que se obtuvieron posiciones con valores medianos de energía estadística de conservación, decidimos calcular las energías estadísticas de acoplamiento (SCA).

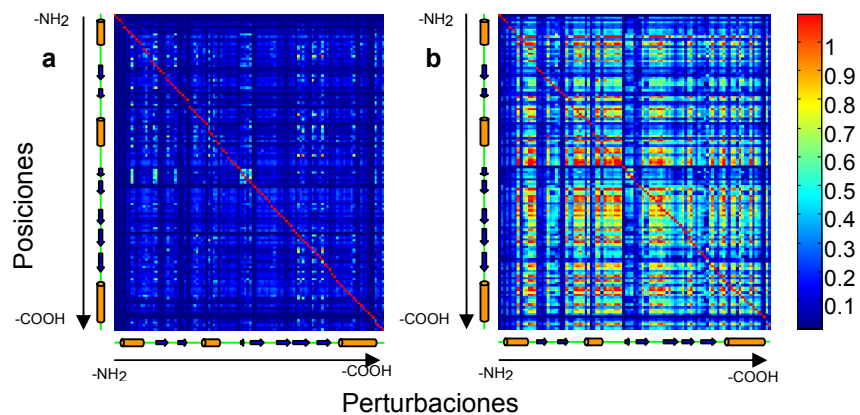


**Figura 6: Energía estadística de conservación en cada una de las posiciones de las profilinas.** Energías de conservación de las profilinas vegetales graficadas en la estructura terciaria de la profilina de *Arabidopsis thaliana* (PDB: 3NUL). Los valores de la energía de conservación están representados con un espectro de colores que va del rojo (residuos más conservados) al azul (residuos menos conservados). La imagen **a** muestra la cara de la profilina que interacciona con la actina (verde) y en la imagen **b** se muestra el sitio de reconocimiento de los PRM (reindicado por el círculo verde). Las imágenes **c** y **d** muestran lo mismo que la **a** y la **b**, respectivamente, para todas las profilinas eucarióticas. Los datos de estructura de la actina se tomaron de la actina cristalizada en complejo con la profilina (verde: PDB: 2BTF).

El SCA se realizó perturbando 112 sitios del alineamiento de las profilinas vegetales y 101 sitios del alineamiento de todas las profilinas. En todo el análisis se utilizó la secuencia de



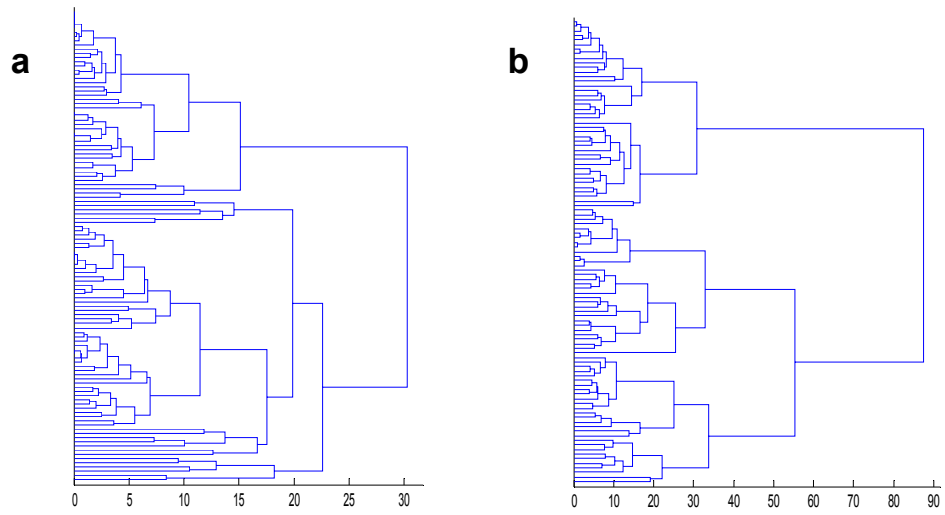
amino ácidos de la profilina de *Arabidopsis thaliana* como secuencia de referencia de las posiciones. Los valores de energía estadística de acoplamiento se graficaron en las matrices que muestra la Figura 7. En esta gráfica podemos observar que la perturbación de ciertas posiciones de la secuencia de las profilinas vegetales (Figura 7 a) produce cambios de frecuencias de amino ácidos en otras posiciones (altos valores de energía de acoplamiento; colores del amarillo al rojo). Por ejemplo, la perturbación de algunas posiciones localizadas en las estructuras  $\beta$  S6 y S7 producen cambios en la distribución de amino ácidos de las posiciones localizadas en el asa que se encuentra entre la  $\alpha$  hélice H2 y la estructura  $\beta$  S3. Esto nos indica que existe una restricción evolutiva entre pares de posiciones debido a una relación funcional de estos dos sitios. Por otro lado, cuando se realizó el análisis con todas las secuencias de profilinas reportadas obtuvimos resultados muy diferentes. En este caso se incrementó considerablemente el número de pares de posiciones que coevolucionan y la mayoría de ellos tienen valores muy altos (rojo) de energía estadística de acoplamiento (Figura 7 b). En esta figura podemos observar que la perturbación de las posiciones localizadas en el amino terminal no producen cambios de frecuencia en otras posiciones, así como la perturbación de otras posiciones no produce cambios de frecuencia de amino ácidos en las posiciones del amino terminal.



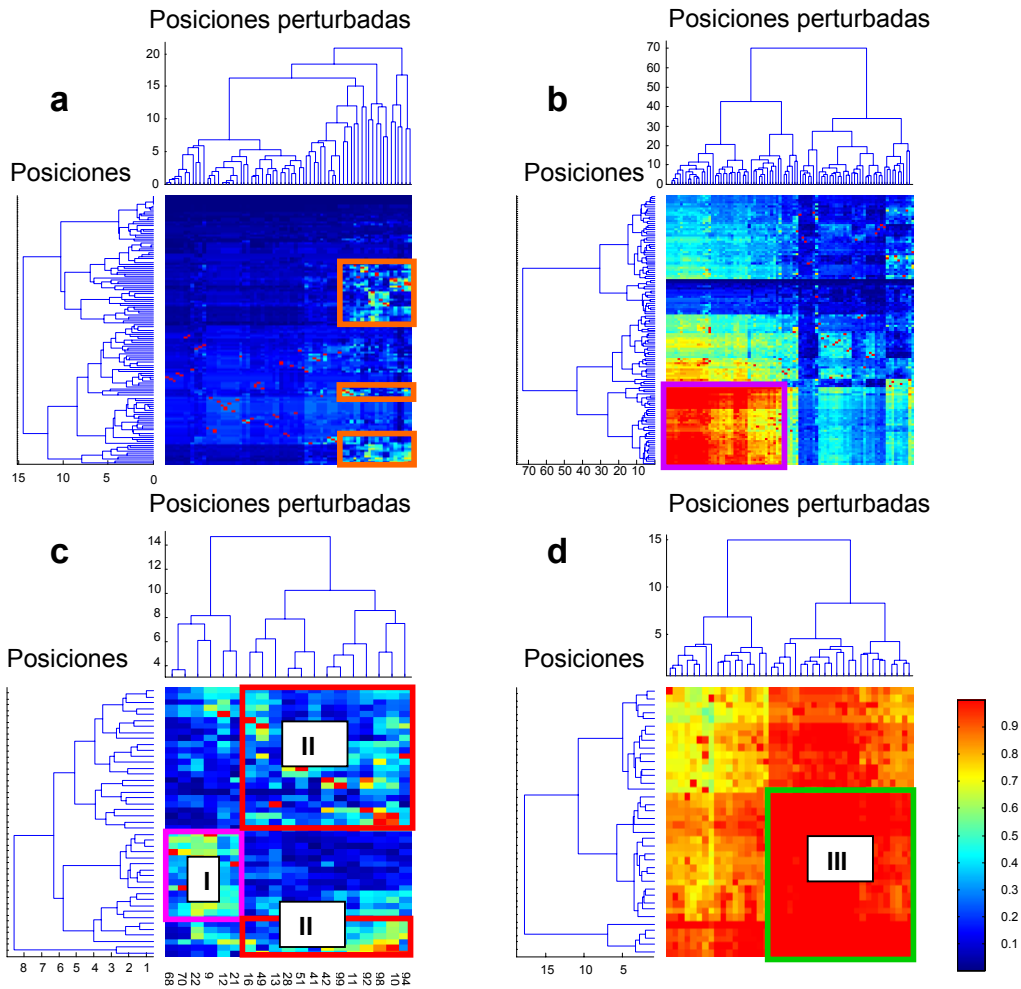
**Figura 7: Energía estadística de coevolución entre las posiciones de la secuencia de amino ácidos de la familia de profilina.** Las energías estadísticas de acoplamiento obtenidas de las **a)** 112 perturbaciones que se hicieron (abscisas) para cada una de las 131 posiciones de la profilina vegetal (ordenadas) y de las **b)** 101 perturbaciones que se hicieron (abscisas) para cada una de las 131 posiciones de todas las profilinas eucarióticas (ordenadas) se muestran en un espectro de color cuya escala está en la barra de la derecha. En cada uno de los ejes se adicionó la representación gráfica de la organización de las estructuras secundarias de la profilina en donde se representa a las  $\alpha$  hélices en amarillo, las  $\beta$  plegadas con flechas azules y las asas con líneas verdes.

Posteriormente, se determinó el grado de relación entre los residuos que se perturbaron y se graficaron en los dendrogramas que se muestran en la Figura 8. En el dendrograma de las profilinas vegetales (Figura 8 a) se puede observar un grupo de posiciones con valores de distancia de cero las cuales ha sido comprobado experimentalmente que son importantes para el reconocimiento de secuencias ricas en prolina. Estos amino ácidos están conservados en todas las isoformas de profilina vegetal y en el 98 % de las isoformas de profilinas reportadas. Por otro lado, en la Figura 8 podemos ver que en ambos dendrogramas se observan entre 5 y 7 grupos que contienen sitios cuya perturbación afecta la frecuencia de amino ácidos de otras posiciones de una manera similar y posiblemente esto se deba a que están funcionalmente relacionados ya sea para pegar una de las moléculas que se unir la profilina a algún ligando o para mantener la estructura terciaria de la misma. Sin embargo, los valores de distancias de relaciones son menores en los grupos obtenidos con las profilinas vegetales (Figura 8 a) comparados los obtenidos utilizando todas las isoformas de profilinas reportadas (Figura 8 b). Esto se debe principalmente a la baja similitud que presentan las secuencias de amino ácidos de las isoformas de profilina entre los diferentes reinos. Con base en estos dendrogramas, se excluyó de los análisis siguientes la rama donde estaban las posiciones perturbadas que tenían una alta energía de conservación y una muy baja energía estadística promedio de acoplamiento con cada una de las posiciones de la profilina.

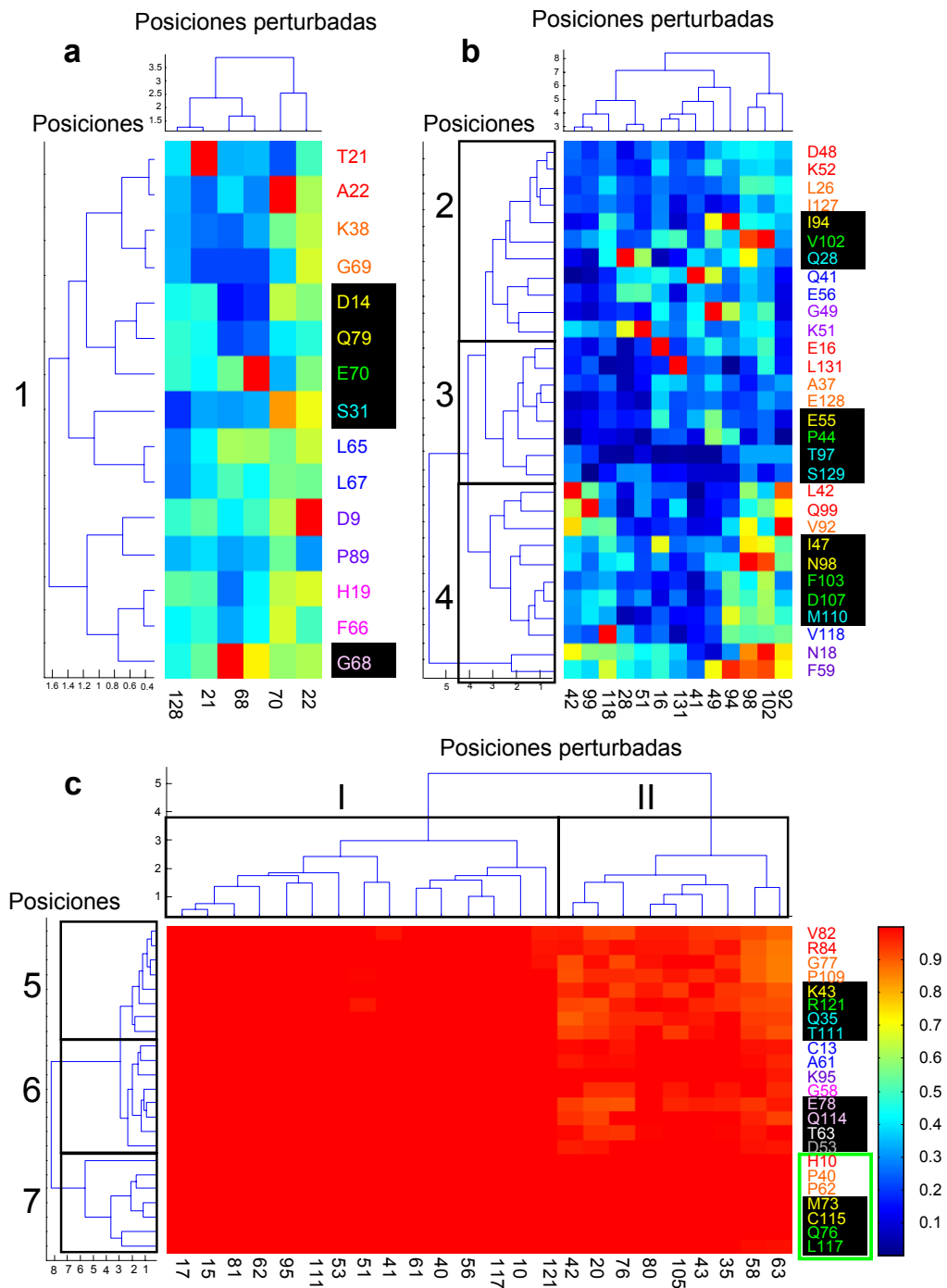
Posteriormente, se organizaron las matrices de la Figura 8 con base en las similitudes de sus energías estadísticas de acoplamiento. La organización de las matrices se realizó de una manera bidimensional en donde se agruparon simultáneamente las posiciones perturbadas de la profilina y cada una de las posiciones de la secuencia de la profilina (Figura 9 a, b). Comparando las matrices resultantes de las isoformas vegetales de la profilina y las de todas las profilinas reportadas podemos observar claramente que utilizando todas las isoformas reportadas (Figura 9 b) obtenemos una zona con altos valores de acoplamiento entre residuos a diferencia del análisis con las isoformas vegetales (Figura 9 a) donde se obtienen tres zonas de acoplamiento con bajos valores de acoplamiento (recuadros). También se puede observar que las ramas que se forman en ambos casos tienen una distribución diferente, posiblemente existió una divergencia evolutiva independiente en cada uno de los reinos eucariontes y esto podría ser debido a sus relaciones en la función biológica.



**Figura 8: Comparación de los perfiles de energías estadísticas de coevolución de los sitios perturbados.** Comparación entre las filas de las matrices de energías estadísticas de coevolución de la Figura 7 (eje X). Las distancias de relaciones entre cada par de filas de la matriz de **a)** las profilinas vegetales y **b)** de la de todas las isoformas de profilina reportadas se calcularon mediante el método "city-block metric", que consiste en sumar las diferencias absolutas de los valores de las dos filas de la matriz para cada columna de ésta. Con base en estas distancias, los grupos se formaron con un método iterativo basado en un algoritmo similar al que se desarrolló para el análisis de micro arreglos de ADN.

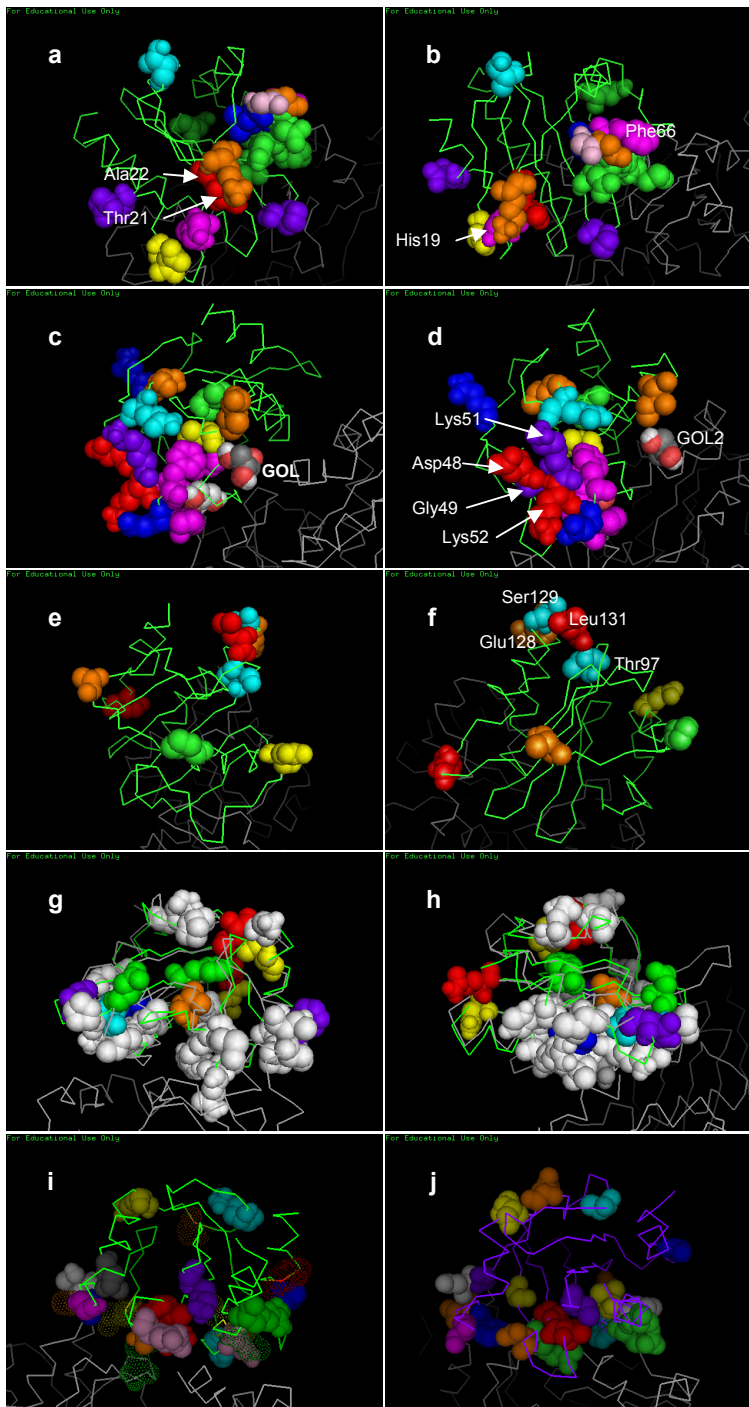


**Figura 9: Energía estadística de acoplamiento entre posiciones de la secuencia de amino ácidos de la familia de la profilina.** Matriz de acoplamiento entre residuos de la secuencia de amino ácidos de las isoformas de **a)** las profilinas vegetales y **b)** la familia completa de la profilina organizada bidimensionalmente por el grado de relación entre los residuos perturbados (cladograma en el eje X) y entre las diferentes posiciones de la secuencia (cladograma en el eje Y). Segunda iteración de las zonas con altos valores de energía de acoplamiento de **c)** las isoformas de profilina vegetal y **d)** de la familia de proteínas de la profilina se organizaron de la misma manera que en el inciso a y b. Los valores de energías estadísticas de acoplamiento se representan con un espectro de colores indicado en la parte inferior derecha de la figura que va del azul ( $\Delta\Delta G = 0kT^*$ ) al rojo ( $0.9 kT^*$ ). Los grupos de posiciones que presentan alto acoplamiento se indican con recuadros.

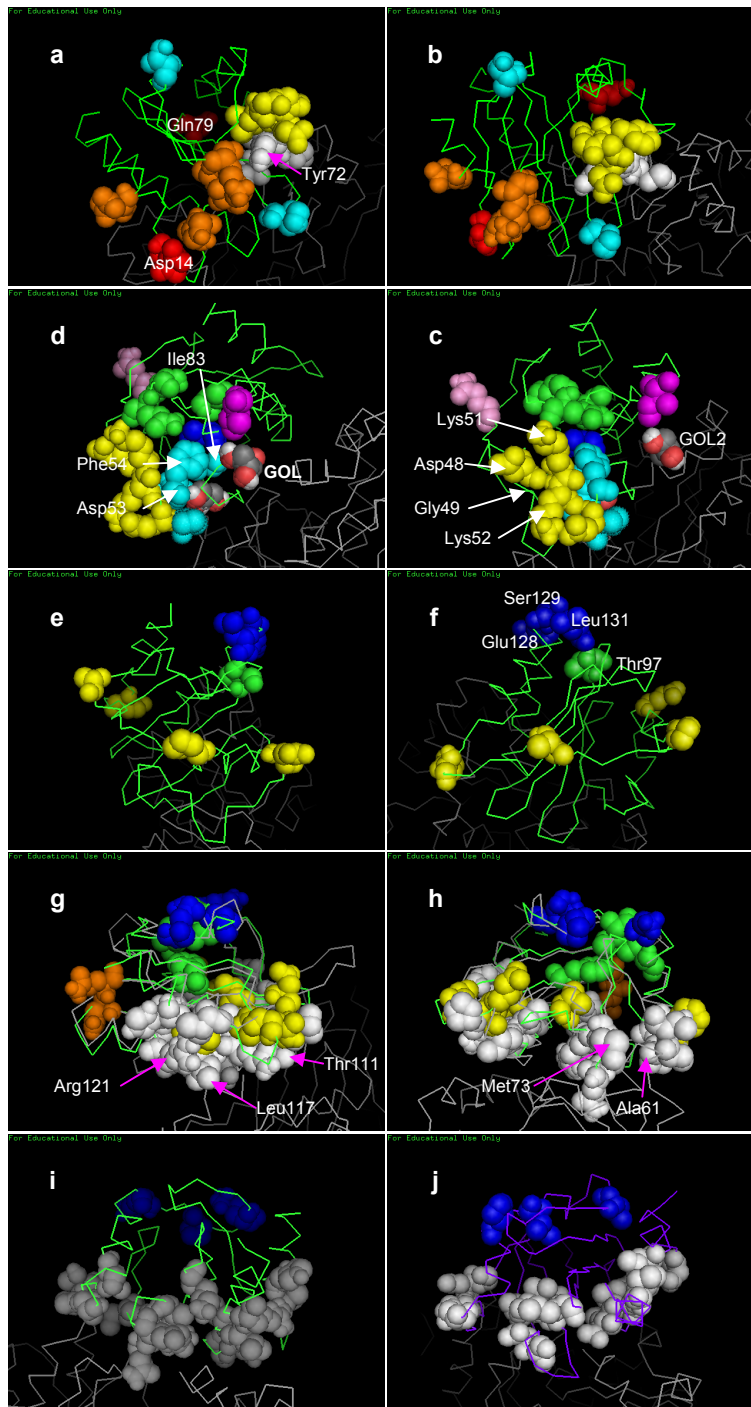


**Figura 10: Zona de mayor energía de acoplamiento entre posiciones de la secuencia de amino ácidos de la familia de la profilina.** Tercera iteración de las zonas de alto acoplamiento entre posiciones de las isoformas de profilina vegetales (a y b) y la familia de proteínas de la profilina (c). La matriz se organizó bidimensionalmente de manera similar a la Figura 9. Los valores de energías estadísticas de acoplamiento se representan con un espectro de colores indicado en la parte inferior derecha de la figura que va del azul ( $\Delta\Delta G = 0\text{kT}^*$ ) al rojo ( $0.9\text{kT}^*$ ). Los grupos de posiciones que presentan alto acoplamiento se indican con recuadros. La coloración de las etiquetas de las posiciones es el color que tendrán los residuos de éstas al representarse gráficamente en la Figura 11, y los residuos de las posiciones de las etiquetas encerradas en el recuadro verde se representan en forma de esferas puntadas en vez de esferas sólidas. Los sombreados de algunas etiquetas son sólo para una mayor claridad de las letras.

Posteriormente, se reorganizó bidimensionalmente la rama de posiciones perturbadas que tiene un alto acoplamiento con tres ramas que agrupan diferentes residuos del alineamiento de las isoformas de profilina vegetales, y la matriz resultante se muestra en la Figura 9 c. Para el caso de la matriz obtenida con todas las isoformas de profilina se reorganizó únicamente la rama con valores muy altos de energía que tienen acoplamiento con una sola rama de posiciones, y la matriz resultante del nuevo agrupamiento se muestra en la Figura 9 d. Como podemos ver en esta figura, la matriz tiene una nueva distribución de las ramas, y podemos observar una zona de la matriz que agrupa a pares de posiciones que tienen valores muy altos de energía de acoplamiento, a diferencia de la matriz con las isoformas de profilinas vegetales que tiene zonas con valores intermedios de energía de acoplamiento. Las zonas de mayor energía de acoplamiento se seleccionaron para una nueva iteración y los resultados se presentan en la Figura 10. En el resultado obtenido en estas nuevas iteraciones para el caso de las profilinas vegetales, podemos ver claramente un grupo de posiciones (grupo 1) que están acopladas con un grupo de posiciones perturbadas (Figura 10 a) y tres grupos de posiciones (grupos 2, 3 y 4) acopladas con otro grupo de posiciones perturbadas (Figura 10 b). Por otro lado, la matriz de energías estadísticas de acoplamiento obtenida para todas las isoformas de la profilina, demuestra que existen dos grupos de posiciones perturbadas (Figura 10: grupos I y II) que están acoplados de manera similar a tres grupos (grupos 5, 6 y 7) de posiciones de la secuencia de amino ácidos de la profilina. Mientras el grupo I de posiciones perturbadas tiene valores de energías de acoplamiento de 0.9 con los grupos 5, 6 y 7 de posiciones, el grupo II de posiciones perturbadas se acopla también de manera similar con los tres grupos de posiciones; con el grupo 7 tiene valores de energía de acoplamiento de 0.9 kT\*, y con los grupos 5 y 6 tiene valores de energía de acoplamiento entre 0.8-0.9 kT\*. Este resultado nos indica que estos pares de posiciones han tenido una restricción evolutiva y esto puede ser debido a sus relaciones en las funciones biológicas de la proteína; como por ejemplo, que participen en la interacción con alguna molécula o que estén acoplados para mantener la estructura terciaria de la proteína. Para corroborar esta hipótesis, se extrapolaron las agrupaciones de posiciones de aminoácidos de la matriz de la Figura 10, a representaciones en colores en la estructura terciaria de la profilina I de *Arabidopsis thaliana* (Figuras 11 y 12).



**Figura 11: Localización estructural de los pares de posiciones con altos valores de energía estadística de acoplamiento en la familia de proteínas de las profilinas.** Representación gráfica de la estructura terciaria de la profilina de *Arabidopsis thaliana* (verde) (PDB: 3NUL) con los residuos de las posiciones coevolucionadas entre sí en las profilinas vegetales de a) y b) la Figura 10 a y en las regiones de la Figura 10 b con etiqueta c) y d) 1, e) y f) 2 y g) y h) 3. Las imágenes b y d son rotaciones de las imágenes moleculares de los paneles a y c 90° hacia la izquierda, y la imagen f es la rotación de la imagen molecular del panel e 90° hacia la derecha. Por otro lado, se representaron en la estructura terciaria de la profilina i) de *Arabidopsis thaliana* y j) bovina (morado) (PDB: 2BTF) los residuos de las posiciones coevolucionadas entre sí en todas las profilinas reportadas de la Figura 10 c. Se muestra la posición de la actina (gris) en su posición predicha a través del complejo profilactina bovino. La coloración de los residuos corresponde a la coloración de las etiquetas de sus posiciones en la Figura 10, y los residuos de posiciones cuyas etiquetas aparecen dentro de un recuadro verde en la Figura 10 se representan como esferas punteadas. Las moléculas etiquetadas como GOL y GOL2 en los paneles c y d son 2 gliceroles que aparecieron al cristalizar la profilina de *Arabidopsis thaliana*.



**Figura 12: Localización estructural de las zonas de posiciones con altos valores de energía estadística de acoplamiento en la familia de proteínas de las profilinas.** Representación gráfica de la estructura terciaria de la profilina de *Arabidopsis thaliana* (verde) (PDB: 3NUL) con los residuos de las posiciones coevolucionadas entre sí en las profilinas vegetales de a) y b) la Figura 10 a y en las regiones de la Figura 10 b con etiqueta c) y d) 1, e) y f) 2 y g) y h) 3. Las imágenes b y d son rotaciones de las imágenes moleculares de los paneles a y c 90° hacia la izquierda, y la imagen f es la rotación de la imagen molecular del panel e 90° hacia la derecha. Por otro lado, se representaron en la estructura terciaria de la profilina i) de *Arabidopsis thaliana* y j) bovina (morado) (PDB: 2BTF) los residuos de las posiciones coevolucionadas entre sí en todas las profilinas reportadas de la Figura 10 c. Se muestra la posición de la actina (gris) en su posición predicha a través del complejo profilactina bovino. La coloración de los residuos se explica en el texto. Las moléculas etiquetadas como GOL y GOL2 en los paneles c y d son 2 gliceroles que aparecieron al cristalizar la profilina de *Arabidopsis thaliana*.



Se vio que la mayoría de los residuos de cada grupo de posiciones se localizaban en la misma zona de la estructura de la profilina I de *Arabidopsis thaliana* (PDB: 3NUL). En los agrupamientos de las posiciones de las profilinas vegetales, los residuos de las posiciones de la gráfica de la Figura 10 b se localizaron en su mayoría en la zona de la profilina que une al PIP<sub>2</sub>, y los correspondientes a las posiciones de la gráfica de la Figura 10 a se localizaron del lado opuesto. En el agrupamiento de las posiciones de todas las profilinas reportadas (Figura 10 c), los residuos de estas posiciones se localizaron en su mayoría en la interfase de la interacción con la actina, y los demás residuos en la cara opuesta de la profilina.

En el caso de ambos análisis, la mayoría de las posiciones con importancia para el pegado de actina, PIP<sub>2</sub> y/o PRM a la profilina que se encontraron en la segunda ronda de agrupamiento permanecieron en la tercera. Algunas de estas posiciones tuvieron una perturbación de frecuencias representada en la segunda ronda, y que se representó también en la tercera. Tal es el caso de las posiciones Thr111, Leu117 y Arg121, cuyos residuos están en la hélice N-terminal de la profilina y participan en el pegado a la actina, a la PLP y al PIP<sub>2</sub>, que se encuentran como posiciones y como sitios de perturbación tanto en la segunda ronda de agrupamiento como en la tercera. En otros casos, la perturbación de frecuencias de la posición se encuentra en la segunda ronda pero no en la tercera, como es el caso de la posición Met73, que es importante para el pegado de la profilina a la actina y está en el motivo que reconoce a la PI3K, o de las posiciones Val82 e Ile84, que son importantes tanto para el plegamiento de la profilina como para su pegado a la actina. Esta diferencia muy posiblemente se deba a que al estar más conservadas estas posiciones que las primeras, las perturbaciones de estas posiciones den una menor energía de acoplamiento que las perturbaciones de las primeras.

Con tal de analizar la posible relevancia biológica de los residuos de las posiciones con alto acoplamiento que se obtuvieron al final del agrupamiento iterativo, éstos se mapearon sobre la estructura terciaria de la profilina de *Arabidopsis thaliana* (PDB: 3NUL).

Los pares de posiciones de la profilina vegetal con alto acoplamiento entre sí que se muestran en la Figura 10 a, se representaron en la estructura de la profilina de *Arabidopsis thaliana* en la Figura 11 a, b. Algunos pares de posiciones tienen sus residuos cercanos entre sí en la estructura terciaria de la profilina, como es el caso de la Thr21 y la Ala22 (en rojo). La posición más acoplada a las dos anteriores, la Lys38 (en naranja), tiene su residuo cercano a los

residuos anteriores. También hay algunos pares de posiciones con alto acoplamiento entre sí en la zona 1 de la Figura 10 b, que tienen a sus residuos cercanos entre sí en la estructura de la profilina de *Arabidopsis thaliana* (Figura 11 c, d), tal es el caso de las posiciones Asp48 y Lys52 (en rojo) y de Gly49 y Lys51 (en violeta), que están cercanas a una cavidad que se encuentra sólo en las profilinas vegetales y que posiblemente sea el sitio de unión de la profilina a una molécula orgánica. El SCA de todas las profilinas reportadas mostró que la posición más acoplada a la Val82 era la Arg84 (Figura 10 b); los residuos de estas posiciones están cercanos entre sí en la estructura de la profilina (Figura 11 i, j) y son importantes para su plegamiento y su función. Sin embargo, en otros casos no es evidente la relación estructural de un par de posiciones altamente acopladas, pero si consideramos la rama del dendrograma a donde pertenece dicho par de posiciones, la relación se hace evidente: se obtiene una red de residuos. Por ejemplo, para el caso de las profilinas vegetales, las posiciones altamente acopladas His 19 y Phe 66 se encuentran lejanas en la estructura de la profilina (Figura 11a), pero están dentro de una red formada por casi todos los residuos del grupo de posiciones al que pertenecen (grupo 1 de la Figura 10 a).

El grupo de posiciones con alto acoplamiento que se muestra en la Figura 10 a se representó en la estructura de la profilina de *Arabidopsis thaliana* en la Figura 12 a, b. Este grupo contiene los residuos que forman una horquilla  $\beta$  localizada cerca de la tirosina 72. Este residuo participa en la unión de la profilina a una región rica en prolinas de la fosfatidil-inositol-3-cinasa [30]. También en este grupo aparecen dos residuos (Asp14 y Gln79, en amarillo) cercanos a las asas L1 y L3 (Figura 2), respectivamente, las cuales son de distinto tamaño en las profilinas vegetales que en las animales, y aparece un grupo de residuos (en anaranjado) cercanos a Asp14. Estos residuos podrían estar involucrados en las diferencias funcionales entre las profilinas animales y las vegetales.

El dendrograma de posiciones coevolucionadas de la Figura 10 b se subdividió para su estudio en 3 regiones. El grupo de residuos de la región 1 se representó gráficamente en la profilina de *Arabidopsis thaliana* (Figura 12 c, d). Algunos de estos residuos (Asp48, Gly49, Lys51, Lys52 y Glu56, en amarillo) forman un agrupamiento en la estructura de la profilina y están cercanos a otros amino ácidos (Asp53, Phe54 e Ile83, en azul claro) que se encuentran en una cavidad que sólo existe en las profilinas vegetales y se cree que puede ser el sitio de unión de estas

profilinas a alguna molécula orgánica [6]. Este sitio putativo se encuentra en dos asas que están muy conservadas entre las profilinas vegetales y son mas largas que las correspondientes en las isoformas de la profilina de los demás eucariontes. En estas asas se encontraron contactos polares con una molécula muy ordenada de glicerol (marcada como GOL) al resolver la estructura de la profilina de *Arabidopsis thaliana* a 100 K (PDB: 3NUL), y se volvieron a dar esos contactos polares al resolver la misma estructura a 25 °C (PDB: 1A0K), por lo que se cree que este podría ser el sitio de pegado de las profilinas vegetales a una molécula orgánica [6]. Otra evidencia de la posible significancia biológica de este agrupamiento de residuos es que la Lys52 y el Glu56 forman un puente de hidrógeno y un enlace iónico. Otros residuos de la región 1 del dendrograma describen una línea de amino ácidos (en verde), que parte del agrupamiento mencionado anteriormente y termina en Ile94 (en azul), residuo que se ha demostrado importante en el plegamiento de la profilina [6] y está en el interior hidrofóbico de ésta. Esto sugiere que los residuos enumerados en la región 1 del dendrograma podrían unir alguna molécula orgánica y transmitir la información hasta la Ile94 para que la profilina pueda responder a la interacción con esta molécula. Cabe mencionar que otro residuo de este grupo, la Ile127 (magenta), resultó cercano a otra molécula de glicerol (GOL2), menos ordenada que la anterior, encontrada al resolver la estructura de la profilina de *Arabidopsis thaliana* (PDB: 3NUL).

Por su parte, los residuos de la región 2 del dendrograma de la Figura 10 b se graficaron contra la estructura terciaria de la profilina de *Arabidopsis thaliana* (Figura 12 e, f). Este grupo de residuos incluye a 3 amino ácidos del final de la  $\alpha$ -hélice C-terminal de la profilina de *Arabidopsis thaliana* (en azul marino), que son Glu128, Ser129 y Leu131. Estos residuos se sabe que son importantes para la unión de la profilina a la PLP [21, 59] y a los motivos ricos en prolina [30]. En el caso de la Ser129, también se ha visto que es importante para el pegado de la profilina al PIP<sub>2</sub>. La importancia de este residuo para el pegado de la profilina tanto a la PLP como al PIP<sub>2</sub> puede explicar la competitividad que se ha visto entre estos dos ligandos para su pegado a la profilina II de ratón [60]. Estos residuos al final de la  $\alpha$ -hélice C-terminal de la profilina inician una curva pequeña que termina en la Thr97, otro residuo del grupo, el cual es importante para el plegamiento de la profilina [6]. Se ha detectado un patrón de coevolución similar a éste mediante el SCA de otras familias de proteínas [4]: los grupos de residuos

acoplados forman una red que es estructuralmente contigua y que conecta al interior de la proteína con sus sitios funcionales. Este grupo de residuos coevolucionados parece ser que tiene la función de participar en el pegado de la profilina a los motivos ricos en prolina.

El grupo de residuos de la zona 3 del dendrograma de la Figura 10 b se graficó contra la estructura de la profilina de *Arabidopsis thaliana* en la Figura 12 g, h. Se puede observar que muchos de ellos (en amarillo) están cercanos a los residuos (en blanco) de la interfase de la interacción con la actina que aparecen como coevolucionados al hacer el SCA de todas las profilinas eucarióticas (véase más abajo). Otro residuo (en verde) se encuentra cerca de los pocos residuos (no mostrados) que resultaron lejanos a la interfase de la interacción con la actina en el SCA de todas las profilinas eucarióticas. Sólo hay dos residuos (en naranja) que no se encuentran cerca de los residuos coevolucionados en todas las profilinas reportadas, sin embargo, estos residuos están cerca del grupo de residuos mostrado en la Figura 12 e, f.

Las 23 posiciones del alineamiento de todas las profilinas reportadas que mostraron altos acoplamientos con otras en los experimentos de perturbación (Figura 10 c), se extrapolaron a la estructura terciaria de la profilina de *Arabidopsis thaliana* (Figura 12 i). La mayoría de estas posiciones tienen sus residuos (en blanco) en la interfase de la interacción de la profilina de *Arabidopsis thaliana* con la actina en su posición predicha a partir del complejo profilactina bovino, y aparecen aún más cerca de la actina al graficarse en las posiciones correspondientes de la profilina de este complejo profilactina (Figura 12 j). Algunos de estos amino ácidos ya han sido demostrados como importantes para la interacción de la profilina con la actina, éstos son: Thr111, Arg121, Val82, Arg84, Leu117, Ala61 y Met73. Uno de estos amino ácidos, la metionina 73, se encuentra en un motivo de tipo SH3 (un tipo de motivos que reconoce motivos ricos en prolina) que se ha visto que es importante para el pegado de la profilina a la región rica en prolina de la PI3K [30]. Sólo unas pocas posiciones tienen a sus residuos (en azul) en otra región de la profilina, sin embargo, también éstos están cercanos unos de otros.

## Discusión.

Se realizó el Análisis de Acoplamiento Estadístico en la familia de proteínas de la profilina, este análisis se hizo de manera independiente para las isoformas de profilina vegetal y para todas las isoformas profilina reportadas en las bases de datos. En casi todos los casos, los grupos de posiciones cuyos residuos se reportaron como coevolucionados en este análisis, se ha reportado que son importantes para la interacción de la profilina con sus ligandos.

El SCA se ha utilizado para analizar otras familias de proteínas, como la familia de proteínas de los dominios PDZ [2], los cuales reconocen regiones ricas en prolina, la familia de las proteínas G [3], los receptores acopladas a proteínas G, las proteasas de serina, las hemoglobinas [4] y los receptores X de retinoides [5]. Para estas familias de proteínas, el SCA detectó redes de amino ácidos que interaccionan entre sí, indistinguibles en la estructura terciaria de las proteínas, importantes en la respuesta de las mismas a la unión con sus ligandos.

Además del SCA, existen otros métodos para medir coevolución entre amino ácidos. Por ejemplo, están el MI (Mutual Information) y el OMES, que miden la ocurrencia de un par de amino ácidos en dos posiciones del alineamiento de secuencias de una familia de proteínas descartando la posibilidad de que ocurra este par al azar [61]. Otro algoritmo es el McBASC (McLahan Based Substitution Correlation) que calcula el coeficiente de correlación entre las puntuaciones según la matriz de McLahan de las sustituciones de los amino ácidos de dos posiciones de un alineamiento de secuencias [61]. Se han hecho comparaciones [61] entre el MI, el OMES, el McBASC y el SCA y se ha visto que el MI tiene una baja efectividad, y el OMES y el McBASC son los más efectivos para detectar coevolución entre posiciones muy conservadas, pero el SCA es mejor en bajos porcentajes de identidad a nivel de secuencias, como son los que existen entre las profilinas vegetales y las animales. Por eso creemos que el SCA puede detectar coevolución entre amino ácidos de todas las profilinas reportadas, y que estos resultados se pueden comparar con los obtenidos mediante el SCA para las profilinas vegetales.

En el caso de las profilinas vegetales, muchas posiciones del alineamiento resultaron ser bastante conservadas, lo que sugiere que ha habido poca divergencia evolutiva entre éstas. Sin embargo, en algunas posiciones la distribución de frecuencias de amino ácidos varía al

hacer una perturbación y en otras no se da ese cambio. Esto podría deberse a que existe una presión selectiva grande sobre toda la estructura primaria de la profilina vegetal. Esto además sugiere que las profilinas vegetales tienen una o varias funciones además de las que son comunes a todas las profilinas eucarióticas.

En el SCA de todas las profilinas reportadas se dan altas energías estadísticas de acoplamiento. La gran magnitud de estas energías puede deberse a que, como las profilinas vegetales son una mayoría y tienen una identidad del 70 al 90% entre sus secuencias pero tienen una identidad del 5 al 30% con las profilinas de otros eucariontes, muchas veces las perturbaciones consisten en escoger casi todas las secuencias de profilina vegetal, y estas perturbaciones inducen cambios en las frecuencias de las posiciones de amino ácidos que son valorados por el SCA con alta energía estadística de acoplamiento porque se dan en posiciones conservadas. Sin embargo, los resultados del SCA para todas las profilinas eucarióticas coinciden en buena parte con la información biológica conocida.

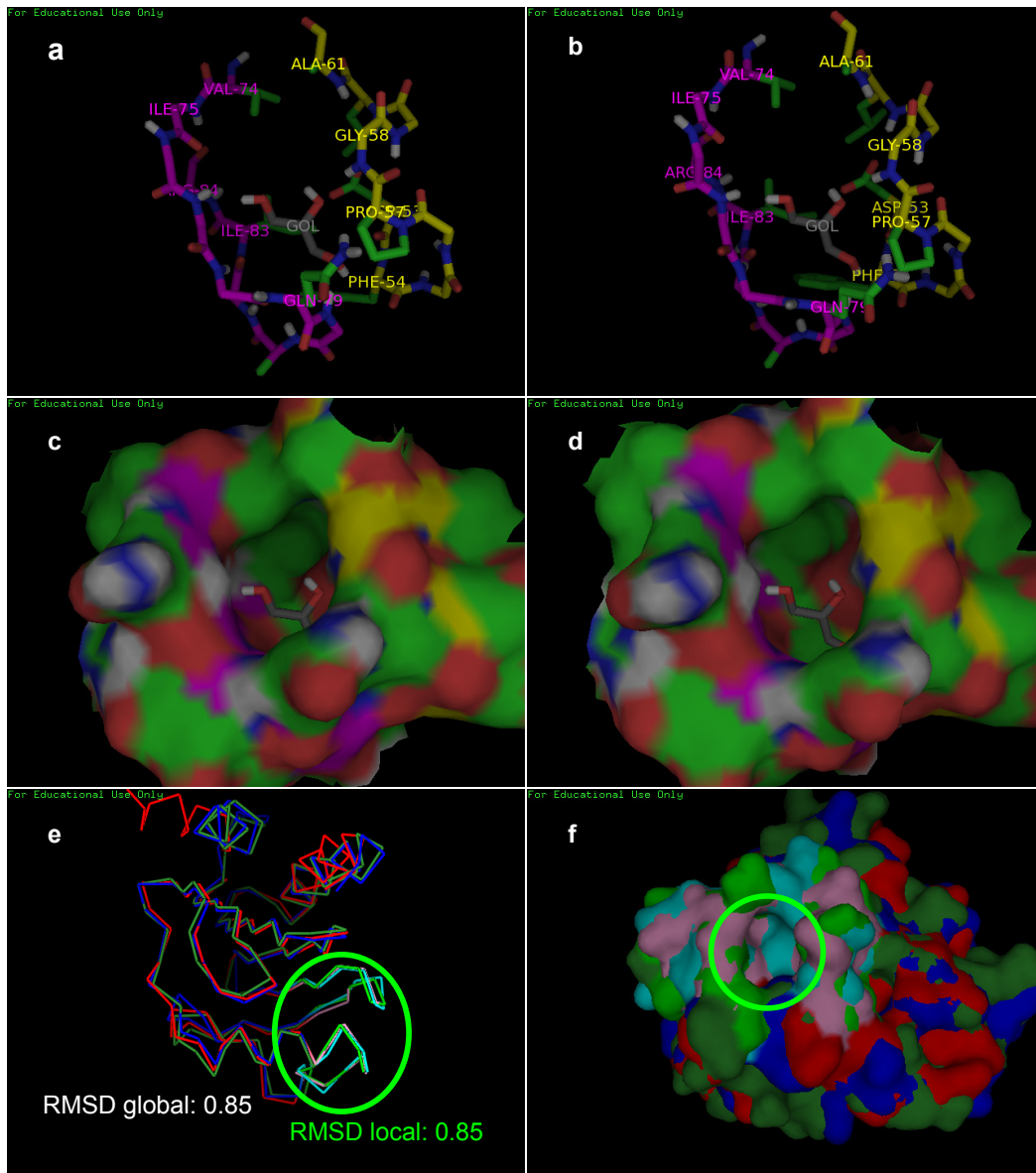
Después de hacer las perturbaciones a las posiciones de amino ácidos del alineamiento múltiple de secuencias de una familia de proteínas, Rama Ranganathan sugiere hacer, con base en la similitud de los patrones de coevolución de las distintas posiciones perturbadas, un dendrograma que las agrupe, a partir del cual se excluirá la rama de posiciones perturbadas que tengan alta conservación, ya que éstas por lo general dan bajas energías estadísticas de acoplamiento con las distintas posiciones de amino ácidos de la familia de proteínas, con tal de que el agrupamiento de las posiciones de la familia de proteínas sea más confiable. Pero nosotros no nos basamos únicamente en la conservación, sino también medimos la energía de acoplamiento promedio de las posiciones perturbadas con todas las posiciones de la familia de la profilina. Escogimos para eliminar la rama de posiciones perturbadas que tuvieran siempre o casi siempre un valor más bajo de este promedio. Estas posiciones no mostraron ningún acoplamiento significativo (datos no mostrados) con las distintas posiciones de la familia de proteínas.

Existen algunos datos que indican que las profilinas vegetales pueden tener otras funciones adicionales a las ya conocidas para todas las profilinas; muy probablemente reguladoras. Uno de estos datos es que un grupo de posiciones coevolucionadas entre sí según el SCA de las profilinas vegetales tiene a algunos de sus residuos en una zona de la profilina cuya

importancia no se conoce, sin embargo, están cerca de una cavidad existente sólo en las profilinas vegetales que Thorn y colaboradores propusieron como un sitio putativo de unión a alguna molécula orgánica [6]. Esta cavidad está formada por dos asas que son muy conservadas en las profilinas vegetales y las cuales son más largas que las correspondientes en las isoformas de la profilina de los otros eucariontes. En esta cavidad se encontraron contactos polares con una molécula muy ordenada de glicerol (Figura 13 a, b) en la estructura de la profilina de *Arabidopsis thaliana* que fue cristalizada por Thorn y colaboradores mediante el método de la gota colgante y resuelta mediante difracción de rayos X a 100 K con 1.6 Å de resolución (PDB: 3NUL), así como contactos con moléculas ordenadas de agua al resolver la estructura de esta profilina a 25 °C con una resolución de 2.2 Å (PDB: 1A0K) [6]. Dicha cavidad existe también en las otras 2 profilinas vegetales de estructura conocida (Figura 13 c, d). Thorn y colaboradores sugieren que esta cavidad podría ser el sitio de reconocimiento de una molécula orgánica [6], y esta hipótesis es reforzada por los resultados del Análisis Estadístico de Acoplamiento (SCA). Por el tamaño, la forma y la distribución de cargas electrostáticas de la cavidad, parece que la molécula orgánica que se podría unir a la profilina vegetal estaría formada por uno o dos anillos aromáticos con átomos polares [6]. Nuestros resultados revelaron que en esta región existen posiciones que están acopladas evolutivamente, además, estos residuos no se han demostrado que sean importantes para el reconocimiento de actina, PIP<sub>2</sub> y PLP; así como tampoco son importantes para mantener la conformación estructural de la profilina, ya que esta región no tiene valores menores de RMSD que toda la profilina al superponer las estructuras conocidas de profilina vegetal (Figura 13 e, f). Se sabe que la expresión del gen de la profilina es inhibido por un tipo de hormonas vegetales, las auxinas, lo cual nos sugirió que el compuesto que se una a la profilina en el sitio sugerido por Thorn sea una citocinina, otro tipo de hormona vegetal, derivada de la adenina, que es antagonista en muchos aspectos con las auxinas. Sin embargo, hicimos un anclaje molecular *in silico* de la profilina con una citocinina, la zeatina, utilizando el software AutoDock 4© [54, 55], mediante el algoritmo genético Lamarckiano, que es el más eficiente de los que usa AutoDock 4© [54], haciendo 25'000,000 evaluaciones, y no se predijo ninguna interacción de la zeatina con el sitio ya mencionado de la profilina. Asimismo, encontramos en la base de datos de PROSITE© (<http://www.expasy.org/prosite>) reportadas como ligandos de la profilina varias moléculas que

se derivan de la adenina por la adición de una cadena hidrocarbonada lineal. Ninguno de ellos tenía esta cadena en el carbono en que la tiene la zeatina, por lo que creemos que quizá la cadena hidrocarbonada lineal de la zeatina le estorba para su unión con la profilina pero la adenina sí podría ser un ligando de la profilina. Existe información que vincula a la profilina con la adenina, como es el hecho de que la adenina es necesaria para la nodulación de las raíces del frijol [62], y la profilina, que constantemente entra y sale del núcleo celular, pudiera entrar a éste cargada de adenina e interactuar con la nodulina 30, proteína por la que tiene una alta afinidad la profilina, que entra al núcleo durante la nodulación (Federico Sánchez, comunicación personal). Ya se ha visto que la profilina interacciona con un factor transcripcional para regular la expresión genética [24], por lo cual hicimos un análisis bioinformático de la nodulina 30 y vimos que esta proteína podría haber evolucionado a partir de un factor transcripcional ya identificado por medio de barajeo de dominios. Por esto creemos que la nodulina 30 podría ser un factor transcripcional que interactúe con la profilina para regular la expresión genética. Por ello, decidimos hacer un anclaje molecular de la adenina, la adenosina, y el mono-, di- y trifosfato de adenosina con la profilina usando el programa AutoDock 4©, y observamos que la adenina podría ser unida a la profilina en la cavidad de las profilinas vegetales, no así la adenosina y sus fosfatos.





**Figura 13: Posible sitio de unión de las profilinas vegetales a una molécula orgánica.** a) y b) Visión estereoscópica en forma de líneas gruesas de este sitio putativo en la profilina de *Arabidopsis thaliana* (PDB: 3NUL), el cual es una cavidad conformada por los amino ácidos de 2 cadenas polipeptídicas en forma de horquilla; la compuesta por los residuos Val74 a Arg84 (lila), y la compuesta por los residuos Asp53 a Ala61 (amarillo). Entre estas 2 cadenas se encontró una molécula de glicerol (cuyos carbonos se muestran en gris) que se cree que puede estar ocupando el lugar de una molécula que se una exclusivamente a las isoformas vegetales de la profilina [6]. Las cadenas laterales se omiten, excepto para los amino ácidos Asp53, Phe54, Pro57, Ala61, Val74, Gln79 e Ile83, cuyas cadenas laterales (que se muestran con los carbonos en verde) forman la cavidad en la cual se aloja la molécula de glicerol. Los átomos de nitrógeno se muestran en azul, los de oxígeno en rojo y los de hidrógeno en blanco. c) y d) Superficie electrostática de esta cavidad en la profilina de *Arabidopsis thaliana*. Los colores de la superficie son los de los átomos que forman ésta, según la coloración usada en los paneles a y b. e) Representación del esqueleto de carbonos  $\alpha$  de la superposición de las profilinas de *Arabidopsis thaliana* (verde oscuro), de látex (PDB: 1G5Ua) (azul marino) y de abedul (PDB: 1CQA) (rojo). Los sitios de las cavidades en estas profilinas (óvalo verde) se muestran en verde claro (profilina de *Arabidopsis thaliana*), en azul claro (profilina de látex) y en rosa (profilina de abedul). En la figura se muestran los RMSD en carbonos  $\alpha$  para esta cavidad (letras verdes) y para toda la profilina (letras blancas). f) Superposición de las superficies electrostáticas de las mismas profilinas que en el panel e) y con los mismos colores que en éste. La cavidad exclusiva de las profilinas vegetales se encierra en un óvalo verde.

Otro grupo de residuos coevolucionados que el SCA detectó para el caso de las profilinas vegetales incluye a algunas posiciones en el extremo carboxilo terminal de la profilina que se ha demostrado que son importantes en el pegado de la profilina a la PLP. En las profilinas vegetales y las isoformas II de las profilinas de mamífero, las cuales tienen alta afinidad por PLP, la posición 128 siempre tiene un residuo negativo, la 130 siempre es una glicina y la 131 es en todos los casos un amino ácido hidrofóbico. Las profilinas I de mamífero tienen distintos amino ácidos en esas posiciones, sin ningún tipo de conservación, y son muy poco afines a la PLP. Esto sugiere que la evolución ha permitido que existan secuencias sin este grupo de residuos del carboxilo terminal para que tengan una función que prácticamente no involucre pegado a los PRM. La posición 130 también es importante para el pegado a PIP<sub>2</sub>, lo que podría explicar que haya competencia entre los pegados a PLP y a PIP<sub>2</sub> en la profilina II de ratón [60].

En el SCA de las todas las profilinas eucarióticas, la mayoría de las posiciones con una alta energía de acoplamiento estadístico se encuentran en la interfase de la interacción con la actina. Una de estas posiciones, la metionina 73, se encuentra en el motivo <sub>72</sub>YMVI, en el cual está la tirosina 72 que se ha demostrado que es importante para el pegado de la PI3K a la profilina [30]. Esto indica que la PI3K podría desplazar a los monómeros de actina unidos a la profilina para incorporarlos a los filamentos en crecimiento. Además, las posiciones altamente coevolucionadas que el SCA reporta para todas las profilinas reportadas cuyos residuos están en la interfase con la actina y que no se han demostrado como importantes en el pegado a ésta, son buenas candidatas para comprobar su importancia en esta interacción por medio de mutagénesis sitio-dirigida.

Hay que notar que las posiciones que muestran coevolución entre sí en el alineamiento de las posiciones de las profilinas eucarióticas pueden tener interrelación entre sí en las profilinas vegetales, pero que estas interrelaciones no han aparecido en este análisis debido a la alta conservación de sus residuos. Se observó que, de hecho, las posiciones que muestran coevolución en el alineamiento de todas las profilinas presentan las conservaciones más altas en el alineamiento de las profilinas vegetales, siendo sólo superadas por las posiciones que unen a la PLP. Esto significa que probablemente existe una exigencia de funcionalidad casi perfecta en estos residuos, de tal forma que en estas posiciones casi no pueden ocurrir

mutaciones, aún cuando estas se compensen por mutaciones en otros sitios del mismo grupo de residuos de alto acoplamiento.

Los acoplamientos entre posiciones en el alineamiento de las profilinas vegetales también se dan, al menos en algunos casos, en el alineamiento de todas las profilinas reportadas, sin embargo, muchas veces este acoplamiento no es de los más altos, y por eso no se destacó en este trabajo.

## Conclusiones.

En esta tesis se demostró mediante el Análisis Estadístico de Acoplamiento que existe coevolución entre las posiciones de la familia de proteínas de la profilina que están en la interfase de la interacción de la profilina con cada uno de sus ligandos: PLP, PI3K, PIP<sub>2</sub> y actina. Cada uno de estos grupos de posiciones coevolucionadas tienen a sus residuos cercanos entre sí en la estructura de la profilina. Estas coevoluciones indican una relación entre los amino ácidos de estas posiciones, la cual puede ser química, física o funcional. Los grupos de posiciones coevolucionadas entre sí en los sitios de interacción con los distintos ligandos de la profilina también coevolucionan con posiciones importantes para el plegamiento de la misma.

## Perspectivas.

El Análisis Estadístico de Acoplamiento de las profilinas vegetales, además de detectar grupos de posiciones coevolucionadas entre sí cuyos residuos están cercanos a la interfase de la interacción de la profilina con sus ligandos conocidos, detectó un grupo de posiciones coevolucionadas cercanas a una cavidad existente sólo en las profilinas vegetales que se cree que podría unir a una molécula orgánica. Debido a la forma de este sitio, a predicciones *in silico* y a distintas observaciones que se han hecho sobre la profilina en este laboratorio, se cree que esta molécula orgánica podría ser la adenina, por lo cual, sería interesante comprobar experimentalmente si la profilina interacciona con la adenina y si lo hace en el sitio mencionado anteriormente.

Por otro lado, durante la elaboración de mi tesis, descubrí que existen muchos métodos de detección de coevolución entre aminoácidos, por lo cual creo que sería interesante ver cuáles dan un resultado que parezca más convincente para la profilina, o si cada uno de ellos detecta un distinto tipo de acoplamiento entre amino ácidos. También sería interesante detectar si existe coevolución entre la profilina y una porción de VASP distinta a sus regiones ricas en prolina que parece ser que está en contacto con la profilina, así como detectar qué aminoácidos participan en la coevolución que ha habido entre las profilinas y las actinas que interactúan específicamente *in vivo*. Además, también se podría investigar si ha habido coevolución entre la profilina y la PI3K.

## Referencias.

1. Socolich, M., Lockless, S. W., Russ, W. P., Lee, H., Gardner, K. H. & Ranganathan, R. (2005) Evolutionary information for specifying a protein fold, *Nature*. 437, 512-518.
2. Lockless, S. W. & Ranganathan, R. (1999) Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families, *Science*. 286, 295-299.
3. Hatley, M. E., Lockless, S. W., Gibson, S. K., Gilman, A. G. & Ranganathan, R. (2003) Allosteric determinants in guanine nucleotide-binding proteins, *Proceedings of the National Academy of Sciences of the United States of America*. 100, 14445-14450.
4. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins, *Nature Structural & Molecular Biology*. 10, 59-69.
5. Shulman, A. I., Larson, C., Mangelsdorf, D. J. & Ranganathan, R. (2004) Structural Determinants of Allosteric Ligand Activation in RXR Heterodimers, *Cell*. 116, 417-429.
6. Thorn, K. S., Christensen, H. E., Shigeta, R., Huddler, D., Shalaby, L., Lindberg, U., Chua, N. H. & Schutt, C. E. (1997) The crystal structure of a major allergen from plants, *Structure*. 5, 19-32.
7. Kandasamy, M. K., Burgos-Rivera, B., McKinney, E. C., Ruzicka, D. R. & Meagher, R. B. (2007) Class-Specific Interaction of Profilin and ADF Isovariants with Actin in the Regulation of Plant Development in pp. tpc.107.052621
8. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science*. 181, 223-230.
9. Flores-Bustos, B. (2006) *Análisis Filogenético de la Secuencia de Aminoácidos de la Profilina*, Universidad Autónoma del Estado de Morelos, Cuernavaca.
10. Ulrike Göbel, C. S. R. S. A. V. (1994) Correlated mutations and residue contacts in proteins, *Proteins: Structure, Function, and Genetics*. 18, 309-317.
11. Olmea, O. & Valencia, A. (1997) Improving contact predictions by the combination of correlated mutations and other sources of sequence information, *Folding and Design*. 2, S25-S32-S25-S32.
12. Neher, E. (1994) How Frequent are Correlated Changes in Families of Protein Sequences?, *Proceedings of the National Academy of Sciences*. 91, 98-102.
13. Pollard, T. D., Blanchoin, L. & Mullins, R. D. (2000) Molecular mechanisms controlling actin filament dynamics in nonmuscle cells, *Annual Review of Biophysics and Biomolecular Structure*. 29, 545-76.
14. Schluter, K., Jockusch, B. M. & Rothkegel, M. (1997) Profilins as regulators of actin dynamics, *Biochimica et Biophysica Acta*. 1359, 97-109.
15. Lassing, I. & Lindberg, U. (1985) Specific interaction between phosphatidylinositol 4,5-bisphosphate and profilactin, *Nature*. 314, 472-4.

16. François Ferron, G. R. S. H. L. a. R. D. (2007) Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP, *The EMBO Journal*. 26, 4597-4606.
17. Bubb, M. R., Baines, I. C. & Korn, E. D. (1998) Localization of actobindin, profilin I, profilin II, and phosphatidylinositol-4,5-bisphosphate (PIP2) in *Acanthamoeba castellanii*, *Cell Motility and the Cytoskeleton*. 39, 134-46.
18. Bubb, M. R., Yarmola, E. G., Gibson, B. G. & Southwick, F. S. (2003) Depolymerization of actin filaments by profilin. Effects of profilin on capping protein function, *The Journal of Biological Chemistry*. 278, 24629-35.
19. Nodelman, I. M., Bowman, G. D., Lindberg, U. & Schutt, C. E. (1999) X-ray structure determination of human profilin II: A comparative structural analysis of human profilins, *J Mol Biol*. 294, 1271-85.
20. Witke, W. (2004) The role of profilin complexes in cell motility and other cellular processes, *Trends in Cell Biology*. 14, 461-9.
21. Metzler, W. J., Bell, A. J., Ernst, E., Lavoie, T. B. & Mueller, L. (1994) Identification of the poly-L-proline-binding site on human profilin, *The Journal of Biological Chemistry*. 269, 4620-5.
22. Witke, W., Podtelejnikov, A. V., Di Nardo, A., Sutherland, J. D., Gurniak, C. B., Dotti, C. & Mann, M. (1998) In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly, *The EMBO Journal*. 17, 967-76.
23. Mahoney, N. M., Janmey, P. A. & Almo, S. C. (1997) Structure of the profilin-poly-L-proline complex involved in morphogenesis and cytoskeletal regulation, *Nature Structural & Molecular Biology*. 4, 953-960.
24. Lederer, M., Jockusch, B. M. & Rothkegel, M. (2005) Profilin regulates the activity of p42POP, a novel Myb-related transcription factor, *Journal of Cell Science*. 118, 331-341.
25. Burke, E., Mahoney, Nicole M., Almo, Steven C., Barik, Sailen. (2000) Profilin Is Required for Optimal Actin-Dependent Transcription of Respiratory Syncytial Virus Genome RNA, *Journal of Virology*. 74, 669-675.
26. Carlsson, L., Nystrom, L. E., Lindberg, U., Kannan, K. K., Cid-Dresdner, H. & Lovgren, S. (1976) Crystallization of a non-muscle actin, *J Mol Biol*. 105, 353-66.
27. Ballweber, E., Giehl, K., Hannappel, E., Huff, T., Jockusch, B. M. & Mannherz, H. G. (1998) Plant profilin induces actin polymerization from actin : beta-thymosin complexes and competes directly with beta-thymosins and with negative co-operativity with DNase I for binding to actin, *FEBS Letters*. 425, 251-5.
28. Eads, J. C., Mahoney, N. M., Vorobiev, S., Bresnick, A. R., Wen, K. K., Rubenstein, P. A., Haarer, B. K. & Almo, S. C. (1998) Structure determination and characterization of *Saccharomyces cerevisiae* profilin, *Biochemistry*. 37, 11171-81.
29. Perelroizen, I., Didry, D., Christensen, H., Chua, N.-H. & Carlier, M.-F. (1996) Role of Nucleotide Exchange and Hydrolysis in the Function of Profilin in Actin Assembly in pp. 12302-12309
30. Aparicio-Fabre, R., Guillen, G., Estrada, G., Olivares-Grajales, J., Gurrola, G. & Sanchez, F. (2006) Profilin tyrosine phosphorylation in poly-L-proline-binding regions inhibits binding to phosphoinositide 3-kinase in *Phaseolus vulgaris*, *The Plant Journal*. 47, 491-500.
31. Almo, A. A. F. K. A. M. M. H. G. E. E. L. T. D. P. S. C. (1994) X-Ray Structures of Isoforms of the Actin-Binding Protein Profilin that Differ in their Affinity for Phosphatidylinositol

Phosphates, *Proceedings of the National Academy of Sciences of the United States of America*. 91, 8636-8640.

32. Sohn, R. H., Chen, J., Koblan, K. S., Bray, P. F. & Goldschmidt-Clermont, P. J. (1995) Localization of a binding site for phosphatidylinositol 4,5-bisphosphate on human profilin, *The Journal of Biological Chemistry*. 270, 21114-20.

33. Bo, M., Håkkan, L. & Uno, L. (1983) The profilin-actin complex: further characterization of profilin and studies on the stability of the complex, *Journal of Muscle Research and Cell Motility*. V4, 569-588.

34. Haarer, B. K., Petzold, A. S. & Brown, S. S. (1993) Mutational analysis of yeast profilin, *Molecular and Cell Biology*. 13, 7864-73.

35. Bjorkegren, C., Rozycki, M., Schutt, C. E., Lindberg, U. & Karlsson, R. (1993) Mutagenesis of human profilin locates its poly(-proline)-binding site to a hydrophobic patch of aromatic amino acids, *FEBS Letters*. 333, 123-126.

36. Vandekerckhove, J. S., Kaiser, D. A. & Pollard, T. D. (1989) Acanthamoeba actin and profilin can be cross-linked between glutamic acid 364 of actin and lysine 115 of profilin, *The Journal of Cell Biology*. 109, 619-626.

37. Schluter, K., Schleicher, M. & Jockusch, B. M. (1998) Effects of single amino acid substitutions in the actin-binding site on the biological activity of bovine profilin I, *Journal of Cell Science*. 111 ( Pt 22), 3261-73.

38. Bjorkegren-Sjogren, C., Korenbaum, E., Nordberg, P., Lindberg, U. & Karlsson, R. (1997) Isolation and characterization of two mutants of human profilin I that do not bind poly(-proline), *FEBS Letters*. 418, 258-264.

39. Hajkova, L., Bjorkegren Sjogren, C., Korenbaum, E., Nordberg, P. & Karlsson, R. (1997) Characterization of a mutant profilin with reduced actin-binding capacity: effects in vitro and in vivo, *Experimental Cell Research*. 234, 66-77.

40. Korenbaum, E., Nordberg, P., Bjorkegren-Sjogren, C., Schutt, C. E., Lindberg, U. & Karlsson, R. (1998) The role of profilin in actin polymerization and nucleotide exchange, *Biochemistry*. 37, 9274-83.

41. Skare, P. & Karlsson, R. (2002) Evidence for two interaction regions for phosphatidylinositol(4,5)-bisphosphate on mammalian profilin I, *FEBS Letters*. 522, 119-24.

42. Huang, S., McDowell, J. M., Weise, M. J. & Meagher, R. B. (1996) The Arabidopsis profilin gene family. Evidence for an ancient split between constitutive and pollen-specific profilin genes, *Plant Physiology*. 111, 115-26.

43. NCBI. (Revised 2006) Entrez Protein

44. Zdobnov, E. M., Lopez, R., Apweiler, R. & Etzold, T. (2002) The EBI SRS server--recent developments, *Bioinformatics*. 18, 368-373.

45. Zdobnov, E. M., Lopez, R., Apweiler, R. & Etzold, T. (2002) The EBI SRS server--new features, *Bioinformatics*. 18, 1149-1150.

46. Wu, R. A. A. B. C. (2002) UniProt

47. H.M. Berman, J. W., Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank, *Nucleic Acids Research*, 28 pp. 235-242.

48. Shindyalov, I. N. & Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Engineering, Design and Selection*. *11*, 739-747.
49. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Research*. *31*, 3497-3500.
50. Gille, C. F., C. (2001) STRAP: editor for STRuctural Alignments of Proteins, *Bioinformatics*. *17*, 377-378.
51. Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symposium Series*. *41*, 95-98.
52. Getz, G., Levine, E. & Domany, E. (2000) Coupled two-way clustering analysis of gene microarray data, *Proceedings of the National Academy of Sciences of the United States of America*. *97*, 12079-12084.
53. DeLano, W. L. (2002) The PyMOL Molecular Graphics System
54. Sanner, M. F. (1999) Python: A Programming Language for Software Integration and Development, *Journal of Molecular Graphics & Modelling*. *17*, 57-61.
55. Garrett M. Morris, D. S. G. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *Journal of Computational Chemistry*. *19*, 1639-1662.
56. Gasteiger, J. & Marsili, M. (1980) Iterative partial equalization of orbital electronegativity--a rapid access to atomic charges, *Tetrahedron*. *36*, 3219-3228.
57. Johann Gasteiger, M. M. (1981) Prediction of proton magnetic resonance shifts: The dependence on hydrogen charges obtained by iterative partial equalization of orbital electronegativity, *Organic Magnetic Resonance*. *15*, 353-360.
58. www.accelrys.com. (2007) Accelrys: Discovery Studio Visualizer in
59. Archer, S. J., Vinson, V. K., Pollard, T. D. & Torchia, D. A. (1994) Elucidation of the poly-proline binding site in Acanthamoeba profilin I by NMR spectroscopy, *FEBS Letters*. *337*, 145-151.
60. A. Lambrechts, J. L. V., V. Jonckheere, M. Goethals, J. Vandekerckhove, and C. Ampe. (1997 February 3 doi: 10.1093/emboj/16.3.484.) The mammalian profilin isoforms display complementary affinities for PIP2 and proline-rich sequences., *The EMBO Journal*. *16*, 484-494.
61. Fodor, A. A. & Aldrich, R. W. (2004) Influence of conservation on calculations of amino acid covariance in multiple sequence alignments, *Proteins*. *56*, 211 - 21-211 - 21.
62. Pankhurst, C. E. & Schwinghamer, E. A. (1974) Adenine requirement for nodulation of pea by an auxotrophic mutant of *Rhizobium leguminosarum*, *Archives of Microbiology*. *100*, 219-238.