



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

ESTUDIO DE LA RELACIÓN ESTRUCTURA-FUNCIÓN DE LAS
PROTEÍNAS LEA DEL GRUPO 4 DE *Arabidopsis thaliana*

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

IBQ CESAR LUIS CUEVAS VELAZQUEZ

TUTOR:

DRA. ALEJANDRA A. COVARRUBIAS ROBLES
INSTITUTO DE BIOTECNOLOGÍA, UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

COMITÉ TUTOR:

FELIPE CRUZ GARCÍA
FACULTAD DE QUÍMICA, UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO
GLORIA SAAB RINCÓN
INSTITUTO DE BIOTECNOLOGÍA, UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

CUERNAVACA, MORELOS. OCTUBRE, 2016



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

Dedicado a la memoria de Osvaldo Jesús Aguilar Suárez

AGRADECIMIENTOS

A la Dra. Alejandra Covarrubias Robles, de quien aprendí innumerables cualidades científicas. Agradezco profundamente que me hayas permitido compartir contigo cientos de anécdotas durante todos estos años. Por ser más que una tutora, una amiga.

A los miembros de mi comité tutorial de maestría y de doctorado. Dr. Mario Rocha Sosa, Dr. Juan Francisco Jimenez Bremont, Dr. Felipe Cruz García, Dra. Gloria Saab Rincón y Dr. Enrique Rudiño. Sus aportaciones en cada una de nuestras reuniones contribuyeron significativamente en el desarrollo de mi tesis.

Al comité de revisión de tesis. Dra. Marcela Ayala, Dr. Luis Cárdenas, Dr. Gerardo Corzo, Dra. Rosario Muñoz Clares y Dr. Lorenzo Segovia. Por su valiosa aportación al mejoramiento del reporte escrito de esta tesis.

Al Programa de Apoyo a los Estudiantes de Posgrado (PAEP) de la UNAM. Por el apoyo económico para asistir a congresos nacionales e internacionales, así como estancias de investigación.

A mis compañeros del grupo de la Dra. Covarrubias. A todos y cada uno de ustedes de quienes aprendí todos los días dentro y fuera del laboratorio. Agradezco especialmente a aquellos con quienes tuve la fortuna de colaborar directamente: Yadira, Pepe, Pancho, Rosy, Inti, Lucero, Coral, David, Paulette, André y Dante.

A Rocio Rodriguez. Por tu apoyo incondicional en todos los aspectos. Sin ti, ninguna de mis metas tendría sentido.

A mi familia, por siempre creer en mi. A mi papá Santiago, mi mamá Luisa, mis hermanas Luisa y Alina, mi familia extendida Ofelia y Fátima.

ÍNDICE

RESUMEN	7
ABSTRACT	10
ABREVIATURAS.	12
I. INTRODUCCIÓN	13
I.I. El agua y el déficit hídrico como factores determinantes en la vida de las plantas	13
I.II. Las respuestas de las plantas al déficit hídrico	16
I.III. Las Proteínas Abundantes en la Embriogénesis Tardía	19
I.IV. Las Proteínas Intrínsecamente Desordenadas	29
I.V. Técnicas utilizadas para caracterizar a las Proteínas Intrínsecamente Desordenadas	34
I.VI. Dicroísmo circular	34
I.VII. Resonancia magnética nuclear	36
II. ANTECEDENTES	39
III. HIPÓTESIS	42
IV. OBJETIVO GENERAL	42
V. OBJETIVOS ESPECÍFICOS	42
VI. MATERIALES Y MÉTODOS	43
VI.I. Análisis <i>in silico</i> de las proteínas LEA del grupo 4 de <i>Arabidopsis thaliana</i>	43
VI.II. Construcción de los plásmidos utilizados con las distintas versiones de las AtLEA4	43
VI.III. Expresión de las proteínas recombinantes	45
VI.IV. Purificación de las proteínas recombinantes	46
VI.V. Expresión y purificación de las proteínas recombinantes marcadas con ¹⁵ N	48
VI.VI. Determinación de la estructura secundaria por dicroísmo circular	51
VI.VII. Determinación de la estructura secundaria por resonancia magnética nuclear	52
VI.VIII. Ensayos de congelamiento/descongelamiento <i>in vitro</i>	52
VI.IX. Ensayos de deshidratación parcial <i>in vitro</i>	53
VI.X. Análisis estadístico	54
VII. RESULTADOS	54
VII.I. Análisis <i>in silico</i> de las proteínas LEA del grupo 4 de <i>Arabidopsis thaliana</i>	54
VII.II. Las proteínas AtLEA4-2 y AtLEA4-5 son proteínas intrínsecamente desordenadas en solución	58
VII.III. Las proteínas AtLEA4-2 y AtLEA4-5 tienen un alto potencial de formar α -hélices	65
VII.IV. La disminución en la cantidad de agua disponible y el amontonamiento macromolecular inducen el plegamiento de las proteínas LEA del grupo 4.	68
VII.V. La región N-terminal conservada de la AtLEA4-5 es necesaria y suficiente para inducir el plegamiento inducido por la disminución en la cantidad de agua disponible y el amontonamiento macromolecular simulados <i>in vitro</i>	72
VII.VI. La región N-terminal conservada de la AtLEA4-5 es necesaria y suficiente para prevenir la pérdida de actividad de la enzima lactato deshidrogenasa causada por deshidratación o por ciclos de congelamiento/descongelamiento <i>in vitro</i>	76

VIII. DISCUSIÓN	78
IX. CONCLUSIONES	85
X. PERSPECTIVAS	85
XI. BIBLIOGRAFÍA	87
PROLOGO A LOS RESULTADOS ANEXOS	103
ANEXO I. Descifrando el mecanismo de protección a nivel molecular de las proteínas abundantes en la embriogénesis tardía.	104
i.i. INTRODUCCIÓN	104
i.i.i. Mecanismos de acción de las chaperonas moleculares y las sHSPs.....	104
i.i.ii. Proteínas condicionalmente desordenadas.....	107
i.i.iii. Plegamiento de las proteínas intrínsecamente desordenadas asociado a la interacción con sus blancos.....	110
i.i.iv. Complejos fuzzy.....	111
i.ii. ANTECEDENTES	112
i.iii. PREGUNTA	114
i.iv. OBJETIVO GENERAL	114
i.v. OBJETIVOS ESPECÍFICOS	114
i.vi. MATERIALES Y MÉTODOS	115
i.vii. RESULTADOS	117
i.vii.i. La AtLEA4-5 se entrecruza con la LDH <i>in vitro</i>	117
i.vii.ii. La AtLEA4-5 no experimenta cambios conformacionales en presencia de la LDH.....	120
i.viii. DISCUSIÓN	123
i.ix. CONCLUSIONES	126
i.x. PERSPECTIVAS	126
ANEXO II. La unión a metales y el efecto sobre la estructura y función de las proteínas abundantes en la embriogénesis tardía.	127
ii.i. INTRODUCCIÓN	127
ii.i.i. Los iones metálicos de protección propuestos para las proteínas LEA.....	127
ii.ii. ANTECEDENTES	128
ii.iii. PREGUNTA	130
ii.iv. OBJETIVO GENERAL	130
ii.v. OBJETIVOS ESPECÍFICOS	130
ii.vi. MATERIALES Y MÉTODOS	130
ii.vi.i. Cromatografía de afinidad a iones metálicos inmovilizados.....	130
ii.vi.ii. Determinación de cambios conformacionales inducidos por la presencia de CuCl ₂	131
ii.vi.iii. Inactivación de la actividad de la LDH causada por CuCl ₂	131
ii.vii. RESULTADOS	132
ii.vii.i. La proteína AtLEA4-5 une metales divalentes.....	132

ii.vii.ii. La proteína AtLEA4-5 no modifica su conformación tridimensional cuando une Cu^{2+}	135
ii.vii.iii. El efecto del Cu^{2+} sobre la función de la proteína AtLEA4-5	136
ii.viii. DISCUSIÓN	137
ii.ix. CONCLUSIONES	139
ii.x. PERSPECTIVAS	139
ANEXO III. Oligonucleótidos.....	140
ANEXO IV. Medios de cultivo utilizados para el crecimiento de bacterias.....	141
ANEXO V. Validación de secuencias nucleotídicas.....	142
ANEXO VI. Identificación de proteínas por espectrometría de masas.....	151
ANEXO VII. Artículos científicos publicados durante los estudios de doctorado.	162

RESUMEN

El déficit hídrico es uno de los factores principales que determinan el desarrollo, productividad y sobrevivencia de las plantas. Debido a esto, las plantas han desarrollado estrategias para contender con dichas condiciones adversas. Una de las respuestas generales al estrés por déficit hídrico es la acumulación de un grupo de proteínas conocidas como proteínas abundantes en la embriogénesis tardía, comúnmente llamadas proteínas LEA (del inglés Late Embryogenesis Abundant proteins). Las proteínas LEA se acumulan en la etapa de desecación natural de todas las semillas ortodoxas (aquellas que requieren pasar por un período de desecación para continuar su ciclo de vida) y también se acumulan en tejidos vegetativos y reproductivos en respuesta a varios tipos de estrés abiótico, principalmente déficit hídrico. La mayoría de estas proteínas se caracterizan por ser pequeñas y estar enriquecidas en aminoácidos hidrofílicos y aminoácidos pequeños (polares y apolares), lo que les permite pertenecer al grupo de las hidrofílicas. Tales características fisicoquímicas predicen que un vasto número de proteínas LEA carece de una estructura tridimensional bien definida y que podrían formar parte de las llamadas proteínas intrínsecamente desordenadas (PIDs). En función de la similitud entre sus secuencias de aminoácidos y dependiendo de la presencia de motivos distintivos, las proteínas LEA se clasifican en siete grupos. En la planta modelo *Arabidopsis thaliana*, existen 51 genes que codifican para proteínas LEA. El grupo 4, que en *Arabidopsis* está representado por tres miembros (*AtLEA4-1*, *AtLEA4-2* y *AtLEA4-5*), es uno de los grupos de proteínas LEA con menor número de genes en esta planta. En estudios previos en nuestro laboratorio, se demostró que las proteínas LEA del grupo 4 participan en la tolerancia de *Arabidopsis* al estrés por falta de agua, pues al anular y/o silenciar la expresión de los genes de esta familia la planta presenta retraso en la germinación, menor recuperación y menor producción de botones florales comparado con la planta silvestre, cuando se somete a condiciones de estrés por déficit hídrico.

Si bien la relación entre la acumulación de las proteínas LEA y la disminución en la cantidad de agua disponible en tejidos vegetales es muy clara, el porqué de dicha acumulación es una pregunta que aún no ha sido resuelta del todo. Una de las hipótesis que aborda esta pregunta considera que las proteínas LEA del grupo 4 previenen los daños a los componentes celulares causados por la deficiencia de agua,

particularmente, protegiendo macromoléculas de forma análoga a la función que tienen las llamadas chaperonas moleculares en condiciones de estrés por calor. Además, dado que se predice que estas proteínas carecen de una estructura bien definida y que se acumulan en condiciones donde el agua es limitante, nos preguntamos: ¿cuál es la estructura de las proteínas LEA del grupo 4 cuando se encuentran en condiciones de déficit hídrico? y ¿cuál es la relevancia de poseer ese tipo de estructura sobre su supuesta función protectora?

En esta tesis, se abordó una estrategia para tratar de entender la relación que existe entre la estructura y la función de las proteínas LEA del grupo 4. Para ello, se purificaron las proteínas recombinantes AtLEA4-2 y AtLEA4-5 y se caracterizó su estructura secundaria usando las técnicas de resonancia magnética nuclear y de diroísmo circular. En concordancia con las predicciones *in silico*, las proteínas LEA del grupo 4 (LEA4) presentan una estructura que muestra giros azarosos (“random coil”) y que dada la ausencia de una estructura estable se le conoce como estructura desordenada. En esta tesis se demuestra que las proteínas LEA4 poseen un alto potencial de formar estructuras tipo α -hélice, capacidad revelada al adicionar cantidades crecientes de trifluoroetanol (agente que induce la formación de α -hélices en proteínas que tienen la capacidad de formar este tipo de estructura). Interesantemente, la AtLEA4-2 y la AtLEA4-5 se pliegan en α -hélices cuando se simula la disminución en la cantidad de agua disponible (usando glicerol) y el amontonamiento macromolecular (usando polietilenglicol, PEG) *in vitro*. Además, la región N-terminal (la cual está conservada entre todas las proteínas LEA4 de plantas) de la AtLEA4-5 es necesaria y suficiente para inducir un cambio conformacional hacia α -hélices, mientras que la región C-terminal permanece desordenada en cualquiera de las condiciones probadas. Nuestros datos sugieren que existe una íntima relación entre la capacidad de las proteínas LEA4 de plegarse y su función, pues tanto la AtLEA4-2, la AtLEA4-5 y la región N-terminal conservada son capaces de evitar la pérdida de actividad de una proteína reportera (lactato deshidrogenasa, LDH) causada por tratamientos de déficit hídrico impuestos *in vitro*, mientras que la región C-terminal no presenta ningún efecto protector. Las relaciones molares a las cuales se observaron altos niveles de protección son bajas (1:1-1:5; LDH:LEA), lo que sugiere que el mecanismo a través del cual las proteínas LEA4 ejercen la protección involucra una interacción proteína-proteína.

Adicionalmente, se incluyen dos secciones de anexos (Anexo I y Anexo II) con resultados que se obtuvieron durante la realización del doctorado. En el Anexo I se presentan los estudios enfocados a dilucidar el mecanismo de protección. En concordancia con la hipótesis, la AtLEA4-5 interacciona con la LDH a bajas relaciones molares; sin embargo, contrario a lo observado para la mayoría de las proteínas intrínsecamente desordenadas, la AtLEA4-5 no se pliega cuando interacciona con su proteína blanco. Finalmente, en el Anexo II se describe la capacidad de la AtLEA4-5 de unir cationes divalentes (Cu^{2+} y Zn^{2+}) y el efecto de éstos sobre la estructura y la función protectora de esta proteína multifacética. En conclusión, nuestros datos muestran que la flexibilidad de las proteínas LEA4 está íntimamente relacionada a su función protectora *in vitro*, por lo que este grupo de proteínas se presenta como un modelo para estudiar la relación estructura-función de las proteínas intrínsecamente desordenadas en plantas.

ABSTRACT

Water deficit is one of the main factors that affect the proper development, productivity and survival of plants. Because of that, plants have developed strategies to cope with such harsh conditions. One of the general responses to water deficit is the accumulation of a group of proteins known as Late Embryogenesis Abundant (LEA) proteins. LEA proteins accumulate during natural desiccation of all orthodox seeds and also in vegetative and reproductive tissues, mostly under water deficit. Most LEA proteins are small and are enriched in hydrophilic amino acids, which allows them to be part of a wider group of proteins known as hydrophilins. These physicochemical characteristics predict that a vast number of LEA proteins lack a well-defined tridimensional structure and could be part of the so-called intrinsically disordered proteins (IDPs). Based on sequence homology and the presence of distinctive motifs, LEA proteins are classified in 7 groups or families. In the model plant *Arabidopsis thaliana* there are 51 genes that code for LEA proteins. Group 4, composed of 3 members in Arabidopsis (AtLEA4-1, AtLEA4-2 and AtLEA4-5), is one of the smallest LEA groups in this plant. Previously in our laboratory, it was demonstrated that group 4 LEA proteins participate in Arabidopsis tolerance to water deficit because knocking out or silencing the genes of this family results in reduced germination, recovery and production of flower buds compared to wild type plants, when subjected to water deficit.

Even though the relationship between LEA proteins accumulation and low water availability is well established, the reason for this accumulation, or what is the function of LEA proteins is still unclear. One of the most accepted hypothesis is that LEA proteins prevent the damage to cellular compartments caused by water scarcity, particularly protecting macromolecules such as other proteins, in a similar manner to the function of molecular chaperones which function under heat stress. Moreover, since LEA proteins are predicted to lack a defined tridimensional structure and accumulate under conditions where water is scarce, we wonder, what is the structure of group 4 LEA proteins when they are under water deficit? And what is the meaning of having such structure on their putative protective function?

In this dissertation, we aim to answer those questions. We employed a strategy aimed to understand the relationship between the structure and function of group 4 LEA proteins. In order to do that, we produced recombinant AtLEA4-2 and AtLEA4-

5 and characterized their secondary structure using nuclear magnetic resonance and circular dichroism. In accordance with *in silico* predictions, group 4 LEA proteins are disordered in aqueous solution however, both proteins have the potential to acquire α -helix structure, as revealed by adding increasing amounts of trifluoroethanol (a typical inductor of α -helix on proteins that have the intrinsic potential to form this kind of structure). Interestingly, AtLEA4-2 and AtLEA4-5 fold into α -helix *in vitro* when we simulated the decrease in water availability (using glycerol) or macromolecular crowding (using polyethylene glycol). In addition, the N-terminal region (which is conserved across all group 4 LEA proteins in plants) of AtLEA4-5 is necessary and sufficient to induce the conformational change into α -helix, while the C-terminal region remains disordered under all the conditions tested. Our data suggest that there is an intimate relationship between the ability to fold into α -helix and their function since AtLEA4-2, AtLEA4-5 and the N-terminal conserved region of AtLEA4-5 are able to prevent the inactivation of a reporter enzyme (lactate dehydrogenase) caused by *in vitro* water deficit treatments, while the C-terminal region of AtLEA4-5 has no protective activity. The low molar ratios needed to attain high protection levels suggest that the mechanism through which group 4 LEA proteins exert their protection involves protein-protein interaction.

Additionally, two adjunct sections were added (Adjunct I and Adjunct II) with additional results obtained during the PhD studies. In Adjunct I, we present the efforts aimed to elucidating the protection mechanism of LEA4 proteins. In accordance with the hypothesis, AtLEA4-5 interacts with LDH at low molar ratios. However, contrary to most intrinsically disordered proteins, it does not fold upon binding to its target protein. Finally, in Adjunct II, we describe the ability of AtLEA4-5 to bind divalent cations (Cu^{2+} and Zn^{2+}) and the effect of these cations on the structure and the protective function of this moonlighting protein.

In conclusion, our data suggest that the flexibility of group 4 LEA proteins is intimately related to their protective function *in vitro*. Therefore, we propose that this group of proteins might be used as a model to study the molecular mechanisms of intrinsically disordered proteins of plants.

ABREVIATURAS.

ADN	Ácido desoxirribonucleico
APX	Ascorbato peroxidasa
ARN	Ácido ribonucleico
ASB	Albúmina sérica bovina
ATP	Adenosín trifosfato
cAND	Ácido desoxirribonucleico complementario
CAR	Catalasa
CRA	Contenido relativo de agua
CS	Citrato sintasa
DC	Dicroísmo circular
DIL	Decaimiento de inducción libre
ENO	Efecto nuclear Overhauser
FRET	Transferencia de energía por resonancia de fluorescencia
FTIR	Espectroscopía de infrarrojo
GR	Glutación reductasa
HSP	Proteínas de choque térmico (Heat Shock Proteins)
IDPs	Intrinsically Disordered Proteins
IPTG	Isopropil β -D-1-tiogalactopiranosido
LDH	Lactato deshidrogenasa
LEA	Proteínas abundantes en la embriogénesis tardía (Late Embryogenesis Abundant proteins)
MDHAR	Monodehidroascorbato reductasa
MPa	Megapascuales
PCR	Reacción en cadena de la polimerasa (Polymerase chain reaction)
PEG	Polietilén glicol
PIDs	Proteínas Intrínsecamente Desordenadas
POD	Peroxidasa
PreSMos	Motivos estructurales pre-formados (Pre-Structured Motifs)
RMN	Resonancia Magnética Nuclear
sARN	Ácido ribonucleico pequeño
SDS	Dodecil sulfato de sodio
SDS-PAGE	Dodecil sulfato de sodio-Electroforésis en geles de poliacrilamida
sHSP	Proteínas de choque térmico pequeñas (small Heat Shock Proteins)
SOD	Superóxido dismutasa
TCA	Ácido tricloroacético
TFE	Trifluoroetanol
TFM	Tubo fotomultiplicador

I. INTRODUCCIÓN

I.I. El agua y el déficit hídrico como factores determinantes en la vida de las plantas

Usualmente, percibimos que el agua es la sustancia más ordinaria en el planeta. Saciamos nuestra sed, nos aseamos, cocinamos e incluso nos divertimos con ella. Es muy difícil concebir nuestro día a día en ausencia de este líquido. Debido a su importancia, el agua es el material más estudiado en el planeta, aunque sorprendentemente, muchas veces se menosprecia su importancia. Todos los organismos están constituidos mayoritariamente por agua. Se sabe que la vida como la conocemos depende de las propiedades fisicoquímicas del agua (Chaplin 2001). La importancia del agua para la vida tiene un especial interés para las plantas debido a que son organismos sésiles, es decir, a diferencia de los animales, las plantas están ancladas a un lugar determinado para toda su vida y no pueden moverse si las cosas se complican en su entorno, por ejemplo cuando existe una carencia de agua. A la carencia de agua o dicho de otra manera, a la disminución en la cantidad de agua disponible para el desarrollo adecuado de un organismo se le conoce como déficit hídrico (Auge 2001). Existen definiciones variadas para el déficit hídrico según las distintas ramas de estudio. Por simplicidad, en esta tesis nos referiremos a la definición establecida con anterioridad. En las plantas, el déficit hídrico puede ser causado por periodos de sequía, suelos con alto contenido de sales y por temperaturas bajas que congelan el agua (Auge 2001). Las plantas que sufren de déficit hídrico usualmente experimentan periodos de deshidratación y si los niveles de deshidratación son muy severos, se puede llegar a una condición crítica conocida como desecación. Dada la importancia que tiene el agua para las plantas, es muy evidente que su desarrollo, reproducción y sobrevivencia dependan del suministro adecuado de los niveles óptimos de agua. El déficit hídrico puede ser letal para las plantas y en el caso de los cultivos de interés comercial, puede conducir a enormes problemas sociales y pérdidas económicas (Fang and Xiong 2015).

El déficit hídrico se puede determinar en función de la cantidad de agua que está contenida dentro de una planta. Los fisiólogos de plantas utilizan el término “potencial hídrico”. El potencial hídrico (Ψ_w) es una medida de la energía libre del agua por unidad de volumen relativo al agua pura y es un reflejo de la salud de una

planta (Jones 2013). Éste se obtiene dividiendo el potencial químico del agua entre su volumen molal parcial ($18 \times 10^{-6} \text{ m}^3 \text{ mol}^{-1}$). El potencial hídrico se expresa en unidades de presión, comúnmente mega-pascales (MPa), aunque se puede expresar en otras unidades (J m^{-3} ; bares; atmósferas y psi). Las células de las plantas tienen un potencial de presión conocido como presión de turgor el cual puede llegar a ser tan alto como 1 MPa, con valores comúnmente alrededor de 0.4 – 0.6 MPa (Jones 2013). En una célula viva, existen tres componentes que contribuyen al potencial hídrico (Ψ_w): la concentración de solutos, la presión y la gravedad y pueden ser representados con la siguiente ecuación:

$$\Psi_w = \Psi_s + \Psi_p + \Psi_g$$

Ψ_s es el potencial osmótico, Ψ_p es el potencial de presión y Ψ_g es el potencial gravitacional. La mayoría de las plantas mueren cuando el Ψ_w de las hojas llega a niveles entre -5 y -10 MPa (Sunkar 2010).

Los solutos disueltos en una solución disminuyen el potencial químico del agua al disolverse efectivamente en ella (Jones 2013). El potencial osmótico (Ψ_s), también conocido como potencial de solutos, es una medida del efecto de los solutos disueltos sobre el potencial hídrico y se puede representar con la siguiente ecuación:

$$\Psi_s = -RTC_s$$

Donde R es la constante de los gases ($8.314 \text{ J mol}^{-1} \text{ K}^{-1}$), T es la temperatura absoluta en °K y C_s es la concentración osmolal de la solución. De esta ecuación se puede observar que el potencial osmótico de una solución siempre es menor o igual a 0.

El incremento en la presión de una solución incrementa el potencial químico del agua al incrementar efectivamente su concentración. A su vez, una disminución en la presión disminuye el potencial químico del agua (Jones 2013). El efecto de la presión en el potencial hídrico se expresa como potencial de presión (Ψ_p). En condiciones estándar, el potencial de presión se define como 0 MPa. Por lo tanto, si la presión es mayor a una atmósfera (aproximadamente 0.1 MPa), el Ψ_p es positivo, y si la presión tiene un valor menor a uno, entonces el Ψ_p es negativo y se refiere como “tensión”. Esta tensión es la responsable de elevar el agua a la copa de los árboles durante la evapo-transpiración.

La gravedad ejerce un efecto similar a la presión sobre el potencial químico del agua. Si una columna de agua incrementa en altura, el potencial hídrico aumenta (Jones 2013). El potencial gravitacional depende de la altura de la columna de agua, su densidad y la gravedad. El efecto de la gravedad en el potencial hídrico es considerable cuando se trata del movimiento de agua en árboles altos sin embargo, si se trata de células adyacentes, el potencial gravitacional se anula porque ambas están a la misma altura.

Dado que la membrana plasmática es semipermeable, el movimiento de agua entre una célula y su ambiente se da a través del fenómeno de ósmosis. La diferencia de potencial hídrico entre una célula y su ambiente determina el movimiento de agua a través de la membrana plasmática de las células vegetales, moviéndose desde regiones con un alto Ψ_w hacia regiones con uno menor. Cuando la toma de agua hacia el interior de una célula causa que la membrana plasmática en el protoplasto se presione hacia la pared celular rígida se genera una presión de turgor (Ψ_p). Cuando el Ψ_p celular es mayor a cero las células están en un estado turgente. Si la célula pierde demasiada agua, el Ψ_p será menor a cero y la célula se plasmoliza (Jones 2013).

Por otra parte, el estatus hídrico o la cantidad de agua dentro de una planta también se puede calcular por medio del contenido relativo de agua (CRA). El CRA se puede calcular restando el peso fresco (húmedo) menos el peso seco comparado con el peso húmedo de un tejido cuyo turgor haya alcanzado la saturación, basándose en la siguiente ecuación:

$$\text{CRA} = [(\text{peso fresco}) - (\text{peso seco}) / (\text{peso turgente}) - (\text{peso seco})] \times 100$$

Por ejemplo, el contenido relativo de agua en hojas de una planta en la cual la toma de agua en la raíz está en equilibrio con la transpiración en las hojas está en un intervalo entre 85 – 95 %. Para cada órgano existe un CRA crítico, de tal forma que si el nivel de CRA resulta menor que valores en este intervalo, el órgano muere. Este valor crítico también varía dependiendo de la especie en cuestión, pero típicamente está por debajo del 50 % (Jones 2013). La gran mayoría de las angiospermas (plantas con flor) no pueden sobrevivir a condiciones de deshidratación en las que el contenido relativo de agua llega al 20% ó 30%, lo que equivale a un intervalo entre -5 y -10 MPa de Ψ_w como se mencionó (Sunkar 2010). La mayoría de los cultivos de interés

comercial son relativamente sensibles a la deshidratación y raramente sobreviven al llegar a un potencial hídrico en las hojas de -4 MPa.

Existe un grupo de plantas que se denominan “tolerantes a la desecación” las cuales pueden tolerar potenciales hídricos de hasta -100 MPa aunque, hasta donde se sabe, son relativamente pocas (aproximadamente 300 especies o el 0.1% de todas las angiospermas) (Porembski and Barthlott 2000).

1.11. Las respuestas de las plantas al déficit hídrico

Las plantas son organismos que debido a su naturaleza sésil están constantemente sometidos a distintas condiciones de estrés, incluyendo la falta de agua. Este tipo de estrés abiótico es tan común que incluso es un componente normal en algunos procesos del desarrollo, como la maduración de la semilla. A lo largo de su historia evolutiva, las plantas han desarrollado complejas respuestas fisiológicas, bioquímicas y moleculares para contender con el déficit hídrico (Figura 1).

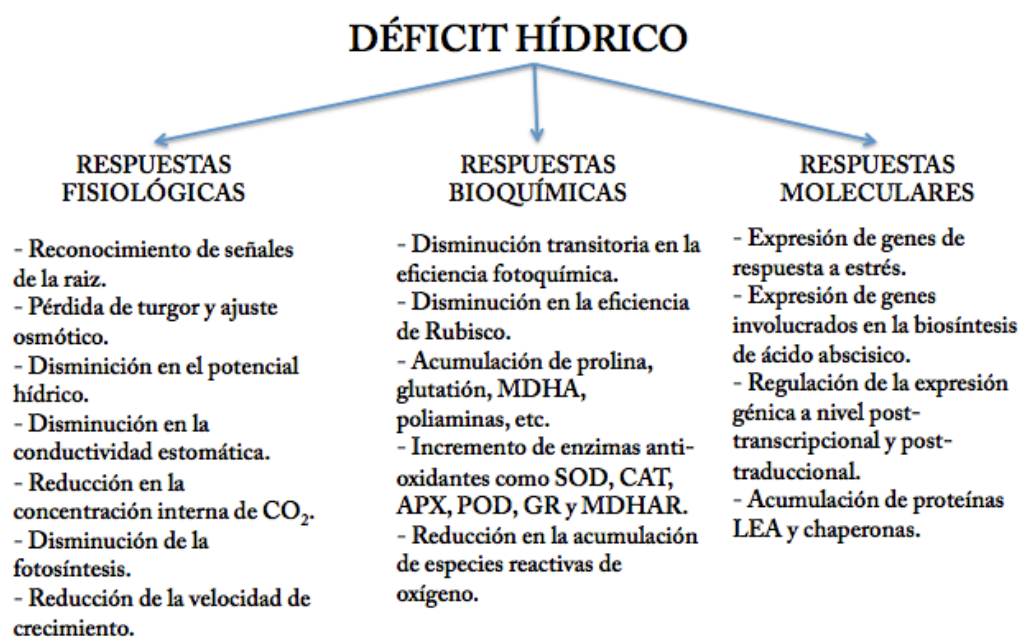


Figura 1. Respuestas a nivel fisiológico, bioquímico y molecular inducidas en respuesta al déficit hídrico. Modificado de Shao *et al.* 2008.

El déficit hídrico celular puede resultar en un aumento en la concentración de solutos, cambios en el volumen celular y en la forma de la membrana, desbalance en los gradientes de potencial osmótico, pérdida de turgor, desestabilización de la

integridad de la membrana y desnaturalización de proteínas (Bray 1997). A la pérdida completa del agua se le conoce como deshidratación o desecación. La capacidad de la planta de responder y sobrevivir al déficit hídrico depende de los mecanismos globales que puedan integrar las respuestas celulares. La respuesta puede ocurrir en pocos segundos (por ejemplo, cambios en el estado de fosforilación de una proteína) o en minutos e incluso horas (por ejemplo, cambios en la expresión génica) (Bray 1997). La resistencia al estrés por déficit hídrico puede ocurrir cuando la planta es capaz de sobrevivir a la condición adversa, lo cual resulta de diversas estrategias que se han seleccionado a lo largo de la evolución de las plantas, de acuerdo al nicho ambiental en el que hayan evolucionado, y que les permiten contender o evitar la situación estresante (Jones 2013). Existen mecanismos durante el ciclo de vida de la planta que involucran adaptaciones evolutivas que les permiten a las plantas *escaparse* o evitar esta condición de estrés. Esta estrategia de *escape* del estrés está acompañada de procesos que involucran cambios a diferentes niveles y que resultan en modificaciones en programas de desarrollo. Tal sería el caso de un ajuste en la temporalidad del ciclo de vida, del tiempo de germinación o floración, etc.. De manera similar ocurre en el caso de la *evitación* del estrés, en donde las plantas reducen la pérdida de agua enrollando las hojas, aumentando la cantidad de cera en la superficie de las mismas; o bien, incrementando la capacidad de acumular agua a través de tejidos especializados, etc.

El estudio de la respuesta a estrés requiere considerar que ésta posee múltiples variables pues depende de la especie y del genotipo, de la longitud y de la severidad de la pérdida de agua, de la edad y del estado de desarrollo de la planta, el órgano y el tipo celular en donde ocurre. Un ejemplo de ajuste o adaptación a nivel celular es el ajuste osmótico que consiste en la acumulación de osmolitos compatibles que generan una disminución del potencial osmótico de la célula con el objetivo de disminuir el potencial hídrico y con ello favorecer la toma de agua del ambiente y, así mantener un turgor adecuado (Bray 1997). Por ejemplo, en el citosol y los cloroplastos de hojas de espinaca estresadas con sal se acumula el osmolito compatible glicina betaína que es un aminoácido modificado que llega a alcanzar una concentración de hasta 250 mM en la vacuola (Jones 2013). La acumulación de osmolitos compatibles puede tener beneficios adicionales; por ejemplo, el manitol (y otros polioles) y la prolina son secuestradores de radicales hidroxilo por lo que moderan los efectos causados por las especies reactivas del oxígeno que se acumulan durante la condición de estrés. Por

otro lado, la disminución en la síntesis de proteínas es otro de los mecanismos generales de respuesta de las plantas. Uno de los procesos más y mejor estudiados entre las respuestas de las plantas a una limitación moderada de agua es el fenómeno que involucra la modulación de la apertura estomática. Una disminución de la humedad ambiental induce el cierre de los estomas, lo cual evita la pérdida de agua de las hojas por transpiración. Sin embargo, también limita la entrada de CO₂ lo que conlleva una disminución en la fotosíntesis y un posterior ajuste metabólico en el que se requieren mecanismos que le permitan a la planta optimizar el uso del CO₂ para su asimilación de carbohidratos (Shao *et al.* 2008).

De forma general, la respuesta al estrés por déficit hídrico a nivel transcripcional se puede dividir en dependientes o independientes del ácido abscísico (ABA) dependiendo de la participación de la llamada hormona del estrés (Bray 1997). El ABA juega un papel primordial en la respuesta, tolerancia y adaptación de las plantas a este tipo de estrés puesto que es una molécula cuya acumulación se induce cuando se percibe el estrés y, que desencadena toda una compleja vía de señalización que culmina en la inducción de genes de respuesta que ayudan a la planta; primero, a tolerar y, posteriormente, a adaptarse a la respuesta a estrés (Yoshida *et al.* 2014). Uno de los procesos mejor estudiados modulados por esta hormona es la apertura y cerrado de los estomas, lo cual a su vez controla la respuesta de la planta a los cambios en la humedad del ambiente, a la concentración de CO₂, a la calidad de la luz, a las concentraciones de especies reactivas de oxígeno, etc. Recientemente, se identificaron algunos de los receptores específicos para esta hormona, los cuales tienen una localización citosólica (REF). Además, existe una vía de señalización que es independiente del ABA (Shinozaki and Yamaguchi-Shinozaki 2007) y que consiste en inducir la expresión de una serie de genes distintos a los controlados por el ABA. Sin embargo, en realidad ambas vías (dependiente e independiente) se encuentran conectadas en distintos puntos, lo cual muestra la complejidad de la respuesta de las plantas a las condiciones de estrés hídrico (Roychoudhury *et al.* 2013).

A nivel molecular, las plantas experimentan grandes cambios en la expresión génica en respuesta al estrés. Los genes que responden a esta condición pueden ser genes cuya expresión disminuye o se apaga, y genes cuya expresión se incrementa o se activa. Los genes y proteínas cuya expresión no se requiere en condiciones de estrés se reprimen transcripcionalmente a nivel genético (a través de represores transcripcionales) y epigenético (a través de modificaciones de proteínas y del ADN).

Esto último puede ocurrir a través de la acción de ARNs pequeños (sARNs; regulación post-transcripcional) o por modificaciones post-traduccionales sobre proteínas que modulan la transcripción y la estructura de la cromatina (Covarrubias and Reyes 2010). Aquellos genes cuyo producto se requiere para lograr un ajuste ante esta condición codifican para una amplia gama de proteínas (como de hecho también es el caso para los genes que se reprimen) que van desde factores transcripcionales y transductores de señales (quinasas, fosfatasa, etc.) hasta proteínas estructurales y enzimas con diferentes actividades. Tal es el caso de los genes que codifican para enzimas antioxidantes como la superóxido dismutasa, la catalasa y ascorbato peroxidasa (APX), entre muchas otras (Shao *et al.* 2008). Además, las condiciones de estrés han llevado a la selección de proteínas con la capacidad de disminuir los daños causados por el estrés, tanto a nivel estructural (sobre membranas) como a nivel proteico. A estas proteínas protectoras se les conoce como las *chaperonas moleculares*, entre las más comunes están las proteínas de choque térmico HSPs (Heat Shock Proteins). Otras proteínas a las cuales más recientemente también se les ha asignado un papel protector ante esta condición de estrés son las llamadas *proteínas abundantes en la embriogénesis tardía*, también conocidas como proteínas LEA (del inglés *Late Embryogenesis Abundant*), que son el motivo de esta tesis y cuyas características se describirán en la siguiente sección.

I.III. Las Proteínas Abundantes en la Embriogénesis Tardía

En la década de los años 80, Leon Dure III encontró un grupo de proteínas que se acumulaban a altos niveles durante la fase de maduración de las semillas de algodón (*Gossypium hirsutum*) (Dure and Chlan 1981; Dure and Galau 1981; Dure *et al.* 1981). Estas proteínas presentaban ciertos motivos distintivos a nivel de secuencia, por lo que las clasificó en diferentes familias a las cuales nombró *proteínas abundantes en la embriogénesis tardía*, o *proteínas LEA* (Dure and Chlan 1981; Dure and Galau 1981; Dure *et al.* 1981). Las proteínas LEA resultaron ser muy interesantes puesto que, además de las distintas firmas a nivel de secuencia, una característica que las unificaba es que contenían una alta proporción de aminoácidos hidrofílicos y cargados y presentaban muy bajos niveles o carecían de residuos muy hidrofóbicos, principalmente fenilalanina y triptófano. Debido a su alta hidrofiliidad y al elevado contenido de aminoácidos pequeños (principalmente glicina y alanina), la mayoría de

los grupos de proteínas LEA se consideraron parte de un nuevo conjunto de proteínas conocido como “*hidrofilinas*”. Las hidrofilinas se definieron en sus orígenes como aquellas proteínas que presentan un índice de hidrofiliidad mayor a 1 y una composición de glicinas mayor al 6% (Garay-Arroyo *et al.* 2000). Si bien, las hidrofilinas se han encontrado principalmente en plantas, también se han detectado en bacterias, levaduras, hongos, crustáceos, artrópodos, nemátodos e incluso en eucariontes superiores (Garay-Arroyo *et al.* 2000; Dang and Hinch 2011; Lopez-Martinez *et al.* 2012; Rodriguez-Porrata *et al.* 2012). Interesantemente, los transcritos y proteínas de estas hidrofilinas se acumulan cuando el organismo en cuestión experimenta condiciones de deshidratación y/o congelamiento, así como otros tipos de estrés. Es de llamar la atención la presencia común de hidrofilinas en organismos anhidrobióticos, aquéllos que pueden resistir condiciones extremas de deshidratación (Tunnacliffe and Wise 2007). De forma similar, las hidrofilinas LEA de plantas fueron encontradas en la etapa tardía de la embriogénesis de las semillas, uno de los estados biológicos con menos agua que se puede encontrar en la naturaleza. Posteriormente, se encontró que las proteínas LEA estaban presentes en todas las plantas en donde se buscaban y, que además de acumularse en la semilla seca, también se expresaban en tejidos vegetativos y reproductivos en respuesta a condiciones de déficit hídrico inducido ya sea por sequía, alta salinidad y/o congelamiento (Bray 1997; Garay-Arroyo *et al.* 2000; Hoekstra *et al.* 2001; Battaglia *et al.* 2008). Debido a las características mencionadas, las proteínas LEA y las hidrofilinas han representado un modelo fundamental para tratar de entender los mecanismos moleculares de la respuesta, tolerancia y adaptación de las plantas a eventos de déficit hídrico.

Como se mencionó anteriormente, las proteínas LEA son ubicuas en el reino vegetal, puesto que se han encontrado en angiospermas, gimnospermas, plantas vasculares sin semilla, briofitas, pteridofitas y algas (Reynolds and Bewley 1993; Honjoh *et al.* 1995; Shinozaki and YamaguchiShinozaki 1996; Oliver *et al.* 2000; Oliver *et al.* 2004; Tanaka *et al.* 2004; Iturriaga *et al.* 2006; Saavedra *et al.* 2006; Proctor *et al.* 2007). Interesantemente, en organismos distintos a las plantas (arqueas, nemátodos, hongos y crustáceos), también existen proteínas tipo LEA que contienen los motivos característicos de algunas de las familias LEA identificadas en plantas (Mtwisha *et al.* 1998; Solomon *et al.* 2000; Browne *et al.* 2004; Abba' *et al.* 2006; Campos *et al.* 2013).

La clasificación de las proteínas LEA se fundamenta en la presencia de motivos de secuencia conservados, los cuales son específicos para cada uno de los grupos y familias. Existe debate y confusión alrededor de esta clasificación. El problema principal es la falta de consenso para nombrar a cada uno de los grupos de proteínas LEA. Dependiendo de diferentes autores, se pueden encontrar distintas formas de nombrar a cada uno de los grupos de proteínas LEA. En la Tabla 1 se muestran las equivalencias entre cada una de las distintas clasificaciones de las proteínas LEA. En esta tesis, utilizaremos la nomenclatura de Battaglia *et al.* (2008) debido a que consideramos que es la clasificación más completa y sencilla que se encuentra en la literatura.

Como ya se mencionó anteriormente, las diferentes familias de proteínas LEA se agrupan dependiendo de la presencia de motivos o “firmas” características para cada grupo. En la Figura 2 se muestra un esquema tomado de Battaglia *et al.* (2008) que representa los diferentes motivos de cada una de los grupos de proteínas LEA.

Los diferentes motivos conservados que identifican a cada grupo no guardan relación a nivel de secuencia entre los distintos grupos; es decir, los grados de similitud entre distintos grupos son muy bajos. Por otro lado, como se mencionó anteriormente, seis de los siete grupos de proteínas LEA comparten las mismas características fisicoquímicas (alta hidrofiliidad, enriquecimiento de aminoácidos pequeños y cargados, ausencia de residuos muy hidrofóbicos).

Precisamente, estas características fisicoquímicas determinan otra propiedad muy importante de las proteínas LEA: la predisposición a carecer de una estructura tridimensional bien definida. A las proteínas que no se pliegan o que no adquieren una estructura estable en el espacio se les conoce como *proteínas intrínsecamente desordenadas* (PIDs). Durante varios años se ha predicho y en algunos casos se ha demostrado experimentalmente que la mayoría de las proteínas LEA son PIDs (Battaglia *et al.* 2008; Hundertmark and Hinch 2008).

(Battaglia <i>et al.</i> 2008)	PFAM (nombre)	(Dure <i>et al.</i> 1989)	(Bray 1997)	(Hundertmark y Hinch 2008)	(Bies-Etheve <i>et al.</i> 2008)	(Tunnacliffe y Wise 2007)
Grupo 1	PF00477 (LEA_5)	D19/D132	Grupo 1	LEA_5	Grupo 1	Grupo 1
Grupo 2	PF00257 (Dehidrina)	D11	Grupo 2	dehidrinas	Grupo 2	Grupo 2
Grupo 3	PF02987 (LEA_4)	D7/D29	Grupo 3 Grupo 5	LEA_4	Grupo 6	Grupo 3
Grupo 4	PF03760 (LEA_1)	D113	Grupo 4	LEA_1	Grupo 4	Grupo 4
Grupo 5	PF04927 (SMP) PF03168 (LEA_2) PF03242 (LEA_3)	D34/D95/D73	Grupo 6	SMP LEA_2 LEA_3	Grupo 5 Grupo 6 Grupo 7	Grupo 6/ Lea5
Grupo 6	PF10714 (LEA_6)	---	---	PvLEA18	Grupo 8	---
Grupo 7	PF02496 (ABA_WDS)	---	---	---	---	---

Tabla 1. Clasificación de los distintos grupos de proteínas LEA con las correspondientes equivalencias entre diferentes autores. En esta tesis se sigue la nomenclatura establecida por Battaglia *et al.*, 2008. PFAM: Identificador de la base de datos PFAM.

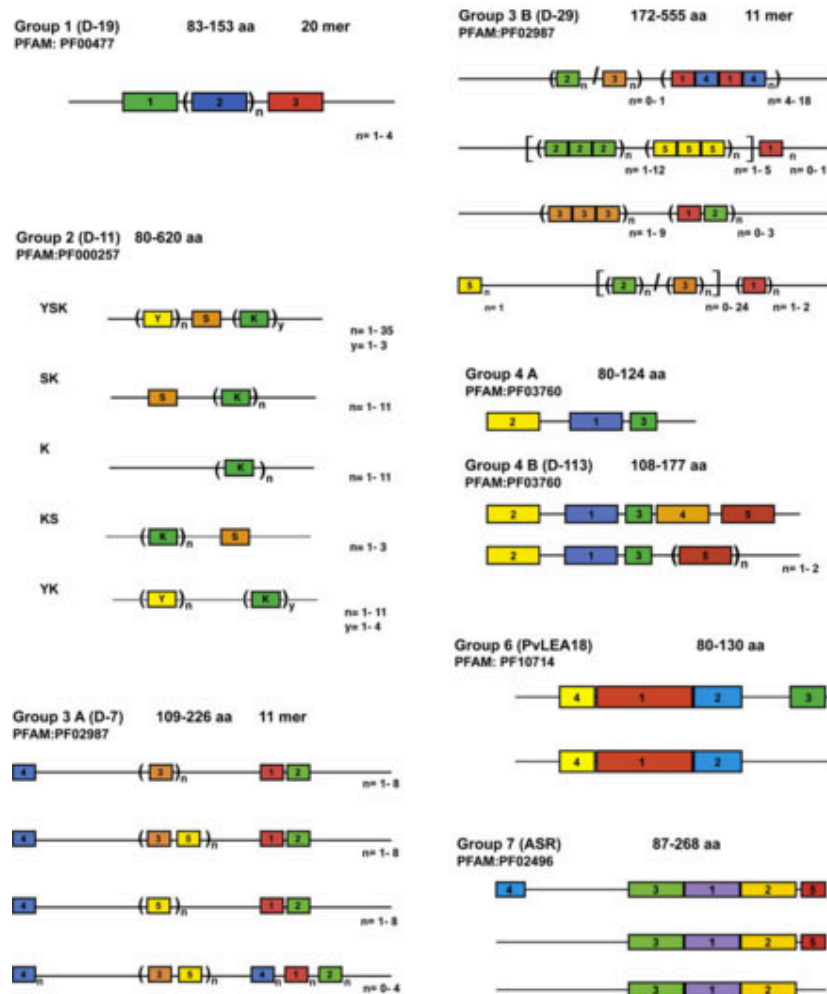


Figura 2. Esquema de los diferentes motivos conservados en cada una de las familias de proteínas LEA hidrofílicas. Tomado de Battaglia *et al.* (2008).

A continuación, se presenta un resumen de las características y hallazgos que consideré más relevantes sobre cada una de las familias de proteínas LEA.

Proteínas LEA del grupo 1:

Las proteínas LEA del grupo 1 se definen por poseer un motivo de 20 aminoácidos (20-mero) con el siguiente consenso: TRKEQ[L/M]G[T/E]EGY[Q/K]EMGRKGG[L/E] (Galau *et al.* 1992; Campos *et al.* 2013). El 20-mero puede estar presente en hasta ocho repeticiones a lo largo de la proteína en cuestión. En plantas, existen otros dos motivos particulares de esta familia (Figura 2). Además de estar presentes en las plantas, se han detectado proteínas que contienen el 20-mero en bacterias, arqueas y en el crustáceo *Artemia franciscana* (Campos *et al.* 2013). Es el único grupo de proteínas LEA que tiene representantes en los tres dominios taxonómicos de la vida (arquea, bacteria y eucariota) (Battaglia *et al.*

2008). Las proteínas LEA1 contienen una alta proporción de aminoácidos cargados y de glicinas. Los transcritos de estas proteínas se acumulan principalmente en el embrión maduro y, en algunos casos, en órganos que experimentan deshidratación, como lo son los granos de polen (Vicent *et al.* 2001). Los transcritos de los genes LEA1 de bacterias también se acumulan bajo condiciones estresantes, como la alta osmolaridad (Stacy and Aalen 1998). Además, las proteínas tipo LEA1 están presentes en proteomas de arqueas extremófilos; así como en quistes de *Artemia*, los cuales pueden permanecer secos e inactivos por muchos años (Campos *et al.* 2013).

Estudios estructurales por difracción circular (DC) y resonancia magnética nuclear (RMN) muestran que algunos miembros del grupo LEA1 tienen altos niveles de desorden (Soulages *et al.* 2002). En relación a su función, se ha observado que una proteína LEA1 recombinante de trigo (*Triticum aestivum*) previene la agregación inducida por procesos de deshidratación y congelamiento *in vitro* de enzimas modelo como la citrato sintasa (CS) y lactato deshidrogenasa (LDH) (Goyal *et al.* 2005). De forma similar, se encontró que dos LEA1 de bacteria (GsiB y PaEM) y una de *Arabidopsis* (AtEM6) previenen la pérdida de actividad de la LDH ante ciclos de congelamiento-descongelamiento, aunque se obtuvieron niveles de protección menores a los observados con una proteína LEA del grupo 4 (Campos *et al.* 2013).

Proteínas LEA del grupo 2:

Mejor conocidas como dehidrinas, las proteínas LEA del grupo 2 son las proteínas LEA más estudiadas. Como la mayoría de las proteínas LEA, las dehidrinas son muy hidrofílicas, tienen altas proporciones de aminoácidos cargados y carecen de residuos hidrofóbicos (principalmente triptófano). El motivo que distingue a este grupo se conoce como *segmento-K* (rico en lisinas) cuya secuencia consenso es EKKGIMDKIKEKLP (Battaglia *et al.* 2008). El segmento-K se puede encontrar repetido hasta 11 veces. Las dehidrinas se pueden clasificar en sub-clases, dependiendo de otros motivos conservados como el *segmento-Y* (rico en tirosinas) o el *segmento-S* (rico en serinas). De esta manera, las dehidrinas se pueden clasificar en cinco sub-familias: $K_{(n)}$, $SK_{(n)}$, $K_{(n)}S$, $Y_{(n)}K$ y $Y_{(n)}SK_{(n)}$ (Battaglia *et al.* 2008). Existen miembros de la familia LEA2 en plantas no vasculares, plantas sin semilla así como en todas las plantas con semilla en donde se han buscado. También se ha encontrado que las proteínas LEA2 son PIDs (Lisse *et al.* 1996; Soulages *et al.* 2003; Mouillon *et al.* 2006). Análisis estructurales *in vitro* indican que tres dehidrinas de *Arabidopsis*

mantienen un estado desplegado después de someterlas a distintos tratamientos (Mouillon *et al.* 2008). Las dehidrinas se acumulan en la semilla seca y en tejidos vegetativos en respuesta a deficiencia de agua (Battaglia *et al.* 2008).

En relación a su función, se ha propuesto que las dehidrinas pueden proteger la integridad de las membranas biológicas debido a que se unen a lípidos cargados negativamente (Koag *et al.* 2009). Además, las proteínas LEA2 son capaces de unir cationes divalentes como el Ca^{2+} , Zn^{2+} , Fe^{3+} , Co^{2+} , Ni^{2+} y Cu^{2+} , por lo que podrían actuar como amortiguadores de estos metales en condiciones de déficit hídrico (Hara *et al.* 2005). Varios esfuerzos se han enfocado en la función protectora de proteínas *in vitro*, donde se ha visto que las dehidrinas previenen la inactivación de enzimas reporteras sometidas a deshidratación y congelamiento (Wisniewski *et al.* 1999; Reyes *et al.* 2008; Hughes and Graether 2011; Hughes *et al.* 2013; Cuevas-Velazquez *et al.* 2014). El segmento-K es fundamental para la función protectora de las proteínas LEA2 puesto que los niveles de protección disminuyen drásticamente si se elimina este segmento en las proteínas ERD10 y RcDhn5 de *Rhododendron catawbiense* y la DHN-5 de trigo (Reyes *et al.* 2008; Drira *et al.* 2013).

Proteínas LEA del grupo 3:

El motivo característico de las proteínas LEA del grupo 3 es un 11-mero (11 aminoácidos), el cual puede estar repetido en múltiples ocasiones, cuya secuencia consenso es FF[E/Q]XFK[E/Q]KFX[E/D/Q] (Battaglia *et al.* 2008). Se han detectado otros motivos menos conservados y que no necesariamente se encuentran en todas las proteínas LEA3 (Figura 2). Muy probablemente, esta familia de proteínas sea la más grande y diversa de todas, debido a la variabilidad del 11-mero, lo que permite agruparlas en dos sub-grupos, el sub-grupo 3A (D-7) y 3B (D-29).

Se predice que el 11-mero posee una tendencia a formar una estructura α -hélice anfipática capaz de dimerizar en un arreglo tipo coil-coiled (Dure 1993). Estudios de DC y espectroscopía de infra rojo (FTIR) de varias LEA3 indican que están mayoritariamente desplegadas en solución acuosa (Wolkers *et al.* 2001; Goyal *et al.* 2003; Tolleter *et al.* 2007). La adición de trifluoroetanol (TFE) o dodecil sulfato de sodio (SDS), así como la deshidratación severa de estas proteínas, induce cambios conformacionales hacia α -hélices (Wolkers *et al.* 2001; Tolleter *et al.* 2007).

Las proteínas LEA3 se han encontrado en algas, plantas no vasculares, plantas con semilla y sin semilla (Joh *et al.* 1995; Hellwege *et al.* 1996; Salmi *et al.* 2005). Junto con la familia LEA1, el grupo 3 de proteínas LEA contiene miembros en otros organismos diferentes a las plantas. Se pueden encontrar en *Deinococcus radiodurans* (Battista *et al.* 2001), *Haemophilus influenzae* (Dure 2001), *Caenorhabditis elegans* (Gal *et al.* 2004), *Philodina roseola* (Tunnacliffe *et al.* 2005), *Polypedilum vanderplanki* (Kikawada *et al.* 2006), organismos anhidrobióticos como *Steinernema feltiae* (Solomon *et al.* 2000) y *Aphelencus avenae* (Browne *et al.* 2004) y en el crustáceo *Artemia franciscana* (Wang *et al.* 2007).

Se ha propuesto ampliamente que las proteínas LEA del grupo 3 podrían estabilizar proteínas actuando como escudos moleculares (Chakrabortee *et al.* 2012). Proteínas de chícharo (PsLEAm), Arabidopsis (AtLEA76 y COR15am), *Chlorella vulgaris* (HIC6) y *Aphelencus avenae* (AavLEA1) protegen enzimas como malato deshidrogenasa, citrato sintasa, fumarasa y LDH de los efectos causados por la deshidratación y por ciclos de congelamiento/descongelamiento impuestos *in vitro* (Honjoh *et al.* 2000; Goyal *et al.* 2005; Grelet *et al.* 2005; Reyes *et al.* 2005; Nakayama *et al.* 2007).

Proteínas LEA del grupo 4:

Las proteínas LEA del grupo 4 se caracterizan por presentar una región N-terminal conservada de entre 70 y 80 residuos (Battaglia *et al.* 2008). Además, poseen un dominio C-terminal variable en secuencia y longitud. El motivo que define a esta familia tiene como consenso la secuencia AQEKAEKMTA[R/H]DPXKEMAHK[E/K][A/E][K/R]. Existen otros nueve motivos conservados que se localizan principalmente en la región N-terminal (Olvera-Carrillo *et al.* 2010). Las proteínas de esta familia están presentes en briofitas, gimnospermas y angiospermas (Battaglia *et al.* 2008; Olvera-Carrillo *et al.* 2010). El análisis filogenético de las proteínas de esta familia permitió dividir las en dos sub-grupos, 4A y 4B. Las proteínas del sub-grupo 4A son pequeñas (80 – 124 aminoácidos) y las del 4B son más largas (108 – 180 aminoácidos). Se ha sugerido que estos sub-grupos emergieron de una duplicación genética ancestral, antes de la separación de las monocotiledóneas y las dicotiledóneas, lo que indica que dicha separación dio origen a una sub-funcionalización de estas proteínas (Olvera-Carrillo *et al.* 2010).

Se ha reportado que los niveles de una proteína LEA4 en embriones de algodón alcanzan encontrado que los transcritos y proteínas de esta familia también se acumulan en concentraciones cercanas a 300 μ M (Roberts *et al.* 1993). Los transcritos de las proteínas LEA del grupo 4 están entre los más altamente acumulados en semillas secas de *Arabidopsis* (Delseny *et al.* 2001). Además, se ha encontrado que los transcritos y proteínas de esta familia también se acumulan en tejidos vegetativos y reproductivos en respuesta a condiciones de déficit hídrico (Schmid *et al.* 2005; Winter *et al.* 2007; Hruz *et al.* 2008; Olvera-Carrillo *et al.* 2010).

En relación a la estructura de estas proteínas, se ha encontrado que una proteína LEA4 de soya (GmPM16) se comporta como una proteína PID en solución acuosa, pero adquiere estructura α -hélice cuando se agregan altas concentraciones de TFE o después de deshidratar la muestra (Shih *et al.* 2004). Por otro lado, una proteína LEA4 de *Arabidopsis* (AtLEA4-5) es capaz de prevenir la inactivación de LDH y malato deshidrogenasa por pérdida parcial de agua y ciclos de congelamiento/descongelamiento *in vitro*, aún con relaciones molares tan bajas como 1:1 (Reyes *et al.* 2005; Reyes *et al.* 2008).

Proteínas LEA del grupo 5:

Este grupo de proteínas se conocen como proteínas *LEA atípicas* puesto que están conformadas por las proteínas LEA que contienen altas proporciones de residuos hidrofóbicos. Debido a que en este grupo se concentran todas las proteínas LEA que no son hidrofílicas, no existen secuencias consenso que definan si son parte de esta familia (Battaglia *et al.* 2008). Estas proteínas son globulares y tienen una estructura tridimensional bien definida (Singh *et al.* 2005). Existe poca información respecto a este grupo de proteínas. Existen reportes que muestran que algunas proteínas de este grupo son capaces de proteger enzimas reporteras de los efectos de la deshidratación (Boucher *et al.* 2010).

Proteínas LEA del grupo 6:

Las proteínas LEA del grupo 6 se descubrieron originalmente en semillas de frijol (*Phaseolus vulgaris*) (Colmenero-Flores *et al.* 1997). Las proteínas de este grupo hasta ahora solo se han detectado en plantas vasculares (Battaglia *et al.* 2008). Se han descrito cuatro motivos, de los cuales dos se encuentran altamente

conservados, a tal grado que la secuencia LEDYK (localizada dentro del motivo 1) se encuentra 100% conservada (Rivera-Najera *et al.* 2014).

En frijol, la proteína PvLEA18 se acumula en altos niveles en semilla seca y en granos de polen. Además, esta proteína responde a condiciones de déficit hídrico y al tratamiento con ABA, como la mayoría de los otros grupos de proteínas LEA (Colmenero-Flores *et al.* 1999).

Para la proteína PvLEA6 de frijol se ha reportado que se acumula no sólo bajo condiciones de déficit hídrico en los tejidos vegetativos sino también en las regiones en elongación o de crecimiento de hipocótilos y raíces, las cuales presentan los potenciales hídricos más bajos, al comparar estas regiones con las regiones maduras o de división de estos órganos en plántulas de frijol sin estrés, crecidas bajo condiciones óptimas (Colmenero-Flores *et al.* 1999). Esta observación sugiere que la expresión de esta proteína es muy sensible a los cambios en el estatus hídrico de los tejidos que ocurren bajo condiciones óptimas de desarrollo. La proteína LEA4 de frijol no presenta este comportamiento.

La PvLEA18 tiene una estructura altamente desordenada en solución acuosa pero se pliega ligeramente al adicionar glicerol o polietilén glicol (Rivera-Najera *et al.* 2014). De forma muy interesante, se sabe que es capaz de formar oligómeros *in vitro* e *in vivo*, por lo que posiblemente sea necesario formar estas especies para ejercer correctamente su función o para mantener su integridad (Rivera-Najera *et al.* 2014).

Contrario a lo observado para las proteínas LEA de los otros grupos, la PvLEA18 no es capaz de evitar la inactivación de LDH o malato deshidrogenasa ante tratamientos de déficit hídrico *in vitro*, lo que sugiere que las proteínas LEA6 tienen una función diferente a las del resto de proteínas LEA (Reyes *et al.* 2005).

Proteínas LEA del grupo 7:

También conocidas como proteínas ASR (ABA [abscisic acid]-stress- ripening-induced), las proteínas LEA del grupo 7 se acumulan abundantemente en la etapa tardía de la embriogénesis de la semilla y en respuesta a déficit hídrico (Maskin *et al.* 2008). Se pueden encontrar miembros del grupo LEA7 en varias especies de dicotiledóneas, monocotiledóneas, gimnospermas y en *Ginkgo biloba* (*Ginkgophyta*) (Battaglia *et al.* 2008). Interesantemente, en la planta modelo *A. thaliana* no existen miembros de este grupo. En este grupo existen hasta cinco motivos conservados (Figura 2), de los cuales el motivo 3 contiene una señal de localización nuclear y el

motivo 5 tiene una actividad de unión a ADN dependiente de Zn^{2+} . Recientemente, se encontró que las proteínas ASR1 y ASR5 pueden funcionar como factores de transcripción que regulan genes de respuesta a aluminio en arroz (Arenhart *et al.* 2016).

Estudios biofísicos y bioquímicos indican que la proteína ASR1 de tomate es una PID en solución acuosa, la cual experimenta cambios hacia una conformación más estable cuando une iones Zn^{2+} o en condiciones de deshidratación.

En ensayos *in vitro*, proteínas de este grupo también muestran un efecto protector sobre enzimas reporteras frente a la deshidratación (Arenhart *et al.* 2016).

1.IV. Las Proteínas Intrínsecamente Desordenadas

El paradigma estructura-función de las proteínas establece que una proteína tiene una función específica dependiendo de la estructura tridimensional que adopte de tal forma que, para una secuencia polipeptídica particular deberá existir una estructura estable que dará origen a una sola función. Durante muchos años este paradigma se aceptó y se fortalecía cada vez que se obtenía la estructura cristalográfica de una proteína. Además, el concepto de llave-cerradura que explica la especificidad de las enzimas y la interacción entre un anticuerpo y un antígeno, sustentaba a la perfección el paradigma. Sin embargo, algunas regiones de las proteínas simplemente no se podían ver usando la cristalografía de rayos X o tenían un patrón muy particular de desplazamientos químicos en los espectros de resonancia magnética nuclear. A estas regiones se les denominó “*asas*” (*loops* en inglés). Durante mucho tiempo se pensó que estas *asas* no tenían relevancia alguna en la función de la proteína y que simplemente funcionaban como conectores entre los distintos motivos plegados. En algunos casos, estas regiones sin una estructura ordenada comprendían prácticamente toda la proteína, por lo que se concluía que dichas proteínas no poseían una estructura tridimensional estable y, por lo tanto, no tenían función alguna. Los biólogos estructurales usualmente renunciaban al estudio de esta clase de proteínas pues contrastaban con todos los principios sobre los cuales habían sido entrenados; por ejemplo, esas proteínas no se podían cristalizar y no presentaban elementos típicos de estructura secundaria en sus espectros de dicroísmo circular (DC) y eran particularmente sensibles a degradación cuando se purificaban. Interesantemente, estas proteínas cumplían a la perfección su función a tal grado que,

se llegó a definir las como “proteínas que puedes hervir toda la noche y al día siguiente siguen ejecutando su función (cualquiera que esta fuera)”. Poco a poco, cerca del año 2000, este fenómeno comenzó a llamar la atención de la biología estructural, cuando diversos estudios tomaron las distintas observaciones y analizaron las generalidades que se mantenían entre cada uno de los ejemplos. Es aquí cuando nace el campo del desorden estructural que se ha encargado de estudiar a las llamadas proteínas *nativamente desnaturalizadas* (*natively unfolded proteins*, NUPs), proteínas *intrínsecamente no estructuradas* (*intrinsically unstructured proteins*, IUPs) o como son llamadas en la actualidad, *proteínas intrínsecamente desordenadas* (*intrinsically disordered proteins*, IDPs). Los diferentes hallazgos que llevaron al descubrimiento y aceptación de las *proteínas intrínsecamente desordenadas* se pueden consultar en las siguientes revisiones sobre el tema: Demchenko 2001; Dunker *et al.* 2001; Dyson and Wright 2002; Tompa 2002; Uversky 2002; Oldfield *et al.* 2005.

De forma muy general, se puede decir que las proteínas intrínsecamente desordenadas son aquellas que carecen de una estructura tridimensional bien definida (Dunker *et al.* 2001). En la Figura 3 se muestran dos tipos de proteínas, una proteína globular y una intrínsecamente desordenada.

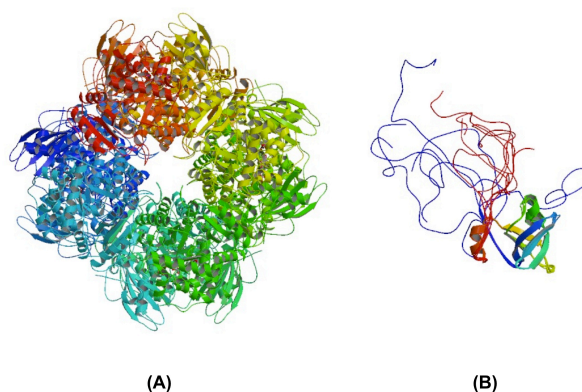


Figura 3. Esquema de A) una proteína globular (ribulosa 1,5-bifosfato-carboxilasa-oxigenasa de espinaca; PDB:1RCX) y B) una proteína intrínsecamente desordenada (eIF1A de humano; PDB:1D7Q). Las líneas en azul y rojo representan las distintas conformaciones que puede adoptar la proteína desordenada. Modificado de Cuevas-Velazquez y Covarrubias (2011).

El concepto de desorden no aplica sólo a proteínas que en su totalidad son desordenadas, sino también a regiones (alrededor de 30 residuos o más) que no tienen una estructura bien definida y que son parte de una proteína globular. A estas zonas se les llama *regiones intrínsecamente desordenadas (RIDs)* (Tompa 2002). A partir de aquí, cuando se mencionen a las PIDs me referiré, ya sea, a las proteínas completamente desordenadas, a las regiones intrínsecamente desordenadas y/o a ambas.

Las PIDs se caracterizan por tener una baja complejidad de secuencia, con un bajo contenido de aminoácidos hidrofóbicos (Trp, Tyr, Phe, Met, Ile, Leu y Val) y una composición enriquecida en aminoácidos cargados y/o polares (Lys, Glu, Gln, Ser, Gly y Ala). Debido al enriquecimiento en estos aminoácidos, ha sido posible desarrollar algoritmos que permiten determinar si una secuencia polipeptídica tiene la tendencia o propensión a ser intrínsecamente desordenada. Entre los predictores de desorden altamente confiables se encuentran PONDR (Romero *et al.* 2001), DISpro (Cheng *et al.* 2005), DISOPRED3 (Jones and Cozzetto 2015), FoldIndex (Uversky *et al.* 2000), DisEMBL (Linding *et al.* 2003) y GLOBPLOT2 (Ward *et al.* 2004).

Valiéndose de estos servidores se ha logrado predecir que las PIDs están ampliamente representadas en los proteomas de los distintos organismos donde se han buscado (Dyson and Wright 2005). Existe una correlación entre la complejidad de los organismos y la proporción de proteínas desordenadas en cada proteoma; es decir, cuanto más complejo es un organismo, mayor es la proporción de proteínas intrínsecamente desordenadas en su proteoma (Ward *et al.* 2004). Además, estos estudios muestran que las PIDs están enriquecidas en los siguientes procesos biológicos: 1) regulación de la transcripción y unión a ácidos nucleicos; 2) transducción de señales y regulación del ciclo celular; 3) biogénesis y mantenimiento de organelos que contienen ácidos nucleicos (cromatina y ribosomas); 4) procesamiento del ARN mensajero y reacciones de “*splicing*”; 5) Biogénesis y organización del citoesqueleto (Ward *et al.* 2004; Tompa *et al.* 2006; Xie *et al.* 2007). Evidencias bioinformáticas muestran la importancia de las proteínas intrínsecamente desordenadas en la regulación transcripcional. Se ha encontrado que entre el 82.6 – 94.1 % de los factores transcripcionales descritos a la fecha contienen amplias RIDs (Liu *et al.* 2006). El nivel de desorden predicho es mayor en los dominios de transactivación (84.2%) que en los dominios de unión a ADN (30.7%) y el nivel de desorden es mayor en los factores transcripcionales de organismos eucariontes que en

los de los procariontes (Liu *et al.* 2006). Se ha descrito ampliamente que las PIDs son capaces de unir una extensa gama de blancos proteicos (Tompa and Fuxreiter 2008; Uversky 2013). Es probable que los dominios de trans-activación de los factores transcripcionales estén enriquecidos en regiones desordenadas para tener una elevada plasticidad en la interacción con sus blancos, característica intrínseca de este tipo de proteínas que deben integrar múltiples señales muy rápidamente. Otra característica de la interacción que involucra proteínas intrínsecamente desordenadas es que tienen una alta especificidad pero baja afinidad (Dyson and Wright 2005). De igual forma, esto es muy apropiado en proteínas implicadas en la transducción de señales.

Las PIDs pueden existir en diferentes estados conformacionales. El estado denominado “*random coil*” o de “*giros azarosos*” consiste en un grupo heterogéneo de conformaciones extendidas que se encuentran fluctuando constantemente entre ellas (Dunker *et al.* 2001). Este grupo de conformaciones se puede visualizar claramente cuando se comparan los “paisajes energéticos” de las proteínas globulares con los de las proteínas intrínsecamente desordenadas (Figura 4). A medida que las proteínas globulares se pliegan, entran a paisajes energéticos cada vez más reducidos pero a su vez más estables. El cambio de energía libre asociado a la proteína plegada es muy negativa y por lo tanto muy estable (Figura 4A). Por el contrario, las PIDs tienen un paisaje energético muy extendido en donde la energía libre entre los distintos conformeros es muy similar, provocando que la estructura tridimensional se encuentre oscilando entre los distintos estados energéticos (Figura 4B).

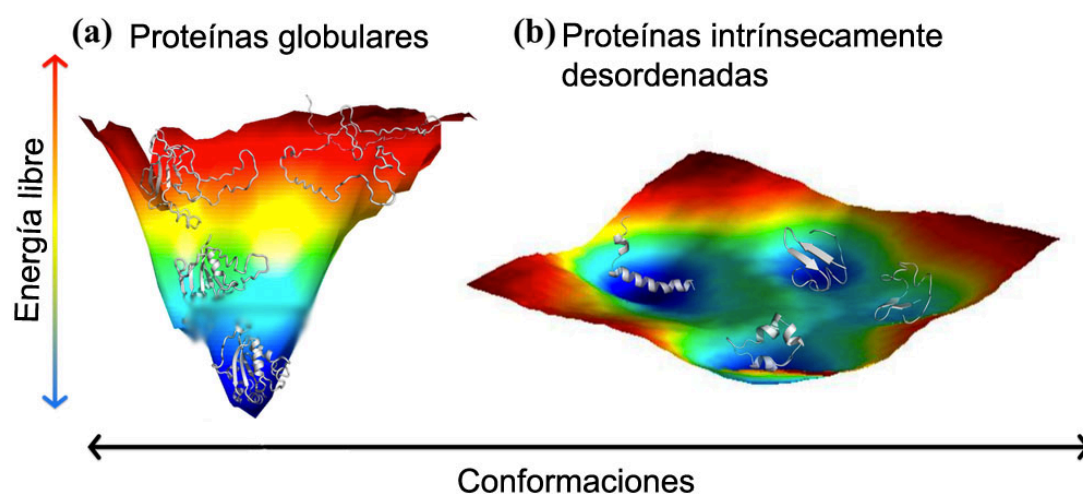


Figura 4. Diagrama de los valles energéticos de a) proteínas globulares y b) proteínas intrínsecamente desordenadas. Entre más baja sea la energía libre del conformero (azul), mayor será la estabilidad del mismo. Usualmente, las proteínas globulares sólo contienen un mínimo de energía libre (una sola conformación) mientras que las proteínas intrínsecamente desordenadas tienen varios

mínimos por lo que tienden a fluctuar entre distintas conformaciones “estables”. Modificada de Burger *et al.* 2014.

En general, las proteínas desordenadas se pueden encontrar altamente extendidas, o bien pueden ser compactas pero desordenadas. A las conformaciones muy extendidas se les conoce como “*giros azarosos*” (“*random coil*”) y a las compactas como “*globo fundido*” (“*molten globules*”) (Dyson and Wright 2005). En algunos casos se pueden producir estructuras conocidas como “*pre-globos fundidos*” (“*pre-molten globules*”) cuando regiones locales de una secuencia tienen una propensión a adquirir elementos aislados y fluctuantes de estructura secundaria (Uversky 2002). Las proteínas intrínsecamente desordenadas rara vez se comportan totalmente como *giros azarosos*; incluso en sus estados más desplegados, aún tienen cierta propensión a formar elementos de estructura secundaria (Dyson and Wright 2002).

Las funciones que ejercen las proteínas intrínsecamente desordenadas son diversas y constantemente se reportan nuevos hallazgos que las involucran en funciones y procesos nuevos (Wright and Dyson 2015). Entre las funciones más estudiadas destacan la participación como cadenas entrópicas (región repetida de Nup2p FG, dominio de proyección MAP2, dominio PEVK de titina, región N-terminal del canal de K), exposición de sitios de modificaciones post-traduccionales (Dominio KID de CREB, dominio N-terminal de la ciclina B), chaperonas (β -sinucleína, ERD10, proteína 7/9 de nucleocapside), efectores (4E-BP1, p27^{Kip1}, FlgM, securina, statmina), proteínas de andamiaje (Dominio CTD de la RNAP II, p21^{Cip}, CREB), secuestradoras de moléculas pequeñas (caseína, ERD14, calsecuestrina) y priones (Ure2p, Sup35p, CPEB) (Wright and Dyson 1999; Dunker *et al.* 2001; Namba 2001; Tompa 2002; Uversky 2002; Wright and Dyson 2015).

La importancia de las proteínas intrínsecamente desordenadas en el funcionamiento óptimo de los organismos se puede observar muy claramente en el caso de los humanos. Algunas proteínas muy importantes implicadas en enfermedades neurodegenerativas, enfermedades cardiovasculares, amiloidosis y cáncer contienen altos niveles de desorden (Dyson and Wright 2005). Proteínas como p53, inhibidores de p21^{Cip1}, p21^{Kip1}, BRCA1 y PTTG son algunas de las PIDs asociadas a cáncer más caracterizadas (Mark *et al.* 2005; Sivakolundu *et al.* 2005; Wells *et al.* 2008). Otras proteínas implicadas en enfermedades, como α -sinucleína y Tau, también presentan

altos niveles de desorden. Se ha propuesto que estas patologías son el resultado de alteraciones en el interactoma de estas proteínas desordenadas, el cual consiste de múltiples proteínas blanco involucradas en diversos y relevantes procesos celulares (Uversky *et al.* 2008).

Las PIDs constituyen un nuevo grupo de proteínas que utilizan su flexibilidad para interactuar con múltiples blancos y realizar más de una función, y emergen como una alternativa nueva para el estudio de procesos bioquímicos y celulares, así como una alternativa para el desarrollo de fármacos dirigidos a evitar enfermedades por lo que estas proteínas continuarán rompiendo las reglas y dogmas de las proteínas.

I.V. Técnicas utilizadas para caracterizar a las Proteínas Intrínsecamente Desordenadas

La estructura de las proteínas intrínsecamente desordenadas se puede estudiar con la mayoría de las técnicas que se utilizan para estudiar a las proteínas globulares. Las técnicas principales son la espectroscopía de fluorescencia UV, blanqueamiento de la fluorescencia, unión de ANS, transferencia de energía por resonancia de fluorescencia (FRET), espectroscopía de correlación de fluorescencia, transformada de Fourier – espectroscopía de resonancia de infrarrojo (FTIR), espectroscopía de actividad óptica de Raman, espectroscopía de resonancia paramagnética electrónica, microscopía electrónica, dicroísmo circular (DC) y resonancia magnética nuclear (RMN). Dependiendo de la técnica utilizada, se puede obtener información estructural en estado estacionario, dinámica e incluso a nivel de residuo. A continuación se describen dos de las técnicas más utilizadas para la caracterización de las proteínas intrínsecamente desordenadas: DC y RMN.

I.VI. Dicroísmo circular

La espectroscopía de dicroísmo circular (DC) se basa en la medición de la diferencia en la absorción de la luz circularmente polarizada que gira hacia la izquierda y la que gira hacia la derecha a longitudes de onda determinadas, la cual resulta de la actividad óptica de moléculas quirales en una muestra (Kelly *et al.* 2005). El DC aplica cuando el vector de campo eléctrico rota alrededor de la dirección de propagación generando ondas de luz polarizadas circularmente (a la derecha o a la izquierda), las cuales al interactuar con moléculas quirales causan la absorción

diferencial de las dos luces polarizadas. La señal que se detecta se puede expresar como una diferencia de absorbancias ($\Delta A = A_L - A_R = (\epsilon_L - \epsilon_R) \times c \times l$ en términos molares), pero usualmente se reporta como la elipticidad de polarización (θ), definida como $\tan \theta = (E_R - E_L) / (E_R + E_L)$, (en donde E_R y E_L son los vectores de campo eléctrico de las dos luces circularmente polarizadas) convertido a un valor de elipticidad molar (Figura 5). Generalmente, la diferencia en la absorbancia de las dos luces circularmente polarizadas es extremadamente pequeña, en el intervalo de $10^{-4} - 10^{-6}$ veces la absorbancia real de la muestra. Las señales de DC se observan en la misma región del espectro en el que la proteína absorbe. Usualmente se analizan dos regiones, el UV-cercano y el UV-lejano, de las cuales se obtiene información distinta acerca de la estructura de la proteína.

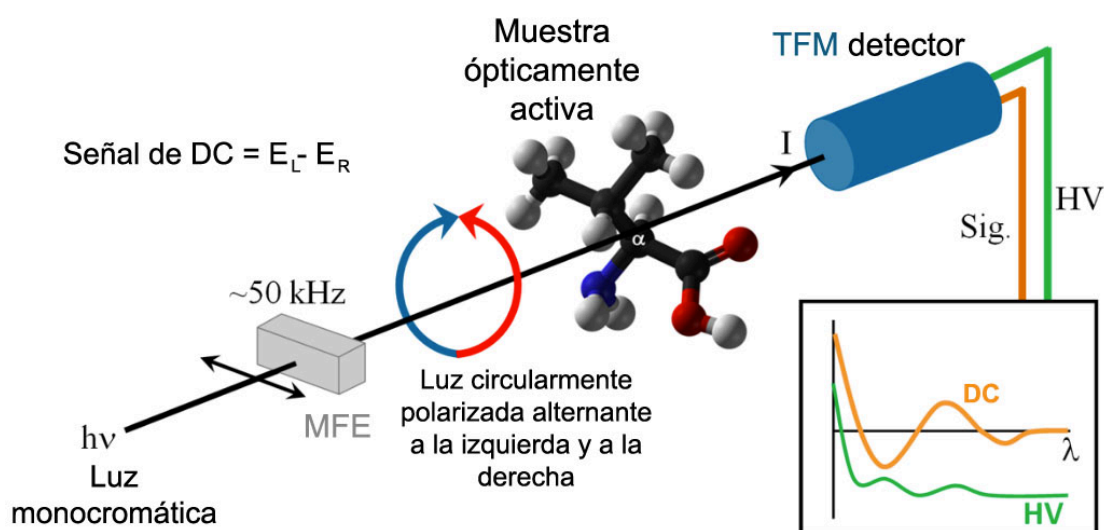


Figura 5. Principio de la espectroscopía de dicroísmo circular (DC). La luz monocromática se pasa a través de un modulador foto-elástico (MFE) el cual convierte la luz linealmente polarizada en luz circularmente polarizada alternante a la izquierda y a la derecha (roja y azul). Las dos polarizaciones se absorben diferencialmente y la diferencia en la absorción se detecta con un tubo foto-multiplicador (TFM) (Modificado de http://www.isa.au.dk/facilities/astrid2/beamlines/AU-cd/AU-CD_3.asp).

El DC de UV-cercano se mide en la región conocida como “aromática”, entre 250-350 nm (Kelly *et al.* 2005). En esta región se obtiene información de la estructura terciaria de las proteínas, ya que los distintos residuos aromáticos tienden a generar distintos perfiles de longitud de onda en este intervalo. Phe contribuye de 250 – 270 nm, Tyr de 270 – 290 y Trp de 280 a 300 nm. El espectro de DC de UV-cercano representa una firma detallada de la estructura terciaria alrededor de los residuos

reporteros, aunque no puede interpretarse en términos de estructura real; la información que ofrece se limita a detectar la(s) estructura(s) predominante(s), sin establecer la localización de las mismas. Las proteínas con una estructura tridimensional estable se caracterizan por un espectro intenso y complejo debido al ambiente asimétrico de sus residuos aromáticos, mientras que el espectro de las proteínas intrínsecamente desordenadas presenta baja intensidad y baja complejidad pues sus residuos aromáticos experimentan ambientes isotópicos. En ciertas proteínas desordenadas que experimentan orden a nivel local, se puede obtener información de la estructura residual puesto que los residuos aromáticos son parte de estructura residual y/o parches hidrofóbicos.

El DC de UV-lejano se mide en el intervalo de 190 – 250 nm y la señal se origina principalmente de los enlaces amida (peptídicos) (Kelly *et al.* 2005). Se utiliza principalmente para determinar la cantidad relativa de diferentes elementos de estructura secundaria puesto que estos presentan espectros característicos (α -hélices, hojas- β , giros, hélices poli-L-prolina tipo II y desordenadas). Un espectro se puede considerar como la combinación lineal de la contribución de los diferentes elementos estructurales aunque esto genera cierta incertidumbre.

El dicroísmo circular ha sido una técnica dominante en la identificación de PIDs (Kelly *et al.* 2005). El espectro tipo *random coil* o *giro desordenado* se ha tomado como un indicador directo del carácter desordenado de una proteína. En la base de datos DisProt que reúne una colección de 500 proteínas desordenadas, 156 de las 500 proteínas se identificaron sólo considerando DC; en tanto que, la mayoría de las PIDs más prominentes se detectaron como desordenadas usando DC de UV-lejano (Sickmeier *et al.* 2007). Sólo por citar algunos ejemplos, el DC se utilizó en el caso de la MAP2 (Hernandez *et al.* 1986), tau (Schweers *et al.* 1994), ProTa (Gast *et al.* 1995), α -sinucleína (Weinreb *et al.* 1996), p21^{Cip1} (Kriwacki *et al.* 1996) y la HMGA (Reeves and Beckerbauer 2001).

I.VII. Resonancia magnética nuclear

La resonancia magnética nuclear (RMN) tiene un estatus especial entre las técnicas espectroscópicas puesto que con ella se puede obtener información a nivel de un residuo de aminoácido en la estructura y la dinámica de las PIDs. En el estudio de estas proteínas, la RMN se utilizó inicialmente para demostrar si una proteína era o no

desordenada (comparando los espectros obtenidos con los de las proteínas globulares, los cuales presentan señales muy distintas) (Figura 6) (Tompa 2010).

Posteriormente, la RMN se utilizó para afinar su caracterización y determinar la estructura residual en las PIDs, lo cual en algunos casos ha permitido relacionar estas propiedades estructurales con su función, principalmente asociada con la interacción con otras moléculas (Radhakrishnan *et al.* 1997).

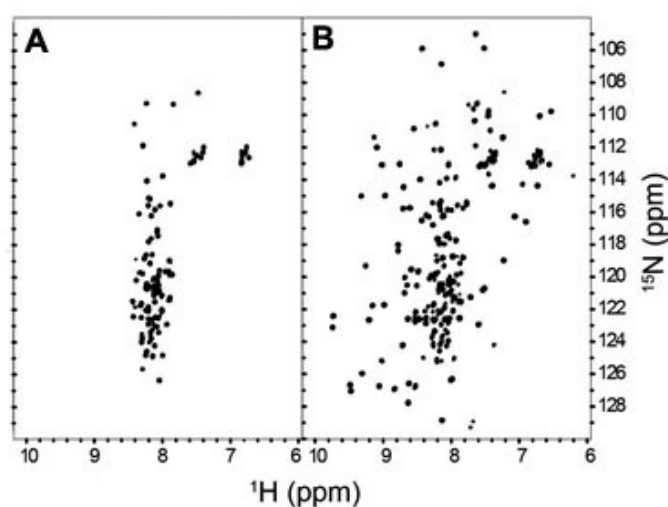


Figura 6. Espectros de resonancia magnética nuclear (RMN). Dispersión de desplazamientos químicos típicos de A) una proteína intrínsecamente desordenada y B) una proteína globular. Nótese que el intervalo de las señales de desplazamiento químico en el dominio del ^1H de las proteínas intrínsecamente desordenadas está entre 6.5 y 8.5 mientras que las proteínas globulares las señales se extienden hasta valores de 10. Modificado de Patel *et al.* 2014.

La RMN se basa en la existencia de los *espines* nucleares y el momento magnético de los núcleos atómicos de átomos como ^1H , ^{13}C y ^{15}N (Barrett *et al.* 2013). Estos núcleos tienen dos estados posibles de *espín*, los cuales están separados cuando se encuentran en un campo magnético externo (B_0). Se pueden inducir transiciones entre los dos estados a través de la absorción de resonancia de la radiación electromagnética a una frecuencia que equipara la diferencia de energía entre los dos estados (Barrett *et al.* 2013). Debido a que dicha diferencia depende del ambiente químico de los núcleos, la señal de RMN contiene mucha información sobre el ambiente local covalente, así como sobre el arreglo espacial de los átomos. En los ensayos de RMN de una dimensión (1-D), dos dimensiones (2-D) o múltiples dimensiones (multi-D), se suelen usar proteínas marcadas en dos (^{15}N , ^{13}C) o tres átomos (^{15}N , ^{13}C , ^2H) (Barrett *et al.* 2013). Incluso se puede usar más de una

frecuencia del campo magnético de excitación, de tal forma que se ensamblan en diferentes combinaciones en el dominio tiempo. Estas combinaciones se diseñan para permitir una transferencia de magnetización entre núcleos (espines) en movimiento con el objeto de detectar sus interacciones. Después de la excitación se determina el regreso al estado basal del vector de magnetización que estuvo fuera del equilibrio. Este proceso se caracteriza por dos eventos de relajación distintos: la relajación del espín de malla longitudinal (caracterizada por una velocidad de relajación R_1) describe la relajación hacia el equilibrio del espín en su componente Z, el cual es paralelo a B_0 ; este evento incluye el intercambio de energía con su ambiente (Barrett *et al.* 2013). La relajación espín-espín transversal (caracterizada por la velocidad de relajación R_2) corresponde a la relajación en los componentes XY del espín, en este caso perpendicular a B_0 , y que ocurre sin intercambio de energía con el ambiente y usualmente es mucho más rápida que R_1 . La señal global obtenida se conoce como el decaimiento de inducción libre (DIL) y contiene la suma de las respuestas de RMN de todos los espines excitados (Barrett *et al.* 2013).

Básicamente, existen dos tipos de interacciones: (i) interacciones a través de un enlace e (ii) interacciones a través del espacio entre los sistemas de espín, siendo estas últimas una consecuencia del *efecto nuclear Overhauser* (Colmenero-Flores *et al.*) (por ejemplo, la transferencia de la polarización del espín entre las poblaciones del espín vía relajación cruzada) (Tompa 2010). La transformación de Fourier del DIL provee lo que se muestra en un espectro de RMN: el desplazamiento de la frecuencia de RMN intrínseca causado por el ambiente químico en el que se encuentra el átomo, el cual se conoce como *desplazamiento químico* (expresado en “partes por millón”; por ejemplo, ppm de la frecuencia de excitación) (Barrett *et al.* 2013). Una dispersión pobre o mala de los desplazamientos químicos provoca que la asignación de la señal (el proceso de identificar cuál de las resonancias corresponde a qué residuo de la proteína) sea muy difícil (Tompa 2010). Si bien este paso es crítico para la obtención de información específica de la estructura, varios enfoques de RMN no requieren que se asignen las señales y más bien arrojan una descripción global del estado estructural de la proteína.

Usualmente, el primer experimento que se realiza con una proteína marcada con el isótopo ^{15}N es la *coherencia sencilla heteronuclear cuántica* (del inglés *Heteronuclear Single Quantum Coherence, HSQC*), la cual correlaciona los nitrógenos amida del esqueleto de la proteína con sus protones unidos directamente

(Barrett *et al.* 2013). En un espectro de HSQC con ^1H - ^{15}N , el enlace amida de cada aminoácido (con excepción del de las prolinas) y los nitrógenos amida de las cadenas laterales de aquellos aminoácidos que contienen amidas arrojarán una señal (Barrett *et al.* 2013). Para las PIDs, los picos en la dimensión del protón tienen muy poca dispersión, con desplazamientos químicos entre 8 y 8.5 ppm (Figura 6A) (Tompa 2010). Este patrón de dispersión es un reflejo claro de que la mayoría de los residuos de las PIDs se encuentran en un ambiente químico muy similar (expuestos al solvente). En la dimensión del nitrógeno las señales son muy amplias, entre 105 – 130 ppm para los grupos amida del esqueleto y de las cadenas laterales. Este patrón de dispersión particular es un indicador confiable del desorden estructural, por lo que los experimentos de HSQC se emplean frecuentemente para caracterizar el estado desordenado de una PID (Tompa 2010).

II. ANTECEDENTES

En el genoma de la planta modelo *Arabidopsis thaliana*, existen 51 genes que codifican para proteínas LEA de los cuales, tres corresponden al grupo 4 (*AtLEA4-1*, *AtLEA4-2* y *AtLEA4-5*). Las proteínas LEA del grupo 4 (LEA 4) se encontraron originalmente en embriones secos de algodón a una concentración cercana a 300 μM (Roberts *et al.* 1993). Se ha encontrado que los transcritos de los miembros del grupo LEA4 son de los que más se acumulan en la semilla seca de *Arabidopsis* (Schmid *et al.* 2005; Winter *et al.* 2007; Hruz *et al.* 2008; Olvera-Carrillo *et al.* 2010). De manera interesante, *AtLEA4-2* y *AtLEA4-5* fueron los transcritos más sobre-regulados en la triple mutante *hab1-labi1-2pp2ca-1* de *Arabidopsis*. Esta mutante carece de tres miembros del clado A de las proteínas PP2Cs (fosfatasas tipo 2C que funcionan como reguladores negativos de la vía de señalización mediada por ABA, y presenta una respuesta constitutiva a ABA (como si la planta percibiera que está constitutivamente en déficit hídrico). Estos datos se suman a otros que indican que la *AtLEA4-2* y la *AtLEA4-5* están reguladas por ABA (Abba' *et al.* 2006; Rubio *et al.* 2009; Olvera-Carrillo *et al.* 2010).

Diversos estudios muestran que las proteínas LEA participan en la respuesta y adaptación de las plantas a condiciones de estrés por falta de agua (Battaglia *et al.* 2008; Hundertmark and Hinch 2008). Una de las evidencias más fuertes que

demuestran la importancia de las proteínas LEA en la respuesta a déficit hídrico se encontró justamente en el grupo 4, donde se demostró que mutantes nulas y silenciadas en cada uno de los genes que codifican para proteínas de este grupo presentan una tasa de germinación reducida cuando se someten a estrés osmótico o salino, comparadas con plantas silvestres (Olvera-Carrillo *et al.* 2010). Además, dichas plantas experimentan una menor recuperación que las plantas silvestres después de una sequía severa progresiva. En ese trabajo se demostró que las proteínas LEA 4 son necesarias para la tolerancia de *Arabidopsis* al déficit hídrico, posiblemente previniendo el daño que causa el estrés a las macromoléculas celulares (Olvera-Carrillo *et al.* 2010). Estos datos muestran claramente que las proteínas LEA del grupo 4 de *Arabidopsis* participan en la tolerancia ante condiciones de déficit hídrico en diferentes estadios de desarrollo; sin embargo, poco se sabe respecto al mecanismo a través del cual realizan este efecto.

La función de las proteínas LEA ha sido un tema difícil de abordar debido a que estas proteínas no comparten homología a nivel de secuencia con dominios de función conocida. Existen varias hipótesis acerca de la función de las proteínas LEA. La gran mayoría de estas hipótesis consideran que estas proteínas son capaces de evitar que los componentes celulares sufran los daños causados por el déficit hídrico (Tunnacliffe and Wise 2007; Battaglia *et al.* 2008; Hundertmark and Hinch 2008; Olvera-Carrillo *et al.* 2011; Chakrabortee *et al.* 2012). Se han propuesto diversas funciones para las proteínas LEA 4 entre las que se encuentran: protectores de membranas, unión de azúcares, unión de metales, secuestradores de radicales libres; así como deshidro-protectores y crio-protectores (Reyes *et al.* 2005; Reyes *et al.* 2008; Hundertmark *et al.* 2011; Liu *et al.* 2011; Dang *et al.* 2014). Sin embargo, a la fecha, no ha sido posible descifrar totalmente su mecanismo de acción a nivel molecular. Una de las hipótesis que se han abordado para el estudio de los mecanismos de acción de las proteínas LEA establece que el desorden estructural de estas proteínas es necesario para realizar apropiadamente su función; por lo tanto, el estudio a nivel estructural es fundamental para tener un mejor entendimiento de la función de las proteínas LEA.

La naturaleza desordenada y flexible predicha para la estructura de las proteínas LEA y el hecho de que se acumulen bajo condiciones en las cuales el agua disponible llega a ser limitante para la célula plantea la pregunta de si estas proteínas sufren cambios en sus organización estructural, los cuales podrían estar relacionados

con su función. Los cambios que ocurren tras la disminución de las moléculas de agua entre las diferentes macromoléculas pueden llevar a que las interacciones entre los átomos de las macromoléculas no se lleven a cabo de la misma manera y, en consecuencia, esto puede causar cambios de diferentes niveles en su organización estructural. En este escenario, es factible pensar que una proteína desordenada podría adquirir un mayor orden y que éste dependiera de los niveles de agua disponible.

Hace algunos años, se demostró que dos proteínas LEA del grupo 4 de soya (GmPM1 y GmPM28) son proteínas intrínsecamente desordenadas cuando se encuentran hidratadas. Ambas proteínas ganaron estructura secundaria tipo α -hélice después de deshidratarlas completamente, por lo que se propuso que GmPM1 y GmPM28 son funcionales en un estado de desecación como componentes del estado vítreo de la semilla y, de esta manera, estabilizan proteínas y membranas cuando se encuentran en dichos estados de deshidratación severa (Shih *et al.* 2010). Diversos estudios sobre PIDs de animales y bacterias han mostrado que la función de estas proteínas se modula por cambios conformacionales (Dyson and Wright 2005; Wright and Dyson 2015). El plegamiento acoplado a la interacción con sus proteínas blanco ha sido el proceso más descrito a la fecha (Borchers *et al.* 2014; Rogers *et al.* 2014). Existen algunos ejemplos en los que el cambio conformacional está mediado por la condición ambiental (pH o estado redox) (Tapley *et al.* 2009; Reichmann *et al.* 2012), sin embargo, se ha visto que el potencial hídrico y el amontonamiento macromolecular solo tienen efectos menores en la estructura de algunas proteínas intrínsecamente desordenadas, incluyendo algunas proteínas LEA (Mouillon *et al.* 2008; Szasz *et al.* 2011; Qin and Zhou 2013; Goldenberg and Argyle 2014). Aún cuando se ha reportado para algunas proteínas LEA la adquisición de estructura secundaria después de una deshidratación completa (Popova *et al.* 2011; Hundertmark *et al.* 2012; Shih *et al.* 2012), no estaba claro si estas proteínas se pueden plegar en condiciones de déficit hídrico menos severo y a través de qué vía de plegamiento ocurre este proceso. Interesantemente, se ha observado que las proteínas LEA 4 tienen cierta tendencia a formar estructura tipo α -hélice *in silico*, por lo que se propone que son un excelente modelo de estudio para determinar si dichas proteínas son capaces de experimentar cambios conformacionales bajo condiciones de limitación del agua disponible, condición en la cual estas proteínas se acumulan en las células vegetales.

Por todo lo anterior, resultaba relevante conocer si las proteínas LEA 4 experimentan cambios conformacionales ante condiciones presentes bajo deficiencia de agua, particularmente en niveles que no alcanzaran una deshidratación total, los cuales reflejarían estados de limitación de agua como los presentes en tejidos vegetativos de la planta bajo ambientes con irrigación o humedad limitada. Además, dada las características y las diferencias en la estructura primaria de estas proteínas resulta relevante entender la participación de los motivos conservados en cada subgrupo y la relación entre su estructura tridimensional y su función protectora en condiciones de déficit hídrico.

III. HIPÓTESIS

Los motivos presentes en la familia de proteínas LEA del grupo 4 de *Arabidopsis thaliana* participan en la adquisición de una conformación funcional en condiciones de déficit hídrico.

IV. OBJETIVO GENERAL

Analizar la relación que existe entre la estructura y la función de las proteínas LEA del grupo 4 de *Arabidopsis thaliana* en condiciones de limitación de agua.

V. OBJETIVOS ESPECÍFICOS

- Predecir los niveles de estructura secundaria y desorden de las proteínas LEA del grupo 4 de *Arabidopsis thaliana* (AtLEA4) *in silico*.
- Analizar la estructura secundaria de las proteínas recombinantes AtLEA4 en solución acuosa *in vitro* mediante dicroísmo circular y resonancia magnética nuclear.
- Determinar si la conformación de las proteínas AtLEA4 se modifica en condiciones de déficit hídrico simulado *in vitro*.
- Encontrar los motivos de la proteína importantes para la estructura de las proteínas AtLEA4.
- Analizar la función de las proteínas AtLEA4 y la relación que guarda con su estructura.

VI. MATERIALES Y MÉTODOS

VI.I. *Análisis in silico de las proteínas LEA del grupo 4 de Arabidopsis thaliana*

Los marcos abiertos de lectura de las proteínas LEA del grupo 4 se obtuvieron del servidor *TAIR* (*The Arabidopsis Information Resource*, <https://www.arabidopsis.org/>), con los identificadores *At1g32560* (AtLEA4-1), *At2g35300* (AtLEA4-2) y *At5g06760* (AtLEA4-5). Los marcos de lectura se tradujeron a la secuencia polipeptídica correspondiente usando *Translate* del *ExpASy Bioinformatics Resource Portal* (<http://web.expasy.org/translate/>). Las secuencias de las proteínas se alinearon con el programa *Multiple Sequence Alignment* de *T-Coffee* (<http://www.ebi.ac.uk/Tools/msa/tcoffee/>) de donde se obtuvieron los porcentajes de similitud. La composición de aminoácidos, tamaño, y punto isoeléctrico se calcularon con el programa *ProtParam tool* de (<http://web.expasy.org/protparam/>). La predicción del contenido de estructura desordenada se calculó con *DISpro* (<http://scratch.proteomics.ics.uci.edu/>), *PONDR* (<http://www.pondr.com/>) y *DISOPRED3* (<http://bioinf.cs.ucl.ac.uk/index.php?id=806>). La predicción del contenido de α -hélices se obtuvo con *AGADIR* (<http://agadir.crg.es/>).

VI.II. *Construcción de los plásmidos utilizados con las distintas versiones de las AtLEA4*

Los marcos abiertos de lectura de la *AtLEA4-1*, *AtLEA4-2* y *AtLEA4-5* se clonaron a partir de cADN obtenido de ARN extraído de semillas secas de *Arabidopsis*. Los tres marcos de lectura se aislaron por PCR usando oligonucleótidos específicos que contenían en sus extremos los sitios de restricción *NcoI* y *Sall*. Esta estrategia originó un cambio puntual de un nucleótido que modificó el segundo codón del marco de lectura, generando ácido glutámico (GAG) en vez de glutamina (CAG). La *AtLEA4-1* se clonó con los oligonucleótidos TRC41Fw y TRC41Rv, la *AtLEA4-2* con los oligonucleótidos TRC42Fw Y TRC42Rv y la *AtLEA4-5* con los oligonucleótidos TRC45Fw Y TRC45Rv. La secuencia de cada oligonucleótido se puede encontrar en la Tabla Anexo I. Los fragmentos de ADN amplificados se insertaron en el vector pJET1.2/blunt para producir los plásmidos pJET1.2:AtLEA4-1; pJET1.2:AtLEA4-2 y pJET1.2:AtLEA4-5. Posteriormente, los marcos abiertos de

lectura fueron digeridos usando las enzimas de restricción *NcoI* y *Sall* y subsecuentemente ligados direccionalmente en el vector pTrc99A (previamente digerido), para obtener los plásmidos pTrc99A:AtLEA4-1; pTrc99A:AtLEA4-2 y pTrc99A:AtLEA4-5. En el caso de la AtLEA4-5, después de purificar la proteína se obtenía una banda de menor peso molecular al esperado, el cual era el resultado de la traducción a partir de un ATG (que codifica para metionina) interno. Con la finalidad de eliminar la producción de esta proteína aberrante se cambiaron estos ATG usando mutagénesis sitio dirigida por medio de la técnica de *mega-primer*. Para ello, se seleccionaron las metioninas (ATG) de la posición 19, 33 y 43 y cada codón independientemente se modificó para codificar leucina (CTG). Para la Met19 se usaron un par de oligonucleótidos que se empalman llamados M19Lfw y M19LRv. Para la Met33 se usaron los oligonucleótidos M33Lfw y M33LRv. Para la Met43 se usaron los oligonucleótidos M43Lfw y M43LRv. Los fragmentos modificados se insertaron en el vector pJET1.2/blunt para producir los plásmidos pJET1.2:M19L; pJET1.2:M33L y pJET1.2:M43L. A partir de aquí, cuando me refiera al plásmido pJET1.2:AtLEA4-5 me estaré refiriendo al plásmido pJET1.2:M33L. Los fragmentos modificados fueron digeridos usando las enzimas de restricción *NcoI* y *Sall* y, subsecuentemente, ligados direccionalmente en el vector pTrc99A (previamente digerido) para obtener los plásmidos pTrc99A:M19L; pTrc99A:M33L y pTrc99A:M43L. A partir de aquí, cuando me refiera al plásmido pTrc99A:AtLEA4-5 me estaré refiriendo al plásmido pTrc99A:M33L. Las regiones N-terminal y C-terminal de la AtLEA4-5 se amplificaron a partir del plásmido pJET1.2:AtLEA4-5 usando los oligonucleótidos TRC45Fw y TRC4HRv (para la AtLEA4-5₁₋₇₇) y TRC4RCFw y TRC45Rv (para la AtLEA4-5₇₈₋₁₅₈). En el caso de la región C-terminal, se adicionó un ATG para iniciar la traducción a partir de ese sitio. Los fragmentos de ADN amplificados se insertaron en el vector pJET1.2/blunt para producir los plásmidos pJET1.2:AtLEA4-5₁₋₇₇ y pJET1.2:AtLEA4-5₇₈₋₁₅₈. Los marcos de lectura para cada una de las regiones fueron digeridos usando las enzimas de restricción *NcoI* y *Sall* y subsecuentemente ligados direccionalmente en el vector pTrc99A (previamente digerido), para obtener los plásmidos pTrc99A:AtLEA4-5₁₋₇₇ y pTrc99A:AtLEA4-5₇₈₋₁₅₈.

La AtLEA4-1, AtLEA4-2 y AtLEA4-5₇₈₋₁₅₈ se clonaron en el vector pTYB11 para obtener las proteínas como fusión a inteína, usando el sistema de expresión IMPACT-CN (Intein Mediated Purification with an Affinity Chitin-binding tag; New

England Biolabs Inc.). Para ello, los marcos de lectura se amplificaron usando como template a los plásmidos pJET1.2:AtLEA4-1, pJET1.2:AtLEA4-2 y pJET1.2:AtLEA4-5₇₈₋₁₅₈ con oligonucleótidos que contuvieran los sitios de restricción *SapI* y *PstI* (para la AtLEA4-1: PTYB41Fw y PTYB41Rv; para la AtLEA4-2: PTYB42Fw y PTYB42Rv y para la AtLEA4-5₇₈₋₁₅₈ PTYBRCFw y PTYBRRCRv). Los fragmentos de ADN amplificados se sub-clonaron en el vector pJET1.2/blunt, se digirieron con las enzimas *SapI* y *PstI* y, subsecuentemente, se ligaron direccionalmente en el vector pTYB11 (previamente digerido) para obtener los plásmidos pTYB11:AtLEA4-1, pTYB11:AtLEA4-2 y pTYB11:AtLEA4-5₇₈₋₁₅₈. En todos los casos se verificó que la secuencia no tuviera mutaciones y que estuviera insertada en la orientación correcta por medio de secuenciación de los ADNs correspondientes, realizada por la Unidad de Síntesis y Secuenciación de ADN del Instituto de Biotecnología.

Para la aislamiento de la región N-terminal conservada de la AtLEA4-5 como fusión a la proteína GB1 y a una etiqueta de histidinas, el marco de lectura que codifica para los primeros 76 aminoácidos de la AtLEA4-5 se amplificó a partir del plásmido pTrc99A:AtLEA4-5₁₋₇₇ con los oligonucleótidos LEA1Fw y LEA76Rv, los cuales contienen los sitios de restricción *NdeI* y *BamHI*, respectivamente. El fragmento de ADN amplificado se digirió con *NdeI* y *BamHI* y, posteriormente, se ligó direccionalmente en el vector GB1 para obtener el plásmido GB1:AtLEA4-5₁₋₇₆. La identidad de la secuencia y la orientación se verificó por secuenciación.

VI.III. Expresión de las proteínas recombinantes

Los plásmidos de expresión pTrc99A:AtLEA4-1; pTrc99A:AtLEA4-2, pTrc99A:AtLEA4-5 (pTrc99A:M19L; pTrc99A:M33L y pTrc99A:M43L), pTrc99A:AtLEA4-5₁₋₇₇, pTrc99A:AtLEA4-5₇₈₋₁₅₈, pTYB11:AtLEA4-1, pTYB11:AtLEA4-2 y pTYB11:AtLEA4-5₇₈₋₁₅₈ se transformaron en la cepa BL21(DE3) pLysS (Promega) de *Escherichia coli* (con esta cepa se obtuvieron los mejores niveles de inducción de la proteína, aunque se utilizaron otras cepas como RosettaTM2 (Merck-Millipore) y ER2566) por el método de transformación por choque térmico. Las colonias transformantes fueron seleccionadas por resistencia a ampicilina y la presencia del plásmido en cuestión se corroboró por PCR. A

continuación se menciona el protocolo que se utilizó para todas las construcciones en donde se pudo inducir exitosamente la proteína. Una colonia se inoculó en 100 mL de medio LB con 100 µg/mL de ampicilina y se creció toda la noche a 37 °C y 200 rpm para llegar a un crecimiento a saturación. Al otro día, se midió la DO₆₀₀ y se mezcló el volumen necesario para preparar 1 litro de medio LB fresco a una DO₆₀₀ de 0.01 con 100 µg/mL de ampicilina. El cultivo se incubó a 37 °C y 200 rpm hasta obtener una DO₆₀₀ 0.5 – 0.8. En este momento, la expresión de la proteína recombinante se indujo con 1 mM de Isopropil β-D-1-tiogalactopiranosido (IPTG) durante por lo menos 6 horas, a 25 °C y 200 rpm. Finalmente, las células se recuperaron por centrifugación a 10,000 rpm por 10 minutos y la pastilla se congeló a -20 °C hasta su uso posterior. Si bien este fue el protocolo que se estableció como el óptimo para la producción de las proteínas recombinantes, se analizaron diferentes condiciones de crecimiento así como concentraciones de IPTG y tiempos de inducción. Para ello, todo el proceso se realizó en 3 mL de medio LB y las diferentes muestras se separaron en los geles de SDS-PAGE para determinar las condiciones óptimas de inducción.

VI.IV. Purificación de las proteínas recombinantes

Para la purificación de las proteínas recombinantes se utilizaron dos métodos diferentes, dependiendo del plásmido de procedencia. Para las proteínas LEA clonadas en el vector pTrc99A se empleó el método de purificación de proteínas desordenadas no-ácidas descrito por Campos y colaboradores (Campos *et al.* 2011). En el caso de las proteínas LEA clonadas en el vector pTYB11 se usó el método descrito por el fabricante (New England BioLabs, Inc). Ambos protocolos se describen a continuación:

Para el protocolo de purificación de proteínas desordenadas no-ácidas, la pastilla de células se resuspendió en 20 mL de amortiguador A (20 mM Tris –HCl pH 8, 10 mM NaCl), se hirvió en baño maría por 10 minutos y se enfrió en hielo por otros 10 minutos. Los restos celulares (incluyendo las proteínas sensibles al hervor) se separaron por centrifugación a 10,000 rpm por 10 minutos. Se tomó el sobrenadante y se agregó gota a gota el volumen necesario de ácido tricloroacético (TCA) al 40% (p/v) para llegar a una concentración final de TCA al 5% (en la fracción enriquecida de proteínas) y se mantuvo en agitación durante 30 minutos en hielo. Las proteínas

sensibles al tratamiento con esta concentración de ácido se separaron por centrifugación a 10,000 rpm durante 10 minutos. El sobrenadante (enriquecido en la proteína de interés) se tomó y se agregó gota a gota el volumen necesario de TCA al 100% (p/v) para llegar a una concentración final de 25% de TCA y se mantuvo en agitación durante 30 minutos. La proteína recombinante se separó por centrifugación a 10,000 rpm durante 10 minutos. La pastilla de proteínas se lavó dos veces con acetona fría al 100% y, posteriormente, se dejó evaporar. La pastilla de proteína recombinante se resuspendió en amortiguador de fosfatos de sodio 10 mM pH 7.5 y se dializó contra ese mismo amortiguador a 4 °C, realizando cambios de amortiguador primero toda la noche y después cada 2 horas al otro día. La solución de proteína se recuperó y se determinó la concentración valiéndonos del coeficiente de extinción molar a 280 nm ($\epsilon=2980 \text{ M}^{-1}\text{cm}^{-1}$ para la AtLEA4-2 y la AtLEA4-5). En el caso de la AtLEA4-5₁₋₇₇, debido a que no contiene aminoácidos aromáticos, la concentración se determinó por cuantificación por Bradford y se ajustó comparando en geles de SDS-PAGE con concentraciones conocidas de la AtLEA4-5. La identidad y la pureza de todas las proteínas se validó por SDS-PAGE (verificando que únicamente se observará la banda de interés con el peso molecular esperado) y por espectrometría de masas, realizada por el Laboratorio Universitario de Proteómica del Instituto de Biotecnología de la UNAM, por métodos estándares para el sistema de ionización por nano-spray con un espectrómetro de masas LQT-Orbitrap Velos (Thermo-Fisher Co).

En el caso del protocolo de purificación mediado por inteína, la pastilla de proteína se resuspendió en 50 mL de amortiguador de lisis (20 mM fosfatos de sodio pH 8, 500 mM NaCl, 0.1% Triton X-100) con una tableta del inhibidor de proteasas cOmplete (Roche). Las células se lisaron por sonicación (intervalos de 10 segundos pulso/10 segundos descanso) en hielo durante 10 minutos (dos veces). Los residuos celulares se separaron por centrifugación a 20,000 x g durante 30 minutos a 4 °C. El extracto clarificado (sobrenadante) se cargó en una columna previamente equilibrada (con el amortiguador de columna) de quitina. La columna se equilibró lavando con 10 volúmenes de amortiguador de columna (20 mM fosfatos de sodio pH 8.5, 500 mM NaCl). El extracto clarificado se pasó por la columna a una velocidad de flujo de 0.5 mL/min. La columna se lavó con 20 volúmenes de amortiguador de columna y, posteriormente, se indujo el auto-corte de la proteína pasando tres volúmenes de amortiguador de corte (20 mM fosfatos de

sodio pH 8.5, 500 mM NaCl, 50 mM β -mercaptoetanol). La columna se cerró y se dejó a temperatura ambiente durante 24 horas. Para eluir la proteína, se abrió la columna y se tomaron 18 muestras de 2 mL cada una. La presencia de la proteína recombinante se determinó por geles de SDS-PAGE. Las fracciones en donde se detectó la proteína se juntaron, se dializaron contra amortiguador de fosfatos 10 mM pH 7.5 primero toda la noche y después haciendo cambios cada dos horas al otro día. A partir de este momento, la proteína se cuantificó y se verificó su pureza e identidad de la misma forma en que está especificado en el protocolo de purificación anterior.

Las proteínas puras recombinantes se mantuvieron a 4 °C para su uso en los distintos experimentos que se realizaron dentro del mes siguiente. Además, otra porción de la proteína se liofilizó o se congeló a – 70 °C.

Finalmente, el polipéptido de la región C-terminal de la AtLEA4-5 (AtLEA4-5₇₈₋₁₅₈) se sintetizó químicamente (síntesis directa) con la compañía BIOMATIK (Cambridge, Canada). Se ordenaron 10 mg del polipéptido el cual se obtuvo con una pureza del 95.79% y la identidad del mismo fue verificada por espectrometría de masas. La AtLEA4-5₇₈₋₁₅₈ se disolvió en amortiguador de fosfatos de sodio 10 mM y se dializó de la misma manera que las proteínas recombinantes. Algunas alícuotas se mantuvieron a 4 °C para su uso inmediato y otras se almacenaron a – 70 °C.

VI.V. Expresión y purificación de las proteínas recombinantes marcadas con ¹⁵N

Para realizar los ensayos de RMN fue necesario marcar las proteínas con ¹⁵N. En el caso de la AtLEA4-5, todo el proceso se realizó utilizando bacterias conteniendo el plásmido pTrc99A:AtLEA4-5. Debido al bajo rendimiento en la purificación de la región N-terminal de la AtLEA4-5, se utilizó la construcción GB1:AtLEA4-5₁₋₇₆ (ver sección “Construcción de los plásmidos utilizados con las distintas versiones de las AtLEA4”), la cual consiste de los primeros 76 aminoácidos de la AtLEA4-5. Se ha observado que la fusión al N-terminal de proteínas intrínsecamente desordenadas a la proteína GB1 (Dominio B1 de la proteína G de streptococco) reduce significativamente la degradación de las proteínas desordenadas expresadas en *E. coli* y, por lo tanto, mejora el rendimiento en la producción de la proteína recombinante (Huth *et al.* 1997; Dames *et al.* 2002; Koenig *et al.* 2003; Sugase *et al.* 2008). Esta construcción además contiene una etiqueta de histidinas, lo

cual facilita su purificación. La proteína recombinante de interés puede escindirse fácilmente de GB1 pues la construcción contiene un sitio de corte de específico para la proteasa TEV.

De esta manera, los plásmidos pTrc99A:AtLEA4-5 y GB1:AtLEA4-5₁₋₇₆ se transformaron en células BL21 (DE3) pLysS. Las colonias positivas fueron validadas por PCR. A continuación se menciona el protocolo de inducción de la proteína para ambas construcciones. Una colonia que contenía esta construcción se creció en medio LB toda la noche a 37 °C y 200 rpm. Al día siguiente, los 20 mL se transfirieron a 1 litro de medio mínimo M9 (Ver anexo) suplementado con (¹⁵NH₄)₂SO₄ y el cultivo se creció a 37 °C y 200 rpm hasta alcanzar una DO₆₀₀ de 0.7. En este momento se indujo la expresión de la proteína con 1 mM de IPTG y el cultivo se creció a 25 °C toda la noche. Las células se recuperaron por centrifugación a 5,500 rpm durante 15 minutos, se descartó el sobrenadante y la pastilla de células se utilizó para la purificación de la proteína recombinante marcada, según sea el caso.

Para la purificación de la ¹⁵N-AtLEA4-5 el protocolo fue el siguiente: la pastilla de células se resuspendió en 80 mL de amortiguador de lisis (20 mM Tris-HCl pH 8, 250 mM NaCl) y la suspensión se sonicó (Sonic dismembrator model 500; Fisher Scientific) en hielo (40% potencia, un segundo encendido/3 segundos apagado, durante dos minutos (de pulso) y repetido cuatro veces). Los residuos celulares se separaron por centrifugación a 14,000 rpm por 25 minutos. El sobrenadante enriquecido en proteínas se cargó en un equipo de FPLC para separar la proteína en una columna de intercambio catiónico (Hi-TRAP SP FF, 5 mL; GE Healthcare Life Sciences) a una velocidad de 3 mL/minuto. Debido a que la ¹⁵N-AtLEA4-5 tiene carga positiva, ésta se pega a la columna. La columna se lavó con 20 mL de amortiguador A (20 mM Tris-HCl pH 8). La ¹⁵N-AtLEA4-5 se eluyó con un gradiente de NaCl (0 – 1 M (0 – 100%)) a una velocidad de 3 mL/minuto (1%/minuto). La presencia de la proteína en cada fracción se determinó por la A₂₈₀ en el equipo de FPLC y por SDS-PAGE. Posteriormente, se procedió a separar la ¹⁵N-AtLEA4-5 por filtración en gel con una columna de 120 mL. La columna se lavó con amortiguador 20 mM Tris-HCl pH 7.5, 100 mM NaCl a 1 mL/minuto. La muestra de proteína se inyectó y se tomaron fracciones de 2 mL. La presencia de la proteína en cada fracción se determinó por la A₂₈₀ en el equipo de FPLC y por SDS-PAGE. Las fracciones en donde se detectó la proteína se juntaron y se les cambió el amortiguador con columnas de desalado PD-10 (GE Healthcare Life Sciences) para que la proteína

quedara en amortiguador 20 mM Tris-HCl pH 6.8, 50 mM NaCl. La muestra de proteína se concentró hasta obtener 500 μ L a una concentración de 155 μ M para su uso posterior.

En el caso de la 15 N-AtLEA4-5₁₋₇₆, el protocolo de purificación fue el siguiente: la pastilla de células que contenía la proteína recombinante se resuspendieron en 80 mL de amortiguador de lisis (20 mM Tris-HCl pH 8, 250 mM NaCl) y se sonicaron en hielo (40% potencia, un segundo encendido/tres segundos apagado, durante 2 minutos (de pulso) y repetido 4 veces). Los residuos celulares se separaron por centrifugación a 14,000 rpm por 20 minutos. Mientras tanto, 5 mL de resina de níquel Roche Complete se cargaron en la columna y se lavó tres veces con 50 mL de agua y luego tres veces más con 50 mL de amortiguador de lisis. El sobrenadante de la centrifugación se filtró por papel Whatman 8 y se cargó en la columna con la resina de níquel dos veces. Posteriormente, la columna se lavó tres veces con 50 mL de amortiguador de lavado #1 (20 mM Tris-HCl pH 8, 250 mM NaCl, 5 mM imidazol). A continuación, la columna se lavó dos veces con 50 mL de amortiguador de lavado #2 (20 mM Tris-HCl pH 8, 250 mM NaCl, 25 mM imidazol). Finalmente, la proteína se eluyó con 5 mL de amortiguador de elución (20 mM Tris-HCl pH 8, 250 mM NaCl, 200 mM imidazol) y este paso se repitió otras cuatro veces. La presencia de las proteínas en cada una de las fracciones se verificó por SDS-PAGE. Las fracciones que contenían la mayor cantidad de proteína se juntaron en una sola. La concentración aproximada de la proteína se determinó por A_{280} . Dependiendo de la concentración obtenida, la muestra se diluyó hasta obtener una concentración de 50 μ M y se agregó la cantidad de enzima TEV necesaria para obtener una relación molar 200:1 (proteína:enzima); así como, 5 mM DTT. La muestra se dializó contra dos litros de amortiguador 20 mM Tris-HCl pH 8, 100 mM NaCl, 5 mM DTT toda la noche a 4 °C. Al otro día, la eficiencia de digestión se detectó por HPLC. Para poder separar la 15 N-AtLEA4-5₁₋₇₆ de la GB1, se ajustó la concentración de NaCl a 250 mM al producto de la digestión para aplicarla en la columna equilibrada de níquel y, se recolectó el volumen que pasó por la columna. La columna se lavó dos veces con 25 mL de amortiguador Tris-HCl pH 8, 200 mM NaCl y se colectó el volumen que pasó por la columna. La presencia de la AtLEA4-5₁₋₇₆ en estas fracciones se detectó por SDS-PAGE. Se determinó la concentración de proteína y se almacenó a 4 °C hasta su uso posterior.

VI.VI. Determinación de la estructura secundaria por dicroísmo circular

Para la determinación de estructura secundaria, muestras de 300 μ L de las proteínas recombinantes a una concentración de 0.3 mg/mL se sometieron a la medición de dicroísmo circular (DC) de UV lejano usando un espectropolarímetro de dicroísmo circular Jasco J-715 (JASCO Analytical Instruments) con una longitud de paso óptico de 0.1 cm. Los espectros se adquirieron desde 190 nm hasta 250 nm. La temperatura deseada se ajustó con un controlador de temperatura Peltier (PTC-4235, JASCO). Por cada tratamiento se promediaron tres espectros y la señal se suavizó para disminuir el ruido. Los espectros se adquirieron cada 1 nm con un tiempo promedio en cada punto de dos segundos. Los experimentos se realizaron en el laboratorio de la Dra. Gloria Saab Rincón en el Instituto de Biotecnología de la UNAM.

Para las termografías, las proteínas recombinantes AtLEA4-2, AtLEA4-5, AtLEA4-5₁₋₇₇ y el polipéptido sintetizado químicamente AtLEA4-5₇₈₋₁₅₈ se diluyeron para obtener una concentración de 0.3 mg/mL en amortiguador de fosfatos de sodio 10 mM pH 7.5. Las muestras se estabilizaron a la temperatura deseada durante 15 minutos antes de realizar la medición del espectro. Los espectros se tomaron con la misma muestra de tal forma que la temperatura siempre fue subiendo cada 10 °C (desde 10 °C hasta 80 °C). Para la medición a diferentes pHs, las proteínas recombinantes se diluyeron a 0.3 mg/mL en amortiguador de fosfatos de sodio 10 mM al pH en cuestión (5.8, 6.8, 7.5 y 8). En el caso de las titulaciones con 2,2,2-Trifluoroetanol (TFE), glicerol y polietilen glicol (PEG-5,000), las muestras se diluyeron a 0.3 mg/mL en amortiguador de fosfatos de sodio 10 mM pH 7.5, solo o con diferentes concentraciones de TFE (4, 8, 12, 16, 20, 40, 60 y 90%), glicerol (10, 20, 30, 40, 50, 60, 70 y 80%) o PEG (45%). Las muestras se incubaron con las distintas concentraciones de los compuestos por lo menos una hora antes de la obtención de los espectros.

Todos los ensayos se repitieron con al menos tres lotes diferentes de proteínas recombinantes purificadas.

La estimación de los niveles de estructura secundaria se calculó con el servidor Dichroweb (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) (Lobley *et al.* 2002; Whitmore and Wallace 2004). Los datos obtenidos se introdujeron al servidor

en donde se utilizó el algoritmo CDSSTR con la base de datos 4, 7 y SP175 para hacer las comparaciones.

VI.VII. Determinación de la estructura secundaria por resonancia magnética nuclear

La ^{15}N -AtLEA4-5 se concentró hasta llegar a 155 μM en amortiguador 20 mM Tris-HCl pH 6.8, 50 mM NaCl. Se tomaron 100 μL de esta solución y se pusieron en una celda para resonancia. La colecta de los datos de resonancia se realizó en el laboratorio de la Dra. Jane Dyson del Instituto Scripps (The Scripps Research Institute) en La Jolla, California. La adquisición de espectros tipo TROSY se realizó en un equipo VNMR-S de 900 MHz a 298 K. Para la AtLEA4-5₁₋₇₆, la proteína se concentró hasta llegar a 267 μM en amortiguador 20 mM Tris-HCl pH 6.8, 50 mM NaCl. Se tomaron 100 μL de esta solución y se pusieron en una celda para resonancia. La colecta de los datos se realizó de la misma manera que para la proteína completa.

VI.VIII. Ensayos de congelamiento/descongelamiento in vitro

Para los ensayos de congelamiento/descongelamiento *in vitro* se tomó como base el protocolo descrito por Reyes y colaboradores (2008), con algunas modificaciones (Reyes *et al.* 2008). Para estos ensayos se utilizó la proteína lactato deshidrogenasa (LDH) de músculo de conejo (Roche). La LDH se diluyó hasta obtener una concentración de 250 nM (en relación al monómero) en amortiguador 25 mM Tris-HCl pH 7.5. Cien μL de esta muestra, que se considera como control, se congelaron en nitrógeno líquido por 30 segundos y posteriormente se descongelaron durante 5 minutos a 25 °C en un *thermomixer*, completando así un ciclo de congelamiento/descongelamiento. El tratamiento se repitió seis veces más (siete ciclos en total) y al finalizar, se tomó una alícuota de 15 μL y se mezcló con 600 μL de amortiguador de reacción (25 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM piruvato, 0.15 mM NADH) para determinar la actividad enzimática. La actividad enzimática de la LDH se siguió como la disminución en la A_{340} como resultado de la conversión de NADH en NAD^+ (NADH absorbe a 340 nm). La actividad de la LDH antes de

someterse al tratamiento se consideró como el 100% y se comparó con las muestras sometidas al tratamiento. El efecto protector de la AtLEA4-2, AtLEA4-5, AtLEA4-5₁₋₇₇, AtLEA4-5₇₈₋₁₅₈ y lisozima se probó adicionando las diferentes proteínas a la muestra de LDH, según la relación molar que se indique (desde 0.5:1 hasta 20:1 de proteína prueba:LDH). La muestra con ambas proteínas se sometió al tratamiento de congelamiento/descongelamiento y, posteriormente, se determinó la actividad enzimática de la LDH.

En todos los casos, cada muestra se midió por triplicado y el experimento global se repitió con muestras de proteínas extraídas de al menos tres diferentes lotes de purificación.

VI.IX. Ensayos de deshidratación parcial in vitro

Los ensayos de deshidratación parcial *in vitro* también se realizaron con la LDH de músculo de conejo. La LDH se diluyó hasta obtener una muestra de 25 μ L a una concentración de 250 nM (monómero), en presencia o ausencia de las diferentes proteínas a probar (AtLEA4-2, AtLEA4-5, AtLEA4-5₁₋₇₇ y AtLEA4-5₇₈₋₁₅₈). Se utilizó una relación molar 5:1 (proteína prueba:LDH). Las muestras se colocaron en un concentrador Speed-Vac (Savant Instruments) para evaporar el agua hasta llegar a un porcentaje de pérdida de agua de más del 98% (El porcentaje de pérdida de agua se determina pesando las muestras antes y después del tratamiento en una balanza analítica). Se monitoreó la temperatura del concentrador para controlar posibles cambios de temperatura, si fuese el caso. Las muestras se re-hidrataron al volumen inicial (asegurándose de resuspender completamente todos los solutos) y se tomaron 15 μ L para medir la actividad enzimática de la LDH de la misma forma descrita en los ensayos de congelamiento/descongelamiento. La actividad de la muestra que contenía sólo a la LDH sin haberse sometido al tratamiento se tomó como el 100% de actividad. En todos los casos, se determinó la actividad de cada muestra por triplicado y el experimento global se repitió con muestras de proteínas extraídas de al menos tres diferentes lotes de purificación.

VI.X. Análisis estadístico

Los análisis estadísticos se realizaron con una prueba de una sola vía de ANOVA. Las diferencias significativas se compararon con el post-test de múltiples comparaciones de Tukey ($P < 0.01$).

VII. RESULTADOS

VII.I. Análisis in silico de las proteínas LEA del grupo 4 de Arabidopsis thaliana

Las proteínas LEA se clasifican en siete grupos o familias, dependiendo de la presencia de motivos conservados específicos para cada una de ellas. La familia de proteínas LEA 4 está ampliamente distribuida en el reino vegetal. En la planta modelo *Arabidopsis thaliana* existen tres genes que codifican para proteínas LEA del grupo 4: AtLEA4-1 (PAP260/LEA6/At1g32560), AtLEA4-2 (PAP80/LEA18/At2g35300) y AtLEA4-5 (PAP51/LEA46/At5g06760) (Hundertmark and Hinch 2008; Olvera-Carrillo *et al.* 2010). La nomenclatura usada en esta tesis es la descrita por Olvera-Carrillo *et al.*, 2010, la cual consiste en las siguientes definiciones. A: *Arabidopsis*; t: *thaliana*; LEA4: proteínas LEA del grupo 4; y el último número representa el cromosoma de Arabidopsis en el cual está codificado cada gen (1, 2 ó 5) (Olvera-Carrillo *et al.* 2010).

Las tres proteínas tienen un alto grado de similitud entre sí. Las proteínas más similares son la AtLEA4-1 y la AtLEA4-2, con 66.7% de similitud, puesto que ambas constan de una región muy similar a la región N-terminal de la proteína AtLEA4-5 y presentan una región C-terminal considerablemente más corta que la misma (Tabla 2).

	AtLEA4-1 vs AtLEA4-2	AtLEA4-1 vs AtLEA4-5	AtLEA4-2 vs AtLEA4-5
Similitud	66.7 %	44.3 %	40.5 %
Identidad	56.3 %	32.9 %	29.7 %

Tabla 2. Comparación entre los porcentajes de similitud e identidad entre las proteínas LEA del grupo 4 de *Arabidopsis thaliana*.

Consistente con lo que se ha observado para otras proteínas LEA del grupo 4, el mayor grado de similitud se localiza en la región N-terminal (entre 55% y 80%), como es evidente del alineamiento de las secuencias de las tres proteínas (Figura 7).

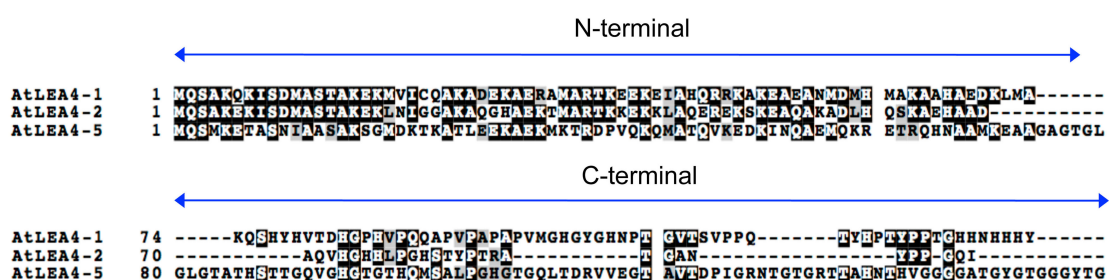


Figura 7. Alineamiento de las proteínas LEA del grupo 4 de *Arabidopsis* que muestra las dos regiones globales que caracterizan este grupo (regiones N-terminal y C-terminal). Los residuos totalmente conservados se muestran sombreados en negro.

Al analizar el contenido de aminoácidos de estas proteínas, encontramos que, en contraste con la composición de aminoácidos promedio de proteínas de plantas y otros organismos (Battaglia *et al.* 2008; Garay-Arroyo *et al.* 2000; Otaki *et al.* 2010), estas proteínas están enriquecidas en residuos hidrofílicos y cargados, como lo son la lisina, la arginina, el ácido aspártico y el ácido glutámico; así como, en aminoácidos pequeños como alanina, treonina y glicina (Tabla 3). Por otro lado, las LEA4 de *Arabidopsis* carecen de aminoácidos hidrofóbicos, los cuales son promotores de plegamiento, como la fenilalanina, el triptofano y la cisteína (Tabla 3).

Las tres proteínas son pequeñas (entre 10.48 kDa y 16.17 kDa) y básicas (punto isoelectrico, pI, entre 8.95 y 9.67) (Tabla 4). Debido a su alto contenido de aminoácidos hidrofílicos y al alto contenido de glicinas, las tres proteínas LEA 4 de *A. thaliana* pertenecen a otro grupo de proteínas conocidas como **hidrofilinas** (Garay-Arroyo *et al.* 2000), las cuales están presentes en otros organismos más allá de las plantas y cuyos transcritos se encuentran acumulados bajo condiciones de estrés, principalmente deshidratación.

	AtLEA4-1 (%)	AtLEA4-2 (%)	AtLEA4-5 (%)
Lys (K)	9.7	14.4	7.6
Arg (R)	3	3.1	3.8
Asp (D)	3.7	3.1	3.2
Glu (E)	6.7	8.2	5.7
Asn (N)	2.2	2.1	3.2
His (H)	11.2	7.2	4.4
Gln (Q)	6	7.2	6.3
Ser (S)	3.7	6.2	3.8
Thr (T)	6	6.2	15.8
Met (M)	6.7	3.1	5.1
Gly (G)	4.5	7.2	17.1
Ala (A)	14.9	17.5	11.4
Tyr (Y)	3.7	2.1	1.3
Val (V)	5.2	1	4.4
Ile (I)	2.2	3.1	1.9
Leu (L)	0.7	4.1	3.2
Pro (P)	9	4.1	1.9
Cys (C)	0.7	0	0
Phe (F)	0	0	0
Trp (W)	0	0	0

Tabla 3. Composición (porcentajes) de aminoácidos de las proteínas LEA del grupo 4 de *Arabidopsis thaliana*. Los aminoácidos se muestran en con código de 3 letras y de una letra (entre paréntesis).

	Longitud (aa)	Masa Molecular (kDa)	Punto Isoeléctrico
AtLEA4-1	134	14.9	8.95
AtLEA4-2	97	10.5	9.67
AtLEA4-5	158	16.2	9.45

Tabla 4. Características generales de las proteínas LEA del grupo 4 de *Arabidopsis thaliana*.

Dadas la composición de aminoácidos de las proteínas LEA4 (así como el de proteínas LEA de otro grupo), se ha predicho que podrían tener una estructura intrínsecamente desordenada (Battaglia *et al.* 2008; Hundertmark and Hinch 2008). Por ello, se analizó la propensión de cada una de las proteínas a formar estructura secundaria o a presentar cierto grado de desorden estructural. Para analizar la predicción de desorden estructural se utilizó el servidor: PONDR (VLXT) (Romero *et al.* 2001). Se observó que las tres proteínas AtLEA4 tienen una alta propensión a ser proteínas intrínsecamente desordenadas (PIDs o IDPs, del inglés Intrinsically Disordered Proteins) (Figura 8). Interesantemente, se observó la presencia de dos regiones con muy baja probabilidad de ser desordenadas las cuales se localizan dentro de la región N-terminal conservada (Figura 8).

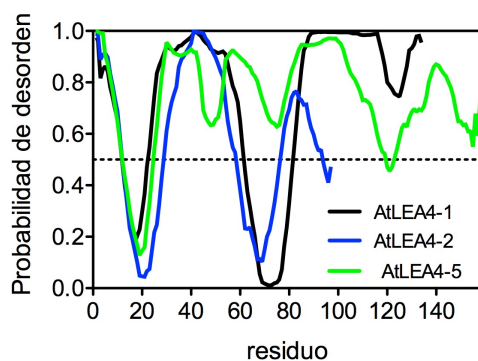


Figura 8. Predicción de los niveles de desorden estructural para la AtLEA4-1 (negro), AtLEA4-2 (azul) y AtLEA4-5 (verde) mediante el servidor PONDR. La línea punteada determina que arriba de ella es muy probable tener regiones intrínsecamente desordenadas.

En el caso de la estructura secundaria, se ha predicho que las proteínas AtLEA4 podrían formar α -hélices (Olvera-Carrillo *et al.* 2010). Con el objeto de analizar la predicción de estructura secundaria, se utilizó el servidor AGADIR (Munoz and Serrano 1994), el cual se ha empleado exitosamente para predecir α -hélices en varias proteínas, las cuales han sido validadas experimentalmente. Este análisis mostró que las tres proteínas poseen predisposición a formar α -hélices y dicha propensión está localizada en la región N-terminal conservada (alrededor de los primeros 80 aminoácidos en las tres proteínas AtLEA4) (Figura 9).

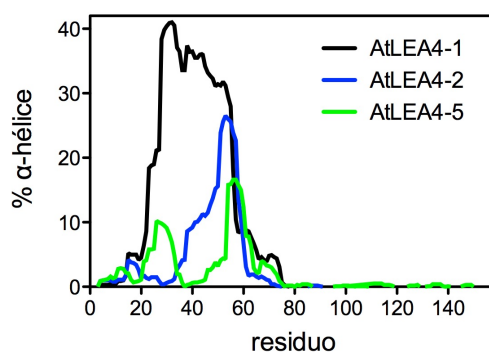


Figura 9. Predicción de α -hélice de la AtLEA4-1 (negro), AtLEA4-2 (azul) y AtLEA4-5 (verde) usando el servidor AGADIR.

En conjunto, las proteínas LEA del grupo 4 tienen un alto grado de similitud, principalmente en la región N-terminal, son hidrofílicas pues están enriquecidas en aminoácidos hidrofílicos y en glicinas, carecen de aminoácidos hidrofóbicos y se

predice que podrían ser proteínas intrínsecamente desordenadas. Se predice que las proteínas LEA4 forman α -hélices en la región N-terminal conservada.

VII.II. Las proteínas AtLEA4-2 y AtLEA4-5 son proteínas intrínsecamente desordenadas en solución

Con el objeto de estudiar las características estructurales de los tres miembros de la familia, se procedió a aislar los marcos de lectura de cada una de las proteínas para insertarlos en los vectores adecuados para su expresión en bacteria. Cada uno de los marcos de lectura de las tres proteínas AtLEA4 se insertaron en el vector pTrc99A (ver la sección Materiales y Métodos), el cual se transformó en células de *E. coli* BL21(DE3) pLysS. Al determinar la inducción de la proteína recombinante por geles de SDS-PAGE y purificarlas, utilizando el método de extracción y purificación de proteínas desordenadas descrito por Campos *et al.* (2011), solo se logró purificar exitosamente la AtLEA4-5. Sorpresivamente, en el patrón de electroforesis por SDS-PAGE se observaron dos bandas discretas en la AtLEA4-5 purificada (Figura 10). Este patrón electroforético ya se había reportado como el resultado de un sitio de inicio de la traducción interno en la AtLEA4-5 (Campos *et al.* 2011). Con el objetivo de eliminar dicho péptido secundario, los primeros tres primeros codones codificantes de metioninas (Met19, Met33 y Met43) internas se mutaron independientemente por leucinas, utilizando la técnica de mutagénesis sitio dirigida conocida como “mega-primer” (Chen *et al.* 2000). De las tres proteínas mutantes obtenidas (Met19Leu, Met33Leu, Met43Leu), sólo la variante Met33Leu mostró una banda discreta con el peso molecular esperado después de purificarse y separarse por SDS-PAGE (Figura 10).

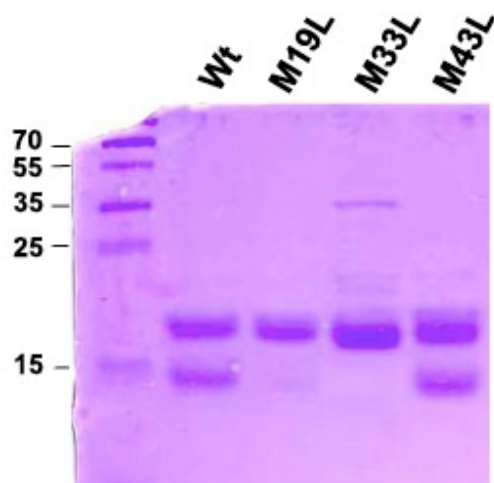


Figura 10. SDS-PAGE de las proteínas recombinantes la AtLEA4-5 (wt) y las mutantes M19L, M33L y M43L. En el primer carril se muestra el marcador de peso molecular.

A partir de este momento, todos los experimentos donde se menciona la AtLEA4-5 fueron realizados con esta mutante. Después del proceso de purificación, la AtLEA4-5 se obtuvo con más del 95% (Figura 11) de pureza y su identidad fue validada usando espectrometría de masas (LC-MS). En la muestra de la proteína purificada se observaron algunas bandas de bajo peso molecular con baja intensidad (Figura 11) y en una proporción aproximada del 5%, las cuales fueron identificadas por LC-MS como productos de degradación de la AtLEA4-5.

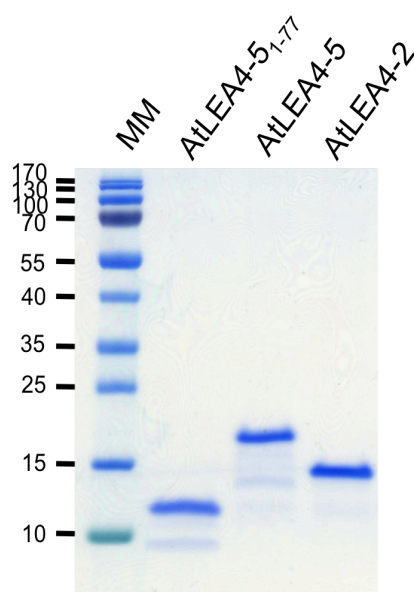


Figura 11. SDS-PAGE de las proteínas recombinantes purificadas que se utilizaron en esta tesis. MM: Marcador de peso molecular.

La obtención de productos de degradación al purificar proteínas intrínsecamente desordenadas es una situación común debido a que éstas, al ser desplegadas, son un blanco fácil y accesible para las proteasas intracelulares, lo cual conduce a un alto nivel de degradación (Tsvetkov *et al.* 2009; Suskiewicz *et al.* 2011). Dicha degradación al parecer fue más prominente en el caso de la AtLEA4-1 y la AtLEA4-2, puesto que nunca se logró observar la inducción ni la purificación de estas proteínas con el mismo método. Por tal motivo, en este caso se utilizó una estrategia diferente, la cual consistió en producir estas proteínas como proteínas recombinantes de fusión con la proteína inteína. La inteína es una proteína que tiene la capacidad de auto-escindir de otra proteína (algo similar a un intrón en un ARN mensajero) cuando se encuentra en un ambiente reductor y, además, tiene un dominio de unión a quitina, la cual se utiliza para purificar la proteína de fusión por cromatografía de afinidad a quitina. Por ello, los marcos de lectura de la AtLEA4-1 y la AtLEA4-2 se clonaron en el vector pTYB11, el cual permite obtener proteínas de fusión a inteína. Las proteínas de fusión se purificaron por medio de cromatografía de afinidad a quitina y se indujo su auto-corte. La AtLEA4-2 se logró obtener con una pureza de más del 95% (Figura 11) pero en el caso de la AtLEA4-1, no fue posible obtener una sola banda discreta. En vez de eso, se obtuvieron una serie de bandas de pesos moleculares similares, las cuales se intentaron separar usando cromatografía de exclusión molecular sin tener éxito (Figura 12).

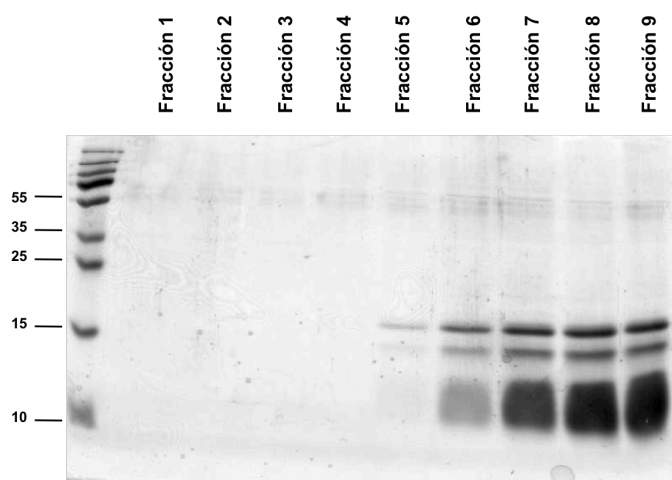


Figura 12. SDS-PAGE de las diferentes fracciones obtenidas después de la purificación de la AtLEA4-1 fusionada a inteína.

Debido a que la AtLEA4-1 y la AtLEA4-2 son las proteínas más conservadas entre sí (Tabla 2) y a que ambas pertenecen al sub-grupo 4A (en contraste con la AtLEA4-5 que pertenece al otro sub-grupo, el 4B), se continuaron todos los experimentos subsecuentes sólo con las proteínas AtLEA4-2 y AtLEA4-5, representantes de los dos sub-grupos.

La cristalografía de rayos X ha sido la técnica más utilizada para estudiar la estructura tridimensional de las macromoléculas, principalmente las proteínas. Dadas las características físicoquímicas de las proteínas LEA hidrofílicas, a la fecha no ha sido posible obtener cristales de proteína de ninguna LEA típica (hidrofílica), por lo que ha sido necesario analizar la estructura con otro tipo de técnicas. Por tal motivo, la estructura secundaria de las proteínas AtLEA4-2 y AtLEA4-5 se analizó usando la técnica de dicroísmo circular (DC) en colaboración con la Dra. Gloria Saab, líder académico del Instituto de Biotecnología de la UNAM.

La técnica de DC consiste en hacer pasar un haz de luz circularmente polarizada por una muestra de proteína pura a diferentes longitudes de onda localizadas en el ultra violeta lejano, de tal forma que, dependiendo de la estructura secundaria de la proteína, se obtendrá un espectro característico de DC. En la Figura 13 se muestran los espectros característicos de estructuras mayoritariamente compuestas por α -hélices, hojas- β y desplegadas (o también conocidas como *giros desordenados* o *random coil*).

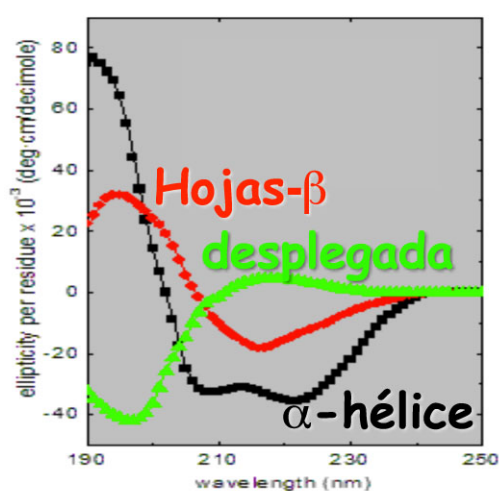


Figura 13. Espectros de dicroísmo circular típicos de estructuras mayoritariamente α -hélices (negro), hojas β (rojo) y desplegadas (verde).

Una vez que se obtuvo el espectro de la proteína de interés, éste se comparó con los espectros tipo de cada una de las estructuras mencionadas anteriormente. A diferencia de la cristalografía de rayos X en donde se obtiene una sola estructura para la proteína en estado sólido, el DC tiene la ventaja de poder analizar la estructura de la proteína en solución y bajo diferentes condiciones aplicadas a dicha solución, por lo que es una técnica adecuada para detectar cambios conformacionales inducidos por cambios por el ambiente de la solución.

Los resultados obtenidos tanto para la AtLEA4-2 como para la AtLEA4-5 demostraron que ambas proteínas son intrínsecamente desordenadas en solución acuosa a diferentes temperaturas (Figura 14A y 14B) y pHs (Figura 14C y 14D); debido a que los espectros obtenidos son muy similares a los de la estructura tipo *random coil* (Comparar Fig 7 con Fig 8). Los espectros tienen el mínimo negativo característico de las PIDs, el cual se localiza alrededor de los 198 nm. Los espectros obtenidos tanto para la AtLEA4-2 como para la AtLEA4-5 tienen un pequeño mínimo con valores negativos alrededor de los 222 nm, el cual es característico de las estructuras tipo α -hélice y además es diferente a lo esperado para las proteínas completamente desplegadas, puesto que éstas presentan valores de elipticidad positivos en dicha región, lo cual sugiere que las proteínas AtLEA4 poseen cierto niveles residuales de estructuras tipo α -hélice. Por otro lado, los espectros diferencia (la sustracción del espectro a menor temperatura del espectro de mayor temperatura) mostraron que la AtLEA4-2 es capaz de formar cierto nivel de α -hélices a una temperatura de 10 °C, en tanto que a temperaturas mayores se mantiene principalmente desordenada.

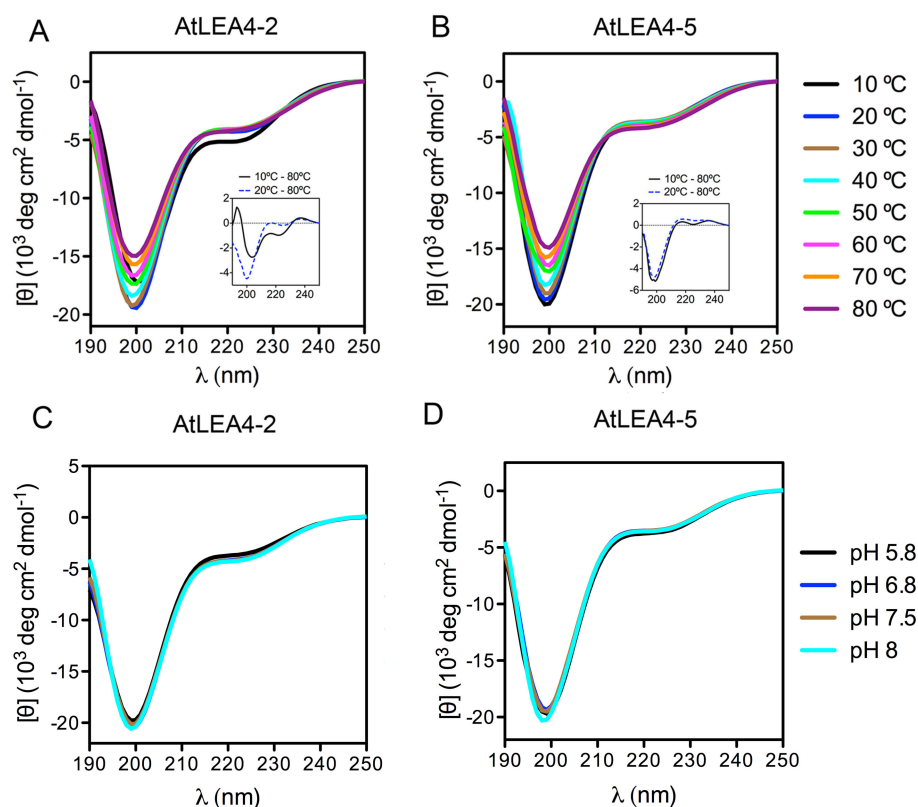


Figura 14. Espectros de DC de la AtLEA4-2 y la AtLEA4-5 a diferentes temperaturas (A y B) y diferentes pHs (C y D). En los recuadros internos en A y B se muestran los espectros diferencia 20°C – 80°C (línea azul discontinua) y 10°C – 80°C (línea negra).

Con el fin de obtener datos cuantitativos de los niveles de estructura de la proteína, se utilizó el servidor DICHROWEB (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>). Este servidor compara los espectros obtenidos con los espectros de otras proteínas cuya estructura se caracterizó empleando otras técnicas. Así, si se ha determinado por cristalografía de proteínas o resonancia magnética nuclear (además del DC) que “X” proteína tiene cierto porcentaje de α -hélice, entonces se conoce el tipo de espectro de DC que tiene la proteína “X” y se puede comparar con la proteína de interés. En el caso de las proteínas AtLEA4, el análisis por DICHROWEB mostró que en solución acuosa la AtLEA4-2 tiene 5% de α -hélices, 33% de hojas β y 61% de desorden. La AtLEA4-5 muestra 5% de α -hélices, 34% de hojas β y 61% de desorden (Tabla 5). Es importante mencionar que en los cálculos de DICHROWEB no se puede diferenciar entre láminas u hojas β verdaderas y estructuras desplegadas tipo poli-L-prolina u otras estructuras desplegadas, por lo que, según los espectros obtenidos en los diferentes resultados de esta tesis, los porcentajes para hojas β se refieren a estructuras

mayoritariamente desplegadas, las cuales se han observado en otras PIDs, incluidas proteínas LEA de otros grupos (Soulages *et al.* 2002; Rivera-Najera *et al.* 2014).

TRATAMIENTO	AtLEA4-2			AtLEA4-5		
	Hélices	Hojas	Desorden	Hélices	Hojas	Desorden
Solución acuosa	5%	33%	61%	5%	34%	61%

Tabla 5. Porcentajes de estructura secundaria de la AtLEA4-2 y la AtLEA4-5 en solución acuosa calculados a partir de los espectros de dicroísmo circular mediante el servidor DICHROWEB.

Además de los estudios por dicroísmo circular, es necesario determinar cuáles son las regiones de la proteína que presentan los distintos arreglos estructurales y relacionarlos con las diferentes zonas conservadas de la proteína. Por ello, utilizamos la técnica de resonancia magnética nuclear (RMN) en colaboración con la Dra. Jane Dyson del Instituto de Investigación Scripps (The Scripps Research Institute, La Jolla, California, USA) y con el Dr. Carlos Amero del Centro de Investigaciones Químicas de la Universidad Autónoma del Estado de Morelos (CIQ, UAEM). La RMN es una técnica muy útil para estudiar y generar modelos de la estructura tridimensional de las proteínas en solución acuosa. La técnica se basa en la capacidad de los espines de los núcleos paramagnéticos (número impar de protones en el núcleo) de resonar en presencia de un campo magnético externo. En el caso de las proteínas, los átomos que se utilizan principalmente son el hidrógeno (^1H), el isótopo estable de nitrógeno (^{15}N) y el isótopo estable de carbono (^{13}C). Para lograr obtener proteínas que contengan dichos isótopos, se deben crecer los cultivos bacterianos en medio mínimos suplementado con fuentes de carbono (glucosa) y nitrógeno (NH_4Cl) marcados con los respectivos isótopos y, posteriormente, inducir la proteína recombinante y purificarla.

Como primer acercamiento, se estudio la AtLEA4-5 por RMN. Para ello, se purificó la proteína recombinante marcada únicamente con ^{15}N (el ^1H es el isótopo natural). En este caso, la proteína recombinante se purificó por cromatografía de intercambio catiónico y, posteriormente, por cromatografía de filtración por tamaño. La proteína se concentró hasta obtener 100 μM y se obtuvo el espectro en dos dimensiones de la AtLEA4-5 utilizando un equipo VNMR-S de 900 MHz a una temperatura de 298 K. El espectro resultante (Figura 15) confirmó que la AtLEA4-5 es una PID puesto que las señales se encuentran en un intervalo muy estrecho de

desplazamiento químico de ^1H (entre 6.8 y 8.6), el cual es característico de las PIDs (Tompa 2012).

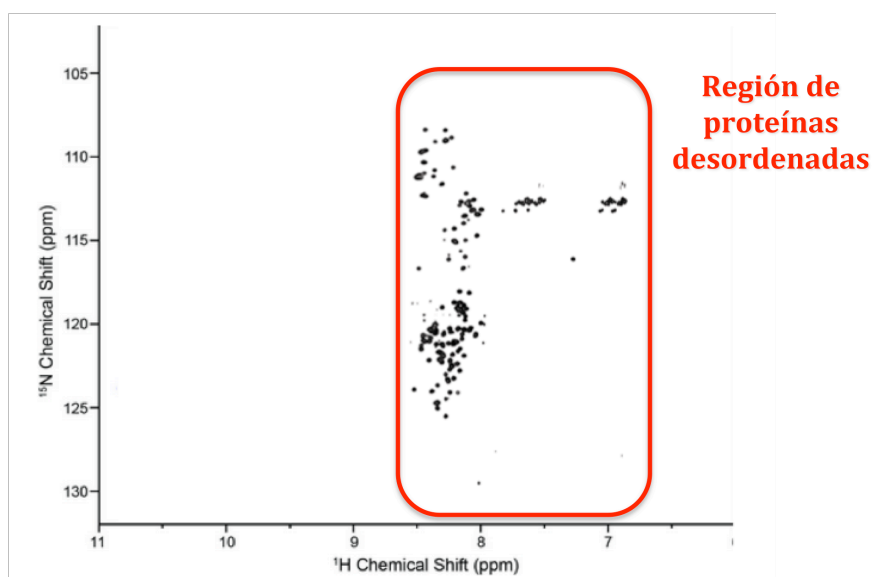


Figura 15. Espectro de resonancia magnética nuclear en dos dimensiones (HSQC) de la AtLEA4-5 marcada con ^{15}N . En el recuadro rojo (donde se localiza la señal) se muestra la región típica donde aparecen las señales de proteínas intrínsecamente desordenadas.

Estos análisis por DC y RMN demostraron que las proteínas LEA del grupo 4 de Arabidopsis son PIDs en solución acuosa y poseen cierto nivel de estructura residual (posiblemente α -hélices).

VII.III. Las proteínas AtLEA4-2 y AtLEA4-5 tienen un alto potencial de formar α -hélices

Una vez demostrado que proteínas LEA del grupo 4 se comportan como proteínas mayoritariamente desordenadas en solución acuosa, con la idea de que las condiciones impuestas por déficit hídrico podrían influenciar la organización estructural de estas proteínas y considerando que el análisis *in silico* predice la formación de α -hélices en estas proteínas; primero, nos preguntamos si estas proteínas realmente tienen la capacidad intrínseca de formar α -hélices. Para ello, se analizó la propensión de formar α -hélices usando el compuesto 2,2,2-trifluoroetanol (TFE) en un análisis por DC. El TFE es un agente que promueve la formación de α -hélices en proteínas que tienen el potencial de formar estas estructuras (Luo and

Baldwin 1997; Buck 1998), por lo que permite obtener información acerca de la capacidad intrínseca de la proteína de adquirir dicha conformación.

Como se puede observar en la Figura 16, el TFE promueve la formación de α -hélices de forma dosis-dependiente (4%, 8%, 12%, 16%, 20%, 40%, 60%, y 90% de TFE) en ambas proteínas, puesto que se observa un incremento progresivo de la $[\theta]_{198}$ hacia valores positivos y un decremento de la $[\theta]_{222}$ hacia valores negativos (Figura 17).

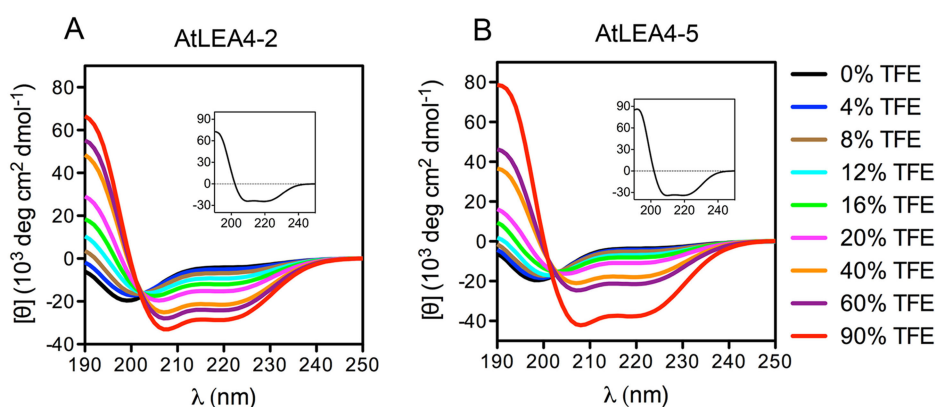


Figura 16. Espectros de dicroísmo circular de la AtLEA4-2 y la AtLEA4-5 a diferentes concentraciones de TFE. En los recuadros internos en A y B se muestran los espectros diferencia ($\Delta 90\% - 0\%$ TFE).

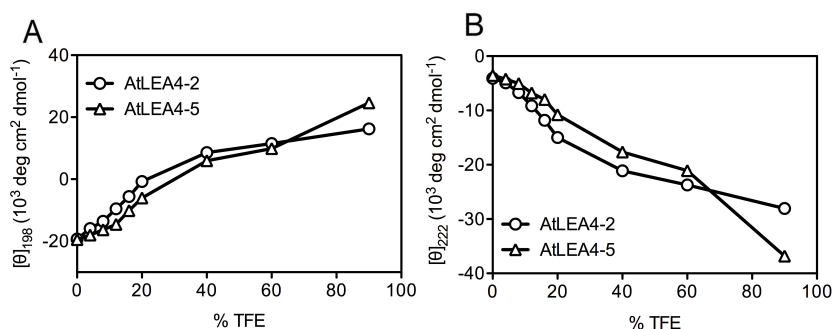


Figura 17. $[\theta]_{198}$ y $[\theta]_{222}$ en función de la concentración de TFE para la AtLEA4-2 (línea con círculos abiertos) y AtLEA4-5 (Línea con triángulos abiertos).

Al comparar la Figura 13 con la Figura 16 se observa claramente como los espectros de ambas proteínas se vuelven progresivamente más parecidos a lo observado para las α -hélices. Si bien los espectros diferencia ($\Delta 90\% - 0\%$ TFE) muestran que ambas proteínas adquieren una mayor proporción de α -hélices, la proteína AtLEA4-5 llega a niveles más altos que la AtLEA4-2 (Figura 16 recuadros). Los cálculos de estructura por DICHROWEB indican que la AtLEA4-2 adopta 7%

de α -hélice en 4% TFE y llega hasta 80% de α -hélice en 90% TFE (Tabla 6) y muestra un decremento de estructura desordenada de 61% en solución acuosa a 16% en 90% TFE (Tabla 6). El contenido de α -hélice de la AtLEA4-5 pasa de 6% a 88% (de 4% a 90% TFE) mientras que los niveles de desorden disminuyen del 61% al 10% (de 4% a 90% TFE).

TRATAMIENTO	AtLEA4-2			AtLEA4-5		
	Hélices	Hojas	Desorden	Hélices	Hojas	Desorden
Solución acuosa	5%	33%	61%	5%	34%	61%
4% TFE	7%	32%	61%	6%	33%	61%
8% TFE	16%	25%	58%	6%	32%	61%
12% TFE	21%	23%	56%	17%	23%	59%
16% TFE	35%	15%	51%	18%	25%	55%
20% TFE	42%	15%	44%	32%	17%	51%
40% TFE	57%	9%	33%	49%	13%	39%
60% TFE	62%	9%	30%	57%	8%	35%
90% TFE	80%	3%	16%	88%	2%	10%

Tabla 6. Porcentajes de estructura secundaria de la AtLEA4-2 y la AtLEA4-5 bajo diferentes concentraciones de TFE calculados a partir de los espectros de dicroísmo circular mediante el servidor DICHROWEB.

De forma interesante, se observó la presencia de un punto isodigráfico (punto sobre el cual todos los espectros cruzan) entre las diferentes concentraciones de TFE para ambas proteínas (Figura 16). El punto isodigráfico de cada proteína se validó al construir diagramas de transición los cuales consisten en graficar la $[\theta]_{198}$ en función de la $[\theta]_{222}$. Al graficar los diferentes puntos de la titulación se obtuvo una regresión lineal lo cual confirma los puntos isodigráficos (Figura 18).

La presencia de estos puntos de cruce indica que existe un equilibrio entre más de un estado conformacional en las proteínas bajo análisis. Si sólo se detecta un punto, como sucedió en este caso, es indicativo de que la proteína se encuentra en una conformación mayoritariamente desordenada, fenómeno que se favorece en solución acuosa (D_{4-2} y D_{4-5}) y, en otra conformación mayoritariamente de tipo α -hélice (H_{4-2} y H_{4-5}), el cual por otro lado, se favorece a altas concentraciones de TFE. El aumento en la concentración de TFE desplaza el equilibrio hacia una conformación helicoidal.

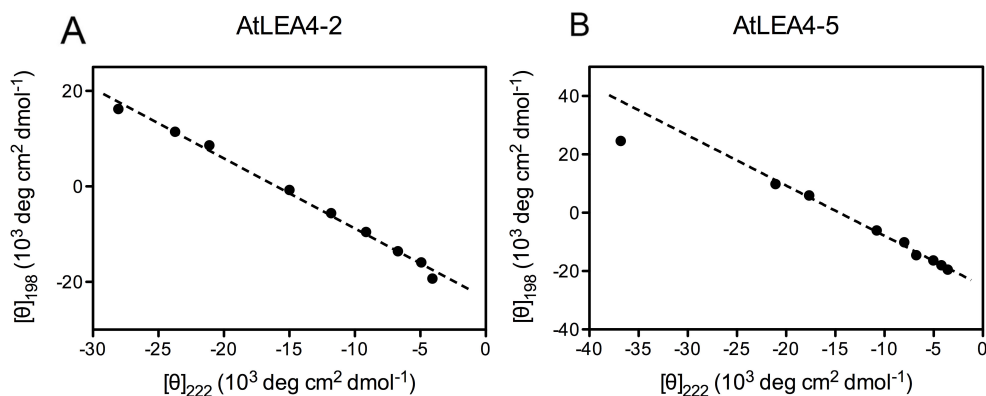


Figura 18. Diagramas de transición (titulación con TFE) para la AtLEA4-2 y la AtLEA4-5. Los diferentes puntos representan los valores obtenidos de $[\theta]_{198}$ y $[\theta]_{222}$ a las diferentes concentraciones de TFE.

Estos resultados demuestran que la AtLEA4-2 y la AtLEA4-5 poseen una habilidad intrínseca a formar α -hélices y que existen dos estados conformacionales que se encuentran en equilibrio, los cuales se favorecen dependiendo de la condición del ambiente ($D_{4-2} \rightleftharpoons H_{4-2}$; $D_{4-5} \rightleftharpoons H_{4-5}$).

VII.IV. La disminución en la cantidad de agua disponible y el amontonamiento macromolecular inducen el plegamiento de las proteínas LEA del grupo 4.

Las proteínas LEA se acumulan abundantemente en la semilla seca así como en tejidos vegetativos en respuesta al déficit hídrico (Delseny *et al.* 2001; Schmid *et al.* 2005; Winter *et al.* 2007; Hruz *et al.* 2008; Olvera-Carrillo *et al.* 2010). En estudios previos con proteínas LEA recombinantes de diferentes grupos y organismos se ha observado que éstas son capaces de adquirir estructura secundaria, principalmente α -hélices, cuando se somete a una deshidratación total (Goyal *et al.* 2003; Tolleter *et al.* 2007; Shih *et al.* 2010; Shimizu *et al.* 2010; Popova *et al.* 2011; Hundertmark *et al.* 2012; Shih *et al.* 2012).

El equilibrio entre los dos estados conformacionales de las proteínas AtLEA4 es un indicio de que probablemente exista una condición en la cual el equilibrio se desplaza hacia la especie más ordenada o estructurada. Además, estas proteínas se acumulan en estados de déficit hídrico menos severos a lo que ocurre en la semilla seca, por lo que nos preguntamos si el déficit hídrico es capaz de inducir su

plegamiento. Para ello, se utilizaron compuestos que, por un lado, simularan la deficiencia de agua disponible y, por otro, de manera independiente, el amontonamiento macromolecular que ocurre bajo una condición de déficit hídrico. De tal manera que, para simular una condición de disminución en el agua disponible, se utilizó un osmolito comúnmente utilizado para tal fin como lo es el glicerol. Este osmolito disminuye considerablemente el potencial hídrico y osmótico de una solución y ha sido exitosamente utilizada para simular esta condición para otras proteínas desordenadas (Mouillon *et al.* 2008; Rivera-Najera *et al.* 2014; Thalhammer *et al.* 2014). Las muestras con cada una de las proteínas se titularon con concentraciones crecientes de glicerol (10%, 20%, 30%, 40%, 50%, 60%, 70% y 80% de glicerol), lo cual disminuyó progresivamente el potencial osmótico y la cantidad de agua disponible. Interesantemente, los espectros resultantes muestran que ambas proteínas adquieren estructura tipo α -hélice cuando se simula la pérdida de agua *in vitro* (Figura 19), puesto que los valores de $[\theta]_{198}$ se vuelven positivos y los valores de $[\theta]_{222}$ se vuelven más negativos (Figura 20).

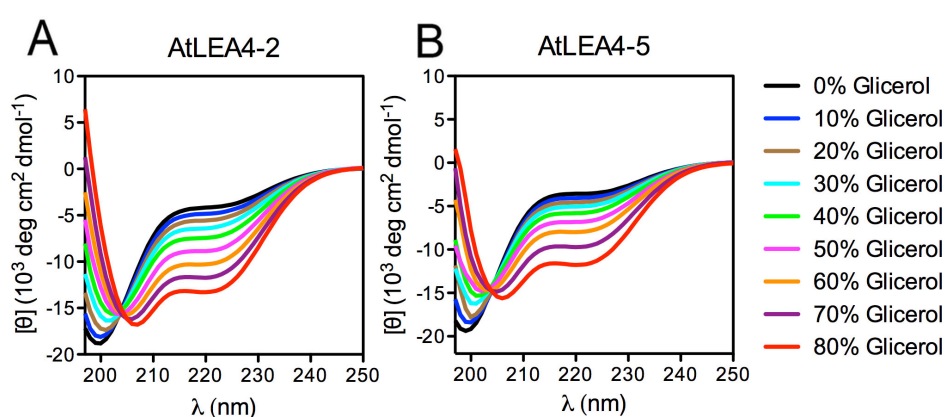


Figura 19. Espectros de dicroísmo circular de la AtLEA4-2 y la AtLEA4-5 a diferentes concentraciones de glicerol.

El análisis por DICHROWEB muestra que existe una pequeña diferencia entre los niveles de α -hélice que adquieren las dos proteínas a la mayor concentración de glicerol (54% para la AtLEA4-2 y 46% para la AtLEA4-5; Tabla 7). En el espectro con la concentración más alta de glicerol (80%) se determinó que la AtLEA4-2 posee 54% de α -hélices, 12% de hojas β y 33% de desorden (Tabla 7), mientras que la AtLEA4-5 posee 46% de α -hélices, 9% de hojas β y 45% de desorden (Tabla 6), lo cual es un aumento notable si se compara con las muestras que no contienen glicerol

(AtLEA4-2: 5% α -hélices, 33% hojas β , 61% desorden; AtLEA4-5: 5% α -hélices, 34% hojas β , 61% desorden).

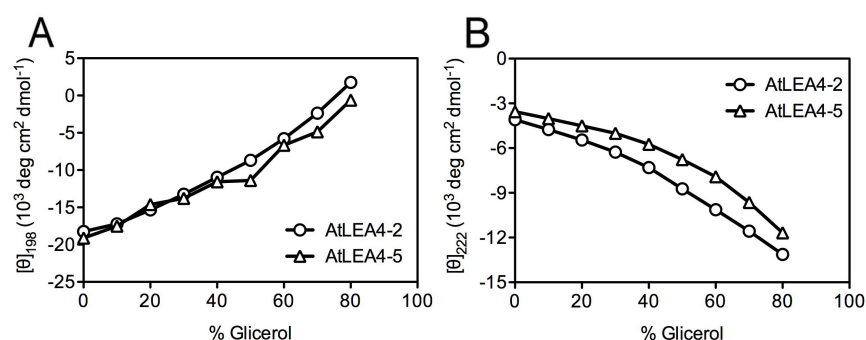


Figura 20. $[\theta]_{198}$ y $[\theta]_{222}$ en función de la concentración de glicerol para la AtLEA4-2 (línea con círculos abiertos) y AtLEA4-5 (Línea con triángulos abiertos).

TRATAMIENTO	AtLEA4-2			AtLEA4-5		
	Hélices	Hojas	Desorden	Hélices	Hojas	Desorden
Solución acuosa	5%	33%	61%	5%	34%	61%
10% Glicerol	5%	34%	60%	4%	34%	61%
20% Glicerol	8%	32%	61%	7%	36%	57%
30% Glicerol	18%	22%	60%	7%	34%	59%
40% Glicerol	21%	23%	57%	18%	24%	57%
50% Glicerol	33%	12%	53%	18%	23%	59%
60% Glicerol	43%	14%	43%	34%	15%	51%
70% Glicerol	49%	9%	42%	42%	11%	48%
80% Glicerol	54%	12%	33%	46%	9%	45%

Tabla 7. Porcentajes de estructura secundaria de la AtLEA4-2 y la AtLEA4-5 bajo diferentes concentraciones de glicerol calculados a partir de los espectros de dicroísmo circular mediante el servidor DICHROWEB.

Interesantemente, la titulación con glicerol también reveló la presencia de un punto isodicroico en ambas proteínas (Figura 21), como sucedió en la titulación con TFE. Además, estos datos nos permiten inferir que la vía de plegamiento que siguen las proteínas AtLEA4 cuando se disminuye la cantidad de agua disponible es la misma que siguen cuando se adiciona TFE, puesto que todos los puntos comparten la misma regresión lineal en el diagrama de transición (Figura 21).

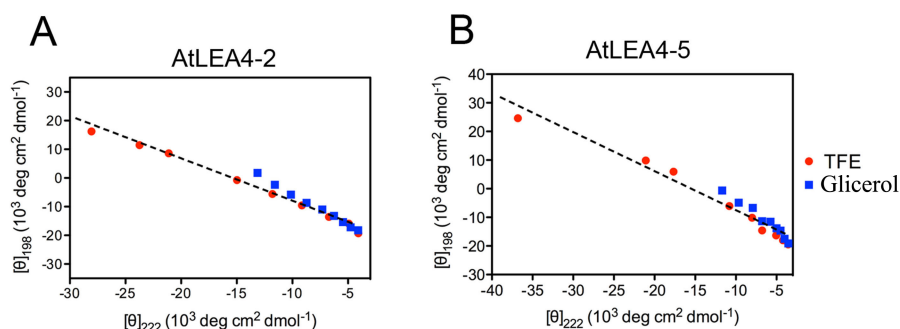


Figura 21. Diagramas de transición (titulación con TFE y glicerol) para la AtLEA4-2 y la AtLEA4-5. Los diferentes puntos representan los valores obtenidos de $[\theta]_{198}$ y $[\theta]_{222}$ a las diferentes concentraciones de TFE (rojo) y glicerol (azul).

En el interior de las células vivas, la concentración de macromoléculas es muy alta (~400 g/L) (Zimmerman and Trach 1991; Zimmerman and Minton 1993; Ellis 2001). A esta condición se le conoce como amontonamiento macromolecular (Minton 1997; Minton 2000). Dicha condición se exagera aún más en condiciones de déficit hídrico, llegando a alcanzar niveles de hasta 900 g/L (Ellis 2001). Con el objetivo de estudiar cual es el efecto del amontonamiento macromolecular sobre la estructura de las proteínas AtLEA4, se utilizó polietilenglicol (PEG-5000), un polímero comúnmente utilizado para simular esta condición *in vitro* (Mouillon *et al.* 2008; Szasz *et al.* 2011; Qin and Zhou 2013; Goldenberg and Argyle 2014). La adición de PEG-5000 al 45% indujo un cambio conformacional hacia α -hélices en ambas proteínas (Figura 22).

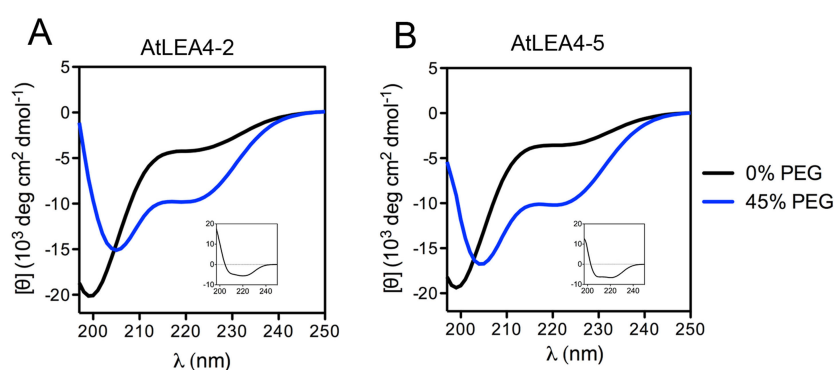


Figura 22. Espectros de dicroísmo circular de la AtLEA4-2 y la AtLEA4-5 en solución acuosa (negro) y bajo 45% de PEG-5000 (azul).

Las estimaciones por DICHROWEB indican que la AtLEA4-2 y la AtLEA4-5 adquieren 37% y 39% de α -hélice, respectivamente (Tabla 8).

TRATAMIENTO	AtLEA4-2			AtLEA4-5		
	Hélices	Hojas	Desorden	Hélices	Hojas	Desorden
Solución acuosa	5%	33%	61%	5%	34%	61%
45% PEG-5000	37%	11%	52%	39%	9%	53%

Tabla 8. Porcentajes de estructura secundaria de la AtLEA4-2 y la AtLEA4-5 bajo 45% PEG calculados a partir de los espectros de dicroísmo circular mediante el servidor DICHROWEB.

En conclusión, las proteínas AtLEA4-2 y AtLEA4-5 se pliegan formando estructuras α -hélices cuando se simula la disminución en la cantidad de agua disponible y el amontonamiento macromolecular; posiblemente, reflejando lo que ocurre en las células vegetales *in vivo*.

VII.V. La región N-terminal conservada de la AtLEA4-5 es necesaria y suficiente para inducir el plegamiento inducido por la disminución en la cantidad de agua disponible y el amontonamiento macromolecular simulados in vitro.

Las proteínas LEA del grupo 4 se caracterizan por la presencia de motivos conservados localizados en la región N-terminal (Battaglia *et al.* 2008; Olvera-Carrillo *et al.* 2010). Además, estas proteínas contienen una región C-terminal que es variable en secuencia y longitud (Figura 7). Como se vio anteriormente, se ha predicho *in silico* que la región N-terminal tiene una alta predisposición a formar estructuras tipo α -hélices, mientras que la región C-terminal se predice como altamente desordenada (Figura 9). Con el objetivo de probar estas predicciones, procedimos a analizar la estructura secundaria de estas regiones. Para ello, utilizamos a la AtLEA4-5, la cual contiene la región C-terminal más larga entre los tres miembros de esta familia. La proteína se separó en dos partes, la región N-terminal (residuos 1-77) y la región C-terminal (78-158). Los marcos de lectura que codifican para cada una de las regiones se clonaron en el vector pTrc99A, se transformaron en cepas de *E. coli* BL21 DE3 (pLysS) y, finalmente, se purificaron las proteínas recombinantes (Ver Materiales y métodos). De esta manera, se produjo de manera exitosa (más del 95% de pureza) la región N-terminal de la AtLEA4-5, la cual se nombró AtLEA4-5₁₋₇₇ (Figura 11). Dado que la región C-terminal no se pudo expresar usando este vector, se procedió a clonar e intentar purificar como fusión a inteína usando el vector pTYB11 (como se realizó exitosamente para la AtLEA4-2). Desafortunadamente, no fue posible purificar la región C-terminal, probablemente

debido a los altos niveles de desorden predichos para este polipéptido lo que la hace un buen sustrato para diferentes proteasas en la célula (similar a lo que pasó con la AtLEA4-1). Por tal motivo, el péptido de 81 residuos de la región C-terminal se sintetizó químicamente utilizando los servicios de una compañía especializada en las metodologías para este proceso (BIOMATIK). Dicho péptido se sintetizó exitosamente, y se purificó por HPLC. Así mismo, la compañía que lo manufacturó corroboró su identidad por espectrometría de masas. A este polipéptido se le denominó AtLEA4-5₇₈₋₁₅₈.

La estructura secundaria de ambas proteínas se analizó por DC. Contrario a las predicciones *in silico*, la estructura de la AtLEA4-5₁₋₇₇ es intrínsecamente desordenada en solución acuosa a diferentes temperaturas con el mínimo negativo alrededor de los 198 nm (Figura 23A), comportándose de manera general como lo que se observó para la AtLEA4-5 completa (ver los espectros diferencia, recuadros de la Figura 23A y Figura 14B). Estas observaciones fueron validadas por resonancia magnética nuclear pues el espectro en dos dimensiones de la AtLEA4-5₁₋₇₇ se localiza en la región de las proteínas intrínsecamente desordenadas (Figura 24).

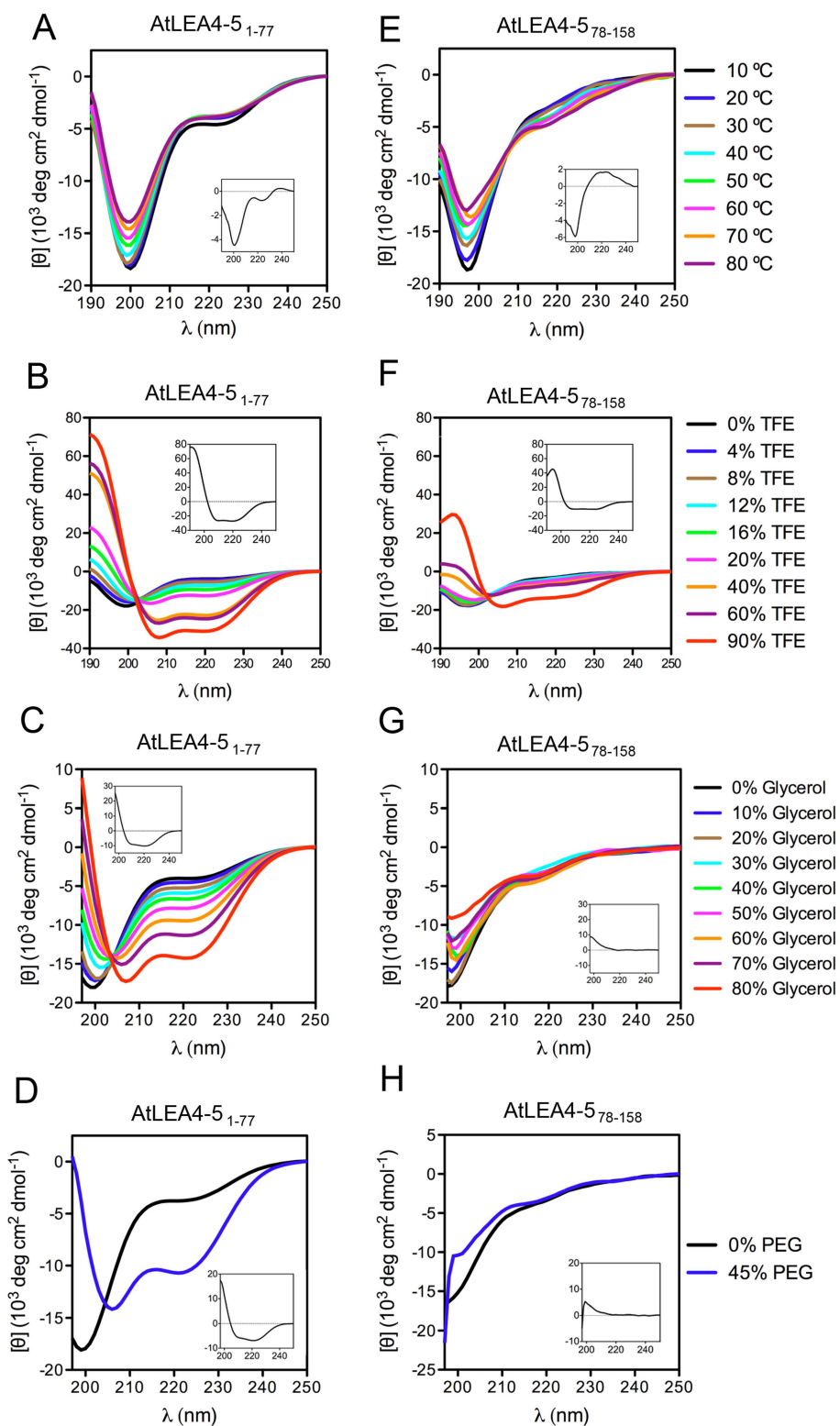


Figura 23. Espectros de dicroísmo circular de la AtLEA4-5₁₋₇₇ y la AtLEA4-5₇₈₋₁₅₈ en solución acuosa (A y E), bajo diferentes concentraciones de TFE (B y F), bajo diferentes concentraciones de glicerol (C y G) y en 45% PEG (D y H). Los recuadros internos indican los espectros diferencia de la mayor concentración del compuesto menos la menor.

Por otro lado, como sucedió para la AtLEA4-5 completa, la adición de TFE, glicerol o PEG promueve el plegamiento en α -hélice de la AtLEA4-5₁₋₇₇ (Figura 23B, 23C y 23D). Además, en las titulaciones con TFE y glicerol se observó la presencia de un punto isodicroico en cada una, sugiriendo que esta región sigue una vía de plegamiento similar al de la proteína completa (Figura 23B, 23C). El punto isodicroico que se evidencia tras la titulación con glicerol fue validado cuando se graficó el diagrama de transición correspondiente (Figura 25).

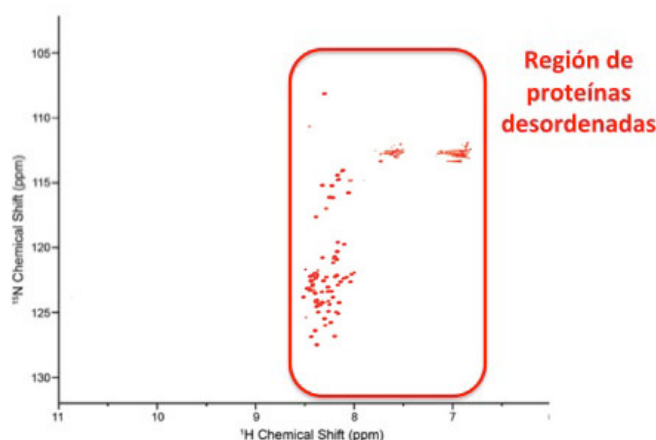


Figura 24. Espectro de resonancia magnética nuclear en dos dimensiones (HSQC) de la AtLEA4-5₁₋₇₇ marcada con ¹⁵N. En el recuadro rojo (donde se localiza la señal) se muestra la región típica donde aparecen las señales de proteínas intrínsecamente desordenadas.

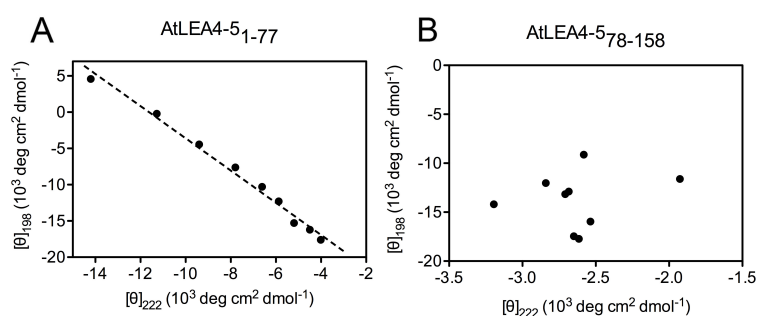


Figura 25. Diagramas de transición (titulación con glicerol) para la AtLEA4-5₁₋₇₇ y la AtLEA4-5₇₈₋₁₅₈. Los diferentes puntos representan los valores obtenidos de $[\theta]_{198}$ y $[\theta]_{222}$ a las diferentes concentraciones de glicerol.

De forma similar, la AtLEA4-5₇₈₋₁₅₈ (región C-terminal) también se comporta como una proteína desordenada en solución acuosa; sin embargo, contrario a la región N-terminal y a la proteína completa, cuando se analiza el comportamiento de los

espectros conforme se aumenta la temperatura, se observa un punto isodicroico, lo cual es una señal típica de las estructuras extendidas poli-L-prolina tipo II (recuadro interno, Figura 23E), que se han observado para otras proteínas LEA de diferentes grupos (Soulages *et al.* 2002; Soulages *et al.* 2003; Abba' *et al.* 2006). Interesantemente, la AtLEA4-5₇₈₋₁₅₈ no se pliega a bajas concentraciones de TFE (sólo hasta que se alcanza 90% de TFE; Figura 23F), ni cuando se adiciona glicerol o PEG-5000 (Figura 23G y 23H). La diferencia en los niveles de plegamiento entre ambas regiones se hace aún más notoria cuando se comparan los espectros de diferencia de cada una de las titulaciones (recuadros internos Figura 23). Además, la AtLEA4-5₇₈₋₁₅₈ no tiene un punto isodicroico en la titulación con glicerol (Figura 23 G). La diferencia se hace evidente cuando se comparan los diagramas de transición de las titulaciones con glicerol de ambas regiones (Figura 25).

En conclusión, con estos resultados se demostró que la región N-terminal conservada de la AtLEA4-5 es necesaria y suficiente para inducir el plegamiento en α -hélice causado por la simulación de la cantidad de agua disponible *in vitro*.

VII.VI. La región N-terminal conservada de la AtLEA4-5 es necesaria y suficiente para prevenir la pérdida de actividad de la enzima lactato deshidrogenasa causada por deshidratación o por ciclos de congelamiento/descongelamiento in vitro.

Las proteínas LEA se descubrieron hace ya más de 30 años; sin embargo, a la fecha, no se ha logrado demostrar contundentemente su función ni su mecanismo de acción molecular. Existen diversas hipótesis fundamentadas en ensayos *in vitro* utilizando proteínas LEA recombinantes. Se ha reportado que las proteínas LEA son capaces de unir cationes divalentes, funcionando como amortiguadores de estos en condiciones de estrés (Battaglia *et al.* 2008). Algunas proteínas LEA pueden unir lípidos y ADN lo cual sugiere que podrían proteger estructuras celulares y macromoléculas cuando se experimenta la pérdida de agua (Tompa and Kovacs 2010). Además, se ha observado que en general, las proteínas LEA pueden proteger a otras proteínas funcionando de manera similar a las chaperonas moleculares (Reyes *et al.* 2005; Reyes *et al.* 2008). Con el objetivo de obtener información referente a la función de las proteínas LEA del grupo 4, se realizaron ensayos de protección de la actividad de la enzima lactato deshidrogenasa (LDH), una enzima sensible a

condiciones de deshidratación y congelamiento/descongelamiento (condición que también representa pérdida de agua pues al congelarse la solución, la cantidad de agua disponible para la enzima disminuye drásticamente) (Reyes *et al.* 2005; Reyes *et al.* 2008). Para los ensayos de congelamiento/descongelamiento, se sometieron muestras de LDH (250 nM) a congelamiento en nitrógeno líquido por 30 segundos y, posteriormente, se descongelaron a 25°C durante 5 minutos. Este tratamiento representa un ciclo, el cual se repitió siete veces. Una vez terminado el tratamiento, se tomó una alícuota de la muestra para medir su actividad enzimática. Como se puede ver en la Figura 26A, el tratamiento abatió prácticamente toda la actividad de la LDH. Sin embargo, en una serie de muestras en donde se agregó (antes de iniciar el tratamiento), ya sea la AtLEA4-2 o la AtLEA4-5 a la muestra con LDH y, posteriormente, se sometieron al tratamiento, se observó que éstas fueron capaces de evitar la pérdida de actividad de la LDH de forma dosis-dependiente (Figura 26A). Cabe señalar que en estos experimentos, las proteínas LEA del grupo 4 fueron capaces de ejercer el efecto protector a relaciones molares muy bajas (alrededor de una relación molar 5:1). Interesantemente, al agregar la AtLEA4-5₁₋₇₇ se obtuvieron niveles de protección similares a los observados para la AtLEA4-5 completa, mientras que la AtLEA4-5₇₈₋₁₅₈ no mostró ningún efecto protector y, más bien, se comportó como la lisozima, una proteína globular de peso molecular similar que no tiene efecto protector reportado (Figura 26A).

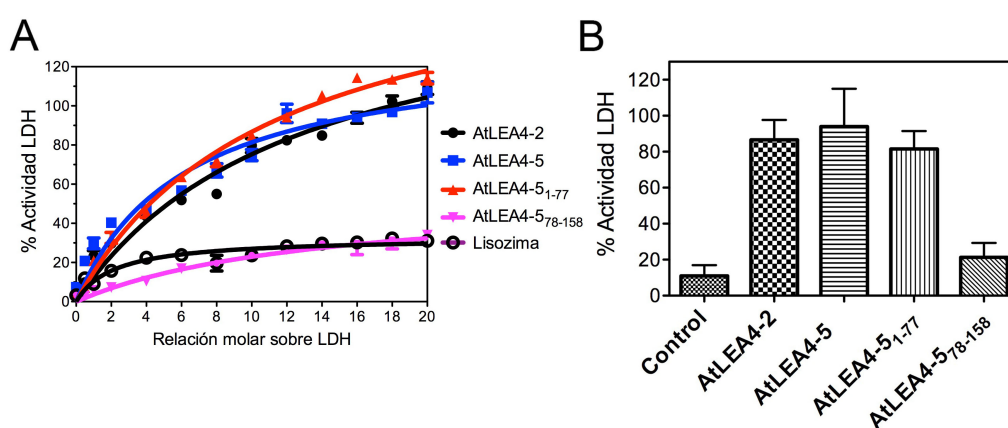


Figura 26. Actividad enzimática de la LDH sometida a (A) 7 ciclos de congelamiento/descongelamiento y (B) más del 98% de deshidratación en ausencia (relación molar sobre LDH = 0 y control) o en presencia las diferentes proteínas analizadas en esta tesis (En el caso de (A), se usaron diferentes relaciones molares de las proteínas sobre LDH y en (B) se usó una relación 5:1 (proteína prueba:LDH)). La lisozima es una proteína sin actividad protectora reportada.

Además, este efecto protector también se observó en muestras sometidas a deshidratación. Brevemente, muestras que contenían LDH, sola o en presencia de las distintas proteínas usadas en el ensayo anterior, se deshidrataron en un concentrador Savant (por evaporación causada por vacío, sin calentar y manteniendo una temperatura constante) hasta alcanzar porcentajes de pérdida de agua de más del 98%. Como se puede ver en la Figura 26B, la actividad de la LDH se anula cuando se llega a esos niveles de pérdida de agua; sin embargo, la adición de la AtLEA4-2, AtLEA4-5 y AtLEA4-5₁₋₇₇ (en una relación molar 5:1) mantienen la actividad cerca del 100%. De la misma forma que en los ensayos de congelamiento/descongelamiento, la AtLEA4-5₇₈₋₁₅₈ no tiene ningún efecto protector.

En conclusión, con estos resultados se demostró que la región N-terminal conservada de la AtLEA4-5 es necesaria y suficiente para ejercer un efecto de protección de la LDH sometida a tratamientos de pérdida de agua *in vitro*, sugiriendo que los elementos que se localizan en la región N-terminal (los motivos conservados y/o la capacidad de formar α -hélices cuando se disminuye la cantidad de agua disponible) son los responsables de la función *in vitro* de estas proteínas.

VIII. DISCUSIÓN

Durante los últimos 20 años, la investigación de las proteínas intrínsecamente desordenadas ha desafiado el paradigma estructura-función que establece que para que una proteína sea funcional, debe plegarse en una forma tridimensional bien definida y estable (Wright and Dyson 2015). Las proteínas intrínsecamente desordenadas (así como las regiones desordenadas contenidas en proteínas mayoritariamente globulares), en contraste con este paradigma son altamente abundantes en los proteomas eucariontes y ejecutan funciones muy importantes en dichos organismos (Tompa 2012). En plantas, se ha visto que las proteínas intrínsecamente desordenadas participan en el control del desarrollo, la percepción de la luz, la regulación transcripcional y la respuesta a estrés abiótico (Sun *et al.* 2013). Las proteínas LEA son el grupo más grande de proteínas intrínsecamente desordenadas de plantas puesto que para muchas de ellas se ha predicho u observado experimentalmente que presentan altos niveles de desorden (Sun *et al.* 2013). Existe evidencia que muestra que las proteínas LEA de diferentes grupos, incluyendo dos del grupo 4 de soya (GmPM1 y GmPM28) adquieren estructuras tipo α -hélice después de

deshidratarlas completamente (Popova *et al.* 2011; Hundertmark *et al.* 2012; Shih *et al.* 2012). Estos resultados sugieren que dichas proteínas podrían formar α -hélice en la semilla seca. Sin embargo, se sabe muy poco acerca de las propiedades estructurales de las proteínas LEA bajo condiciones de deshidratación menos severa. Mouillon *et al.* (2008) reportaron que tres proteínas LEA del grupo 2 (Cor47, Lti29 y Lti30) son proteínas intrínsecamente desordenadas en solución acuosa y se mantienen desordenadas aún cuando se someten *in vitro* a un bajo potencial de agua o a un amontonamiento macromolecular (simulados con glicerol y PEG, respectivamente) (Mouillon *et al.* 2008). Aún cuando se observó un ligero plegamiento en α -hélices en altas concentraciones de glicerol y PEG, los autores proponen que las deshidrinas (LEAs del grupo 2) han evolucionado para permanecer desordenadas en las condiciones celulares y que esta conformación puede ser necesaria para realizar su función (Mouillon *et al.* 2008). Se han encontrado resultados similares experimentalmente para tres diferentes IDPs de animales (α -caseína, MAP2c y p21^{Cip1}), puesto que no se observó ningún cambio conformacional al simular amontonamiento macromolecular, sustentando la idea de que el estado fisiológico de las proteínas intrínsecamente desordenadas sigue siendo mayoritariamente desplegado en este tipo de condiciones (Szasz *et al.* 2011). Contrario a lo observado en esos reportes, las proteínas LEA del grupo 4 de Arabidopsis analizadas en este trabajo (AtLEA4-2 y AtLEA4-5) son capaces de plegarse significativamente en condiciones de baja disponibilidad de agua y amontonamiento macromolecular simulado *in vitro*. Es importante remarcar que ambas proteínas alcanzaron altos niveles de α -hélice (hasta 54%), considerablemente más alto que lo observado para otras proteínas analizadas bajo tratamientos similares, o incluso más alto que lo reportado después de deshidratarlas completamente (Mouillon *et al.* 2008; Popova *et al.* 2011; Sun *et al.* 2013; Rivera-Najera *et al.* 2014). El hecho de que tres proteínas LEA del grupo 4 de soya también adquieren niveles elevados de α -hélice con la adición de SDS y TFE; así como, después de deshidratarlas completamente, sustenta los hallazgos de este trabajo (Shih *et al.* 2004; Shih *et al.* 2010). Además, la vía de plegamiento observada en la titulación con glicerol para la AtLEA4-2 y la AtLEA4-5 se asemeja a la obtenida a concentraciones intermedias de TFE (hasta 20%), donde se observa la presencia de un punto isodicroico que revela la existencia de un equilibrio entre dos estados conformacionales bien definidos. El uso de titulaciones de TFE para encontrar

evidencia de los distintos estados conformacionales de las proteínas intrínsecamente desordenadas no es nuevo. Uno de los ejemplos es la α -sinucleína, una proteína intrínsecamente desordenada de humanos muy bien caracterizada. Para esta proteína, la titulación con TFE reveló la presencia de tres estados conformacionales (desplegado, intermediario y plegado) en equilibrio (Anderson *et al.* 2010). De manera interesante, se observó que la formación de fibrillas enriquecidas en hojas β se iniciaba específicamente por la especie conocida como intermediario (Anderson *et al.* 2010). En la mismo orden de ideas, recientemente se ha reconocido que algunas proteínas intrínsecamente desordenadas usan elementos de estructura secundaria pre-existentes o pre-formados, conocidos como motivos pre-estructurados (PreSMos, por su nombre en inglés), para reconocer y unir sus blancos de interacción (Lee *et al.* 2012). Las proteínas intrínsecamente desordenadas están compuestas por diferentes estados conformacionales en equilibrio que podrían contener PreSMos particulares y, por lo tanto, modular la unión con blancos específicos dependiendo de la presencia de cada uno de estos motivos. La plasticidad conformacional de esta clase de proteínas ofrece muchas ventajas comparada con las proteínas globulares; por ejemplo la promiscuidad de unión que es particularmente deseable para proteínas con capacidad tipo chaperonas, justo como las proteínas LEA. Por lo tanto, es posible que las proteínas LEA del grupo 4 necesiten formar PreSMos específicos para poder encontrar ciertos blancos y así ejecutar adecuadamente su función. Esta hipótesis lleva a pensar que la presencia o ausencia de los diferentes PreSMos en las proteínas LEA estaría modulada por el estatus hídrico o alguna otra condición de estrés que se presente en las células. El determinar si realmente se requiere la presencia de diferentes conformeros (desordenados o plegados) de las proteínas LEA del grupo 4 para desarrollar diferentes funciones en diferentes niveles de déficit hídrico es algo que desconoce y que representa un reto que habrá que abordar para conocer más sobre la relación entre la estructura y la función de estas proteínas.

Las proteínas LEA del grupo 4 se caracterizan por la presencia de una región N-terminal conservada de entre 70 y 80 aminoácidos aproximadamente. En este trabajo, se demostró que la región N-terminal conservada es necesaria y suficiente para inducir el plegamiento en α -hélices en respuesta a déficit hídrico impuesto *in vitro*, sugiriendo que la plasticidad estructural de esta región se ha seleccionado a través de la evolución en este tipo de proteínas. La región N-terminal conservada está

enriquecida en aminoácidos cargados (lisina y ácido glutámico) y pequeños no polares (alanina y metionina). Esta composición de aminoácidos es, por lo tanto, apropiada para la formación de α -hélices anfipáticas. En un artículo reciente, Thalhammer *et al.* (2014) encontraron que cuando dos proteínas tipo LEA de *Arabidopsis*, COR15A y COR15B, se tratan con 50% glicerol, éstas se pliegan en α -hélices anfipáticas y, que la formación de estructura secundaria se incrementa cuando se adicionan liposomas de membranas, aparentemente a través de un plegamiento acoplado a la interacción con membranas (Thalhammer *et al.* 2014). Dado que las membranas por sí solas no indujeron ningún cambio conformacional, se sugirió que la orientación de algunos aminoácidos hidrofóbicos en la α -hélice anfipática (pre-formada por la adición de glicerol) es necesaria para la interacción proteína-membrana lo que incrementa y estabiliza el plegamiento (Thalhammer *et al.* 2014). Dado que las proteínas LEA del grupo 4 podrían formar este tipo de α -hélices, no se debe descartar la posible interacción con membranas y un plegamiento acoplado a esta unión. De hecho, previamente se reportó que la AtLEA4-2 es capaz de interactuar con membranas en el estado deshidratado (Hundertmark *et al.* 2011). Será necesario demostrar si la región N-terminal conservada de la AtLEA4-2 es la responsable de dicha posible interacción con las membranas. En otros grupos de proteínas LEA se ha observado que los motivos conservados son en realidad los que tienen la capacidad de plegarse en respuesta a la adición de distintos tratamientos. Por ejemplo, dentro de las proteínas LEA del grupo 2 (o dehidrinas), los segmentos conservados K son los responsables del plegamiento cuando se unen membranas (Mouillon *et al.* 2008). En el caso de las proteínas LEA del grupo 3, el 11-amero que define a la familia es el que se pliega cuando se deshidrata completamente (Chakrabortee *et al.* 2012). En este trabajo demostramos que la región conservada en el grupo 4 de proteínas LEA también es la responsable del plegamiento asociado a un tratamiento, en este caso al déficit hídrico. La importancia de la región N-terminal conservada cobra más fuerza cuando se analiza su función *in vitro*. En estudios previos se había demostrado que la AtLEA4-5 era capaz de prevenir la inactivación de enzimas modelo como lactado deshidrogenasa (LDH) y malato deshidrogenasa (MDH) ante condiciones de deshidratación y en ciclos de congelamiento/descongelamiento *in vitro* a una relación molar tan baja como 1:1 (Reyes *et al.* 2005; Reyes *et al.* 2008). En el presente trabajo, se profundizó aún más

en el mecanismo molecular de este grupo de proteínas. El hallazgo más importante fue que la región N-terminal conservada de la familia es necesaria y suficiente para ejercer el efecto protector en los dos tipos de ensayos *in vitro*, lo que sugiere que el plegamiento en α -hélices es fundamental para la función de estas proteínas. Por otro lado, el hecho de que la protección se dio a relaciones molares muy bajas favorece un mecanismo de protección tipo chaperona y va en contra de la propuesta de protección por escudo molecular (Chakrabortee *et al.* 2012). Recientemente, se propuso que las proteínas LEA ejercen su efecto protector a través del escudo molecular y que sólo es necesario tener polipéptidos hidrofílicos, desordenados y con un radio hidrodinámico grande para proteger adecuadamente (Hughes *et al.* 2013). Sin embargo, en un estudio más exhaustivo se encontró que no existe una correlación directa entre el radio hidrodinámico y los niveles de protección (Cuevas-Velazquez *et al.* 2014). En este trabajo se encontró que la AtLEA4-5 (158 residuos), la AtLEA4-2 (97 residuos) y la AtLEA4-5₁₋₇₇ (77) tienen niveles similares de protección a pesar de su diferencia en tamaño. Además se observó que la hidrofílicidad y el desorden no fueron suficientes para la adecuada protección *in vitro* pues la AtLEA1 y la AtLEA4-5₇₈₋₁₅₈ (proteínas más hidrofílicas y desordenadas que las que contienen a la región N-terminal conservada) tuvieron efectos protectores muy bajos o nulos comparados con las proteínas que sí protegen (Campos *et al.* 2013). Así pues, aunque en general, los datos favorecen la idea de que éstas proteínas actúan por interacción con sus proteínas *blanco* o *clientes*, como ocurre para otras chaperonas, no es posible descartar totalmente la posibilidad de que tanto este mecanismo como la formación de un escudo proteico alrededor de la proteína *cliente* funcionen como mecanismos protectores alternativos dependiendo de condiciones como la concentración local, tanto de la proteína LEA como de las o las proteínas clientes, del nivel de agua disponible, del amontonamiento macro-molecular, o bien de otras condiciones físicas o químicas del medio. En el caso de las proteínas LEA del grupo 4 se han reunido diferentes evidencias que indican que éstas protegen a través de un mecanismo de acción tipo chaperona que involucra una interacción proteína-proteína.

Cabe mencionar que las evidencias obtenidas en este trabajo fueron encontradas en un sistema *in vitro*, lo cual podría ser muy diferente a lo que sucede en las células vegetales. La obtención de datos de la forma (estructura) de las proteínas a una escala atómica siempre ha mostrado un enfoque reduccionista, en donde una proteína “X” se estudia de forma aislada; ya sea, en un cristal (cristalografía de rayos

X) o en solución (RMN), pero siempre en sistemas *in vitro*. En esta tesis, tratamos de aproximarnos lo más posible al ambiente celular en donde se acumulan estas proteínas LEA; es decir, amontonamiento macromolecular y bajo potencial hídrico. Por otro lado, existen otros enfoques cuya intención es observar que es lo que sucede con la estructura de las proteínas directamente dentro de las células. Así, la técnica de fluorescencia por transferencia de energía (FRET, *Fluorescence Resonance Energy Transfer*) es capaz de detectar cambios conformacionales en una proteína que está fusionada en sus extremos a dos proteínas fluorescentes (Jones *et al.* 2014). Si la estructura de la proteína de interés se modifica, lo mismo ocurre con la eficiencia de transferencia de energía (fluorescencia). La desventaja principal de esta técnica es que sólo se puede determinar que la proteína en cuestión experimentó cambios conformacionales, pero no nos dice el tipo de cambio que ocurrió y, mucho menos, da detalles a nivel atómico. Una técnica que tiene la capacidad de dar detalles a nivel atómico sobre la estructura de las proteínas es la RMN. Como se ha mencionado anteriormente, la RMN requiere que la proteína de estudio esté marcada con isótopos estables de nitrógeno y carbono. Esto representa una gran limitación para utilizarla como una técnica para observar a las proteínas dentro de una célula. Sin embargo, se han reportado experimentos exitosos en donde al sobre-expresar a la proteína de interés en células de *E. coli* es posible observar señales que corresponden a dicha proteína y que sobrepasan a las de otras proteínas marcadas (Selenko and Wagner 2007; Takaoka *et al.* 2013; Freedberg and Selenko 2014). A esta técnica se le conoce como RMN-In cell o *Live Cell NMR* (Freedberg and Selenko 2014). Así pues, a pesar de la complejidad y las complicaciones de los enfoque *in vivo*, será importante encaminar los esfuerzos en sobre el estudio de la estructura de las proteínas LEA en un ambiente que sea lo más parecido a su “realidad celular”.

Otro detalle a tener en cuenta para el estudio de la función de las proteínas LEA del grupo 4 es su perfil de localización sub-celular. Se sabe que las proteínas LEA del grupo 4 se acumulan abundantemente en núcleo y citosol en respuesta a estrés por déficit hídrico (Martinez, 2016; (Candat *et al.* 2014). Además, se ha observado que la proteína AtLEA4-5 se acumula en estos mismos compartimentos cuando se expresa en la etapa de semilla, lo que sustenta las evidencias de que esta proteína ejerce su función en dichos compartimentos. Así pues, los blancos naturales de las proteínas LEA del grupo 4 deben ser moléculas contenidas en alguno de estos compartimentos celulares. Por otro lado, no se puede descartar el hecho de que las

proteínas LEA podrían ejecutar su función protegiendo a otras macromoléculas como ácidos nucleicos y/o lípidos, funciones que se han propuesto para proteínas LEA de otros grupos (Hara *et al.* 2009; Koag *et al.* 2009; Rahman *et al.* 2011). Por otra parte, dada su acumulación abundante en núcleos, no debe descartarse el hecho de que estas proteínas pudieran tener otra función, como en el caso de las proteínas LEA conocidas como ASR, las cuales han sido caracterizadas como factores transcripcionales que regulan genes de respuesta a aluminio (Arenhart *et al.* 2016). Otro factor a considerar es la acumulación diferencial de los tres miembros de la familia AtLEA4. Estudios previos muestran que la AtLEA4-1, AtLEA4-2 y AtLEA4-5 se acumulan en distintos tejidos en respuesta a estrés (Schmid *et al.* 2005; Olvera-Carrillo *et al.* 2010). Estos patrones de acumulación diferenciales aunados a los fenotipos observados en mutantes (por inserción y silenciadas) en los diferentes miembros y a la separación de las distintas proteínas LEA4 de plantas en dos cladas bien definidas (sub-grupos 4A y 4B) sugieren que cada una de estas proteínas ejercen funciones diferentes y podrían tener módulos de regulación distintos (Olvera-Carrillo *et al.* 2010; Olvera-Carrillo *et al.* 2011).

Cabe hacer notar que los resultados obtenidos en esta tesis sugieren que la función protectora de las proteínas LEA del grupo 4 está modulada por el estado conformacional en el que se encuentre la proteína, dependiendo del ambiente. Sin embargo, no debemos descartar que existan otras señales capaces de modular el estatus conformacional y/o la función de las proteínas LEA del grupo 4. Tal es el caso de las modificaciones post-traduccionales, por ejemplo la fosforilación. Se sabe que la fosforilación de proteínas es capaz de modular el plegamiento de algunas proteínas intrínsecamente desordenadas (Dyson and Wright 2005; Rahman *et al.* 2011). De hecho, se ha detectado la fosforilación de la AtLEA4-5 en el polen maduro de *Arabidopsis* (Mayank *et al.* 2012). Cabe señalar que, así como la semilla, el polen es un estadio del desarrollo que se encuentra en un estado de desecación, por lo que la regulación de la AtLEA4-5 mediada por la fosforilación podría tener un significado funcional en este órgano de la planta.

IX. CONCLUSIONES

Las proteínas AtLEA4-2 y AtLEA4-5 son hidrofílicas que poseen una estructura intrínsecamente desordenada en solución acuosa en un intervalo amplio de temperaturas y pHs. Ambas proteínas poseen una alta capacidad para formar α -hélices, presentando dos estados conformacionales en equilibrio, uno mayoritariamente desplegado y el otro principalmente α -hélice. El déficit hídrico y el amontonamiento molecular (condiciones prevalentes en las células expuestas a estrés hídrico) simulados *in vitro* modulan el equilibrio entre estos conformeros, favoreciendo el plegamiento en especies conformadas por α -hélices. La región N-terminal conservada de la AtLEA4-5 es necesaria y suficiente para inducir el cambio conformacional en α -hélices y para ejercer la función tipo chaperona de las proteínas AtLEA4. Los altos niveles de protección así como la presencia de motivos distintivos con funciones diferenciales permite proponer a las proteínas LEA del grupo 4 como un modelo para el estudio de la estructura y la función de las proteínas intrínsecamente desordenadas de plantas.

X. PERSPECTIVAS

Los resultados obtenidos en esta tesis sugieren que el plegamiento en α -hélices se requiere para ejercer la función tipo chaperona de las proteínas AtLEA4. Por tal motivo, será necesario diseñar mutaciones específicas que interfieran con el plegamiento en α -hélices en condiciones de déficit hídrico y probar la función de estas mutantes. Además, se deberán generar otras mutantes donde se modifiquen los motivos conservados para determinar la participación de éstos en el plegamiento y en la función protectora. Por otro lado, se podría delimitar la región mínima de la proteína necesaria para ejercer el efecto protector, con el fin de diseñar estrategias para facilitar el estudio de las AtLEA4 y para buscar posibles aplicaciones en el campo de la biotecnología. Actualmente, estos experimentos se abordan en las tesis del estudiante de doctorado David Rendón Luna y del estudiante de licenciatura André Purcel Ontrup.

La técnica de resonancia magnética nuclear fue exitosamente utilizada para detectar la estructura secundaria de las proteínas LEA4 por lo que se planteará dilucidar la estructura tridimensional de las proteínas AtLEA4-2 y AtLEA4-5 cuando

están plegadas en α -hélices. Esto se puede lograr titulando con TFE puesto que los datos obtenidos en esta tesis muestran que la vía de plegamiento en déficit hídrico (glicerol) y en TFE es la misma, por lo que la estructura tridimensional nos daría información sobre la posición y el papel de los residuos conservados en cada uno de los motivos. Esto se realizará como parte de la colaboración con la Dra. Jane Dyson del Scripps Research Institute en San Diego.

En esta tesis se caracterizó la estructura de las proteínas AtLEA4 *in vitro* pero queda pendiente determinar si la proteína es desordenada y se pliega cuando disminuye el agua disponible *in vivo*. Para ello, se puede utilizar la técnica de fluorescencia por transferencia de energía (FRET, *Fluorescence Resonance Energy Transfer*) para detectar si existen cambios conformacionales de la proteína *in vivo*. Este proyecto se encuentra en proceso y constituye la tesis de maestría de Dante Cosío y es parte de una colaboración con los doctores Wolf Frommer y José Dinneny del Instituto Carnegie en Stanford, California.

XI. BIBLIOGRAFÍA

- Abba, S; Ghignone, S and Bonfante, P (2006). A dehydration-inducible gene in the truffle *Tuber borchii* identifies a novel group of dehydrins. *BMC Genomics* 7. Artn 39
10.1186/1471-2164-7-39.
- Ahmad, MS and Ashraf, M (2011). Essential roles and hazardous effects of nickel in plants. *Reviews of Environmental Contamination and Toxicology* 214: 125-167. Doi:10.1007/978-1-4614-0668-6_6.
- Ahmed, MA; Bamm, VV; Shi, L; Steiner-Mosonyi, M; Dawson, JF; Brown, L; Harauz, G and Ladizhansky, V (2009). Induced secondary structure and polymorphism in an intrinsically disordered structural linker of the CNS: Solid-state NMR and FTIR spectroscopy of myelin basic protein bound to actin. *Biophysical Journal* 96(1): 180-191. Doi:10.1016/j.bpj.2008.10.003.
- Anderson, VL; Ramlall, TF; Rospigliosi, CC; Webb, WW and Eliezer, D (2010). Identification of a helical intermediate in trifluoroethanol-induced alpha-synuclein aggregation. *Proceedings of the National Academy of Sciences of the United States of America* 107(44): 18850-18855. Doi 10.1073/Pnas.1012336107.
- Arai, M; Sugase, K; Dyson, HJ and Wright, PE (2015). Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proceedings of the National Academy of Sciences of the United States of America* 112(31): 9614-9619. Doi:10.1073/pnas.1512799112.
- Arenhart, RA; Schunemann, M; Neto, LB; Margis, R; Wang, ZY and Margis-Pinheiro, M (2016). Rice ASR1 and ASR5 are complementary transcription factors regulating aluminium responsive genes. *Plant Cell and Environment* 39(3): 645-651. Doi:10.1111/pce.12655.
- Auge, RM (2001). Water relations, drought and vesicular-arbuscular mycorrhizal symbiosis. *Mycorrhiza* 11(1): 3-42. Doi 10.1007/S005720100097.
- Bardwell, JC and Jakob, U (2012). Conditional disorder in chaperone action. *Trends in Biochemical Sciences* 37(12): 517-525. Doi10.1016/j.tibs.2012.08.006.
- Barrett, PJ; Chen, J; Cho, MK; Kim, JH; Lu, ZW; Mathew, S; Peng, DG; Song, YL, *et al.* (2013). The quiet renaissance of protein nuclear magnetic resonance. *Biochemistry* 52(8): 1303-1320. Doi:10.1021/bi4000436.
- Basha, E; O'Neill, H and Vierling, E (2012). Small heat shock proteins and alpha-crystallins: Dynamic proteins with flexible functions. *Trends in Biochemical Sciences* 37(3): 106-117. Doi10.1016/j.tibs.2011.11.005.
- Battaglia, M; Olvera-Carrillo, Y; Garcíarrubio, A; Campos, F and Covarrubias, AA (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiology* 148(1): 6-24. Doi 10.1104/Pp.108.120725.
- Battista, JR; Park, MJ and McLemore, AE (2001). Inactivation of two homologues of proteins presumed to be involved in the desiccation tolerance of plants sensitizes *Deinococcus radiodurans* r1 to desiccation. *Cryobiology* 43(2): 133-139. Doi:10.1006/cryo.2001.2357.
- Bepperling, A; Alte, F; Kriehuber, T; Braun, N; Weinkauff, S; Groll, M; Haslbeck, M and Buchner, J (2012). Alternative bacterial two-component small heat shock protein systems. *Proceedings of the National Academy of Sciences of the United States of America* 109(50): 20407-20412. Doi:10.1073/pnas.1209565109.

- Berlow, RB; Dyson, HJ and Wright, PE (2015). Functional advantages of dynamic protein disorder. *FEBS Letters* 589(19 Pt A): 2433-2440. Doi: 10.1016/j.febslet.2015.06.003.
- Bies-Etheve, N; Gaubier-Comella, P; Debures, A; Lasserre, E; Jobet, E; Raynal, M; Cooke, R and Delseny, M (2008). Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Molecular Biology* 67(1-2): 107-124. Doi:10.1007/s11103-008-9304-x.
- Borcherds, W; Theillet, FX; Katzer, A; Finzel, A; Mishall, KM; Powell, AT; Wu, H; Manieri, W, *et al.* (2014). Disorder and residual helicity alter p53-MDM2 binding affinity and signaling in cells. *Nature Chemical Biology* 10(12): 1000-1002. Doi:10.1038/nchembio.1668.
- Boucher, V; Buitink, J; Lin, X; Boudet, J; Hoekstra, FA; Hundertmark, M; Renard, D and Leprince, O (2010). MtPM25 is an atypical hydrophobic late embryogenesis-abundant protein that dissociates cold and desiccation-aggregated proteins. *Plant Cell and Environment* 33(3): 418-430. Doi:10.1111/j.1365-3040.2009.02093.x.
- Bowler, C; Slooten, L; Vandenbranden, S; De Rycke, R; Botterman, J; Sybesma, C; Van Montagu, M and Inze, D (1991). Manganese superoxide dismutase can reduce cellular damage mediated by oxygen radicals in transgenic plants. *EMBO Journal* 10(7): 1723-1732.
- Bray, EA (1997). Plant responses to water deficit. *Trends in Plant Sciences* 2(2): 48-54. Doi:10.1016/S1360-1385(97)82562-9.
- Browne, JA; Dolan, KM; Tyson, T; Goyal, K; Tunnacliffe, A and Burnell, AM (2004). Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode *aphelenchus avenae*. *Eukaryotic Cell* 3(4): 966-975. Doi:10.1128/EC.3.4.966-975.2004.
- Buck, M (1998). Trifluoroethanol and colleagues: Cosolvents come of age. Recent studies with peptides and proteins. *Quarterly Reviews of Biophysics* 31(3): 297-355. Doi:10.1017/S003358359800345x.
- Bukau, B; Weissman, J and Horwich, A (2006). Molecular chaperones and protein quality control. *Cell* 125(3): 443-451. Doi:10.1016/j.cell.2006.04.014.
- Burger, VM; Gurry, T and Stultz, CM (2014). Intrinsically disordered proteins: Where computation meets experiment. *Polymers* 6(10): 2684-2719. Doi:10.3390/polym6102684.
- Campos, F; Cuevas-Velazquez, C; Fares, MA; Reyes, JL and Covarrubias, AA (2013). Group 1 LEA proteins, an ancestral plant protein group, are also present in other eukaryotes, and in the archaea and bacteria domains. *Molecular Genetics and Genomics* 288(10): 503-517. Doi:10.1007/s00438-013-0768-2.
- Campos, F; Guillen, G; Reyes, JL and Covarrubias, AA (2011). A general method of protein purification for recombinant unstructured non-acidic proteins. *Protein Expression and Purification* 80(1): 47-51. Doi:10.1016/J.Pep.2011.06.007.
- Chakrabortee, S; Boschetti, C; Walton, LJ; Sarkar, S; Rubinsztein, DC and Tunnacliffe, A (2007). Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceedings of the National Academy of Sciences of the United States of America* 104(46): 18073-18078. Doi:10.1073/pnas.0706964104.
- Chakrabortee, S; Tripathi, R; Watson, M; Schierle, GSK; Kurniawan, DP; Kaminski, CF; Wise, MJ and Tunnacliffe, A (2012). Intrinsically disordered proteins as

- molecular shields. *Molecular Biosystems* 8(1): 210-219. Doi 10.1039/C1mb05263b.
- Chaplin, MF (2001). Water: Its importance to life. *Biochemistry and Molecular Biology Education* 29(2): 54-59. Doi:10.1111/J.1539-3429.2001.Tb00070.X.
- Chen, GJ; Qiu, N; Karrer, C; Caspers, P and Page, MG (2000). Restriction site-free insertion of PCR products directionally into vectors. *Biotechniques* 28(3): 498-500, 504-495.
- Cheng, JL; Sweredoski, MJ and Baldi, P (2005). Accurate prediction of protein disordered regions by mining protein structure data. *Data Mining and Knowledge Discovery* 11(3): 213-222. Doi:10.1007/S10618-005-0001-Y.
- Clare, DK; Vasishtan, D; Stagg, S; Quispe, J; Farr, GW; Topf, M; Horwich, AL and Saibil, HR (2012). ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the groel chaperonin. *Cell* 149(1): 113-123. Doi:10.1016/j.cell.2012.02.047.
- Colmenero-Flores, JM; Campos, F; Garcarrubio, A and Covarrubias, AA (1997). Characterization of *Phaseolus vulgaris* cDNA clones responsive to water deficit: Identification of a novel late embryogenesis abundant-like protein. *Plant Molecular Biology* 35(4): 393-405. Doi:10.1023/A:1005802505731.
- Colmenero-Flores, JM; Moreno, LP; Smith, CE and Covarrubias, AA (1999). Pvlea-18, a member of a new late-embryogenesis-abundant protein family that accumulates during water stress and in the growing regions of well-irrigated bean seedlings. *Plant Physiology* 120(1): 93-103. Doi:10.1104/Pp.120.1.93.
- Covarrubias, AA and Reyes, JL (2010). Post-transcriptional gene regulation of salinity and drought responses by plant micromas. *Plant Cell and Environment* 33(4): 481-489. Doi:10.1111/j.1365-3040.2009.02048.x.
- Cremers, CM; Reichmann, D; Hausmann, J; Ilbert, M and Jakob, U (2010). Unfolding of metastable linker region is at the core of HSP33 activation as a redox-regulated chaperone. *Journal of Biological Chemistry* 285(15): 11243-11251. Doi:10.1074/jbc.M109.084350.
- Cuevas-Velazquez, CL; Rendon-Luna, DF and Covarrubias, AA (2014). Dissecting the cryoprotection mechanisms for dehydrins. *Frontiers in Plant Science* 5: 583. Doi:10.3389/fpls.2014.00583.
- Dames, SA; Martinez-Yamout, M; De Guzman, RN; Dyson, HJ and Wright, PE (2002). Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. *Proceedings of the National Academy of Sciences of the United States of America* 99(8): 5271-5276. Doi:10.1073/pnas.082121399.
- Dang, NX and Hinch, DK (2011). Identification of two hydrophilins that contribute to the desiccation and freezing tolerance of yeast (*Saccharomyces cerevisiae*) cells. *Cryobiology* 62(3): 188-193. Doi:10.1016/j.cryobiol.2011.03.002.
- Dang, NX; Popova, AV; Hundertmark, M and Hinch, DK (2014). Functional characterization of selected LEA proteins from *Arabidopsis thaliana* in yeast and *in vitro*. *Planta* 240(2): 325-336. Doi:10.1007/S00425-014-2089-Z.
- Delseny, M; Bies-Etheve, N; Carles, C; Hull, G; Vicient, C; Raynal, M; Grellet, F and Aspart, L (2001). Late embryogenesis abundant (LEA) protein gene regulation during *Arabidopsis* seed maturation. *Journal of Plant Physiology* 158(4): 419-427. Doi:10.1078/0176-1617-00353.
- Demchenko, AP (2001). Recognition between flexible protein molecules: Induced and assisted folding. *Journal of Molecular Recognition* 14(1): 42-61. Doi 10.1002/1099-1352(200101/02)14:1<42::Aid-Jmr518>3.0.Co;2-8.

- Drira, M; Saibi, W; Brini, F; Gargouri, A; Masmoudi, K and Hanin, M (2013). The K-segments of the wheat dehydrin dh5 are essential for the protection of lactate dehydrogenase and beta-glucosidase activities *in vitro*. *Molecular Biotechnology* 54(2): 643-650. Doi:10.1007/s12033-012-9606-8.
- Dunker, AK; Lawson, JD; Brown, CJ; Williams, RM; Romero, P; Oh, JS; Oldfield, CJ; Campen, AM, *et al.* (2001). Intrinsically disordered protein. *Journal of Molecular Graphics & Modelling* 19(1): 26-59. Doi:10.1016/S1093-3263(00)00138-8.
- Dure, L (1993). A repeating 11-mer amino-acid motif and plant desiccation. *Plant Journal* 3(3): 363-369. Doi:10.1046/J.1365-313x.1993.T01-18-00999.X.
- Dure, L (2001). Occurrence of a repeating 11-mer amino acid sequence motif in diverse organisms. *Protein and Peptide Letters* 8(2): 115-122. Doi: 10.2174/0929866013409643.
- Dure, L, 3rd; Greenway, SC and Galau, GA (1981). Developmental biochemistry of cottonseed embryogenesis and germination: Changing messenger ribonucleic acid populations as shown by *in vitro* and *in vivo* protein synthesis. *Biochemistry* 20(14): 4162-4168.
- Dure, L and Chlan, C (1981). Developmental biochemistry of cottonseed embryogenesis and germination : Xii. Purification and properties of principal storage proteins. *Plant Physiology* 68(1): 180-186.
- Dure, L; Crouch, M; Harada, J; Ho, THD; Mundy, J; Quatrano, R; Thomas, T and Sung, ZR (1989). Common amino-acid sequence domains among the LEA proteins of higher-plants. *Plant Molecular Biology* 12(5): 475-486. Doi:10.1007/Bf00036962.
- Dure, L and Galau, GA (1981). Developmental biochemistry of cottonseed embryogenesis and germination : Xiii. Regulation of biosynthesis of principal storage proteins. *Plant Physiology* 68(1): 187-194.
- Dyson, HJ and Wright, PE (2002). Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* 12(1): 54-60.
- Dyson, HJ and Wright, PE (2002). Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Unfolded Proteins* 62: 311-340.
- Dyson, HJ and Wright, PE (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* 6(3): 197-208. Doi:10.1038/Nrm1589.
- Ellis, RJ (2001). Macromolecular crowding: An important but neglected aspect of the intracellular environment (vol 11, pg 114, 2001). *Current Opinion in Structural Biology* 11(4): 500-500. Doi:10.1016/S0959-440x(00)00239-6.
- Fancy, DA and Kodadek, T (1999). Chemistry for the analysis of protein-protein interactions: Rapid and efficient cross-linking triggered by long wavelength light. *Proceedings of the National Academy of Sciences of the United States of America* 96(11): 6020-6024. Doi:10.1073/Pnas.96.11.6020.
- Fang, YJ and Xiong, LZ (2015). General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and Molecular Life Sciences* 72(4): 673-689. Doi:10.1007/s00018-014-1767-0.
- Freedberg, DI and Selenko, P (2014). Live cell NMR. *Annual Review of Biophysics* 43: 171-192. Doi:10.1146/annurev-biophys-051013-023136.
- French, L (2015). Caracterización espectroscópica de la interacción con metales de la familia de proteínas AtLEA4 (Late Embryogenesis Abundant) de Arabidopsis

- thaliana (Brassicales:Brassicaceae). Tesis de Maestría. Universidad Autónoma del Estado de Morelos.
- Gal, TZ; Glazer, I and Koltai, H (2004). An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress. *FEBS Letters* 577(1-2): 21-26. Doi:10.1016/j.febslet.2004.09.049.
- Galau, GA; Wang, HYC and Hughes, DW (1992). Cotton lea4 (D19) and leaA2 (D132) group-1 lea genes encoding water stress-related proteins containing a 20-amino acid motif. *Plant Physiology* 99(2): 783-788. Doi:10.1104/Pp.99.2.783.
- Garay-Arroyo, A; Colmenero-Flores, JM; Garcarrubio, A and Covarrubias, AA (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *Journal Biological Chemistry* 275(8): 5668-5674. Doi:10.1074/Jbc.275.8.5668.
- Gast, K; Damaschun, H; Eckert, K; Schulzeforster, K; Maurer, HR; Mullerfrohne, M; Zirwer, D; Czarnecki, J, *et al.* (1995). Prothymosin .alpha.: A biologically-active protein with random coil conformation. *Biochemistry* 34(40): 13211-13218. Doi:10.1021/Bi00040a037.
- Giese, KC and Vierling, E (2002). Changes in oligomerization are essential for the chaperone activity of a small heat shock protein *in vivo* and *in vitro*. *Journal of Biological Chemistry* 277(48): 46310-46318. Doi:10.1074/jbc.M208926200.
- Goldenberg, DP and Argyle, B (2014). Minimal effects of macromolecular crowding on an intrinsically disordered protein: A small-angle neutron scattering study. *Biophysical Journal* 106(4): 905-914. Doi:10.1016/J.Bpj.2013.12.003.
- Goyal, K; Tisi, L; Basran, A; Browne, J; Burnell, A; Zurdo, J and Tunnacliffe, A (2003). Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *Journal of Biological Chemistry* 278(15): 12977-12984. Doi:10.1074/Jbc.M212007200.
- Goyal, K; Walton, LJ and Tunnacliffe, A (2005). LEA proteins prevent protein aggregation due to water stress. *Biochemical Journal* 388: 151-157. Doi:10.1042/BJ20041931.
- Grelet, J; Benamar, A; Teyssier, E; Avelange-Macherel, MH; Grunwald, D and Macherel, D (2005). Identification in pea seed mitochondria of a late embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiology* 137(1): 157-167. Doi:10.1104/pp.104.052480.
- Hara, M; Fujinaga, M and Kuboi, T (2005). Metal binding by citrus dehydrin with histidine-rich domains. *Journal of Experimental Botany* 56(420): 2695-2703. Doi:10.1093/jxb/eri262.
- Hara, M; Kondo, M and Kato, T (2013). A KS-type dehydrin and its related domains reduce Cu-promoted radical generation and the histidine residues contribute to the radical-reducing activities. *Journal of Experimental Botany* 64(6): 1615-1624. Doi:10.1093/jxb/ert016.
- Hara, M; Monna, S; Murata, T; Nakano, T; Amano, S; Nachbar, M and Watzig, H (2016). The arabidopsis KS-type dehydrin recovers lactate dehydrogenase activity inhibited by copper with the contribution of his residues. *Plant Science* 245: 135-142. Doi:10.1016/j.plantsci.2016.02.006.
- Hara, M; Shinoda, Y; Kubo, M; Kashima, D; Takahashi, I; Kato, T; Horiike, T and Kuboi, T (2011). Biochemical characterization of the Arabidopsis KS-type dehydrin protein, whose gene expression is constitutively abundant rather than stress dependent. *Acta Physiologiae Plantarum* 33(6): 2103-2116. Doi:10.1007/s11738-011-0749-1.

- Hara, M; Shinoda, Y; Tanaka, Y and Kuboi, T (2009). DNA binding of citrus dehydrin promoted by zinc ion. *Plant Cell and Environment* 32(5): 532-541. Doi:10.1111/j.1365-3040.2009.01947.x.
- Hartl, FU; Bracher, A and Hayer-Hartl, M (2011). Molecular chaperones in protein folding and proteostasis. *Nature* 475(7356): 324-332. Doi:10.1038/nature10317.
- Hartl, FU and Hayer-Hartl, M (2009). Converging concepts of protein folding *in vitro* and *in vivo*. *Nature Structural & Molecular Biology* 16(6): 574-581. Doi:10.1038/nsmb.1591.
- Haslbeck, M and Vierling, E (2015). A first line of stress defense: Small heat shock proteins and their function in protein homeostasis. *Journal of Molecular Biology* 427(7): 1537-1548. Doi:10.1016/j.jmb.2015.02.002.
- Hellwege, EM; Dietz, KJ and Hartung, W (1996). Abscisic acid causes changes in gene expression involved in the induction of the landform of the liverwort *Riccia fluitans* l. *Planta* 198(3): 423-432. Doi:10.1007/Bf00620059.
- Hernandez, MA; Avila, J and Andreu, JM (1986). Physicochemical characterization of the heat-stable microtubule-associated protein MAP2. *European Journal of Biochemistry* 154(1): 41-48. Doi:10.1111/J.1432-1033.1986.Tb09356.X.
- Hernandez-Sanchez, IE; Martynowicz, DM; Rodriguez-Hernandez, AA; Perez-Morales, MB; Graether, SP and Jimenez-Bremont, JF (2014). A dehydrin-dehydrin interaction: The case of SK3 from *Opuntia streptacantha*. *Frontiers in Plant Sciences* 5: 520. Doi:10.3389/fpls.2014.00520.
- Hernandez-Sanchez, IE; Maruri-Lopez, I; Ferrando, A; Carbonell, J; Graether, SP and Jimenez-Bremont, JF (2015). Nuclear localization of the dehydrin opsdhn1 is determined by histidine-rich motif. *Frontiers in Plant Sciences* 6: 702. Doi:10.3389/fpls.2015.00702.
- Hoekstra, FA; Golovina, EA; Tetteroo, FAA and Wolkers, WF (2001). Induction of desiccation tolerance in plant somatic embryos: How exclusive is the protective role of sugars? *Cryobiology* 43(2): 140-150. Doi:10.1006/cryo.2001.2358.
- Honjoh, K; Matsumoto, H; Shimizu, H; Ooyama, K; Tanaka, K; Oda, Y; Takata, R; Joh, T, *et al.* (2000). Cryoprotective activities of group 3 late embryogenesis abundant proteins from *Chlorella vulgaris* c-27. *Bioscience Biotechnology and Biochemistry* 64(8): 1656-1663. Doi:10.1271/Bbb.64.1656.
- Honjoh, K; Yoshimoto, M; Joh, T; Kajiwara, T; Miyamoto, T and Hatano, S (1995). Isolation and characterization of hardening-induced proteins in *Chlorella vulgaris* c-27: Identification of late embryogenesis abundant proteins. *Plant and Cell Physiology* 36(8): 1421-1430.
- Hornshaw, MP; McDermott, JR; Candy, JM and Lakey, JH (1995). Copper-binding to the N-terminal tandem repeat region of mammalian and avian prion protein - structural studies using synthetic peptides. *Biochemical and Biophysical Research Communications* 214(3): 993-999. Doi:10.1006/Bbrc.1995.2384.
- Hruz, T; Laule, O; Szabo, G; Wessendorp, F; Bleuler, S; Oertle, L; Widmayer, P; Gruissem, W, *et al.* (2008). Genevestigator v3: A reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics* 2008: 420747. Doi:10.1155/2008/420747.
- Hughes, S and Graether, SP (2011). Cryoprotective mechanism of a small intrinsically disordered dehydrin protein. *Protein Science* 20(1): 42-50. Doi:10.1002/pro.534.

- Hughes, SL; Schart, V; Malcolmson, J; Hogarth, KA; Martynowicz, DM; Tralman-Baker, E; Patel, SN and Graether, SP (2013). The importance of size and disorder in the cryoprotective effects of dehydrins. *Plant Physiology* 163(3): 1376-1386. Doi:10.1104/pp.113.226803.
- Hundertmark, M; Dimova, R; Lengefeld, J; Seckler, R and Hinch, DK (2011). The intrinsically disordered late embryogenesis abundant protein lea18 from *arabidopsis thaliana* modulates membrane stability through binding and folding. *BBA Biomembranes* 1808(1): 446-453. Doi:10.1016/J.Bbamem.2010.09.010.
- Hundertmark, M and Hinch, DK (2008). Lea (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9. Doi:10.1186/1471-2164-9-118.
- Hundertmark, M; Popova, AV; Rausch, S; Seckler, R and Hinch, DK (2012). Influence of drying on the secondary structure of intrinsically disordered and globular proteins. *Biochemical and Biophysical Research Communications* 417(1): 122-128. Doi:10.1016/J.Bbrc.2011.11.067.
- Huth, JR; Bewley, CA; Jackson, BM; Hinnebusch, AG; Clore, GM and Gronenborn, AM (1997). Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by nmr. *Protein Science* 6(11): 2359-2364. Doi:10.1002/pro.5560061109.
- Iakoucheva, LM; Brown, CJ; Lawson, JD; Obradovic, Z and Dunker, AK (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology* 323(3): 573-584. Doi:10.1016/S0022-2836(02)00969-5.
- Iturriaga, G; Cushman, MAF and Cushman, JC (2006). An est catalogue from the resurrection plant *Selaginella lepidophylla* reveals abiotic stress-adaptive genes. *Plant Science* 170(6): 1173-1184. Doi:10.1016/j.plantsci.2006.02.004.
- Joh, T; Honjoh, K; Yoshimoto, M; Funabashi, J; Miyamoto, T and Hatano, S (1995). Molecular-cloning and expression of hardening-induced genes in *Chlorella-vulgaris* c-27 - the most abundant clone encodes a late embryogenesis abundant protein. *Plant and Cell Physiology* 36(1): 85-93.
- Jones, AM; Danielson, JA; Manojkumar, SN; Lanquar, V; Grossmann, G and Frommer, WB (2014). Abscisic acid dynamics in roots detected with genetically encoded FRET sensors. *eLIFE* 3: e01741. Doi:10.7554/eLife.01741.
- Jones, DT and Cozzetto, D (2015). Disopred3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 31(6): 857-863. Doi:10.1093/Bioinformatics/Btu744.
- Jones, RO, H.; Thomas, H.; Waaland, S. (2013). *The molecular life of plants*, Wiley-Blackwell.
- Kelly, SM; Jess, TJ and Price, NC (2005). How to study proteins by circular dichroism. *Biochimica Et Biophysica Acta-Proteins and Proteomics* 1751(2): 119-139. Doi:10.1016/j.bbapap.2005.06.005.
- Kikawada, T; Nakahara, Y; Kanamori, Y; Iwata, KI; Watanabe, M; Mcgee, B; Tunnacliffe, A and Okuda, T (2006). Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochemical and Biophysical Research Communications* 348(1): 56-61. Doi:10.1016/j.bbrc.2006.07.003.
- Koag, MC; Wilkens, S; Fenton, RD; Resnik, J; Vo, E and Close, TJ (2009). The K-segment of maize dhn1 mediates binding to anionic phospholipid vesicles and concomitant structural changes. *Plant Physiology* 150(3): 1503-1514. Doi:10.1104/pp.109.136697.

- Koenig, BW; Rogowski, M and Louis, JM (2003). A rapid method to attain isotope labeled small soluble peptides for NMR studies. *Journal of Biomolecular NMR* 26(3): 193-202.
- Kriehuber, T; Rattei, T; Weinmaier, T; Bepperling, A; Haslbeck, M and Buchner, J (2010). Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *FASEB Journal* 24(10): 3633-3642. Doi:10.1096/fj.10-156992.
- Kriwacki, RW; Hengst, L; Tennant, L; Reed, SI and Wright, PE (1996). Structural studies of p21(Waf1/Cip1/Sdi1) in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences of the United States of America* 93(21): 11504-11509. Doi:10.1073/Pnas.93.21.11504.
- Kruger, C; Berkowitz, O; Stephan, UW and Hell, R (2002). A metal-binding member of the late embryogenesis abundant protein family transports iron in the phloem of *Ricinus communis* L. *Journal of Biological Chemistry* 277(28): 25062-25069. Doi:10.1074/jbc.M201896200.
- Lee, GJ and Vierling, E (2000). A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. *Plant Physiology* 122(1): 189-198.
- Lee, SH; Kim, DH; Han, JJ; Cha, EJ; Lim, JE; Cho, YJ; Lee, C and Han, KH (2012). Understanding pre-structured motifs (presmos) in intrinsically unfolded proteins. *Current Protein and Peptide Science* 13(1): 34-54.
- Linding, R; Jensen, LJ; Diella, F; Bork, P; Gibson, TJ and Russell, RB (2003). Protein disorder prediction: Implications for structural proteomics. *Structure* 11(11): 1453-1459. Doi:10.1016/j.str.2003.10.002.
- Lisse, T; Bartels, D; Kalbitzer, HR and Jaenicke, R (1996). The recombinant dehydrin-like desiccation stress protein from the resurrection plant *Cratogeomys plantagineum* displays no defined three-dimensional structure in its native state. *Journal of Biological Chemistry* 271(9): 555-561.
- Liu, G; Xu, H; Zhang, L and Zheng, Y (2011). Fe binding properties of two soybean (*Glycine max* L.) *lea4* proteins associated with antioxidant activity. *Plant Cell and Physiology* 52(6): 994-1002. Doi:10.1093/pcp/pcr052.
- Liu, J; Xing, Y; Hinds, TR; Zheng, J and Xu, W (2006). The third 20 amino acid repeat is the tightest binding site of APC for beta-catenin. *Journal of Molecular Biology* 360(1): 133-144. Doi:10.1016/j.jmb.2006.04.064.
- Lobley, A; Whitmore, L and Wallace, BA (2002). Dichroweb: An interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 18(1): 211-212. Doi 10.1093/Bioinformatics/18.1.211.
- Lopez-Martinez, G; Rodriguez-Porrata, B; Margalef-Catala, M and Cordero-Otero, R (2012). The STF2p hydrophilin from *Saccharomyces cerevisiae* is required for dehydration stress tolerance. *Plos One* 7(3). ARTN e33324 Doi:10.1371/journal.pone.0033324.
- Luo, PZ and Baldwin, RL (1997). Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* 36(27): 8413-8421. Doi:10.1021/Bi9707133.
- Mark, WY; Liao, JCC; Lu, Y; Ayed, A; Laister, R; Szymczyzna, B; Chakrabarty, A and Arrowsmith, CH (2005). Characterization of segments from the central region of BRCA1: An intrinsically disordered scaffold for multiple protein-

- protein and protein-DNA interactions. *Journal of Molecular Biology* 345(2): 275-287. Doi:10.1016/j.jmb.2004.10.045.
- Martinez-Martinez, C (2016). Localización del transcrito y la proteína del gen AtLEA4-5 que participa en la respuesta al déficit hídrico en *Arabidopsis thaliana*. Tesis de Maestría. Universidad Nacional Autónoma de México.
- Maskin, L; Maldonado, S and Iusem, N (2008). Tomato leaf spatial expression of stress-induced ASR genes. *Molecular Biology Reports* 35(4): 501-505. Doi:10.1007/s11033-007-9114-2.
- Mayank, P; Grossman, J; Wuest, S; Boisson-Dernier, A; Roschitzki, B; Nanni, P; Nuhse, T and Grossniklaus, U (2012). Characterization of the phosphoproteome of mature *Arabidopsis* pollen. *Plant Journal* 72(1): 89-101. Doi:10.1111/j.1365-313X.2012.05061.x.
- Minton, AP (1997). Influence of excluded volume upon macromolecular structure and associations in 'crowded' media. *Current Opinion in Biotechnology* 8(1): 65-69. Doi:10.1016/S0958-1669(97)80159-0.
- Minton, AP (2000). Implications of macromolecular crowding for protein assembly. *Current Opinion in Structural Biology* 10(1): 34-39. Doi:10.1016/S0959-440x(99)00045-7.
- Mittag, T; Orlicky, S; Choy, WY; Tang, X; Lin, H; Sicheri, F; Kay, LE; Tyers, M, *et al.* (2008). Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proceedings of the National Academy of Sciences of the United States of America* 105(46): 17772-17777. Doi:10.1073/pnas.0809222105.
- Mouillon, JM; Eriksson, SK and Harryson, P (2008). Mimicking the plant cell interior under water stress by macromolecular crowding: Disordered dehydrin proteins are highly resistant to structural collapse. *Plant Physiology* 148(4): 1925-1937. Doi:10.1104/Pp.108.124099.
- Mouillon, JM; Gustafsson, P and Harryson, P (2006). Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiology* 141(2): 638-650. Doi:10.1104/Pp.106.079848.
- Mtwisha, L; Brandt, W; McCready, S and Lindsey, GG (1998). HSP 12 is a LEA-like protein in *Saccharomyces cerevisiae*. *Plant Molecular Biology* 37(3): 513-521. Doi:10.1023/A:1005904219201.
- Mu, P; Feng, D; Su, J; Zhang, Y; Dai, J; Jin, H; Liu, B; He, Y, *et al.* (2011). Cu²⁺ triggers reversible aggregation of a disordered His-rich dehydrin MpDhn12 from *Musa paradisiaca*. *Journal of Biochemistry* 150(5): 491-499. Doi:10.1093/jb/mvr082.
- Munoz, V and Serrano, L (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nature Structural and Molecular Biology* 1(6): 399-409. Doi:10.1038/Nsb0694-399.
- Nakayama, K; Okawa, K; Kakizaki, T; Honma, T; Itoh, H and Inaba, T (2007). *Arabidopsis* Cor15am is a chloroplast stromal protein that has cryoprotective activity and forms oligomers. *Plant Physiology* 144(1): 513-523. Doi:10.1104/pp.106.094581.
- Namba, K (2001). Roles of partly unfolded conformations in macromolecular self-assembly. *Genes to Cells* 6(1): 1-12. Doi:10.1046/J.1365-2443.2001.00384.X.
- Oldfield, CJ; Cheng, YG; Cortese, MS; Romero, P; Uversky, VN and Dunker, AK (2005). Coupled folding and binding with alpha-helix-forming molecular

- recognition elements. *Biochemistry* 44(37): 12454-12470. Doi:10.1021/Bi050736e.
- Oliver, MJ; Dowd, SE; Zaragoza, J; Mauget, SA and Payton, PR (2004). The rehydration transcriptome of the desiccation-tolerant bryophyte *Tortula ruralis*: Transcript classification and analysis. *BMC Genomics* 5. Artn 89 Doi:10.1186/1471-2164-5-89.
- Oliver, MJ; Tuba, Z and Mishler, BD (2000). The evolution of vegetative desiccation tolerance in land plants. *Plant Ecology* 151(1): 85-100. Doi:10.1023/A:1026550808557.
- Olvera-Carrillo, Y; Campos, F; Reyes, JL; Garcarrubio, A and Covarrubias, AA (2010). Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in arabidopsis. *Plant Physiology* 154(1): 373-390. Doi:10.1104/Pp.110.158964.
- Olvera-Carrillo, Y; Luis Reyes, J and Covarrubias, AA (2011). Late embryogenesis abundant proteins: Versatile players in the plant adaptation to water limiting environments. *Plant Signaling and Behavior* 6(4): 586-589.
- Otaki, JM; Tsutsumi, M; Gotoh, T and Yamamoto, H (201). Secondary structure characterization based on amino acid composition and availability in proteins. *Journal of Chemical Information and Modeling* 50(4): 690-700. Doi:10.1021/ci900452z
- Patel, S; Ramanujam, V; Srivastava, AK and Chary, KVR (2014). Conformational propensities and dynamics of a beta gamma-crystallin, an intrinsically disordered protein. *Physical Chemistry Chemical Physics* 16(25): 12703-12718. Doi:10.1039/c3cp53558d.
- Polverini, E; Rangaraj, G; Libich, DS; Boggs, JM and Harauz, G (2008). Binding of the proline-rich segment of myelin basic protein to SH3 domains: Spectroscopic, microarray, and modeling studies of ligand conformation and effects of posttranslational modifications. *Biochemistry* 47(1): 267-282. Doi:10.1021/bi701336n.
- Popova, AV; Hundertmark, M; Seckler, R and Hincha, DK (2011). Structural transitions in the intrinsically disordered plant dehydration stress protein LEA7 upon drying are modulated by the presence of membranes. *BBA Biomembranes* 1808(7): 1879-1887. Doi:10.1016/J.Bbamem.2011.03.009.
- Porembski, S and Barthlott, W (2000). Granitic and gneissic outcrops (inselbergs) as centers of diversity for desiccation-tolerant vascular plants. *Plant Ecology* 151(1): 19-28. Doi:10.1023/A:1026565817218.
- Proctor, MCF; Oliver, MJ; Wood, AJ; Alpert, P; Stark, LR; Cleavitt, NL and Mishler, BD (2007). Desiccation-tolerance in bryophytes: A review. *Bryologist* 110(4): 595-621. Doi:10.1639/0007-2745(2007)110[595:Dibar]2.0.Co;2.
- Qin, SB and Zhou, HX (2013). Effects of macromolecular crowding on the conformational ensembles of disordered proteins. *Journal of Physical Chemistry Letters* 4(20): 3429-3434. Doi:10.1021/Jz401817x.
- Radhakrishnan, I; Perez-Alvarado, GC; Parker, D; Dyson, HJ; Montminy, MR and Wright, PE (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:Coactivator interactions. *Cell* 91(6): 741-752.
- Ragsdale, SW (1998). Nickel biochemistry. *Current Opinion in Chemical Biology* 2(2): 208-215.
- Rahman, LN; Smith, GS; Bamm, VV; Voyer-Grant, JA; Moffatt, BA; Dutcher, JR and Harauz, G (2011). Phosphorylation of *Thellungiella salsuginea* dehydrins

- TsDHN-1 and TsDHN-2 facilitates cation-induced conformational changes and actin assembly. *Biochemistry* 50(44): 9587-9604. Doi:10.1021/bi201205m.
- Ratajczak, E; Zietkiewicz, S and Liberek, K (2009). Distinct activities of *Escherichia coli* small heat shock proteins ibpA and ibpB promote efficient protein disaggregation. *Journal of Molecular Biology* 386(1): 178-189. Doi:10.1016/j.jmb.2008.12.009.
- Reeves, R and Beckerbauer, L (2001). HMGI/Y proteins: Flexible regulators of transcription and chromatin structure. *Biochimica Et Biophysica Acta- Gene Structure and Expression* 1519(1-2): 13-29. Doi:10.1016/S0167-4781(01)00215-9.
- Reichmann, D; Xu, Y; Cremers, CM; Ilbert, M; Mittelman, R; Fitzgerald, MC and Jakob, U (2012). Order out of disorder: Working cycle of an intrinsically unfolded chaperone. *Cell* 148(5): 947-957. Doi:10.1016/J.Cell.2012.01.045.
- Reyes, JL; Campos, F; Wei, H; Arora, R; Yang, YI; Karlson, DT and Covarrubias, AA (2008). Functional dissection of hydrophilins during *in vitro* freeze protection. *Plant Cell and Environment* 31(12): 1781-1790. Doi:10.1111/J.1365-3040.2008.01879.X.
- Reyes, JL; Rodrigo, MJ; Colmenero-Flores, JM; Gil, JV; Garay-Arroyo, A; Campos, F; Salamini, F; Bartels, D, *et al.* (2005). Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects *in vitro*. *Plant Cell and Environment* 28(6): 709-718. Doi:10.1111/J.1365-3040.2005.01317.X.
- Reynolds, TL and Bewley, JD (1993). Characterization of protein synthetic changes in a desiccation-tolerant fern, *Polypodium virginianum* - comparison of the effects of drying, rehydration and abscisic-acid. *Journal of Experimental Botany* 44(262): 921-928. Doi:10.1093/Jxb/44.5.921.
- Rivera-Najera, LY; Saab-Rincon, G; Battaglia, M; Amero, C; Pulido, NO; Garcia-Hernandez, E; Solorzano, RM; Reyes, JL, *et al.* (2014). A group 6 late embryogenesis abundant protein from common bean is a disordered protein with extended helical structure and oligomer-forming properties. *Journal of Biological Chemistry* 289(46): 31995-32009. Doi:10.1074/jbc.M114.583369.
- Roberts, JK; Desimone, NA; Lingle, WL and Dure, L (1993). Cellular concentrations and uniformity of cell-type accumulation of 2 LEA proteins in cotton embryos. *Plant Cell* 5(7): 769-780.
- Rodriguez-Porrata, B; Carmona-Gutierrez, D; Reisenbichler, A; Bauer, M; Lopez, G; Escote, X; Mas, A; Madeo, F, *et al.* (2012). Sip18 hydrophilin prevents yeast cell death during desiccation stress. *Journal of Applied Microbiology* 112(3): 512-525. Doi:10.1111/j.1365-2672.2011.05219.x.
- Rogers, JM; Wong, CT and Clarke, J (2014). Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *Journal of the American Chemical Society* 136(14): 5197-5200. Doi:10.1021/ja4125065.
- Romero, P; Obradovic, Z; Li, XH; Garner, EC; Brown, CJ and Dunker, AK (2001). Sequence complexity of disordered protein. *Proteins-Structure Function and Genetics* 42(1): 38-48. Doi:10.1002/1097-0134(20010101)42:1<38::Aid-Prot50>3.0.Co;2-3.
- Rosano, GL and Ceccarelli, EA (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology* 5: 172. Doi:10.3389/fmicb.2014.00172.

- Roychoudhury, A; Paul, S and Basu, S (2013). Cross-talk between abscisic acid-dependent and abscisic acid-independent pathways during abiotic stress. *Plant Cell Reports* 32(7): 985-1006. Doi:10.1007/s00299-013-1414-5.
- Rozhdestvenskii, VI; Vil'iams, MV; Tsvetkova, IV; Lebedeva, EV and Alekhina, TP (1975). [control of mineral nutrition of higher plants in biological life support systems]. *Kosm Biol Aviakosm Med* 9(6): 30-35.
- Rubio, S; Rodrigues, A; Saez, A; Dizon, MB; Galle, A; Kim, TH; Santiago, J; Flexas, J, *et al.* (2009). Triple loss of function of protein phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. *Plant Physiology* 150(3): 1345-1355. Doi:10.1104/Pp.109.137174.
- Saavedra, L; Svensson, J; Carballo, V; Izmendi, D; Welin, B and Vidal, S (2006). A dehydrin gene in *Physcomitrella patens* is required for salt and osmotic stress tolerance. *Plant Journal* 45(2): 237-249. Doi:10.1111/j.1365-3113X.2005.02603.x.
- Salmi, ML; Bushart, TJ; Stout, SC and Roux, SJ (2005). Profile and analysis of gene expression changes during early development in germinating spores of *Ceratopteris richardii*. *Plant Physiology* 138(3): 1734-1745. Doi:10.1104/pp.105.062851.
- Schmid, M; Davison, TS; Henz, SR; Pape, UJ; Demar, M; Vingron, M; Scholkopf, B; Weigel, D, *et al.* (2005). A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics* 37(5): 501-506. Doi:10.1038/Ng1543.
- Schweers, O; Schonbrunnhanebeck, E; Marx, A and Mandelkow, E (1994). Structural studies of tau-protein and alzheimer paired helical filaments show no evidence for beta-structure. *Journal of Biological Chemistry* 269(39): 24290-24297.
- Selenko, P and Wagner, G (2007). Looking into live cells with in-cell NMR spectroscopy. *Journal of Structural Biology* 158(2): 244-253. Doi:10.1016/j.jsb.2007.04.001.
- Shao, HB; Chu, LY; Jaleel, CA and Zhao, CX (2008). Water-deficit stress-induced anatomical changes in higher plants. *Comptes Rendus Biologies* 331(3): 215-225. Doi:10.1016/j.crv.2008.01.002.
- Shih, MD; Hsieh, TY; Jian, WT; Wu, MT; Yang, SJ; Hoekstra, FA and Hsing, YIC (2012). Functional studies of soybean (*Glycine max* L.) seed LEA proteins GmPM6, GmPM11, and GmPM30 by CD and FTIR spectroscopy. *Plant Science* 196: 152-159. Doi:10.1016/J.Plantsci.2012.07.012.
- Shih, MD; Hsieh, TY; Lin, TP; Hsing, YIC and Hoekstra, FA (2010). Characterization of two soybean (*Glycine max* L.) lea IV proteins by circular dichroism and fourier transform infrared spectrometry. *Plant Cell and Physiology* 51(3): 395-407. Doi:10.1093/Pcp/Pcq005.
- Shih, MD; Lin, SD; Hsieh, JS; Tsou, CH; Chow, TY; Lin, TP and Hsing, YIC (2004). Gene cloning and characterization of a soybean (*Glycine max* L.) LEA protein, GmPM16. *Plant Molecular Biology* 56(5): 689-703. Doi:10.1007/S11103-004-4680-3.
- Shimizu, T; Kanamori, Y; Furuki, T; Kikawada, T; Okuda, T; Takahashi, T; Mihara, H and Sakurai, M (2010). Desiccation-induced structuralization and glass formation of group 3 late embryogenesis abundant protein model peptides. *Biochemistry* 49(6): 1093-1104. Doi:10.1021/Bi901745f.
- Shinozaki, K and Yamaguchi-Shinozaki, K (2007). Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany* 58(2): 221-227. Doi:10.1093/jxb/erl164.

- Shinozaki, K and YamaguchiShinozaki, K (1996). Molecular responses to drought and cold stress. *Current Opinion in Biotechnology* 7(2): 161-167. Doi:10.1016/S0958-1669(96)80007-3.
- Sickmeier, M; Hamilton, JA; LeGall, T; Vacic, V; Cortese, MS; Tantos, A; Szabo, B; Tompa, P, *et al.* (2007). Disprot: The database of disordered proteins. *Nucleic Acids Research* 35: D786-D793. Doi:10.1093/nar/gkl893.
- Singh, S; Cornilescu, CC; Tyler, RC; Cornilescu, G; Tonelli, M; Lee, MS and Markley, JL (2005). Solution structure of a late embryogenesis abundant protein (LEA14) from *Arabidopsis thaliana*, a cellular stress-related protein. *Protein Science* 14(10): 2601-2609. Doi:10.1110/Ps.051579205.
- Sivakolundu, SG; Bashford, D and Kriwacki, RW (2005). Disordered p27(kip1) exhibits intrinsic structure resembling the cdk2/cyclin a-bound conformation. *Journal of Molecular Biology* 353(5): 1118-1128. Doi:10.1016/j.jmb.2005.08.074.
- Solomon, A; Salomon, R; Paperna, I and Glazer, I (2000). Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat-stable protein. *Parasitology* 121: 409-416. Doi:10.1017/S0031182099006563.
- Soroka, J; Wandinger, SK; Mausbacher, N; Schreiber, T; Richter, K; Daub, H and Buchner, J (2012). Conformational switching of the molecular chaperone HSP90 via regulated phosphorylation. *Molecular Cell* 45(4): 517-528. Doi:10.1016/j.molcel.2011.12.031.
- Soulages, JL; Kim, K; Arrese, EL; Walters, C and Cushman, JC (2003). Conformation of a group 2 late embryogenesis abundant protein from soybean. Evidence of poly (L-proline)-type II structure. *Plant Physiology* 131(3): 963-975. Doi:10.1104/pp.015891.
- Soulages, JL; Kim, K; Walters, C and Cushman, JC (2002). Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. *Plant Physiology* 128(3): 822-832. Doi:10.1104/Pp.010521.
- Stacy, RAP and Aalen, RB (1998). Identification of sequence homology between the internal hydrophilic repeated motifs of group 1 late-embryogenesis-abundant proteins in plants and hydrophilic repeats of the general stress protein Gsib of *Bacillus subtilis*. *Planta* 206(3): 476-478. Doi:10.1007/S004250050424.
- Sugase, K; Dyson, HJ and Wright, PE (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447(7147): 1021-1025. Doi:10.1038/nature05858.
- Sugase, K; Landes, MA; Wright, PE and Martinez-Yamout, M (2008). Overexpression of post-translationally modified peptides in *Escherichia coli* by co-expression with modifying enzymes. *Protein Expression and Purification* 57(2): 108-115. 10.1016/j.pep.2007.10.018.
- Sun, XL; Rikkerink, EHA; Jones, WT and Uversky, VN (2013). Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* 25(1): 38-55. Doi:10.1105/Tpc.112.106062.
- Sunkar (2010). Plant stress tolerance: Methods and protocols. *Plant Stress Tolerance: Methods and Protocols* 639: 1-386. Doi:10.1007/978-1-60761-702-0.
- Suskiewicz, MJ; Sussman, JL; Silman, I and Shaul, Y (2011). Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein Science* 20(8): 1285-1297. Doi:10.1002/Pro.657.

- Svensson, J; Palva, ET and Welin, B (2000). Purification of recombinant *Arabidopsis thaliana* dehydrins by metal ion affinity chromatography. *Protein Expression and Purification* 20(2): 169-178. Doi:10.1006/Prep.2000.1297.
- Szasz, C; Alexa, A; Toth, K; Rakacs, M; Langowski, J and Tompa, P (2011). Protein disorder prevails under crowded conditions. *Biochemistry* 50(26): 5834-5844. Doi:10.1021/Bi200365j.
- Takaoka, Y; Kioi, Y; Morito, A; Otani, J; Arita, K; Ashihara, E; Ariyoshi, M; Tochio, H, *et al.* (2013). Quantitative comparison of protein dynamics in live cells and *in vitro* by in-cell (19)F-NMR. *Chemical Communications (Camb)* 49(27): 2801-2803. Doi:10.1039/c3cc39205h.
- Tanaka, S; Ikeda, K and Miyasaka, H (2004). Isolation of a new member of group 3 late embryogenesis abundant protein gene from a halotolerant green alga by a functional expression screening with cyanobacterial cells. *FEMS Microbiology Letters* 236(1): 41-45. Doi:10.1016/j.femsle.2004.05.019.
- Tapley, TL; Korner, JL; Barge, MT; Hupfeld, J; Schauerte, JA; Gafni, A; Jakob, U and Bardwell, JCA (2009). Structural plasticity of an acid-activated chaperone allows promiscuous substrate binding. *Proceedings of the National Academy of Sciences of the United States of America* 106(14): 5557-5562. Doi:10.1073/Pnas.0811811106.
- Thalhammer, A; Bryant, G; Sulpice, R and Hinch, DK (2014). Disordered cold regulated 15 proteins protect chloroplast membranes during freezing through binding and folding, but do not stabilize chloroplast enzymes *in vivo*. *Plant Physiology* 166(1): 190-201. Doi:10.1104/Pp.114.245399.
- Tolte, D; Jaquinod, M; Mangavel, C; Passirani, C; Saulnier, P; Manon, S; Teyssier, E; Payet, N, *et al.* (2007). Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *Plant Cell* 19(5): 1580-1589. Doi:10.1105/Tpc.107.050104.
- Tompa, P (2002). Intrinsically unstructured proteins. *Trends in Biochemical Sciences* 27(10): 527-533. Pii S0968-0004(02)02169-2 Doi:10.1016/S0968-0004(02)02169-2.
- Tompa, P (2010). Structure and function of intrinsically disordered proteins. Boca Raton, FL, Chapman & Hall/CRC Press.
- Tompa, P (2012). Intrinsically disordered proteins: A 10-year recap. *Trends in Biochemical Sciences* 37(12): 509-516. Doi:10.1016/J.Tibs.2012.08.004.
- Tompa, P; Dosztanyi, Z and Simon, I (2006). Prevalent structural disorder in *E coli* and *S cerevisiae* proteomes. *Journal of Proteome Research* 5(8): 1996-2000. Doi:10.1021/pr0600881.
- Tompa, P and Fuxreiter, M (2008). Fuzzy complexes: Polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences* 33(1): 2-8. Doi:10.1016/j.tibs.2007.10.003.
- Tompa, P and Kovacs, D (2010). Intrinsically disordered chaperones in plants and animals. *Biochemistry and Cell Biology* 88(2): 167-174. Doi:10.1139/O09-163.
- Tsvetkov, P; Reuven, N and Shaul, Y (2009). The nanny model for IDPs. *Nature Chemical Biology* 5(11): 778-781. Doi:10.1038/Nchembio.233.
- Tunnacliffe, A; Lapinski, J and McGee, B (2005). A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers. *Hydrobiologia* 546: 315-321. Doi:10.1007/s10750-005-4239-6.
- Tunnacliffe, A and Wise, MJ (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94(10): 791-812. Doi:10.1007/S00114-007-0254-Y.

- Uversky, VN (2002). Natively unfolded proteins: A point where biology waits for physics. *Protein Science* 11(4): 739-756. Doi:10.1110/ps.4210102.
- Uversky, VN (2013). Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta* 1834(5): 932-951. Doi:10.1016/j.bbapap.2012.12.008.
- Uversky, VN; Gillespie, JR and Fink, AL (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins-Structure Function and Genetics* 41(3): 415-427. Doi:10.1002/1097-0134(20001115)41:3<415::Aid-Prot130>3.0.Co;2-7.
- Uversky, VN; Oldfield, CJ and Dunker, AK (2008). Intrinsically disordered proteins in human diseases: Introducing the d(2) concept. *Annual Review of Biophysics* 37: 215-246. Doi:10.1146/annurev.biophys.37.032807.125924.
- Vicient, CM; Gruber, V and Delseny, M (2001). The Arabidopsis AtEM1 promoter is active in *Brassica napus* and is temporally and spatially regulated. *Journal of Experimental Botany* 52(360): 1587-1591. Doi:10.1093/Jexbot/52.360.1587.
- Wang, WW; Meng, B; Chen, WH; Ge, XM; Liu, S and Yu, J (2007). A proteomic study on postdiapaused embryonic development of brine shrimp (*Artemia franciscana*). *Proteomics* 7(19): 3580-3591. Doi:10.1002/pmic.200700259.
- Ward, JJ; Sodhi, JS; McGuffin, LJ; Buxton, BF and Jones, DT (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* 337(3): 635-645. Doi:10.1016/j.jmb.2004.02.002.
- Weinreb, PH; Zhen, WG; Poon, AW; Conway, KA and Lansbury, PT (1996). NACP, a protein implicated in alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35(43): 13709-13715. Doi:10.1021/Bi961799n.
- Wells, M; Tidow, H; Rutherford, TJ; Markwick, P; Jensen, MR; Mylonas, E; Svergun, DI; Blackledge, M, *et al.* (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proceedings of the National Academy of Sciences of the United States of America* 105(15): 5762-5767. Doi:10.1073/pnas.0801353105.
- Whitmore, L and Wallace, BA (2004). Dichroweb, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research* 32: W668-W673. Doi:10.1093/Nar/Gkh371.
- Winter, D; Vinegar, B; Nahal, H; Ammar, R; Wilson, GV and Provart, NJ (2007). An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *Plos One* 2(8). Doi:10.1371/journal.pone.0000718.
- Wisniewski, M; Webb, R; Balsamo, R; Close, TJ; Yu, XM and Griffith, M (1999). Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: A dehydrin from peach (*Prunus persica*). *Physiologia Plantarum* 105(4): 600-608. Doi:10.1034/J.1399-3054.1999.105402.X.
- Wolkers, WF; McCreedy, S; Brandt, WF; Lindsey, GG and Hoekstra, FA (2001). Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses *in vitro*. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* 1544(1-2): 196-206. Doi:10.1016/S0167-4838(00)00220-X.
- Wright, PE and Dyson, HJ (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* 293(2): 321-331. Doi:10.1006/Jmbi.1999.3110.

- Wright, PE and Dyson, HJ (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* 16(1): 18-29. Doi:10.1038/Nrm3920.
- Xie, HB; Vucetic, S; Iakoucheva, LM; Oldfield, CJ; Dunker, AK; Uversky, VN and Obradovic, Z (2007). Functional anthology of intrinsic disorder. Biological processes and functions of proteins with long disordered regions. *Journal of Proteome Research* 6(5): 1882-1898. Doi:10.1021/pr060392u.
- Yamasaki, H; Pilon, M and Shikanai, T (2008). How do plants respond to copper deficiency? *Plant Signal Behav* 3(4): 231-232.
- Yoshida, T; Mogami, J and Yamaguchi-Shinozaki, K (2014). ABA-dependent and aba-independent signaling in response to osmotic stress in plants. *Current Opinion in Plant Biology* 21: 133-139. Doi:10.1016/j.pbi.2014.07.009.
- Zhuravleva, A and Gierasch, LM (2011). Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (HSP70) molecular chaperones. *Proceedings of the National Academy of Sciences of the United States of America* 108(17): 6987-6992. Doi:10.1073/pnas.1014448108.
- Zimmerman, SB and Minton, AP (1993). Macromolecular crowding - biochemical, biophysical, and physiological consequences. *Annual Review of Biophysics and Biomolecular Structure* 22: 27-65. Doi:10.1146/Annurev.Bb.22.060193.000331.
- Zimmerman, SB and Trach, SO (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *Journal of Molecular Biology* 222(3): 599-620. Doi:10.1016/0022-2836(91)90499-V.

PROLOGO A LOS RESULTADOS ANEXOS

Durante la realización de la presente tesis se obtuvieron una serie de resultados que nos ayudan a comprender mejor la función molecular de las proteínas LEA del grupo 4. Los resultados más completos culminaron en la publicación de un artículo científico: Cuevas-Velazquez CL, Saab-Rincón G, Reyes JL, Covarrubias AA. The unstructured N-terminal region of Arabidopsis Group 4 Late Embryogenesis Abundant (LEA) proteins is required for folding and for chaperone-like activity under water deficit. *Journal of Biological Chemistry* 291, 10893-10903 (2016). Estos resultados constituyen la parte central de esta tesis.

Adicionalmente, en las siguientes secciones anexas I y II (Anexo I y Anexo II) presento aquellos resultados que se obtuvieron durante mis estudios de doctorado pero que aún no culminan en la publicación de un artículo científico. En el Anexo I se muestran los mecanismos de protección de la AtLEA4-5, en donde se buscó caracterizar la posible interacción entre dicha proteína con su sustrato modelo, la lactato deshidrogenasa (LH). En el Anexo II se describe la caracterización de otra función propuesta para las proteínas LEA: la interacción con iones metálicos. Para el mejor entendimiento de dichas secciones anexas, ambas contienen introducción, antecedentes, pregunta, objetivo general, objetivos específicos, materiales y métodos, resultados, discusión, conclusiones y perspectivas. Es importante hacer notar que, en ambos casos, es indispensable realizar experimentos adicionales para validar las observaciones presentadas en estas secciones. Así pues, al momento de la publicación de esta tesis, se están desarrollando dos proyectos de doctorado que se fundamentan directamente en cada uno de dichos Anexos.

ANEXO I. Descifrando el mecanismo de protección a nivel molecular de las proteínas abundantes en la embriogénesis tardía.

i.i. INTRODUCCIÓN

i.i.i. Mecanismos de acción de las chaperonas moleculares y las sHSPs

Durante la evolución, los organismos vivos han desarrollado la capacidad de acumular proteínas que les permiten proteger a sus células de los daños causados por diferentes condiciones de estrés (calor, congelamiento, estrés por pH, oxidativo, etc.). Un grupo, amplio y diverso, particularmente interesante es el que conforman las chaperonas moleculares, las cuales están clasificadas en varias familias de acuerdo a su masa molecular y a su historia evolutiva (Bukau *et al.* 2006). Estas proteínas usan mecanismos ingeniosos para prevenir la agregación y para promover un plegamiento eficiente. Las chaperonas moleculares son necesarias para el mantenimiento de la homeostasis del *proteoma* (*proteostasis*), lo cual es crucial para una función celular adecuada del organismo (Hartl *et al.* 2011). Algunos autores definen a las chaperonas moleculares como cualquier proteínas que interacciona, estabiliza o ayuda a otra proteína a adquirir su conformación funcional (activa), sin que permanezca presente en la estructura final de dicha proteína (Hartl and Hayer-Hartl 2009; Hartl *et al.* 2011). Existen diversas clases de chaperonas las cuales forman una red de vías cooperativas. Los miembros de estas familias de proteínas se conocen usualmente como proteínas de estrés o proteínas de choque térmico (HSPs, del inglés Heat Shock Proteins) debido a que éstas se acumulan bajo dichas condiciones, cuando los niveles de las proteínas susceptibles a agregarse aumentan. Usualmente se clasifican según su peso molecular: HSP40, HSP60, HSP70, HSP90, HSP100 y las HSPs pequeñas (o sHSPs) (Hartl *et al.* 2011). Estas proteínas están involucradas en una multitud de funciones del mantenimiento del *proteoma*, entre las que destacan el plegamiento *de novo*, re-plegamiento de proteínas desnaturalizadas a causa del estrés, ensamblaje de oligómeros y transporte de proteínas (Hartl and Hayer-Hartl 2009). Las proteínas que participan en el plegamiento *de novo* y en el re-plegamiento, como la HSP70, HSP90 y las chaperoninas (HSP60s) son grandes máquinas moleculares que promueven el plegamiento a través de ciclos de unión y liberación regulados por ATP. Usualmente, reconocen residuos hidrofóbicos expuestos en las proteínas mal plegadas y en ocasiones pueden cooperar con otras chaperonas que no requieren ATP para su

función como las sHSPs, las cuales previenen la agregación de proteínas (Bukau *et al.* 2006).

En el mecanismo de acción de las chaperonas dependientes de ATP, el plegamiento *de novo* y el re-plegamiento se promueven a través de una *partición dinámica*. Este fenómeno está implícito en el funcionamiento de estas proteínas: primero, la unión de la chaperona a las regiones hidrofóbicas de una proteína mal plegada bloquea transitoriamente la agregación, en tanto que la liberación inducida por la hidrólisis de ATP permite que el plegamiento continúe (Bukau *et al.* 2006). Si bien, la HSP70 y las chaperoninas funcionan con este mecanismo básico, la diferencia fundamental es que la HSP70 (así como las otras chaperonas dependientes de ATP) libera al sustrato a la solución acuosa, mientras que las chaperoninas cilíndricas permiten que el plegamiento ocurra dentro de una especie de “jaula” (Hartl *et al.* 2011). Los dos sistemas actúan de forma secuencial. La HSP70 interacciona primero con el polipéptido naciente para que posteriormente la chaperonina funcione en el paso final del plegamiento, en particular de aquellas proteínas que no se plegaron completamente con la ayuda de HSP70 (Hara *et al.* 2009).

Las sHSPs están presentes en los tres dominios de la vida; sin embargo, están mucho menos conservadas que las chaperonas moleculares (Kriehuber *et al.* 2010). Las dos sHSPs más estudiadas son las α -cristalinas (α A y α B), las cuales constituyen el 30% de las proteínas del lente en los ojos de los vertebrados. En el lente del ojo, su función es mantener la solubilidad de otras proteínas lenticulares y así prevenir la fuga de luz (Haslbeck and Vierling 2015). Las sHSPs están compuestas de una secuencia N-terminal no conservada y variable en longitud (SNT) y que posee una estructura intrínsecamente desordenada, un dominio α -cristalino globular conservado (DAC) y una secuencia corta en el C-terminal (SCT). En varias bacterias (por ejemplo, *Synechocystis sp* 6803), existe sólo una sHSP (Hsp 16.6) que es esencial para la termotolerancia (Giese and Vierling 2002). En otras bacterias como *E. coli*, existen dos o más sHSPs (IbpA e IbpB) las cuales forman hetero-oligómeros que funcionan de manera cooperativa (Ratajczak *et al.* 2009). En otras bacterias como *D. radiodurans*, existen sHSPs que funcionan en paralelo, independientemente una de la otra (Bepperling *et al.* 2012). En levaduras, también hay dos sHSPs oligoméricas que actúan de forma paralela e independiente. En plantas superiores existen al menos 12

familias génicas que codifican para sHSPs. La mayoría de las sHSPs forman oligómeros de 12 hasta 32 monómeros (Basha *et al.* 2012).

Las sHSPs de diferentes especies pueden actuar como chaperonas moleculares uniendo proteínas parcialmente desnaturalizadas y evitando su subsecuente agregación sin necesidad de usar ATP (Haslbeck and Vierling 2015). Las sHSPs realizan sus función estabilizando intermediarios en el proceso de desplegamiento de proteínas causado por situaciones de estrés como altas temperaturas o estrés oxidativo. Para que sean capaces de ejercer esta función, las sHSPs deben estar presentes durante el proceso de desplegamiento del sustrato puesto que no pueden rescatar sustratos que ya fueron desnaturalizados o que ya experimentaron agregación (Basha *et al.* 2012). Para lograr ésto, se ha observado que los oligómeros de las sHSPs se disocian en dímeros los cuales son los responsables de unir a los sustratos parcialmente desplegados. Posteriormente, estas estructuras se asocian para formar “agregados” solubles de alto peso molecular (Haslbeck and Vierling 2015). De esta manera, en la *proteostasis* de la célula, las sHSPs son una especie de amortiguador que une proteínas parcialmente desplegadas cuando las células perciben el estrés, previniendo así que se formen agregados irreversibles. Experimentos *in vitro* han revelado que las proteínas atrapadas en los complejos sHSPs/sustrato son liberadas y re-plegadas en presencia de las chaperonas moleculares dependientes de ATP (Lee and Vierling 2000; Haslbeck and Vierling 2015).

El espectro de proteínas encontradas como sustratos potenciales de las sHSPs durante el estrés indica que éstas son muy promiscuas, pero tienen cierta preferencia por proteínas relacionadas al proceso de traducción; por ejemplo, proteínas ribosomales, factores de traducción y amino-acil tRNA sintetasas; así como, algunas enzimas metabólicas (Haslbeck and Vierling 2015). En términos de la relación molar de sHSP requerida para proteger a los sustratos, se ha encontrado que las sHSPs son menos efectivas previniendo la agregación de proteínas muy grandes, lo cual sugiere que la interacción depende de la relación por masa en vez de la relación molar a través de una interacción mediada por cargas o por interacciones hidrofóbicas (Bukau *et al.* 2006). Distintos estudios sugieren que la región desordenada SNT es la responsable de la interacción con las regiones hidrofóbicas de los sustratos parcialmente desplegados (Basha *et al.* 2012), lo cual podría indicar que el carácter desordenado de la región SNT es importante para el acoplamiento adecuado de distintos sustratos.

i.i.ii. Proteínas condicionalmente desordenadas

Las proteínas condicionalmente desordenadas son proteínas que pueden existir en al menos dos estados, uno que contienen altos niveles de flexibilidad y otro donde la proteína está más ordenada (Bardwell and Jakob 2012). Las proteínas que tienen alto nivel de flexibilidad o desorden usualmente participan en múltiples interacciones mutuamente excluyentes; tal es el caso de las chaperonas, las cuales actúan a través de la interacción con una amplia variedad de proteínas blanco (sustratos) en diferentes estados de plegamiento y así evitan su agregación inespecífica, promoviendo a su vez su re-plegamiento. Los niveles de desorden predichos para las chaperonas moleculares oscila entre el 24% y el 100% (Tompa and Kovacs 2010). Las chaperonas que pliegan (conocidas como “foldasas”) como la Hsp70, Hsp60 y Hsp90 experimentan grandes rearrreglos conformacionales que modulan la interacción con el sustrato. Algunos de estos cambios conformacionales han sido asociados con regiones intrínsecamente desordenadas (Bardwell and Jakob 2012). En Hsp70, las regiones que contienen el dominio de unión al nucleótido y el de unión al *cliente* es altamente flexible, lo que permite que existan grandes cambios conformacionales inter-dominio (Zhuravleva and Gierasch 2011). Hsp60 contiene un 14-mero en su región C-terminal, el cual es desordenado cuando la proteína no une ATP pero se encuentra plegado en la isoforma que une ATP (Clare *et al.* 2012). Hsp90 tiene múltiples regiones desordenadas que probablemente ejecuten papeles importantes en su actividad como chaperona; por ejemplo, confiriendo solubilidad al complejo Hsp90-cliente. Además, muchos de los sitios de fosforilación en Hsp90 están localizados en sus regiones desordenadas que están potencialmente involucradas en transiciones de orden a desorden durante el ciclo de unión e hidrólisis de ATP (Soroka *et al.* 2012).

Si bien, las chaperonas moleculares dependientes de ATP han sido ampliamente estudiadas en los últimos 25 años, ha sido muy complicado establecer con claridad cual es el papel de las regiones intrínsecamente desordenadas en la función de este tipo de chaperonas, probablemente debido a que estas proteínas poseen estructuras cuaternarias grandes y complejas y las interacciones entre las chaperonas y sus clientes parcialmente desplegados son técnicamente muy difíciles de resolver. Sin embargo, recientemente se han encontrado un grupo de chaperonas

pequeñas, altamente promiscuas e independientes de ATP, las cuales se activan específicamente en condiciones de estrés. Es justamente el estrés el que convierte a estas proteínas globulares y bien plegadas en chaperonas intrínsecamente desordenadas, o también conocidas como *chaperonas condicionalmente desordenadas* (Bardwell and Jakob 2012).

Existen evidencias recientes que demuestran que algunas chaperonas independientes de ATP experimentan grandes cambios conformacionales conocidos como transiciones de orden a desorden lo cual activa a la chaperona y permite la interacción con sus clientes y a su vez, usan transiciones de desorden a orden para controlar la liberación y entrega del cliente a las chaperonas moleculares dependientes de ATP. Dos ejemplos muy bien caracterizados de este tipo de chaperonas son una que es activada por estrés ácido conocida como HdeA (Tapley *et al.* 2009) y otra que está regulada por el estatus redox de la célula llamada Hsp33 (Reichmann *et al.* 2012). Las condiciones de estrés que activan a estas chaperonas (bajo pH y estrés oxidativo) generan la desnaturalización masiva de proteínas, lo cual es una grave amenaza para la integridad de la célula. HdeA y Hsp33 también se desnaturalizan en estas condiciones de estrés pero en su caso, esta desnaturalización promueve su activación como chaperonas. Este tipo de regulación post-traducciona es órdenes de magnitud más rápida que la transcripción y la traducción, las cuales de hecho se reducen durante dichas condiciones de estrés (Bardwell and Jakob 2012).

HdeA es una chaperona pequeña (10 kDa) de *E. coli* que se activa por estrés ácido (por ejemplo, cuando algún mamífero ingiere a la bacteria y ésta se enfrenta a los bajos niveles de acidez del estómago). A pH 7, HdeA conforma un dímero plegado y sin actividad de chaperona pero cuando el pH se reduce a 2, HdeA se desnaturaliza rápidamente (0.125 segundos) y se separa en monómeros, lo cual activa la función de chaperona (Tapley *et al.* 2009). Esta separación induce que la zona involucrada en la dimerización y que además es la responsable de interactuar con los *clientes* se convierta en una región intrínsecamente desordenada. La naturaleza parcialmente desplegada de HdeA a pH bajos es vital para su habilidad de interactuar con una amplia variedad de sustratos. Estudios de FRET enfocados en monitorear el complejo HdeA-cliente muestran que HdeA adopta múltiples conformaciones cuando interactúa con diferentes sustratos (Tapley *et al.* 2009). Se ha propuesto que después de que la bacteria pasa por el ambiente ácido del estómago y transita al pH neutro del intestino delgado, entonces HdeA libera lentamente a sus

clientes. La liberación lenta de éstos ayuda a mantener baja la concentración de intermediarios sensibles a agregarse, permitiendo que los clientes se re-plieguen pasivamente y adquieran nuevamente su conformación funcional.

Hsp33 es otro ejemplo de una proteína condicionalmente desordenada que se activa en condiciones de estrés, en este caso, por estrés oxidativo. De forma similar a HdeA, Hsp33 es una proteína compacta y plegada en condiciones normales. Hsp33 es monomérica y contiene cuatro cisteínas absolutamente conservadas que están encargadas de unir un ion de zinc (Reichmann *et al.* 2012). La unión del zinc estabiliza la estructura plegada de la proteína. La exposición a condiciones oxidantes causa que la proteína se desnaturalice lo que induce la formación de dos enlaces disulfuro intra-moleculares así como la liberación del zinc, lo que desestabiliza el dominio de unión a zinc, despliega la zona adyacente conocida como región “*linker*” y, finalmente, la dimerización de la molécula. El desplegamiento de la región *linker* es crucial para la activación pues una mutación puntual que altera exclusivamente los niveles de plegamiento de esta región es suficiente para activar constitutivamente a la chaperona (Cremers *et al.* 2010). Una vez activa, la Hsp33 protege cientos de *clientes* citosólicos de la agregación producida por el estrés. Estudios recientes sugieren que la Hsp33 interacciona con intermediarios de plegamiento que contienen ciertos niveles de estructura secundaria, a través de su región *linker* intrínsecamente desordenada (Reichmann *et al.* 2012). La estabilización de la región intrínsecamente desordenada de Hsp33 después de unirse a los elementos de estructura secundaria de sus proteínas cliente contribuye a la alta afinidad por los mismos y provee información valiosa para entender los mecanismo que utilizan las chaperonas para seleccionar a sus proteínas *cliente*. Análisis estructurales del complejo Hsp33-*cliente* después del estrés, cuando vuelven a condiciones reductoras no estresantes, muestra que existen cambios conformacionales considerables tanto en la chaperona como en la proteína *cliente*, consistente con el replegamiento parcial del dominio de unión a zinc de Hsp33 y una mayor desestabilización de la proteína cliente (Bardwell and Jakob 2012). Estos resultados son consistentes con el modelo de “*transferencia entrópica*”, el cual propone que el costo entrópico del replegamiento de las regiones intrínsecamente desordenadas de las chaperonas al unir a sus proteínas *cliente*, se utiliza para desnaturalizar completamente a las proteínas *cliente*. El desplegamiento de las proteínas *cliente* disminuye la afinidad de éstas por Hsp33, y facilita su liberación, preparándolas para transferirlas a las chaperonas moleculares dependientes de ATP,

las cuales reconocen proteínas *cliente* desplegadas a las que eventualmente replegarán.

Es muy claro que el desorden estructural juega un papel fundamental en la función de las chaperonas independientes de ATP. Existen otros ejemplos de este comportamiento, entre los cuales destaca la actividad de chaperona de las sHSPs; así como, la actividad tipo chaperona de algunas proteínas globalmente desordenadas, como las caseínas, la α -sinucleína y las proteínas LEA.

i.i.iii. Plegamiento de las proteínas intrínsecamente desordenadas asociado a la interacción con sus blancos

La transición de un estado desordenado en solución a un estado ordenado cuando está implicado en la interacción con otras moléculas es una característica muy común en la función de algunas proteínas intrínsecamente desordenadas (Wright and Dyson 1999). Si bien, no existe un consenso para este mecanismo conocido como *plegamiento asociado a la interacción*, múltiples evidencias sugieren que existen dos mecanismos generales. En un extremo se encuentra el mecanismo del “*acoplamiento inducido*” el cual asume que el plegamiento ocurre después de que la proteína intrínsecamente desordenada interacciona con su blanco (Dyson and Wright 2002). En el otro extremo se encuentra el mecanismo de “*selección conformacional*” en el cual, la estructura final (plegada) de la proteína desordenada, cuando ésta ya está unida a su ligando, se selecciona de un conjunto de conformeros que ya existen para esa proteína en solución, los cuales están en equilibrio con sus conformaciones desordenadas (Berlow *et al.* 2015). Otra de las posibilidades es que la mayoría de las proteínas intrínsecamente desordenadas se unan a su blanco a través de una combinación de estos dos mecanismos. De hecho, un estudio reciente muestra que dos proteínas intrínsecamente desordenadas diferentes pueden interaccionar con el mismo blanco en el mismo sitio de interacción pero a través de diferentes mecanismos de plegamiento asociado a la interacción (Arai *et al.* 2015).

i.i.iv. Complejos fuzzy

Las proteínas intrínsecamente desordenadas consisten de un conjunto de conformaciones altamente flexibles en el estado libre; sin embargo, durante muchos años se pensó que éstas se comportaban como proteínas “normales” cuando interactuaban con un blanco, pues en este estado se encontraban de manera plegada (Dyson and Wright 2002). Esto realmente representaba un sesgo en las interpretaciones de los resultados puesto que la imagen estructural de las PIDs unidas a sus moléculas blanco refleja las dificultades experimentales en la caracterización de complejos proteicos flexibles. De hecho, existen algunas proteínas que mantienen estas estructuras desordenadas cuando interactúan con otras moléculas. Estas zonas que se mantienen desordenadas coinciden con regiones funcionalmente críticas, lo que sugiere que este desorden estructural en el complejo es de relevancia biológica. A las PIDs que mantienen el estado desordenado, aún al interactuar con su blanco, se les conoce como proteínas *fuzzy* o complejos *fuzzy* (Tompa and Fuxreiter 2008). De la misma forma que el desorden estructural en el estado libre, el desorden en los complejos *fuzzy* está compuesto de distintos estados conformacionales, desde desorden local hasta desorden global; o bien, desde un estado compacto a un estado extendido. Los complejos *fuzzy* han sido clasificados en cuatro categorías: Complejos polimórficos, complejos abrazadera, complejos flanco y complejos aleatorios. En los complejos polimórficos, por lo menos uno de los interactores adopta varias conformaciones. En los complejos flanco, como en los complejos abrazadera, el segmento desordenado se encuentra adyacente o conecta regiones de interacción ordenadas. Los complejos aleatorios son un caso extremo de los complejos *fuzzy* y no se observa ningún plegamiento asociado a la interacción (Tompa and Fuxreiter 2008).

La heterogeneidad conformacional de los complejos *fuzzy* puede resultar en contactos alternativos, permanentes o transitorios con las moléculas blanco. Variaciones en las interacciones permanentes pueden provocar morfologías diferentes en complejos grandes. Los contactos alternativos están formados usualmente por interacciones electrostáticas. El inhibidor de las quinasas dependientes de ciclinas Sic1 tienen nueve sitios de reconocimiento de la proteínas Cdc4. Cualquiera de los nueve sitios de reconocimiento es capaz de interactuar con un solo sitio en Cdc4. Las interacciones se inducen por la fosforilación de ciertos motivos lo que genera orden estructural transitorio alrededor de los mismos. Sin embargo, el complejo en su

totalidad permanece dinámico, lo que hace que los sitios de fosforilación sean intercambiables (Mittag *et al.* 2008).

La plasticidad de las regiones intrínsecamente desordenadas usualmente permite la adaptación de un mismo motivo para interactuar con distintas proteínas blanco. Por ejemplo, el segmento rico en prolina de la proteína MBP puede interactuar con las proteínas Yes1, PSD95, cortactina, PexD, Abl, Fyn, c-Src e Itk (Polverini *et al.* 2008). Además del contacto directo entre los residuos de prolina, todos estos complejos son estabilizados por interacciones electrostáticas con regiones desordenadas de MBP, lejanas al sitio de interacción con las proteínas (Ahmed *et al.* 2009).

i.ii. ANTECEDENTES

En esta tesis se demostró que las proteínas LEA del grupo 4 (AtLEA4-2 y AtLEA4-5) son PIDs en solución acuosa y que cambian su conformación plegándose en α -hélice cuando se encuentran en condiciones de déficit hídrico simulado *in vitro*. Además, se encontró que la región N-terminal conservada de la AtLEA4-5 es necesaria y suficiente para inducir el cambio conformacional y para prevenir la pérdida de actividad de la lactado deshidrogenasa (LDH) causada por tratamientos de disminución en la cantidad de agua disponible *in vitro*. Si bien, los niveles de protección de las dos proteínas LEA del grupo 4 y de la región N-terminal de la AtLEA4-5 son muy evidentes, aún queda por identificar el mecanismo molecular a través del cual ejercen dicha protección. Existen dos hipótesis principales que tratan de explicar estos mecanismos de acción. La hipótesis del escudo molecular establece que las proteínas LEA son escudos o barreras físicas que impiden el contacto entre los parches hidrofóbicos de las proteínas cliente que se generan como consecuencia del estrés. De esta manera se estaría evitando no sólo la pérdida de función si no también la eventual agregación de las proteínas cliente. En este modelo no ocurre una interacción física entre las proteínas LEA y sus proteínas cliente y se requiere que la proteína LEA esté altamente desplegada con un radio hidrodinámico grande (Chakrabortee *et al.* 2012; Hughes *et al.* 2013). La otra hipótesis propone que las proteínas LEA son chaperonas que reconocen a las proteínas cliente a través de una interacción proteína-proteína, manteniendo así su viabilidad funcional mediante la transferencia de entropía de la chaperona a la proteína cliente. En este caso, debe

haber motivos específicos responsables del reconocimiento molecular entre ambas proteínas y no se necesitan cantidades muy elevadas de la chaperona respecto a la proteína cliente (Tompa and Kovacs 2010; Olvera-Carrillo *et al.* 2011). En un estudio reciente, en donde se compararon los datos obtenidos en distintos trabajos se propuso que los dos modelos mencionados anteriormente no son mutuamente excluyentes y que algunas proteínas LEA pudieran proteger de acuerdo al modelo del escudo molecular y otras protegen a través de mecanismos análogos a los descritos para otras chaperonas (Cuevas-Velazquez *et al.* 2014).

Una de las principales observaciones obtenidas en esta tesis es que las proteínas LEA del grupo 4 son capaces de proteger a la LDH a relaciones molares bajas (entre 1:1 y 20:1 respecto a la LDH). Lo anterior sugiere que estas proteínas protegen a través de un mecanismo tipo chaperona que involucra la interacción entre la proteína chaperona y la proteína cliente (Bardwell y Jakob, 2012). A la fecha sólo un estudio ha reportado la búsqueda de la interacción entre las dos proteínas. Hughes y colaboradores (2011) reportaron que una proteína LEA del grupo 2 no interacciona con la enzima reportera, en ese caso, la LDH. Dicho resultado es coherente con la hipótesis del escudo molecular pues para lograr proteger a la LDH se requirieron utilizar relaciones molares muy altas de la dehidrina (Hughes and Graether 2011).

La interacción entre las proteínas LEA y sus proteínas cliente también tiene relevancia desde el punto de vista de las PIDs. A la fecha parece haber coincidencia en que la mayoría de las PIDs de animales se mantienen desplegadas cuando están en su estado nativo y se pliegan cuando interaccionan con su proteína blanco (Dyson and Wright 2002; Dyson and Wright 2005; Sugase *et al.* 2007). El plegamiento asociado a la interacción con un blanco consiste en que una PID (o una región desordenada dentro de una proteína globular) se pliega y adquiere una estructura más estable al mismo tiempo que interacciona con una proteína blanco. Existe un costo entrópico cuando una proteína desordenada se pliega, el cual es cubierto por la entalpía de unión de las dos proteínas. Existen varias proteínas desordenadas de animales que realizan dicho mecanismo (Dyson and Wright 2002; Dyson and Wright 2005). Por ejemplo, el dominio de cinasa inducible (KID) en su estado fosforilado (pKID) del factor de transcripción llamado “proteína de unión al elemento de respuesta para AMP cíclico” (CREB) se encuentra desordenado en la proteína en solución acuosa; sin embargo, se pliega en α -hélice cuando interacciona y forma un complejo con el dominio de unión a KID (conocido como KIX) de la proteína de unión a CREB (CBP) (Dyson and

Wright 2005). Cabe la posibilidad de que las proteínas LEA del grupo 4 se comporten de la misma manera que las PIDs de animales y que se plieguen cuando interaccionen con sus proteínas *cliente*. Por otro lado, existe otro mecanismo de interacción para las PIDs en el cual éstas no se pliegan cuando interaccionan con su blanco y permanecen mayoritariamente desordenadas (Tompa and Fuxreiter 2008). Esta posibilidad es consistente para el caso de las proteínas LEA, pues les permitiría interaccionar con un catálogo de blancos aún mayor y, además, proteger una mayor área superficial de la proteína cliente.

Debido a lo expuesto hasta ahora, resulta relevante dilucidar el mecanismo de acción de las proteínas LEA del grupo 4 y determinar si protegen a través de una interacción proteína-proteína (modelo de protección tipo chaperona) o si funcionan como escudos moleculares, además de investigar cuál es la estructura que adquieren estas proteínas en el proceso de interacción.

i.iii. PREGUNTA

¿La proteína AtLEA4-5 interaccionan con el sustrato modelo (lactato deshidrogenasa)?

i.iv. OBJETIVO GENERAL

Determinar si la AtLEA4-5 interacciona con la lactato deshidrogenasa *in vitro*.

i.v. OBJETIVOS ESPECÍFICOS

- Determinar si la AtLEA4-5 interacciona con la enzima reportera lactato deshidrogenasa (LDH), utilizando un método de entrecruzamiento de proteínas que se encuentran a longitud cero.
- Analizar el efecto de la LDH sobre la estructura de la AtLEA4-5 usando la técnica de dicroísmo circular.
- Analizar el efecto de la LDH sobre la estructura de la AtLEA4-5 usando la técnica de resonancia magnética nuclear.

i.vi. MATERIALES Y MÉTODOS

i.vi.i. Entrecruzamiento foto-inducido in vitro.

Se realizaron ensayos de entrecruzamiento para determinar si existe una interacción entre la AtLEA4-5 y la lactato deshidrogenasa (LDH). Para ello, se siguió el método de entrecruzamiento descrito por Fancy y Kodadek (1999), con algunas modificaciones, el cual se denomina “entrecruzamiento foto-inducido de proteínas no modificadas” (Fancy and Kodadek 1999). Este método consiste en la generación de un enlace covalente entre residuos tirosina-tirosina, tirosina-cisteína y tirosina-lisina, si estos se encuentran a una longitud cero. La AtLEA4-5 recombinante (Ver materiales y métodos de esta tesis) se fijó a una concentración de 1.88 μM en un volumen total de 8 μL . En oscuridad, se agregó 1 μL de 12.5 mM Ru(II)bpy₃²⁺ y 1 μL de persulfato de amonio 25 mM. La mezcla se expuso a un pulso de luz de ocho segundos con un foco de 75 W a través de un filtro de agua de 2.5 cm. La reacción de entrecruzamiento se detuvo con 10 μL de amortiguador Laemmli 2X (62.25 mM Tris-HCl pH 6.8, 4% SDS, glicerol 10%, β -mercaptoetanol 10%, azul de bromofenol 0.02%). Una muestra con todos los componentes pero sin exponer al pulso de luz se estableció como la muestra sin entrecruzamiento. Para determinar la interacción, a la muestra de AtLEA4-5 (1.88 μM) se agregaron distintas cantidades de LDH (ver Materiales y métodos de esta tesis) para guardar relaciones molares de 0.5:1, 1:1, 5:1 y 10:1 (LDH:AtLEA4-5). Además, se preparó una muestra con sólo la LDH (sin la AtLEA4-5) a la relación molar más alta (10:1). Las muestras entrecruzadas y no entrecruzadas se separaron por SDS-PAGE. Las bandas correspondientes a la proteína AtLEA4-5 se detectaron por experimentos tipo Western usando un anticuerpo específico para la AtLEA4-5, descrito por Olvera-Carrillo y colaboradores (2010) (Olvera-Carrillo *et al.* 2010).

Para la albúmina sérica bovina (ASB) se siguió el mismo procedimiento pero en este caso se añadieron relaciones molares 1:1, 5:1, 7:1, 10:1, 15:1, 20:1 y 50:1 de la ASB (Sigma). Al igual que en el caso anterior, las bandas correspondientes a la proteína AtLEA4-5 se detectaron por experimentos tipo Western usando un anticuerpo específico para la AtLEA4-5,

i.vi.ii. Detección de cambios conformacionales por dicroísmo circular

La AtLEA4-5 y la LDH se diluyeron (por separado) a una concentración de 18.5 μM en amortiguador de fosfatos de sodio 10 mM pH 7.5 en presencia de 0%, 20% ó 60% (v/v) de glicerol. Los espectros de dicroísmo circular se adquirieron a 25 °C con un espectropolarímetro de dicroísmo circular Jasco J-715 (JASCO Analytical Instruments) con una longitud de paso óptico de 0.1 cm desde 190 nm hasta 250 nm. Los espectros de cada una de las proteínas (AtLEA4-5 o LDH) se midieron por separado y los espectros resultantes en cada condición se sumaron para obtener los espectros correspondientes a las sumas aritméticas. En otros tubos, se mezclaron las dos proteínas juntas a la misma concentración (18.5 μM) en cada una de los tratamientos (0%, 20% y 60% de glicerol) y se midieron los espectros de dicroísmo circular. Las muestras se incubaron durante una hora antes de la adquisición de los espectros. Los parámetros de medición fueron los mismos que se utilizaron en la sección de Materiales y métodos de esta tesis. Las mediciones se realizaron por lo menos tres veces con distintos lotes de purificación de la AtLEA4-5.

i.vi.iii. Detección de cambios conformacionales por resonancia magnética nuclear

Para la determinación de los posibles cambios conformacionales en la AtLEA4-5 generados por la LDH, se utilizó la ^{15}N -AtLEA4-5 obtenida en la sección “Expresión y purificación de proteínas marcadas con ^{15}N ” de esta tesis. Se tomaron 100 μL de una solución de ^{15}N -AtLEA4-5 a una concentración de 155 μM en amortiguador 20 mM Tris-HCl pH 6.8, 50 mM NaCl. y se pusieron en una celda para resonancia. Se adquirieron espectros tipo TROSY en un equipo VNMR-S de 900 MHz a 298 K. Esta muestra representa el espectro de la AtLEA4-5 en ausencia de la LDH. Por otro lado, 5 mL de LDH (5 mg/mL) se dializaron toda la noche contra amortiguador Tris-HCl 20 mM pH 6.8, 50 mM NaCl. La muestra se concentró hasta llegar a una concentración de 540 μM . La ^{15}N -AtLEA4-5 y la LDH se mezclaron a 70 μM y 280 μM , respectivamente, 100 μL de esta muestra se pusieron en una celda para resonancia y se adquirieron los espectros tipo TROSY. Finalmente, en otra muestra se combinaron la ^{15}N -AtLEA4-5 y la LDH a 30 μM y 410 μM , respectivamente, y 100

μL de esta muestra se pusieron en una celda para resonancia y se adquirieron los espectros tipo TROSY, de la misma forma que en los experimentos anteriores.

i.vii. RESULTADOS

i.vii.i. La AtLEA4-5 se entrecruza con la LDH in vitro

Con el objetivo de determinar la posible interacción entre la proteína LEA y la proteína *cliente* (LDH) se realizaron experimentos de entrecruzamiento, utilizando como agente entrecruzador al tris-bipiridil rutenio dicación (Ru(II)bpy_3^{2+}), el cual es un reactivo foto-activable que induce entrecruzamiento sin la necesidad de tener cambios químicos previos en las proteínas. En este procedimiento, también conocido como *entrecruzamiento foto-inducido de proteínas no modificadas* (PICUP del inglés *Photo Induced Crosslinking of Unmodified Proteins*), el entrecruzamiento ocurre a través de una reacción dada por la proteólisis rápida del complejo tris-bipiridil rutenio inducida por luz visible en presencia de un aceptor de electrones (Fancy and Kodadek 1999). Se seleccionó a la AtLEA4-5 como modelo para monitorear la posible interacción. Así, la AtLEA4-5, sola o en presencia de la LDH (a diferentes relaciones molares), se sometió a entrecruzamiento seguido de la separación de las muestras por SDS-PAGE. Con el fin de detectar la proteína AtLEA4-5 entrecruzada se utilizaron anticuerpos específicos contra esta proteína en experimentos tipo Western (Olvera-Carrillo *et al.* 2010). Se validó la especificidad del anticuerpo pues no detectó a la LDH, aún en la muestra que contenía la mayor cantidad de ésta (Figura 1, carril 1).

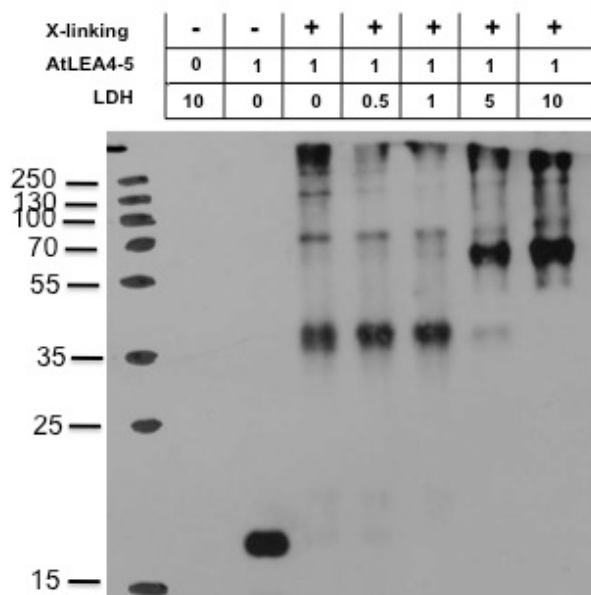


Figura 1. Entrecruzamiento de la AtLEA4-5 en ausencia o presencia de diferentes cantidades de LDH. La relación molar utilizada está representada por números (0, 0.5, 1, 5 y 10) en la parte superior de la figura. Se indica cuales fueron las muestras entrecruzadas (+) y no entrecruzadas (-).

De forma interesante, se observó que cuando la AtLEA4-5 (sola) se somete a entrecruzamiento, la señal del monómero (Figura 1, carril 2) desaparece y se detectan bandas de mayor peso molecular, las cuales coinciden con los pesos calculados para una especie dimérica, tetramérica, hexamérica y octamérica (Figura 1, carril 3). La formación de oligómeros es una característica que se ha encontrado en proteínas LEA de otros grupos (Hernandez-Sanchez *et al.* 2014; Rivera-Najera *et al.* 2014); así como, en chaperonas de bacterias y de plantas que funcionan en otros tipos de estrés (Basha *et al.* 2012; Reichmann *et al.* 2012; Tapley *et al.* 2009). En el caso de la AtLEA4-5, las especies oligoméricas solo se pueden observar cuando se entrecruza la proteína.

Una vez que se agrega la LDH a la muestra con AtLEA4-5 a relaciones molares bajas de LDH:AtLEA4-5 (0.5:1 y 1:1) no se detectó ninguna señal diferente a lo que se observa cuando se entrecruza la AtLEA4-5 sola (comparar carriles 3, 4 y 5 de la Figura 1); sin embargo, cuando se aumentó la cantidad de LDH, llegando a relaciones de 5:1 y 10:1, el dímero de la AtLEA4-5 desaparece y en cambio se observa una nueva banda con masa molecular de aproximadamente 69 kDa (Figura 1, carriles 6 y 7), la cual corresponde al tamaño esperado entre el dímero de la AtLEA4-5 (32.4 kDa) y un monómero de la LDH (36.4 kDa).

La interacción entre la AtLEA4-5 y la LDH, detectada por entrecruzamiento, se podría deber a que, dado el desorden estructural de la AtLEA4-5, ésta podría interaccionar inespecíficamente con muchos blancos. Por ello, se analizó la posible interacción entre la AtLEA4-5 con otra proteína. Se seleccionó la albúmina sérica bovina (ASB) para tal efecto. Como se observa en la Figura 2, no se encontró ninguna banda diferente al comparar la AtLEA4-5 entrecruzada sola o en presencia de distintas concentraciones de ASB (hasta relaciones molares de 50:1, ASB:AtLEA4-5).

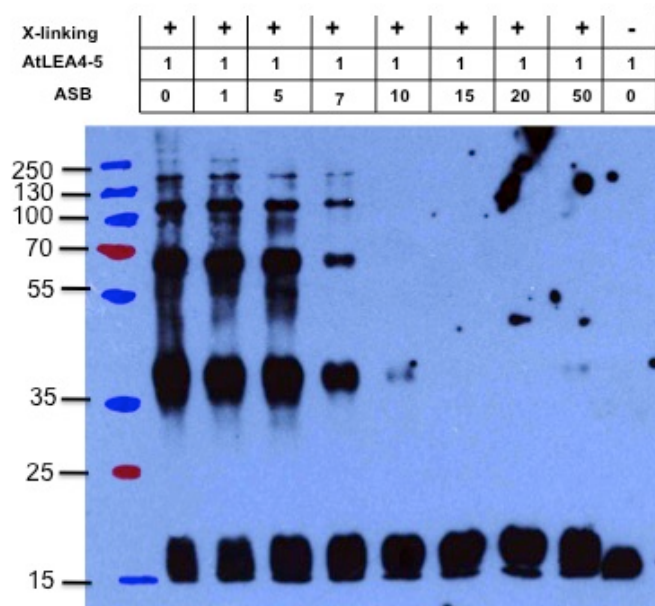


Figura 2. Entrecruzamiento de la AtLEA4-5 en ausencia o presencia de diferentes cantidades de ASB. La relación molar utilizada está representada por números (0, 1, 5, 7, 10, 15, 20 y 50) en la parte superior de la figura. Se indica cuales fueron las muestras entrecruzadas (+) y no entrecruzadas (-).

Interesantemente, las especies oligoméricas de la AtLEA4-5 desaparecieron conforme se aumentaban los niveles de ASB puesto que, probablemente, el aumento en el amontonamiento macromolecular generado por la ASB evitó que se dieran los contactos necesarios para la formación de los oligómeros de la AtLEA4-5.

En conjunto, estos datos sugieren que la AtLEA4-5 es capaz de interaccionar con la LDH (y no con ASB), al menos *in vitro*.

i.vii.ii. La AtLEA4-5 no experimenta cambios conformacionales en presencia de la LDH

La importancia del desorden estructural en las proteínas intrínsecamente desordenadas radica en que, al poseer altos niveles de flexibilidad, son capaces de interactuar con una amplia variedad de proteínas blanco lo cual tiene como consecuencia la multifuncionalidad en estas proteínas. Los mecanismos de interacción han sido ampliamente estudiados en las proteínas intrínsecamente desordenadas de animales, en donde se ha descrito que el plegamiento acoplado a la interacción con la proteína blanco es el mecanismo principal de interacción (Dyson and Wright 2002; Dyson and Wright 2005; Tompa 2012; Wright and Dyson 2015).

Con el objeto de determinar si la AtLEA4-5 se pliega y/o experimenta cambios conformacionales acoplados a la interacción con LDH (su proteína cliente), se analizó la estructura tridimensional de la AtLEA4-5 en presencia de LDH utilizando dos metodologías: dicroísmo circular y resonancia magnética nuclear. El primer acercamiento fue determinar si la estructura secundaria de cualquiera de los dos proteínas cambia en presencia de la otra, comparando los espectros de DC. Para ello, los espectros de la AtLEA4-5 y la LDH se obtuvieron por separado. Los espectros resultantes de estos ensayos se sumaron aritméticamente y se compararon con el espectro resultante de mezclar ambas proteínas experimentalmente. Si existen cambios en la estructura tridimensional de alguna de las dos proteínas, el espectro resultante de la mezcla será diferente al obtenido al sumar aritméticamente los espectros de cada proteínas por separado. Cabe hacer notar que estos experimentos fueron realizados en solución acuosa. Congruente con los resultados obtenidos en esta tesis, la AtLEA4-5 presentó un espectro típico de proteína intrínsecamente desordenada, mientras que la LDH tiene una estructura mayoritariamente de α -hélice (Figura 3). Interesantemente, el espectro de la mezcla de las dos proteínas es prácticamente idéntico al espectro obtenido al sumar los de las proteínas por separado (Figura 3).

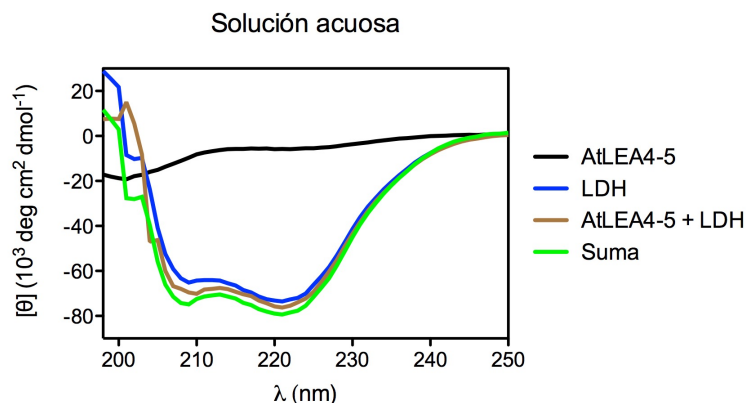


Figura 3. Espectros de DC de la AtLEA4-5 (negro), la LDH (azul) y de ambas proteínas mezcladas en la misma muestra (café), en solución acuosa. Se muestra la suma aritmética de los espectros de la AtLEA4-5 y la LDH adquiridos por separados (verde).

Dado que la AtLEA4-5 modifica su conformación cuando se disminuye la cantidad de agua disponible (por ejemplo, en presencia de glicerol), se utilizó el mismo enfoque pero ahora las proteínas fueron colocadas en soluciones que contenían 20% o 60% de glicerol. De la misma forma que en solución acuosa, los espectros obtenidos al mezclar ambas proteínas son muy parecidos a los calculados a partir de la suma aritmética (Figura 4). Sólo cuando se agregó 60% de glicerol, el espectro de las proteínas mezcladas tiene un poco menos de α -hélice (Figura 4B), lo cual indica que una de las dos proteínas en la mezcla perdió un poco de este tipo de estructura.

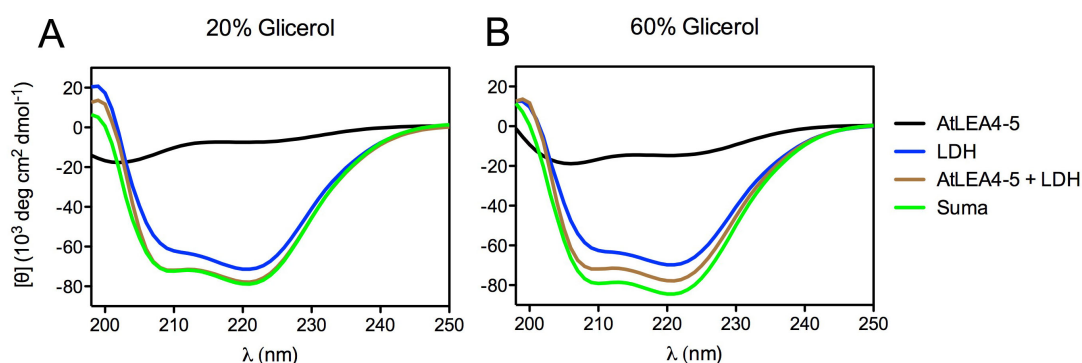


Figura 4. Espectros de DC de la AtLEA4-5 (negro), la LDH (azul) y de ambas proteínas mezcladas en la misma muestra (café) que contenía (A) 20% y (B) 60% de glicerol. Se muestra la suma aritmética de los espectros de la AtLEA4-5 y la LDH adquiridos por separados (verde).

Como se puede notar en los resultados obtenidos por DC, la gran desventaja de esa técnica radica en que no es posible saber cuál de las dos proteínas está cambiando su conformación (si fuera el caso). Por tal motivo, se decidió seguir los

cambios conformacionales experimentados por la AtLEA4-5 por RMN. Para ello, se obtuvieron los espectros en dos dimensiones de la AtLEA4-5 marcada con ^{15}N (como la utilizada en esta tesis) sola o mezclada con dos cantidades diferentes de LDH. Como la LDH no está marcada, sólo podremos seguir los cambios conformacionales que ocurren en la AtLEA4-5 causados específicamente por la interacción con su proteína blanco. El espectro de la AtLEA4-5 presenta desplazamientos químicos característicos de una proteína intrínsecamente desordenada (Figura 5 A). Consistente con lo encontrado en los ensayos por DC, no se detectaron cambios en la posición ni en la intensidad de las señales de desplazamiento químico al agregar una relación molar 4:1 (Figura 5B) ni al agregar una relación 13:1 (LDH:AtLEA4-5) (Figura 5C), lo cual resulta claramente evidente al comparar los 3 espectros (Figura 5D).

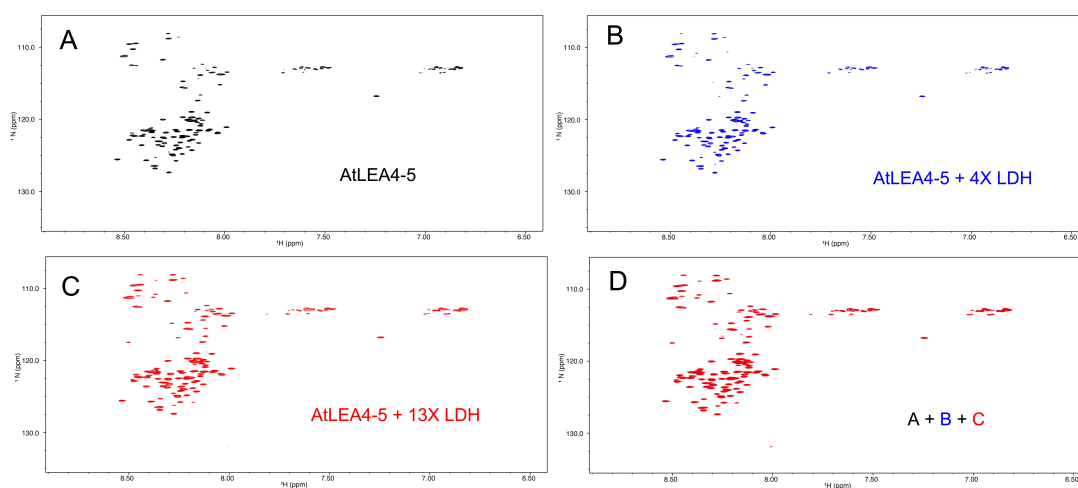


Figura 5. Espectros de resonancia magnética nuclear en dos dimensiones de la AtLEA4-5 sola (A), en presencia de 4 veces más (relación molar) de LDH (B) y en presencia de 4 veces más (relación molar) de LDH (C). (D) Superposición de los espectros A, B y C. Las señales negras y azules en D no se distinguen pues todas están en la misma posición por lo que se solo se distingue la señal roja.

Estos datos sugieren que la AtLEA4-5 no se pliega al interactuar con la LDH, contrario a lo que se ha observado para la mayoría de las proteínas intrínsecamente desordenadas de animales. Aunque los resultados obtenidos por DC y por RMN indican que la proteína AtLEA4-5 no sufre cambios detectables por estas técnicas, en una mezcla con LDH, estos datos no permiten concluir que esta proteína no interactúe con la LDH; o bien, que en caso de que interactúe, que no cambia su conformación al interactuar. La razón de ello es que las condiciones de la mezcla no necesariamente son las óptimas para que ocurra tal evento; ya que, al estar en

solución acuosa, la proteína se encuentra mayoritariamente desordenada, de tal forma que es posible que en esta conformación no sea capaz de reconocer y unirse a la LDH.

i.viii. DISCUSIÓN

Una de las funciones propuestas para las proteínas LEA es proteger a otras macromoléculas y/o compartimentos celulares de los efectos adversos causados por la falta de agua (Battaglia *et al.* 2008). Las macromoléculas “cliente” de las proteínas LEA que se han estudiado más extensamente son las proteínas. Se ha encontrado en diversos estudios *in vitro* que las proteínas LEA de distintos grupos tienen la capacidad de evitar que otras proteínas pierdan su función e incluso que lleguen a experimentar agregación (Goyal *et al.* 2005; Reyes *et al.* 2005; Chakrabortee *et al.* 2007; Reyes *et al.* 2008; Campos *et al.* 2013; Hughes *et al.* 2013). El mecanismo de acción molecular a través del cual estas proteínas protegen a sus clientes aún se encuentra en debate. Existen dos hipótesis principales que intentan explicar este mecanismo de protección, sin que estas sean mutuamente excluyentes (Reyes *et al.* 2005; Tunnacliffe and Wise 2007; Tompa and Kovacs 2010; Olvera-Carrillo *et al.* 2011; Chakrabortee *et al.* 2012; Hughes *et al.* 2013; Cuevas-Velazquez *et al.* 2014). La primera hipótesis considera a las proteínas LEA como “escudos moleculares” que tienen la capacidad de evitar los contactos inespecíficos entre los parches hidrofóbicos de las proteínas cliente que se generan cuando comienza el proceso de desnaturalización inducido por el tratamiento de estrés (Tunnacliffe and Wise 2007; Chakrabortee *et al.* 2012; Hughes *et al.* 2013). Es muy común encontrar que la cantidad de proteína LEA necesaria para ejercer el efecto protector a través de este tipo de mecanismo sea muy elevada respecto a la proteína cliente; es decir, se requieren relaciones molares altas (>100:1, proteína LEA:proteína cliente) (Hughes *et al.* 2013). Existen dos requisitos principales para que una proteína LEA funcione como escudo molecular: que sea muy hidrofílica y que tengan un radio hidrodinámico grande (Chakrabortee *et al.* 2012; Hughes *et al.* 2013). Como se puede notar de las características descritas anteriormente, la secuencia primaria de las proteínas LEA no es un factor primordial para ejercer el efecto protector (sólo se necesita que exista cierta composición de aminoácidos), por lo que esta hipótesis no explica la alta conservación en secuencia de los motivos característicos de cada grupo de proteínas LEA. Además, puesto que se requiere que exista un radio hidrodinámico amplio, las

proteínas deben estar lo más desplegadas que sea posible, lo cual excluye la capacidad de plegarse en estructuras secundarias más estables y compactas de algunas proteínas LEA. Existe otra hipótesis propuesta que se fundamenta en el mecanismo de protección de las chaperonas moleculares y las proteínas pequeñas de choque térmico. En este caso, las proteínas LEA utilizan sus motivos conservados para encontrar proteínas *cliente* con características particulares. Consecuentemente, debe ocurrir una interacción proteína-proteína para que ocurra el fenómeno de protección (Reyes *et al.* 2005; Tompa and Kovacs 2010; Olvera-Carrillo *et al.* 2011; Cuevas-Velazquez *et al.* 2014). Probablemente, el plegamiento (por ejemplo en α -hélice) podría acomodar los residuos conservados de cierta forma, lo que facilitaría el reconocimiento molecular y la subsecuente protección. En este modelo, se necesitarían relaciones molares bajas (<20:1, proteína LEA:proteína cliente) para obtener altos niveles de protección (Reyes *et al.* 2005; Cuevas-Velazquez *et al.* 2014). Además de proteger a bajas relaciones molares, la AtLEA4-5 se entrecruza directamente con la LDH incluso en solución acuosa, lo cual sustenta que esta proteína protege a través del mecanismo tipo chaperona. Cabe aclarar que el que la interacción se detecte por entrecruzamiento, aún en solución acuosa, no indica necesariamente que la proteína que interaccionó está en una conformación desordenada, pues es posible que una pequeña fracción de la misma posea estructura secundaria y sean estas pocas moléculas las responsables de la interacción, para lo cual el método de PICUP tiene la sensibilidad necesaria pero no así la RMN o el DC.

Recientemente se ha encontrado que existe un nuevo grupo de chaperonas que utilizan el desorden estructural para llevar a cabo su función protectora. A estas proteínas se les conoce como *condicionalmente desordenadas* debido a que en su estado nativo son proteínas globulares pero cuando se encuentran en la condición de estrés en la cual protegen, algunas zonas de la proteína se desordenan y son justamente estas regiones las encargadas del reconocimiento molecular y el posterior efecto protector (Bardwell and Jakob 2012). En el caso de las proteínas LEA ocurre lo contrario, cuando se encuentran bajo la condición de estrés, la proteína adquiere estructura secundaria por lo que posiblemente el mecanismo de reconocimiento molecular y protección sea muy diferente cuando se trata del estrés por deshidratación que cuando ocurra bajo los tipos de estrés estudiados para otras chaperonas (estrés oxidativo, estrés por pH, estrés por temperatura). Existe la posibilidad de que las proteínas LEA del grupo 4 reconozcan a la proteína cliente cuando se encuentran en

su estado desplegado y que el plegamiento generado por la condición de déficit hídrico estabilice la unión con el blanco. Finalmente, los datos hasta ahora sugieren que la interacción de la AtLEA-5 con la proteína cliente no ocurre como se ha descrito para la mayoría de las interacciones en las que participan PIDs; es decir, no parece presentarse un plegamiento asociado a la interacción (Dyson and Wright 2002)..

Existen algunas proteínas intrínsecamente desordenadas que no se pliegan cuando interactúan con su proteína blanco. Este tipo de interacciones forman “*complejos dinámicos*”; es decir, la proteína desordenada es capaz de interactuar con la misma proteína *cliente* de formas distintas y a veces, con diferentes regiones de la misma proteína. A este fenómeno de interacción dinámica que implica que la proteína se mantenga desordenada (no necesariamente desplegada) en el complejo de interacción se le conoce como “fuzzines” o “complejos difusos” (Tompa and Fuxreiter 2008).

El espectro de resonancia magnética nuclear de la AtLEA4-5 en presencia de la LDH no mostró cambios en la posición ni en la intensidad de las señales de desplazamiento químico. Este fenómeno se ha reportado en la interacción de *complejos difusos* de otras proteínas intrínsecamente desordenadas (Tompa and Fuxreiter 2008). De hecho, es probable que este tipo de interacciones dinámicas sea especialmente importante para proteínas con función tipo chaperona puesto que de esta manera podrían interactuar con varias regiones de la proteína, expandiendo así la capacidad protectora sobre las proteínas *cliente*.

Si bien los resultados obtenidos en esta sección anexa permiten acercarnos a dilucidar el mecanismo de protección a nivel molecular, aún será necesario responder una serie de preguntas esenciales. Los resultados del entrecruzamiento sugieren que una especie dimérica de la AtLEA4-5 es la que interactúa con un monómero de la LDH. Si esto es así, será necesario caracterizar a detalle si esta especie dimérica se forma y/o es más estable en alguna condición; por ejemplo en déficit hídrico. Además, se deberá determinar la estequiometría, la termodinámica y la constante de afinidad para la interacción de ambas proteínas.

El análisis de las interacciones con otras posibles proteínas blanco (e.g. GFP o luciferasa) nos permitirá determinar si los mecanismos de protección son similares para los diferentes blanco o no. Así mismo, reporteros como proteínas fluorescentes o

la luciferasa permitirán obtener información *in vivo* en cuanto a los niveles de protección y posibles cambios estructurales implicados.

i.ix. CONCLUSIONES

La AtLEA4-5 protege a su enzima blanco a través de un mecanismo tipo chaperona puesto que protege a relaciones molares bajas y entrecruza con la LDH con una estequiometría de 2:1 (un dímero de la AtLEA4-5 con un monómero de la LDH). Este entrecruzamiento ocurre incluso en solución acuosa. Contrario a lo que sucede en la mayoría de las proteínas intrínsecamente desordenadas, la AtLEA4-5 no se pliega o experimenta cambios conformacionales en presencia de su proteína modelo, al menos bajo las condiciones analizadas. Este tipo de interacción sugiere que la AtLEA4-5 forma un complejo *dinámico o difuso (Fuzzy complex)* con la LDH.

i.x. PERSPECTIVAS

Será necesario determinar a detalle el mecanismo de protección a nivel molecular de las proteínas LEA del grupo 4. Para ello se plantea caracterizar la interacción a profundidad con otras técnicas como la calorimetría de titulación isotérmica (ITC), lo que nos permitiría encontrar la constante de afinidad y la estequiometría de la interacción, además de los parámetros fisicoquímicos asociados a ésta. Por otro lado, se deberán encontrar las zonas de la proteína responsables del reconocimiento molecular (interacción con la proteína blanco) lo cual se puede realizar con la técnica de espectrometría de masas, valiéndonos del método de entrecruzamiento. De esta forma podremos detectar cuáles son los residuos de la proteína importantes para ello y correlacionar esta información con los motivos conservados de las proteínas AtLEA4. Es fundamental realizar los experimentos con los otros miembros de la familia así como las regiones N-terminal y C-terminal de la AtLEA4-5. Finalmente, es primordial ampliar el abanico de proteínas cliente de las AtLEA4 con otras proteínas reporteras y, mejor aún, si se logran encontrar sus proteínas blanco *in vivo*. Estas estrategias representan el marco experimental de la tesis de doctorado de Paulette Romero.

ANEXO II. La unión a metales y el efecto sobre la estructura y función de las proteínas abundantes en la embriogénesis tardía.

ii.i. INTRODUCCIÓN

ii.i.i. Los iones metálicos de protección propuestos para las proteínas LEA

Los metales son componentes naturales de los sistemas terrestres. Algunos de ellos como el cobre (Cu^{2+}), el manganeso (Mn^{2+}) y el zinc (Zn^{2+}) son esenciales para el desarrollo adecuado de las plantas pero solo en cantidades muy bajas (Yamasaki *et al.* 2008). Por ejemplo, el Cu^{2+} está involucrado en muchos procesos metabólicos. En plantas superiores, la proteína que posee la mayor cantidad de Cu^{2+} es la plastocianina, la cual participa en la cadena de transporte de electrones en los cloroplastos. Otra proteína que une metales, en este caso Cu^{2+} ó Zn^{2+} , es la superóxido dismutasa, la cual tiene variantes que se localizan en el citoplasma, en el estroma de los cloroplastos y en los peroxisomas, en donde su función es el secuestro de especies reactivas de oxígeno (Bowler *et al.* 1991). El contenido promedio de Cu^{2+} en los tejidos vegetales es de 10 mg/g de peso seco (Rozhdestvenskii *et al.* 1975). La concentración crítica de Cu^{2+} que pueden tolerar las plantas en un medio externo se encuentra entre 10^{-14} y 10^{-16} M (Rozhdestvenskii *et al.* 1975). El Ni^{2+} ejerce una variedad de roles esenciales en las plantas. La deficiencia de Ni^{2+} produce una serie de defectos en el crecimiento y metabolismo, incluyendo reducción del crecimiento, inducción de senescencia, clorosis de los meristemos y las hojas, alteraciones en el metabolismo del nitrógeno y reducción en la captura de Fe^{3+} (Ahmad and Ashraf 2011). La concentración de Ni en la mayoría de las plantas es muy baja (0.05 – 10 mg/Kg de peso seco) (Ragsdale 1998). El Ni^{2+} es un constituyente de muchas metaloenzimas como la ureasa, superóxido dismutasa, hidrogenasas de Ni-Fe, metil coenzima M reductasa, monóxido de carbono deshidrogenasa, acetil coenzima A sintasa y RNasa A (Ahmad and Ashraf 2011). Por tal motivo, la deficiencia de Ni^{2+} reduce la actividad de la ureasa, altera la asimilación de nitrógeno y reduce el secuestro de radicales libres de oxígeno.

A pesar de la importancia fisiológica de los iones metálicos, el exceso de estos genera riesgos altos para las plantas puesto que éstos pueden generar especies reactivas de oxígeno a través de la reacción de Fenton (Yamasaki *et al.* 2008). La toxicidad de los metales conlleva varios síntomas. La toxicidad por Cu^{2+} usualmente

causa clorosis en las venaciones de las hojas, eventualmente generando necrosis tras una exposición prolongada. Los síntomas de toxicidad por Mn^{2+} incluyen clorosis de hojas viejas, puntos necróticos y “arrugamiento” de las hojas nuevas. La toxicidad por Zn^{2+} causa clorosis en hojas jóvenes así como necrosis en casos severos (Yamasaki *et al.* 2008).

Altas concentraciones de Ni^{2+} retardan la germinación de muchas especies de plantas. El efecto del Ni^{2+} en este caso es directo porque afecta directamente la actividad de amilasas, proteasas y ribonucleasas y, de esta manera, altera la digestión y movilización de reservas en la semilla durante la germinación (Ahmad and Ashraf 2011). En los tejidos vegetativos, altas concentraciones de Ni^{2+} retardan el crecimiento tanto de la parte aérea como del sistema radicular lo que genera una alteración en la producción de nuevos brotes, deformación de distintas partes de la planta, producción de flores con formas anormales, disminución en la producción de biomasa, inducción de manchas en las hojas y, a su vez, deficiencia de Fe^{3+} , produciéndose por ello una clorosis y necrosis foliar. Adicionalmente, el exceso de Ni^{2+} afecta la absorción de nutrientes, altera el metabolismo e inhibe la fotosíntesis y la transpiración (Ahmad and Ashraf 2011).

Los efectos de la toxicidad por metales a nivel de planta completa consisten principalmente en la reducción del crecimiento y otros síntomas que son evidentes en las hojas. En una escala más pequeña, se pueden observar efectos o síntomas a nivel celular. Los metales tienen un efecto directo sobre el metabolismo de la célula. Cada metal tiene un modo de acción particular, pero en general se puede observar reducción en la fotosíntesis, problemas en la producción y utilización de enzimas, transporte alterado de nutrientes y afectación en las funciones celulares básicas (Yamasaki *et al.* 2008).

ii.ii. ANTECEDENTES

Existen diversas propuestas sobre la función de las proteínas LEA. Como se ha mencionado en esta tesis, la protección de proteínas frente a condiciones de déficit hídrico es la función que más se ha estudiado a la fecha; sin embargo, existen evidencias de otras posibles funciones. Se he propuesto que las proteínas LEA podrían ser secuestradores o amortiguadores de metales pesados, los cuales se

acumulan durante condiciones de déficit hídrico, y que en altas concentraciones son tóxicos para la célula vegetal. Diversos estudios han reportado que algunas proteínas LEA, principalmente del grupo 2 (dehidrininas) son capaces de unir iones metálicos (Svensson *et al.* 2000; Kruger *et al.* 2002; Hara *et al.* 2005). En *Arabidopsis*, se ha reportado que la proteína LTI30 une Cu^{2+} , Ni^{2+} , Co^{2+} y Zn^{2+} mientras que RAB18, LTI29 y COR47 une Cu^{2+} y Ni^{2+} pero no unen Co^{2+} ni Zn^{2+} (Svensson *et al.* 2000). La proteína ITP de *Ricinus communis* une Fe^{3+} , Cu^{2+} , Zn^{2+} y Mn^{2+} pero no une Fe^{2+} (Kruger *et al.* 2002). La proteína CuCOR15 de *Citrus unshiu* une Cu^{2+} , Ni^{2+} , Co^{2+} , Zn^{2+} y Fe^{3+} (Hara *et al.* 2005) a través del motivo rico en histidinas (HKGEHHS GDHH). La proteína AtHIRD11 une Cu^{2+} , Ni^{2+} , Co^{2+} y Zn^{2+} pero no une Mg^{2+} , Ca^{2+} , Mn^{2+} y Cd^{2+} . La K_a de la proteína AtHIRD11 para Zn^{2+} es de 1.4 μM , para Cu^{2+} de 3.3 μM , para Ni^{2+} de 2.8 μM y para Co^{2+} de 4.8 μM (Hara *et al.* 2011). La proteína AtHIRD11 es capaz de inhibir eficientemente la generación de especies reactivas de oxígeno generadas *in vitro* por Cu^{2+} , con una actividad más alta que el glutatión o la albúmina sérica bovina (Hara *et al.* 2013).

En el caso de las proteínas LEA del grupo 4, se ha encontrado que las proteínas GmPM1 y GmPM9 de soya unen Fe^{2+} , Ni^{2+} , Cu^{2+} y Zn^{2+} pero no unen Ca^{2+} , Mg^{2+} o Mn^{2+} (Liu *et al.* 2011). La K_a reportada para la proteína GmPM1 por Fe^{3+} es de 19.4 μM y la correspondiente a la proteína GmPM9 es de 14.2 μM . Ambas proteínas son capaces de secuestrar especies reactivas de oxígeno generadas por el Fe^{3+} (Liu *et al.* 2011).

Se ha observado que la unión a iones metálicos induce cambios en la estructura secundaria de proteínas desordenadas, por ejemplo, en el prion PrP (Hornshaw *et al.* 1995). En el caso de las proteínas LEA, se han observado cambios conformacionales ligeros en respuesta a la unión de los iones metálicos. La proteína CuCOR15 se pliega ligeramente en α -hélice en presencia de Zn^{2+} (Hara *et al.* 2009). MpDhn12 se vuelve menos desordenada tras la adición de Cu^{2+} (Mu *et al.* 2011). Las proteínas TsDHN-1 y TsDHN-2 de *Thellungiella salsuginea* experimentan cambios conformacionales inducidos por Zn^{2+} y Ca^{2+} (Rahman *et al.* 2011). Finalmente, la proteína AtHIRD11 se vuelve menos desordenada al unir Cu^{2+} , Co^{2+} , Ni^{2+} y Zn^{2+} mientras que no se observó ningún cambio conformacional en presencia de Ca^{2+} , Mg^{2+} ni Mn^{2+} (metales que no une) (Hara *et al.* 2011). Interesantemente, las proteínas

LEA del grupo 4 de soya (GmPM1 y GmPM9) no parecen experimentar cambios conformacionales al unir Fe^{3+} (Liu *et al.* 2011).

Las proteínas LEA del grupo 4 contienen altos niveles del aminoácido histidina el cual, como se ha descrito anteriormente, está contenido en los motivos caracterizados de unión a metales en las proteínas LEA caracterizadas a la fecha (Otaki *et al.* 2010). La proteína AtLEA4-1 contiene 11.2% de histidinas mientras que la proteína AtLEA4-2 tiene 7.2% y la proteína AtLEA4-5 4.4%. Interesantemente, las histidinas están contenidas casi exclusivamente en la región C-terminal de las tres proteínas, lo que sugiere que la posible función de la región C-terminal de las proteínas LEA del grupo 4 es unir iones metálicos.

ii.iii. PREGUNTA

¿La proteína AtLEA4-5 unen cationes metálicos?

ii.iv. OBJETIVO GENERAL

Determinar la capacidad de la proteína AtLEA4-5 de unir metales divalentes.

ii.v. OBJETIVOS ESPECÍFICOS

- Determinar si la proteína AtLEA4-5 interacciona con Cu^{2+} y Zn^{2+} .
- Analizar el efecto de la unión de Cu^{2+} sobre la estructura secundaria de la proteína AtLEA4-5.
- Analizar el efecto del Cu^{2+} sobre la función protectora de la proteína AtLEA4-5.

ii.vi. MATERIALES Y MÉTODOS

ii.vi.i. Cromatografía de afinidad a iones metálicos inmovilizados

Para determinar la unión de la proteína AtLEA4-5 a cationes divalentes se utilizó cromatografía de afinidad a iones metálicos inmovilizados con las columnas de 1 mL Hi-TRAP Chelating HP (GE Healthcare). Estas columnas tienen la capacidad de

cargarse con diferentes cationes divalentes según sea necesario. Se utilizó la proteína AtLEA4-5 recombinante obtenida como se describe en esta tesis (Ver Materiales y métodos). Primero se verificó que la proteína no se uniera a la columna a la que no se le había cargado ningún metal. Para esto, primero se equilibró la columna con 10 mL de amortiguador de unión (20 mM fosfatos de sodio pH 7.2, 500 mM NaCl); una vez equilibrada, se cargaron 200 μ L de proteína con una concentración de 50 μ M. La columna se lavó con 8 mL de amortiguador de unión y se recolectaron 4 muestras de 2 mL cada una. Posteriormente, se eluyó la proteína con 4 mL de amortiguador de elución (20 mM fosfatos de sodio pH 7.2, 500 mM NaCl, 50 mM EDTA) y se recolectaron cuatro muestras de 1 mL cada una. Posteriormente, alícuotas de cada una de las fracciones de lavado y de las eluidas se separaron en SDS-PAGE y la proteína se detectó en ensayo tipo Western, con un anticuerpo específico para la proteína AtLEA4-5 (Materiales y Métodos, Anexo I). El mismo procedimiento se repitió con las columnas cargadas ya sea con CuCl_2 o ZnCl_2 antes de pasar la proteína.

ii.vi.ii. Determinación de cambios conformacionales inducidos por la presencia de CuCl_2

La proteína AtLEA4-5 se diluyó a una concentración de 18.5 μ M en amortiguador de fosfatos de sodio 10 mM pH 7.5 en presencia de 0 μ M, 10 μ M ó 50 μ M de CuCl_2 , cada una en na soluciones con 0%, 60% u 80% de glicerol. Las muestras se incubaron durante 1 hora antes de la adquisición de los espectros. Los espectros de dicroísmo circular se adquirieron a 25 °C con un espectropolarímetro de dicroísmo circular Jasco J-715 (JASCO Analytical Instruments) con una longitud de paso óptico de 0.1 cm desde 190 nm hasta 250 nm. Las mediciones se realizaron por lo menos tres veces con distintos lotes de purificación de la proteína AtLEA4-5.

ii.vi.iii. Inactivación de la actividad de la LDH causada por CuCl_2

Con el objeto de determinar si la unión a cobre tiene algún efecto funcional, se utilizó el ensayo de inactivación de la LDH causado por Cu^{2+} . Se siguió el método descrito por Hara y colaboradores (2016) con algunas modificaciones (Hara *et al.*

2016). La LDH se diluyó a una concentración de 47 nM en amortiguador 25 mM Tris-HCl pH 7.5, en presencia o ausencia de 9.3 μM de CuCl_2 . Para probar la capacidad de secuestrar iones Cu^{2+} y así evitar la inactivación de la LDH, a la mezcla de LDH y CuCl_2 se le agregó la proteína AtLEA4-5 a una concentración de 37.2 μM (relación molar 4:1 respecto a los iones Cu^{2+}). La albúmina sérica bovina y la lisozima se utilizaron a la misma concentración como controles. La LDH, el CuCl_2 y las proteínas (90 μL totales) se homogenizaron durante 30 segundos y la actividad enzimática se midió inmediatamente después. La mezcla se adicionó a 510 μL de amortiguador de reacción (25 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM piruvato, 0.15 mM NADH) para determinar la actividad enzimática. La actividad enzimática de la LDH se siguió como la disminución en la A_{340} como resultado de la conversión de NADH en NAD^+ (NADH absorbe a 340 nm). La actividad de la LDH a la que no se adicionó CuCl_2 se consideró como el 100%.

ii.vii. RESULTADOS

ii.vii.i. La proteína AtLEA4-5 une metales divalentes

Debido a la participación de las proteínas LEA en los procesos de tolerancia y adaptación de las plantas a la pérdida de agua, diversos estudios se han enfocado en tratar de dilucidar la función de este enigmático grupo de proteínas. Como ya se mostró en esta tesis, la hipótesis más estudiada a la fecha es la que propone que las proteínas LEAs son proteínas con actividad tipo chaperona que evitan la desnaturalización y agregación de proteínas importantes para la correcta función celular, cuando la célula se encuentra sometida a condiciones de déficit hídrico. Sin embargo, existen otras hipótesis sobre la función de las proteínas LEA que no necesariamente son excluyentes. Una de ellas se sustenta en la capacidad de algunas proteínas LEA de unir cationes divalentes. En este caso, las proteínas LEA podrían funcionar como amortiguadores y/o secuestradores de estos metales (los cuales aumentan peligrosamente cuando la célula experimenta déficit hídrico), evitando así que se generen especies reactivas de oxígeno, las cuales representan un estrés sumamente severo para la planta.

De esta manera, se abordó esta hipótesis para las proteínas LEA del grupo 4, tomando como modelo la proteína AtLEA4-5. Como se puede ver en la Tabla 3 de esta tesis, las proteínas LEA del grupo 4 tienen un alto porcentaje de histidinas (entre 4.4% y 11.2%), las cuales se localizan mayoritariamente en la región C-terminal variable de este grupo de proteína sugiriendo que podrían unir cationes divalentes (Figura 1).

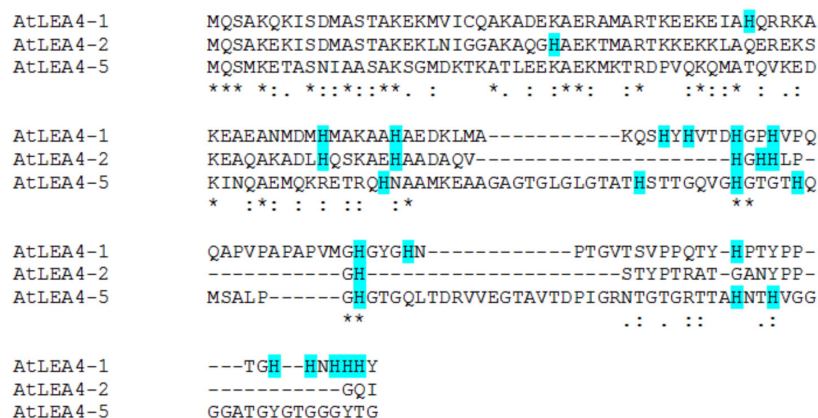


Figura 1. Alineamiento de las proteínas LEA del grupo 4 de Arabidopsis. Se muestran resaltadas en azul las histidinas de de cada una de las proteínas.

Para determinar si las proteínas LEA del grupo 4 son capaces de unir estas moléculas se utilizó la técnica de cromatografía de afinidad a iones metálicos inmovilizados, valiéndose para ello de columnas HiTrap (GE Healthcare life sciences). Estas columnas tienen la ventaja de que pueden cargarse con el metal que se necesite. Como primer acercamiento, se verificó que la proteína recombinante AtLEA4-5 no se uniera a la columna *per se*. La proteína AtLEA4-5 se cargó en la columna la cual no se había inmovilizado con ningún metal. Posteriormente, la columna se lavó con amortiguador y se tomaron las fracciones de lavado. Finalmente, la columna se pasó EDTA con el fin de eluir aquello que se pudiera haber unido a la columna, en caso de que hubiese residuos de metal presentes y se recolectaron las fracciones eluidas. Las distintas fracciones se separaron por SDS-PAGE. Las proteínas se transfirieron a una membrana y la proteína AtLEA4-5 se detectó por ensayos tipo Western , utilizando un anticuerpo específico contra la proteína AtLEA4-5. Como se puede ver en la Figura 2A, la proteína AtLEA4-5 no se pega a la columna a la cual no se le unió algún metal; ya que, se detecta en la primera fracción del lavado. Posteriormente, la columna se cargó con Cu^{2+} , un catión que muestra

afinidad por la regiones ricas en histidinas y que posee la capacidad de actuar como oxidante. Los resultados en la Figura 2B muestran que la proteína AtLEA4-5 se une al Cu^{2+} puesto que se requiere el paso de EDTA para que ésta se separe de la resina, detectándose en la primera fracción del volumen eluido con EDTA.

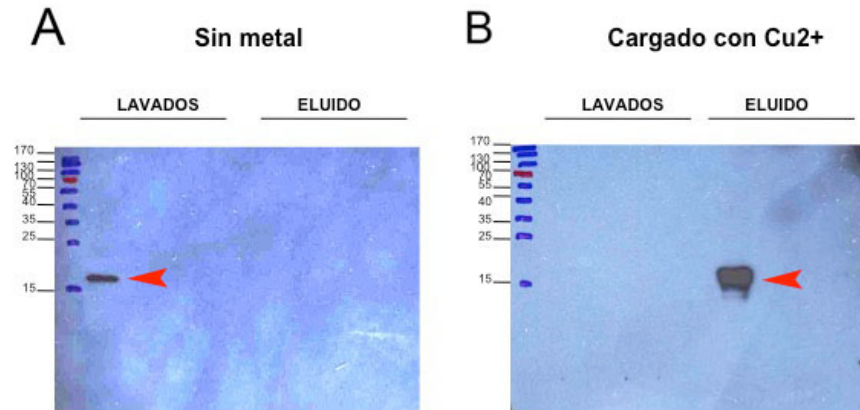


Figura 2. Prueba de unión de la AtLEA4-5 a Cu^{2+} . Ensayo tipo Western con un anticuerpo específico contra la proteína AtLEA4-5. Se muestran las fracciones obtenidas al lavar y, posteriormente, al eluir las muestras pasadas en columna con resina a la que no se le unió metal alguno (A), o a la que se le unió Cu^{2+} (B).

Por otro lado, se expandió el estudio a otros metales. Cuando se cargó la proteína AtLEA4-5 en una columna con resina a la que se le unió Zn^{2+} , se encontró que la proteína AtLEA4-5 también puede unir a ese metal, pues sólo se separa de la columna cuando se eluyó con EDTA (Figura 3B).

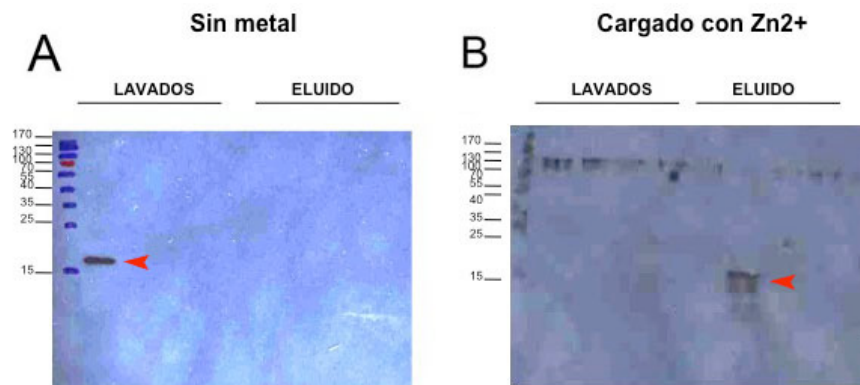


Figura 3. Prueba de unión de la AtLEA4-5 a Zn^{2+} . Los ensayos tipo Western se realizaron utilizando un anticuerpo específico contra la proteína AtLEA4-5. Se muestran las fracciones obtenidas al lavar y, posteriormente, al eluir las muestras pasadas en columna con resina a la que no se le unió metal alguno (A), o a la que se le unió Zn^{2+} (B).

Finalmente, para probar la unión de otros metales y caracterizar la química de la interacción de la proteína AtLEA4-5 y la proteína AtLEA4-2 con estos metales, se emprendió una colaboración muy cercana con la M. en C. Leidys French Pacheco y el Dr. Carlos Amero del Centro de Investigaciones Químicas de la Universidad Autónoma del Estado de Morelos. Durante la tesis de maestría de Leidys, se encontró que la proteína AtLEA4-2 y la proteína AtLEA4-5 también unen Ni^{2+} (además de Cu^{2+} y Zn^{2+}), mientras que, al parecer, ambas son incapaces de unir Ca^{2+} , Mn^{2+} y Fe^{3+} . Ambas proteínas mostraron una afinidad moderada al Ni^{2+} , Cu^{2+} y Zn^{2+} . Los datos obtenidos por French y colaboradores (2015) indicaron que la proteína AtLEA4-5 tiene tres sitios de unión a Cu a pH fisiológico (French, 2015). La proteína AtLEA4-2 tiene una K_a de 14.5 μM para Cu^{2+} , de 190 μM para Zn^{2+} y 720 μM para Ni^{2+} mientras que la proteína AtLEA4-5 tiene una K_a de 660 μM para Cu^{2+} , 99 μM para Zn^{2+} y 290 μM para Ni^{2+} (French, 2015).

ii.vii.ii. La proteína AtLEA4-5 no modifica su conformación tridimensional cuando une Cu^{2+}

Si bien en la sección anterior se estableció que la proteína AtLEA4-5 es capaz de unir Ni^{2+} , Cu^{2+} y Zn^{2+} , aún es necesario determinar si dicha interacción afecta las propiedades de la proteína. Para ello se analizó el efecto de la unión del metal sobre la estructura secundaria de la proteína. Se analizó el espectro de la proteína AtLEA4-5 recombinante, sola o en presencia de dos concentraciones distintas de CuSO_4 (10 μM y 50 μM). Como se había detectado en los resultados de esta tesis, la proteína AtLEA4-5 presenta un espectro de DC que mostró que esta proteína es mayoritariamente desordenada (Figura 4A). La adición de 10 μM y 50 μM de Cu^{2+} aumentó el valor del mínimo a 198 nm (se volvió menos negativo), pero el valor a 222 nm no disminuyó (Figura 4A). Estos datos indican que la proteína AtLEA4-5 en estas condiciones podría estar cambiando ligeramente su organización estructural, volviéndose menos desordenada.

Puesto que la proteína AtLEA4-5 se pliega en α -hélice cuando se simula la disminución en la cantidad de agua disponible *in vitro*, nos preguntamos si la unión a Cu^{2+} tendría un efecto sobre la conformación de la proteína previamente plegada. Para ello se agregaron dos concentraciones de Cu^{2+} (10 μM y 50 μM de Cu^{2+}) a muestras

de proteína AtLEA4-5 con 60% u 80% de glicerol. De forma similar a lo que ocurre en solución acuosa, en ambas condiciones el espectro de la proteína AtLEA4-5 sólo aumenta ligeramente (se hace menos negativo) el mínimo a 202 nm (en 60% glicerol) y a 208 nm (en 80% glicerol), lo cual indica que la α -hélice se encuentra ligeramente restringida cuando une Cu^{2+} bajo esas condiciones, manteniéndose un poco más desordenada (Figura 4B y 4C).

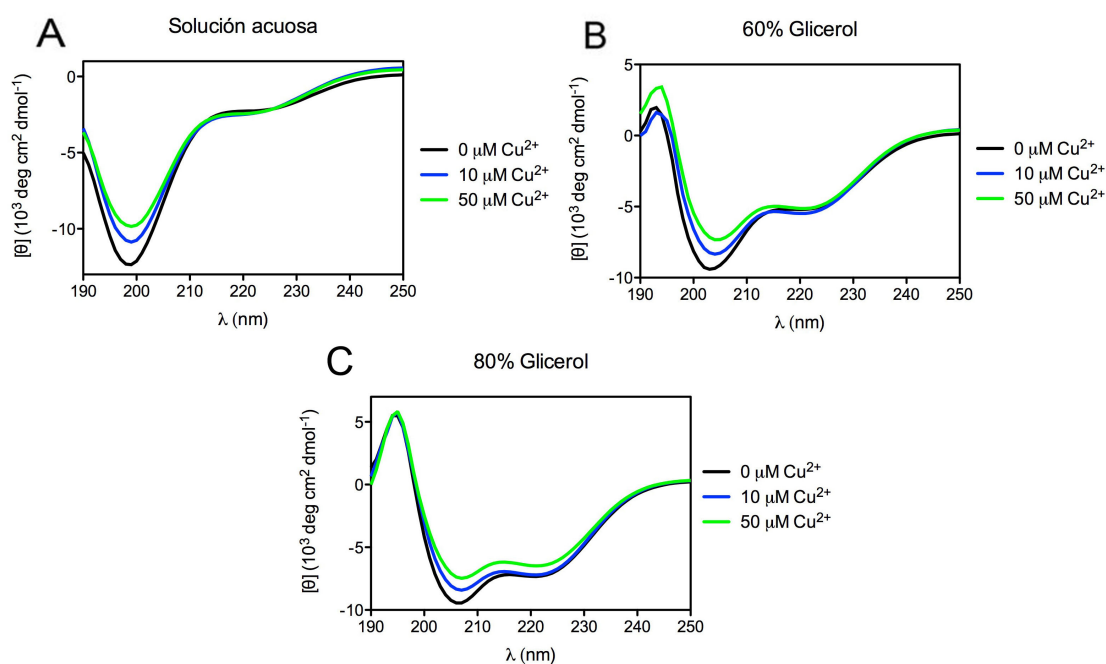


Figura 4. Efecto de la unión a Cu^{2+} sobre la estructura secundaria de la proteína AtLEA4-5. Espectros de DC en solución acuosa (A), 60% glicerol (B) y 80% glicerol (C). Se analizó la estructura sin Cu^{2+} (negro), 10 μM (azul) y 50 μM (verde) de Cu^{2+} .

ii.vii.iii. El efecto del Cu^{2+} sobre la función de la proteína AtLEA4-5

La proteína AtLEA4-5 une metales divalentes pero se debe determinar si esta unión tiene un significado funcional. Con el objeto de estudiar si la proteína AtLEA4-5 es capaz de secuestrar los metales en un sistema *in vitro*, utilizamos un ensayo que se fundamenta en el hecho de que los iones Cu^{2+} abaten la actividad enzimática de la LDH (Hara *et al.* 2016). Como se puede ver en la Figura 5, la presencia de una concentración de 9.3 μM de CuCl_2 inhibe completamente la actividad enzimática de la LDH. Si la AtLEA4-5 puede secuestrar a los iones Cu^{2+} , la LDH no se inactivará por la presencia de estos iones, como se ha observado para una proteína LEA del grupo 2 (Hara *et al.* 2016). Contrario a lo esperado, la adición de la proteína AtLEA4-

5 en una relación molar 4:1 respecto al metal no evitó la inactivación de la LDH causada por el Cu^{2+} . La adición de otras proteínas como la albúmina sérica bovina o la lisozima tampoco tuvo un efecto protector sobre la actividad de LDH. Estos datos muestran que, si bien la proteína AtLEA4-5 es capaz de unir Cu^{2+} , la unión no es tan fuerte como para secuestrar efectivamente estos metales y así evitar la toxicidad causada por los mismos.

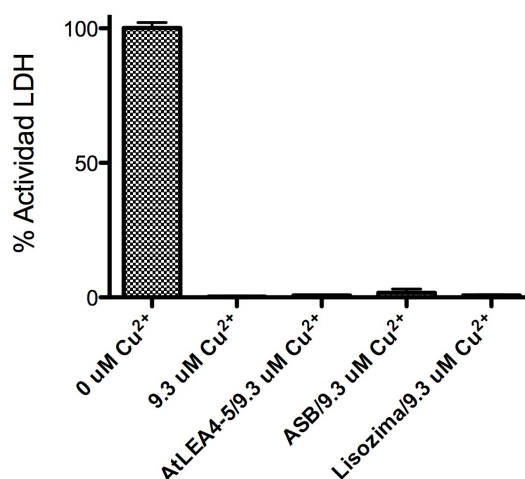


Figura 5. La AtLEA4-5 no previene la inactivación de LDH causada por iones Cu^{2+} *in vitro*. Se muestra el efecto sobre la actividad enzimática de la LDH en ausencia (0 μM) o en presencia (9.3 μM) de CuCl_2 . La presencia de la proteína AtLEA4-5 no previno la inactivación de la LDH. La ASB o la lisozima tampoco mostraron algún efecto protector de la actividad incluso a una relación molar 4:1, de las proteínas respecto al Cu^{2+} .

ii.viii. DISCUSIÓN

Las proteínas intrínsecamente desordenadas son proteínas que tienen la capacidad de adoptar un vasto número de conformaciones (a nivel de estructura secundaria). Esta propiedad ha permitido proponer que las proteínas intrínsecamente desordenadas pueden realizar más de una función, interaccionando con más de una proteína blanco. A este tipo de proteínas que realizan más de una función se les conoce como proteínas multi-funcionales (o en inglés “*moonlighting proteins*”). Como se puede leer en esta tesis, las proteínas LEA del grupo 4 tienen una función tipo chaperona en condiciones de déficit hídrico; sin embargo, la plasticidad de su estructura permite proponer que podrían tener más de una función. Por ejemplo, se ha

observado que la proteína AtLEA4-2 puede prevenir la fusión de membranas artificiales, proceso que puede ocurrir en las células vegetales en estados de deshidratación (Hundertmark *et al.* 2011). Por otro lado, la proteína AtLEA4-5 es capaz de unir Cu^{2+} , Zn^{2+} y Ni^{2+} , metales pesados que se pueden llegar a acumular en las células vegetales que experimentan déficit hídrico y que provocan grandes daños a través de promover la formación de especies reactivas de oxígeno. La proteína AtLEA4-5 une estos cationes divalentes a través de la coordinación con sus residuos de histidina que están localizados en la región C-terminal variable (French 2015). Las otras dos proteínas LEA del grupo 4 de Arabidopsis contienen altos niveles de histidinas sólo en el C-terminal. Estas observaciones podrían indicar que la región C-terminal tiene como objetivo capturar estos iones y para ello necesita mantener un estado conformacional desplegado. La unión de los cationes divalentes podría tener un efecto a diferentes niveles. Se ha propuesto que las proteínas LEA podrían ser secuestradores de los metales pesados, evitando que se desencadenen los efectos tóxicos generados por estos metales (Liu *et al.* 2011; Hara *et al.* 2013). Los datos obtenidos en este anexo muestran que la AtLEA4-5 es incapaz de prevenir la inactivación de LDH causada por la presencia de iones Cu^{2+} en un sistema *in vitro*. Estos experimentos se realizaron a una concentración de Cu^{2+} de $9.3 \mu\text{M}$. Dado que la K_a de la AtLEA4-5 para Cu^{2+} es de $660 \mu\text{M}$, es posible que la cantidad de Cu^{2+} utilizado en el ensayo no sea el adecuado para evaluar su capacidad de secuestrar iones; sin embargo, el uso de concentraciones en el intervalo de cientos de micromolar no es viable para realizar este tipo de experimentos (Hara *et al.* 2016). Es posible que la proteína AtLEA4-5 necesite de otras proteínas que la auxilien en esta función, como proteínas LEA de otros grupos (Hara *et al.* 2016). Como ya se ha discutido, otra de las posibilidades es que las proteínas LEA4 sólo modulen los niveles celulares de los cationes sin ser necesariamente secuestradores. Por otro lado, la unión podría modular de alguna forma la actividad de las proteínas LEA4. Si bien la unión de Cu^{2+} , sólo ejerce un efecto mínimo sobre la estructura secundaria de la proteína AtLEA4-5, posiblemente esta unión provoque que esta tenga una estructura ligeramente más rígida, lo cual consecuentemente induciría la interacción con alguna otra macromolécula. También es posible que la unión a los metales no module la estructura ni la función protectora pero que sí tenga un efecto sobre otra función aún desconocida de estas proteínas. Por ejemplo, se ha reportado que un motivo de unión

a metales (motivo rico en histidinas) de una proteína LEA del grupo 2 es necesario para la translocación al núcleo de esta proteína (Hernandez-Sanchez *et al.* 2015). Estudios de localización subcelular de la proteína AtLEA4-5 (Martinez, 2016) indican que la proteína AtLEA4-5 tiene una localización citoplasmática y nuclear. Existe la posibilidad de que la unión a metales module la translocación a núcleo de la AtLEA4-5, aunque se deberán obtener otros datos que sustenten esta hipótesis. Además resulta de especial importancia la identificación de los motivos localizados en la región C-terminal responsables de dicha interacción.

ii.ix. CONCLUSIONES

La proteína AtLEA4-5 es capaz de unir Ni^{2+} , Cu^{2+} y Zn^{2+} con una afinidad moderada. La unión a Cu^{2+} tiene un efecto mínimo en la estructura secundaria de la proteína AtLEA4-5 puesto que solamente pierde un poco de orden pero este cambio no está acompañado de una ganancia significativa de α -hélice o alguna otra estructura secundaria estable. Este comportamiento se conserva tanto en solución acuosa (estado desordenado) como en dos condiciones de glicerol (estado plegado, mayoritariamente α -hélice).

ii.x. PERSPECTIVAS

Será necesario caracterizar a detalle cuáles son las regiones de la proteína involucradas en la unión de los metales divalentes. Una estrategia es generar mutantes carentes de histidinas, las cuales son los residuos probables de la unión a estos cationes. Se deberán probar los otros miembros de la familia y las regiones N-terminal y C-terminal en estos ensayos para delimitar los mecanismos de interacción. También es necesario determinar si las proteínas AtLEA4 son secuestradoras de metales probando Ni^{2+} y Zn^{2+} en otro ensayo funcional; por ejemplo, la generación de especies reactivas de oxígeno mediada por metales. Finalmente, se debe analizar si las proteínas AtLEA4 presentan la función de unir a los metales *in vivo*, probando mutantes en estos genes y midiendo la tolerancia a estrés por metales pesados.

ANEXO III. Oligonucleótidos.

Nombre	Secuencia (5' → 3')
TRC41Fw	AAACCATGGAATCGGCGAAACAGAAGATAAGCG
TCR41Rv	GGTCGACCTTAGTAGTGATGATGATTATGATGTCCC
TRC42Fw	AAACCATGGAGTCGGCGAAGG
TRC42Rv	GGTCGACCTTAGATCTGTCCCG
TRC45Fw	AAACCATGGAGTCGATGAAAGAAAC
TRC45Rv	GCGGTCGACCCGTTTATCCAGTATATCC
M19LFw	GCAAAATCTGGCCTGGACAAAACC
M19LRv	GGTTTTGTCCAGGCCAGATTTTGC
M33LFw	GGAGGAAAAGGCGGAGAAGCTGAAGAC
M33LRv	GTCTTCAGCTTCTCCGCCTTTTCCTCC
M43LFw	GTTCAGAAACAGCTGGCTACACAGGTT
M43LRv	AACCTGTGTAGCCAGCTGTTTCTGAAC
TRC4HRv	CGCGTCGACTCAGGTTCCGGCTCCAGCCGC
TRC4RCFw	AAACCATGGCCGGTTTAGGTTTGGGGA
PTYB41Fw	GGTGGTTGCTCTTCCAACATGCAATCGGCGAAACAGAAG
PTYB41Rv	GGTGGTCTGCAGTCATTAGTAGTGATGATGATTATGATGTCC
PTYB42Fw	GGTGGTTGCTCTTCCAACATGCAGTCGGCGAAGG
PTYB42Rv	GGTGGTCTGCAGTCATTAGATCTGTCCCGGCG
PTYBRCFw	GGTGGTTGCTCTTCCAACACCGGTTTAGGTTTGGGGAC
PTYBRv	GGTGGTCTGCAGTCATTATCCAGTATATCCCCCGC
LEA1Fw	GGACGTCTCATATGCAGTCGATGAAAGAAACAGCTTCG
LEA76Rv	GGCTACTGGATCCTATTATCCGGCTCCAGCCGCTTCTTTC

ANEXO IV. Medios de cultivo utilizados para el crecimiento de bacterias.

Medio LB (Luria-Bertani)

- 10 g Triptona
- 5 g Extracto de levadura
- 5 g NaCl

Ajustar el volumen a un litro con agua.

Esterilizar en autoclave.

Medio mínimo M9 (1X)

- 100 mL Sales M9 10X
- 1 mL MgCl₂ 1M
- 1 mL CaCl₂ 0.1M
- 10 mL Vitaminas 100X
- 4 g glucosa
- 1 g (NH₄)₂SO₄

Ajustar el volumen a 1 litro con agua.

Nota: En el caso del marcaje con ¹⁵N, se deberá usar 1 g de (NH₄)₂SO₄. Para marcar con ¹³C, se deberán usar 2-3 g de ¹³C glucosa.

La glucosa y el sulfato de amonio se deben esterilizar pasando por filtros de 0.2 μM. Los otros componentes se pueden esterilizar con autoclave.

Sales M9 10X

- 68 g Na₂HPO₄ (o 72 g Na₂HPO₄(H₂O) ó 128 g Na₂HPO₄(7H₂O))
- 30 g KH₂PO₄
- 5 g NaCl

Ajustar el volumen a 1 litro con agua.

Esterilizar en autoclave.

ANEXO V. Validación de secuencias nucleotídicas

Los diferentes marcos abiertos de lectura (ORF) que codifican para las proteínas utilizadas en este trabajo se clonaron y se enviaron a secuenciar a la unidad de síntesis y secuenciación de ADN de Instituto de Biotecnología (ver materiales y métodos de esta tesis). En todos los casos, la secuenciación se realizó con oligonucleótidos en los extremos del marco de lectura denominados “Forward” y “Reverse”. A manera de ejemplo, se muestran los alineamientos del marco de lectura con las secuencias obtenidas con los oligonucleótidos “Forward” y “Reverse” para la AtLEA4-2 en pTrc99A así como el de las mutantes de la AtLEA4-5 (M19L, M33L y M43L). En los alineamientos “Reverse” se observa el cambio de una C por una G causado por la adición del sitio de restricción *NcoI*. En el alineamiento “Forward”, la secuencia es de mala calidad al inicio pues es donde se alinea el oligonucleótido. El mismo problema ocurre con la secuencia “Reverse”, en este caso al final de la secuencia.

En el caso de las mutantes de la AtLEA4-5, se muestra en amarillo el cambio realizado a través de la mutagénesis. Se observa además una mutación sinónima de una A por una T que no genera un cambio de aminoácido.

* AtLEA4-2 en pTrc99A

- Alineamiento con la secuencia “Forward”

```

Forward      -----GSYYKRKKYMWKTCCAGT
ORF          ATGCAGTCGGCGAAGGAAAAGATCAGTGACATG-----GCCAGT
                                     *           *****

Forward      ACGGCCA-GGAGA-ACTCAACATCGGTGGCGCAAAGGCACAAGGTCATGC
ORF          ACGGCCAAGGAGAACTCAACATCGGTGGCGCAAAGGCACAAGGTCATGC
          ***** ***** *****

Forward      GGAGAAGACGATGGCAAGGACCAAAAAGAGAAGAAGTTGGCCCAAGAGC
ORF          GGAGAAGACGATGGCAAGGACCAAAAAGAGAAGAAGTTGGCCCAAGAGC
          *****

Forward      GAGAGAAGTCTAAGGAGGCGCAGGCCAAAGCTGACCTCCATCAATCCAAG
ORF          GAGAGAAGTCTAAGGAGGCGCAGGCCAAAGCTGACCTCCATCAATCCAAG
          *****

Forward      GCTGAGCATGCTGCGGACGCTCAGGTTACGGCCACCATCTTCCCGGTCA
ORF          GCTGAGCATGCTGCGGACGCTCAGGTTACGGCCACCATCTTCCCGGTCA
          *****

Forward      CTCCACCTACCCTACCCGAGCCACCGGAGCTAATTACCCGCCGGGACAGA
ORF          CTCCACCTACCCTACCCGAGCCACCGGAGCTAATTACCCGCCGGGACAGA
          *****

Forward      TCTAA
ORF          TCTAA
          *****

```

- Alineamiento con la secuencia “Reverse”

```

ORF      ATGCAGTCGGCGAAGGAAAAGATCAGTGACATGGCCAGTACGGCCAAGGA
Reverse  ATGGAGTCGGCGAAGGAAAAGATCAGTGACATGGCCAGTACGGCCAAGGA
          ***
ORF      GAAACTCAACATCGGTGGCGCAAAGGCACAAGGTCATGCGGAGAAGACGA
Reverse  GAAACTCAACATCGGTGGCGCAAAGGCACAAGGTCATGCGGAGAAGACGA
          *****
ORF      TGGCAAGGACCAAAAAAGAGAAGAAGTTGGCCCAAGAGCGAGAGAAGTCT
Reverse  TGGCAAGGACCAAAAAAGAGAAGAAGTTGGCCCAAGAGCGAGAGAAGTCT
          *****
ORF      AAGGAGGCGCAGGCCAAAGCTGACCTCCATCAATCCAAGGCTGAGCATGC
Reverse  AAGGAGGCGCAGGCCAAAGCTGACCTCCATCAATCCAAGGCTGAGCATGC
          *****
ORF      TCGGACGCTCAGGTTACGGCCACCATCTTCCCGGTCACTCCACCTACC
Reverse  TCGGACGCTCAGGTTACGGCCACCATCTTCCCGGTCACTCCACCTACC
          *****
ORF      -CTACCCGAGCCACCGAGCTAATTACCCGCCGGGACAGATCT-----AA
Reverse  YA---CCGAGC-ACCGA-----G---GAATCGSRRGRAK
          *****   ***
ORF      -----
Reverse  KRKWRRWWWRT

```

* AtLEA4-5 (M19L) en pTrc99A

- Alineamiento con la secuencia “Forward”

```

Forward  -----C-T---CATTGC-GCTTCTGCAA-ATC
ORF      ATGCAGTCGATGAAAGAAACAGCTTCGAATATTGCAGCTTCTGCAAAATC
          * *      *****
Forward  TGGCCTGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
ORF      TGGCAATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
          ****
Forward  AGACACGAGACCCGTGTTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
ORF      AGACACGAGACCCGTGTTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
          *****
Forward  AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAAACGC
ORF      AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAAACGC
          *****
Forward  GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
ORF      GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
          *****
Forward  CCACTCACTCGACCAC'TGGACAAGTCGGACACGGCACTGGGACACATCAG
ORF      CCACTCACTCGACCAC'TGGACAAGTCGGACACGGCACTGGGACACATCAG
          *****
Forward  ATGTCGGCTCTGCC'TGGTCACGGAACGGGACAAC'TGACCGACCGGCTTGT
ORF      ATGTCGGCTCTGCC'TGGTCACGGAACGGGACAAC'TGACCGACCGGCTTGT
          *****
Forward  GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAAC'TGGTC
ORF      GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAAC'TGGTC
          *****
Forward  GGACAACCGCTCATAACACACACGTTGGTGGTGGTGGTGGCCACCGGGTAC
ORF      GGACAACCGCTCATAACACTCACGTTGGTGGTGGTGGTGGCCACCGGGTAC
          *****

```



```

Forward      GGAACCGCGGGGGATATACTGGATAA
ORF          GGAACCGCGGGGGATATACTGGATAA
*****

```

- Alineamiento con la secuencia "Reverse"

```

ORF          ATGCAGTCGATGAAAGAAACAGCTTCGAATATTGCAGCTTCTGCAAATC
Reverse      ATGGAGTCGATGAAAGAAACAGCTTCGAATATTGCAGCTTCTGCAAATC
***          *****

ORF          TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
Reverse      TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
***          *****

ORF          AGACACGAGACCCGTGTTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
Reverse      AGACACGAGACCCGTGTTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
***          *****

ORF          AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
Reverse      AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
***          *****

ORF          GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
Reverse      GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
***          *****

ORF          CCACTCACTCGACCAC'TGGACAAGTCGGACACGGCACTGGGACACATCAG
Reverse      CCACTCACTCGACCAC'TGGACAAGTCGGACACGGCACTGGGACACATCAG
***          *****

ORF          ATGTCGGCTCTGCC'TGGTACGGAACGGGACAAC'TGACCGACCGC'GTTGT
Reverse      ATGTCGGCTCTGCC'TGGTACGGAACGGGACAAC'TGACCGACCGC'GTTGT
***          *****

ORF          GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACTGGTTC
Reverse      GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACTGGTTC
***          *****

ORF          GGACAACCGCTCATAACACTCACGTTGGTGGTGGTGGTCCACCGGTAC
Reverse      GGACAACCGCTCATAACACACACGTTGGTGGTGGTGG--CCACCGGTAC
***          *****

ORF          GGAACCGCGGGGGATATACTGGATAA
Reverse      GA-AGCGGC-----C

```

* AtLEA4-5 (M33L) en pTrc99A

- Alineamiento con la secuencia "Forward"

```

Forward      --GCCATC-----CGAG-----GGGTTTTGG-AAATC
ORF          ATGCAGTCGATGAAAGAAACAGCTTCGAATATTGCAGCTTCTGCAAATC
***          **          ***          * * * * *

Forward      TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGCTGA
ORF          TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
***          *****

Forward      AGACACGAGACCCGTGTTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
ORF          AGACACGAGACCCGTGTTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
***          *****

Forward      AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
ORF          AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
***          *****

```

```

Forward      GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
ORF          GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
*****
Forward      CCACTCACTCGACCACCTGGACAAGTCGGACACGGCACTGGGACACATCAG
ORF          CCACTCACTCGACCACCTGGACAAGTCGGACACGGCACTGGGACACATCAG
*****
Forward      ATGTCGGCTCTGCCCTGGTCACGGAACGGGACAACCTGACCGACCGCGTTGT
ORF          ATGTCGGCTCTGCCCTGGTCACGGAACGGGACAACCTGACCGACCGCGTTGT
*****
Forward      GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACTGGTC
ORF          GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACTGGTC
*****
Forward      GGACAACCGCTCATAACACACACGTTGGTGGTGGTGGTGCCACCGGGTAC
ORF          GGACAACCGCTCATAACACTCACGTTGGTGGTGGTGGTGCCACCGGGTAC
*****
Forward      GGAACCGCGGGGGATATACTGGATAA
ORF          GGAACCGCGGGGGATATACTGGATAA
*****

```

- Alineamiento con la secuencia “Reverse”

```

ORF          ATGCAGTCGATGAAAGAAACAGCTTCGAATATTGCAGCTTCTGCAAATC
Reverse      ATGGAGTCGATGAAAGAAACAGCTTCGAATATTGCAGCTTCTGCAAATC
*** *****
ORF          TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
Reverse      TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGCTGA
*****
ORF          AGACACGAGACCCGTGTTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
Reverse      AGACACGAGACCCGTGTTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
*****
ORF          AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
Reverse      AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
*****
ORF          GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
Reverse      GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
*****
ORF          CCACTCACTCGACCACCTGGACAAGTCGGACACGGCACTGGGACACATCAG
Reverse      CCACTCACTCGACCACCTGGACAAGTCGGACACGGCACTGGGACACATCAG
*****
ORF          ATGTCGGCTCTGCCCTGGTCACGGAACGGGACAACCTGACCGACCGCGTTGT
Reverse      ATGTCGGCTCTGCCCTGGTCACGGAACGGGACAACCTGACCGACCGCGTTGT
*****
ORF          GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACTGGTC
Reverse      GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACTGGTC
*****
ORF          GGACAACCGCTCATAACACTCACGTTGGTGGTGGTGGTGCCACCGGGTAC
Reverse      GGACAACCGCTCATAACACACACGTTGGTGGTGGTGGTGG-GCCACCGGGTAC
*****
ORF          GGAACCGCGGGGGATATACTGGATAA
Reverse      GAAGCCG-----G
* * **

```

* AtLEA4-5 (M43L) en pTrc99A

- Alineamiento con la secuencia “Forward”

```

Forward      -----C-----CATTGC-GCTTCTGCAA-ATC
ORF          ATGCAGTCGATGAAAGAAACAGCTTTCGAATATTGCAGCTTCTGCAAAATC
              *                               *****
Forward      TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
ORF          TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
              *****
Forward      AGACACGAGACCCGTTCAGAAACAGCTGGCTACACAGGTTAAAGAAGAT
ORF          AGACACGAGACCCGTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
              *****
Forward      AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
ORF          AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
              *****
Forward      GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
ORF          GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
              *****
Forward      CCACTCACTCGACCCTGGACAAGTCGGACACGGCACTGGGACACATCAG
ORF          CCACTCACTCGACCCTGGACAAGTCGGACACGGCACTGGGACACATCAG
              *****
Forward      ATGTCGGCTCTGCCGGTTCACGGAACGGGACAACCTGACCGACCGCGTTGT
ORF          ATGTCGGCTCTGCCGGTTCACGGAACGGGACAACCTGACCGACCGCGTTGT
              *****
Forward      GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACCTGGTC
ORF          GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACCTGGTC
              *****
Forward      GGACAACCGCTCATAACACACACGTTGGTGGTGGTGGTGGCCACCGGGTAC
ORF          GGACAACCGCTCATAACACTCACGTTGGTGGTGGTGGTGGCCACCGGGTAC
              *****
Forward      GGAACCGCGGGGGATATACTGGATAA
ORF          GGAACCGCGGGGGATATACTGGATAA
              *****

```

- Alineamiento con la secuencia “Reverse”

```

ORF          ATGCAGTCGATGAAAGAAACAGCTTTCGAATATTGCAGCTTCTGCAAAATC
Reverse      ATGGAGTCGATGAAAGAAACAGCTTTCGAATATTGCAGCTTCTGCAAAATC
              *** *****
ORF          TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
Reverse      TGGCATGGACAAAACCAAAGCTACCTTGGAGGAAAAGGCGGAGAAGATGA
              *****
ORF          AGACACGAGACCCGTTCAGAAACAGATGGCTACACAGGTTAAAGAAGAT
Reverse      AGACACGAGACCCGTTCAGAAACAGCTGGCTACACAGGTTAAAGAAGAT
              *****
ORF          AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
Reverse      AAGATCAATCAAGCTGAGATGCAGAAGAGAGAAAACGCGTCAGCACAACGC
              *****
ORF          GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
Reverse      GGCCATGAAAGAAGCGGCTGGAGCCGGAACCGGTTTAGGTTTGGGGACGG
              *****
ORF          CCACTCACTCGACCCTGGACAAGTCGGACACGGCACTGGGACACATCAG
Reverse      CCACTCACTCGACCCTGGACAAGTCGGACACGGCACTGGGACACATCAG

```

```

*****
ORF      ATGTCGGCTCTGCCTGGTCACGGAACGGGACAACCTGACCGACCGCGTTGT
Reverse  ATGTCGGCTCTGCCTGGTCACGGAACGGGACAACCTGACCGACCGCGTTGT
*****
ORF      GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACCTGGTC
Reverse  GGAGGGCACGGCTGTGACCGACCCGATTGGAAGGAACACTGGAACCTGGTC
*****
ORF      GGACAACCGCTCATAACACTCACGTTGGTGGTGGTGGTGCCACCGGGTAC
Reverse  GGACAACCGCTCATAACACACACGTTGGTGGTGGTGG--CCACCG-GTAC
*****
ORF      GGAACCGCGGGGGATATACTGGATAA
Reverse  G-AACCG-----C
* *****

```

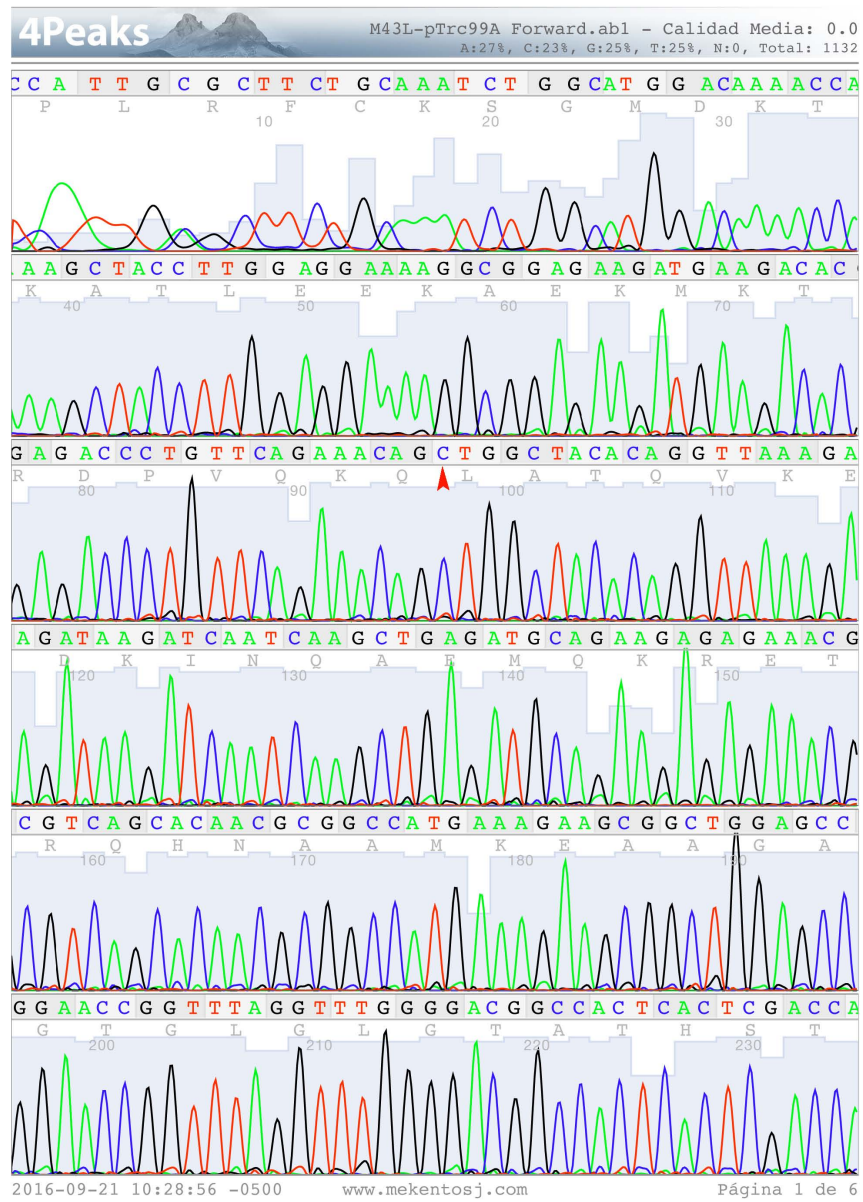
- M19L



- M33L



- M43L



ANEXO VI. Identificación de proteínas por espectrometría de masas.

AtLEA4-2

No. de folio: 54-2014

Laboratorio Universitario de Proteómica
IBT/UNAM

Servicios de identificación, caracterización estructural y cuantificación de proteínas a través de la Espectrometría de Masas de alta resolución.

Usuario: Cesar Luis Cuervas Velázquez
Teléfono: 27668
E. mail: ccuevas@ibt.unam.mx
Jefe de Grupo: Dra. Alejandra Covarrubias
Institución: IBT
Número de Muestras: 01
Organismo: *Arabidopsis thaliana*
Tipo de servicio: Identificación de proteínas
Fecha de entrega de la muestra al LUP: 13/08/2014

RESULTADOS

Los datos espectrométricos fueron sometidos; i) búsqueda general utilizando la base de datos del NCBIInr a través Protein Prospector, ii) búsqueda restringida contra UniProt de *Arabidopsis thaliana* a través Protein Prospector iii) búsqueda en Proteome Discoverer 1.4 utilizando la secuencia proporcionada por el usuario NP_181073. En color verde se indica la secuencias encontradas experimentalmente. Todas las secuencias reportadas fueron verificadas de forma manual.

Accession	Unique identifier assigned to the protein by the FASTA database used to generate the report.
Description	Name of the protein exclusive of the identifier that appears in the Accession column.
Score	Protein score, which is the sum of the scores of the individual peptides.
Coverage	Percentage of the protein sequence covered by identified peptides.
# Peptides	Number of distinct peptide sequences in the protein group.
MW [kDa]	Calculated molecular weight of the protein.
MH+ [Da]	Protonated monoisotopic mass of the peptides, in daltons.
Modifications	Static and dynamic modifications identified in the peptide.
Sequence	Sequence of amino acids that compose the peptide.

MUESTRA-1

Accession	Description	Score	Coverage	# Peptides	MW [kDa]	calc. pI
15226951	late embryogenesis abundant protein 4-2 [Arabidopsis thaliana]	6450.72	80.41%	11	10.5	9.66

MQSAKEKISD MASTAKEKLN IGGAKAQQHA EKT MARTKKE K KLAQER EK S
KEAQAKADLH QSKAEHAADA QVHGHHLP GH STYPTRATGA NYPPGQI

Sequence	Modifications	Charge	MH+ [Da]	RT [min]	# Missed Cleavages
AEHAADAQVHGHHLP GHSTYPTR		4	2489.18667	110.77	0
ATGANYP PGQI		2	1088.54460	5.57	0
EKLNIGGAK		1	929.54303	3.78	1
EKISDMASTAK		3	1180.59024	8.03	1
LNIGGAK		1	672.40552	3.76	0
EKISDMASTAK	M6(Oxidation)	2	1196.58537	121.88	1
KLAQER		3	744.43755	33.22	1
ADLHQSK		3	798.41095	79.25	0
ISDMASTAKEK	M4(Oxidation)	2	1196.58354	23.66	1
MQSAK	M1(Oxidation)	2	564.28000	66.59	0
SKEAQAK		2	761.4152	25.18	1

METODOLOGÍA

Las muestras fueron previamente reducidas con ditioneitol (DTT), alquiladas con iodoacetamida (obtenida con Sigma-Aldrichy) digerida "in gel" con tripsina (Promega Sequencing Grade Modified Trypsin). Los péptidos resultantes fueron aplicados en un sistema LC-MS bomba de nanoflujo EASY-nLC II (Thermo-Fischer Co. San Jose, CA) acoplado a un espectrómetro de masas LTQ-Orbitrap Velos (Thermo-Fischer Co., San Jose, CA) con sistema de ionización tipo nano-electrospray (ESI). La calibración del espectrómetro fue realizado con una solución Calmix (N-butilamina, cafeína, Met-Arg-Phe-Ala (MRFA) y Ultramark 1621 se utilizan para calibrar Thermo Scientific LTQ Velos de trampa de iones y Orbitrap espectrómetros de masas en el modo de ionización positiva ESI. La N-butilamina se incluye para extender a menor rango de calibración de masas) esto permite determinaciones con exactitudes menores a 5 ppm (partes/millón). En el sistema de cromatografía de líquidos se utilizó un sistema gradiente de 10-80% de solvente B (agua/acetronitrilo con 0.1% de ácido fórmico) en 120 minutos utilizando una columna capilar hecha en casa (ID 0.75 µm y 10cm largo RP-C18) . El flujo del sistema LC fue de 400 nanolitros/minuto. Para la fragmentación de los péptidos se utilizó los métodos de CID (Collision-Induced Dissociation) y HCD (High energy Collision Dissociation) donde solamente los iones con carga 2⁺, 3⁺ y 4⁺ fueron seleccionados para los eventos de fragmentación. Fueron desconsiderados los iones con cargas 1⁺, superiores a 5⁺ y de cargas indefinidas. Todos los espectros fueron adquiridos en modo de detección positivo. La ejecución y captura de los datos de fragmentación fueron realizados de forma dependiente del

escaneo total de iones según las cargas pre-determinadas con un ancho de aislamiento de 3.0 (m/z), energía de colisión normalizada de 35 unidades arbitrarias, activación Q de 0.250, tiempo de activación de 10 milisegundos y tiempo máximo de inyección de 10 milisegundos por micro-escaneo. Durante la captura automática de los datos fue utilizado la exclusión dinámica de iones: (i) lista de exclusión de 400 iones, (ii) tiempo de pre-exclusión de 30 segundos y (iii) tiempo de exclusión de 300 segundos.

No. de folio: 60-2014

MUESTRA-1

Found Modifications:

O Oxidation (M)

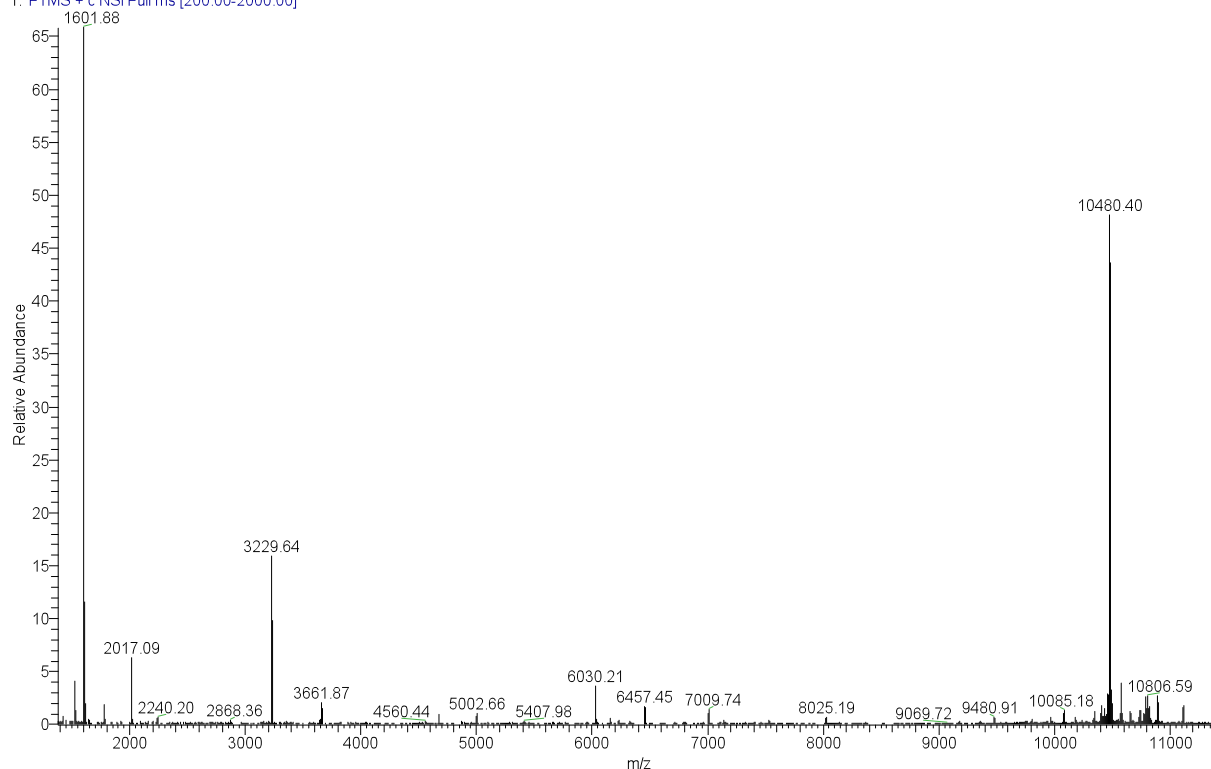
MQSAK**E**KISD **M**ASTAKEKLN **I**GGAN**A**QGHAEK**T**MARTK**E** **K**KL**A**Q**E**REK**S** **K**EAQAKADL**H** **Q**SK**A**EHAADA QVHGHLPGH STYPTATGA NYPPGQI

DETERMINACION DE MASA MOLECULAR

60-2014Cesarmasas_XT_00001_M_

9/10/2014 11:45:42 AM

60-2014Cesarmasas_XT_00001_M_#1 RT: 1.00 AV: 1 NL: 3.87E6
T: FTMS + c NSI Full ms [200.00-2000.00]



AtLEA4-5

No. de folio:68 -2015

Laboratorio Universitario de Proteómica
IBT/UNAM

*Servicios de identificación, caracterización estructural y cuantificación de proteínas
a través de la Espectrometría de Masas de alta resolución.*

Usuario: Cesar Luis Cuevas Velázquez
Teléfono: 27668
Correo electrónico: ccuevas@ibt.unam.mx
Jefe de Grupo: Dra. Alejandra Covarrubias
Institución: IBt
Número de Muestras: 1
Organismo: *A. thaliana*
Tipo de servicio: Determinación de masa molecular
Fecha de entrega de la muestra al LUP: 28 /10/2015

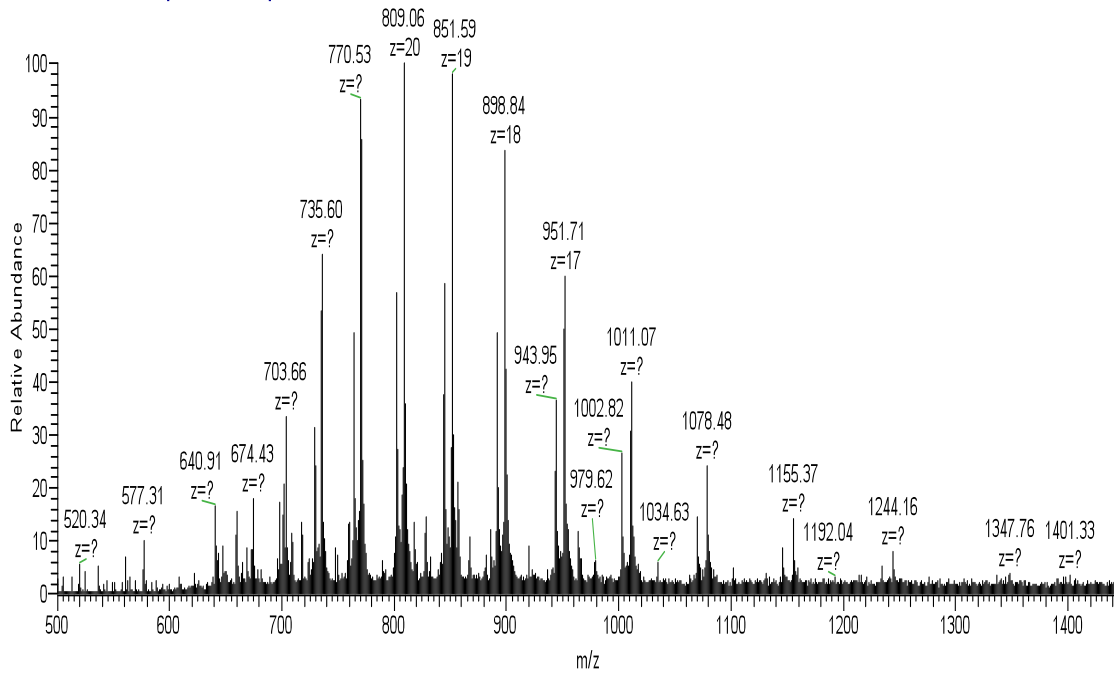
RESULTADOS

- En la tabla de abajo se presenta la determinación de masa molecular realizada de forma automática por el programa Xcalibur. La masa molecular es reportada como “average mass”.

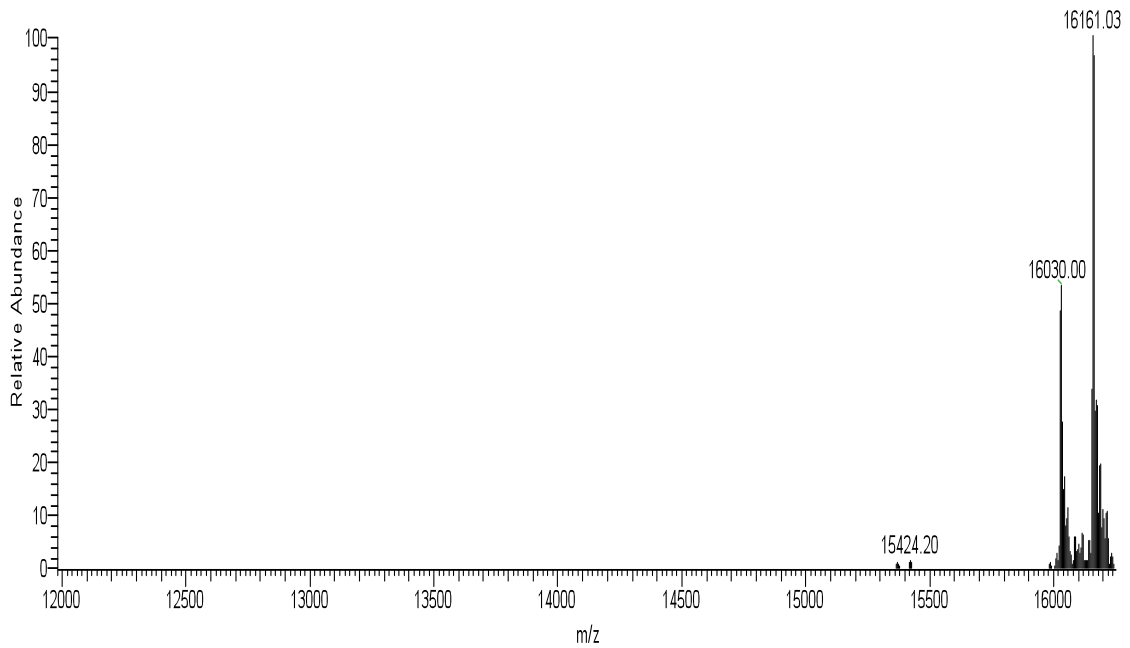
MUESTRA	MASA MOLECULAR PROMEDIO (Da)
68-2015-WT	16161.03 *,16030.00

68-2015-Wt

68-2015-MW-WT #538-611 RT: 4.01-5.02 AV: 74 NL: 1.30E5
T: FTMS + c NSI Full ms [500.00-2000.00]



68-2015-MW-WT XT_00001_M_#1 RT: 1.00 AV: 1 NL: 4.12E4
T: FTMS + c NSI Full ms [500.00-2000.00]



Laboratorio Universitario de Proteómica
IBT/UNAM

*Servicios de identificación, caracterización estructural y cuantificación de proteínas
a través de la Espectrometría de Masas de alta resolución.*

Usuario: Cesar Luis Cuervas Velazquez
Teléfono: 27668
E. mail: ccuevas@ibt.unam.mx
Jefe de Grupo: Dra. Alejandra Covarrubias
Institución: IBT
Número de Muestras: 01
Organismo: *Arabidopsis thaliana*
Tipo de servicio: Identificación de proteínas
Fecha de entrega de la muestra al LUP: 31/07/2014

RESULTADOS

Accession	Unique identifier assigned to the protein by the FASTA database used to generate the report.
Description	Name of the protein exclusive of the identifier that appears in the Accession column.
Score	Protein score, which is the sum of the scores of the individual peptides.
Coverage	Percentage of the protein sequence covered by identified peptides.
# Peptides	Number of distinct peptide sequences in the protein group.
MW [kDa]	Calculated molecular weight of the protein.
MH+ [Da]	Protonated monoisotopic mass of the peptides, in daltons.
Modifications	Static and dynamic modifications identified in the peptide.
Sequence	Sequence of amino acids that compose the peptide.

MUESTRA-1

Found Modifications:

MESMKETASN IAASAKSGMD KTKATLEEK EKMKTRDPVQ KQMATQVKED KINQAEMQKR
ETRQHNAAMK EAAGAGT

Accession	Description	Score	Coverage	# Peptides	MW [kDa]
15240144	late embryogenesis abundant protein 4-5 [Arabidopsis thaliana]	3425.48	44.30	14	16.2
	Sequence	Modifications	Charge	MH+ [Da]	
	MESMKETASNIA ASAK		2	1668.80217	
	mESmKETASNIA ASAK	M1(Oxidation); M4(Oxidation)	3	1700.79013	
	INQAEMQK		2	961.48137	
	EDKINQAEmQK	M9(Oxidation)	2	1349.64238	
	ETASNIAASAK		2	1062.54766	
	EDKINQAEMQK		2	1333.64958	
	QMATQVKEDK		2	1177.59392	
	TRDPVQK		2	843.47069	
	TKATLEEK		2	919.51329	
	MESmKETASNIA ASAK	M4(Oxidation)	2	1684.79717	
	QMATQVK		2	805.42790	
	QHNAAMK		2	799.39006	
	EDKINQAEMQKR		3	1489.75605	
	ATLEEK		2	690.37065	
	QmATQVK	M2(Oxidation)	2	821.42046	
	QmATQVKEDK	M2(Oxidation)	2	1193.59111	
	SGmDKTK	M3(Oxidation)	2	782.36571	
	INQAEmQK	M6(Oxidation)	2	977.47722	
	INQAEmQKRETR	M6(Oxidation)	3	1519.74863	
	EGTAVTDPI		2	902.44725	
	AEKMKTR		2	863.47276	

METODOLOGÍA

Las muestras fueron previamente reducidas con ditioneitol (DTT), alquiladas con iodoacetamida (obtenida con Sigma-Aldrich) digerida “in gel” con tripsina (Promega Sequencing Grade Modified Trypsin). Los péptidos resultantes fueron aplicados en un sistema LC-MS bomba de nanoflujo EASY-nLC II (Thermo-Fischer Co. San Jose, CA) acoplado a un espectrómetro de masas LTQ-Orbitrap Velos (Thermo-Fischer Co., San Jose, CA) con sistema de ionización tipo nano-electrospray (ESI). La calibración del espectrómetro fue realizado con una solución Calmix (N-butilamina, cafeína, Met-Arg-Phe-Ala (MRFA) y Ultramark 1621 se utilizan para calibrar Thermo Scientific LTQ Velos de trampa de iones y Orbitrap espectrómetros de masas en el modo de ionización positiva ESI. La N-butilamina se incluye para extender a menor rango de calibración de masas) esto permite determinaciones con exactitudes menores a 5 ppm (partes/millón). En el sistema de cromatografía de líquidos se utilizó un sistema gradiente de 10-80% de solvente B (agua/acetoneitrilo con 0.1% de ácido fórmico) en 120 minutos utilizando una columna capilar hecha en casa (ID 0.75 μ m y 10cm largo RP-C18) . El flujo del sistema LC fue de 400 nanolitros/minuto. Para la fragmentación de los péptidos se utilizó los métodos de CID (Collision-Induced Dissociation) y HCD (High energy Collision Dissociation) donde solamente los iones con carga 2⁺, 3⁺ y 4⁺ fueron seleccionados para los eventos de fragmentación. Fueron desconsiderados los iones con cargas 1⁺, superiores a 5⁺ y de cargas indefinidas. Todos los espectros fueron adquiridos en modo de detección positivo. La ejecución y

captura de los datos de fragmentación fueron realizados de forma dependiente del escaneo total de iones según las cargas pre-determinadas con un ancho de aislamiento de 3.0 (m/z), energía de colisión normalizada de 35 unidades arbitrarias, activación Q de 0.250, tiempo de activación de 10 milisegundos y tiempo máximo de inyección de 10 milisegundos por micro-escaneo. Durante la captura automática de los datos fue utilizado la exclusión dinámica de iones: (i) lista de exclusión de 400 iones, (ii) tiempo de pre-exclusión de 30 segundos y (iii) tiempo de exclusión de 300 segundos. Los datos espectrométricos fueron sometidos a la búsqueda contra la base de datos del NCBIInr a través del programa ProteinProspector, así como también fueron sometidos a la búsqueda restringida contra PDB de UniProt de *Arabidopsis thalia* y sprot en el Proteome Discoverer1.4 (datos reportados).

Quality Control Report

Date:	2015-01-15
Order Number:	#SP140990
Catalog Number:	423213
Peptide Name:	45RC(GG-81)
Sequence:	GLGLGTATHSTTGQVGHGTGTHQMSALPGHGT GQLTDRVVEGTAVTDPIGRNTGTGRRTAHNTHV GGGGATGYGTGGGYTG
Lot Number:	P141128-DG423213
Molecular Formula:	C ₃₁₉ H ₅₀₇ N ₁₀₇ O ₁₁₅ S ₁
Molecular Weight:	7713.31
Mass Spectral Analysis:	Consistent - See attached
Purity (by HPLC):	95.79% (95% Requested) - see attached
Salt Form:	Trifluoroacetate (TFA Salt)
Quantity:	10.3mg
Appearance:	White to off White Powder
Reconstitution Condition Used in Purification:	H ₂ O:Acetonitrile=4:1
Solubility Test:	Not Requested. For Reconstitution Guideline, please visit: http://biomatik.com/linkfiles/PeptideHandlingGuideline.pdf

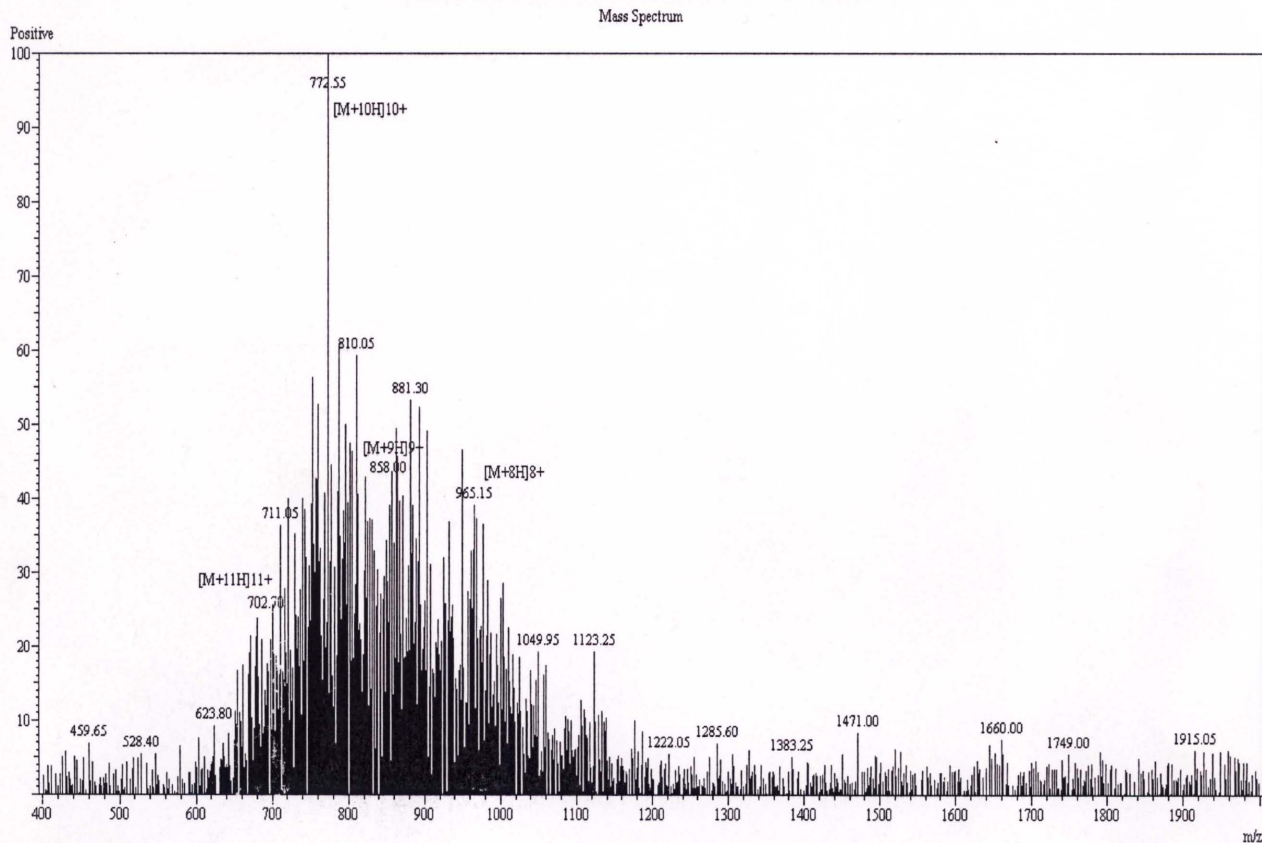
Storage, Reconstitution Condition and Handling

-20°C Upon Arrival. Please refer to the *Peptide Handling Guideline* for reconstitution condition and storage information. It is not recommended that peptides be stored in solution. The shelf life of peptides in solution is rather limited, especially for sequences containing Cysteine, Methionine, Tryptophan, Asparagine and Glutamine. Generally, the best approach is to reconstitute sufficient peptide for immediate usage only. For long-term storage please store at -20°C.

Limited Usage and Warranty

This product is for laboratory research use only. Not for diagnostic or in vivo use. Not for human or drug use. Biomatik warrants material of said quality at the time of delivery. It is the sole responsibility of the customers to determine the adequacy of all materials for any intended or specific purpose or use. Biomatik's sole obligation shall be the replacement of the material without charge or refund of purchase price. This warranty does not apply to a product which has been tampered with or altered in any way or which has been misused or damaged by accident or negligence. All claims must be received in writing (by fax or email) within 30 days upon successful delivery and failure to do so shall constitute a waiver by customers for any and all such claims.

Mass Spectrometry Report

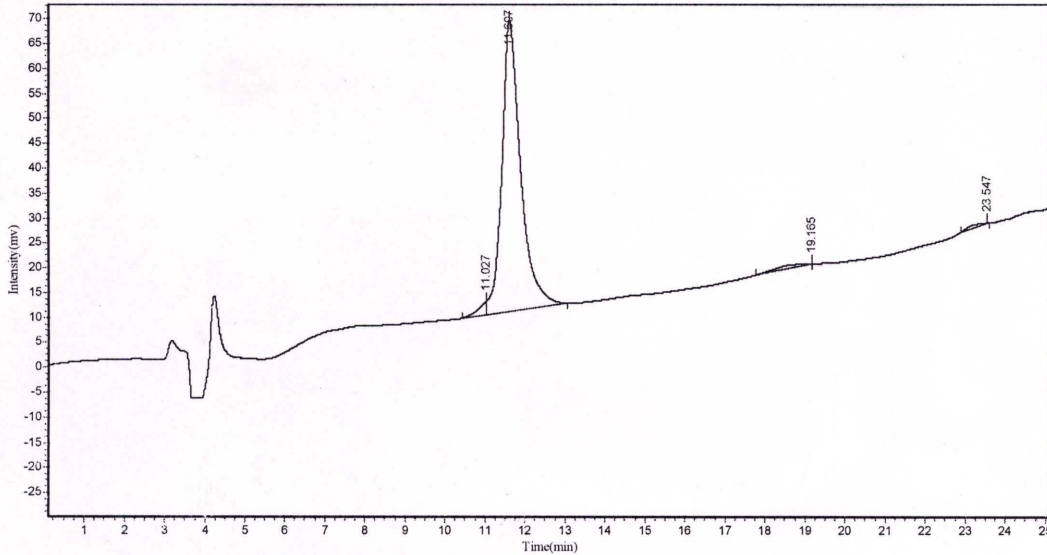


Sample Information
 Date and Time : 2014-12-15 13:56:52
 User : CHAO
 Sample : 45RC GG-81
 Inj. Volume : 1
 MW : 7713.31
 Lot No. : P141128-DG423213

Probe: ESI Probe bias: -3.5kv
 Nebulizer Gas Flow: 1.5L/min Detector: 1.0kv
 DL: -20.0v T. Flow: 0.2ml/min
 DL Temp: 250°C B. conc: 50%H₂O/50%ACN
 Block Temp: 200°C

HPLC Report

Structure : 45RC GG-81
 Lot No : P141128-DG423213
 Column : 4.6×250mm,PLRP-S 100A
 Solvent A : 0.1% trifluoroacetic in 100% acetonitrile
 Solvent B : 0.1% trifluoroacetic in 100% water
 Gradient : A B
 0.01min 20% 80%
 25min 70% 30%
 25.1min 100% 0%
 30min STOP
 Flow rate : 1.0 mL/min
 Wavelength : 220nm
 Volumn : 5ul



Peak No.	Ret Time	Height	Area	Conc
1	11.027	2606.414	36281.719	1.8228
2	11.607	58551.680	1906699.125	95.7944
3	19.165	5.455	30916.805	1.5533
4	23.547	155.923	16510.320	0.8295
Total				100.0000

ANEXO VII. Artículos científicos publicados durante los estudios de doctorado.

The Unstructured N-terminal Region of *Arabidopsis* Group 4 Late Embryogenesis Abundant (LEA) Proteins Is Required for Folding and for Chaperone-like Activity under Water Deficit*

Received for publication, February 6, 2016, and in revised form, March 21, 2016. Published, JBC Papers in Press, March 22, 2016, DOI 10.1074/jbc.M116.720318

Cesar L. Cuevas-Velazquez^{†1}, Gloria Saab-Rincón[§], José Luis Reyes[‡], and Alejandra A. Covarrubias^{‡2}

From the Departamentos de [†]Biología Molecular de Plantas and [§]Ingeniería Celular y Biotecnología, Instituto de Biotecnología, Universidad Nacional Autónoma de México, 62250 Cuernavaca, México

Late embryogenesis abundant (LEA) proteins are a conserved group of proteins widely distributed in the plant kingdom that participate in the tolerance to water deficit of different plant species. *In silico* analyses indicate that most LEA proteins are structurally disordered. The structural plasticity of these proteins opens the question of whether water deficit modulates their conformation and whether these possible changes are related to their function. In this work, we characterized the secondary structure of *Arabidopsis* group 4 LEA proteins. We found that they are disordered in aqueous solution, with high intrinsic potential to fold into α -helix. We demonstrate that complete dehydration is not required for these proteins to sample ordered structures because milder water deficit and macromolecular crowding induce high α -helix levels *in vitro*, suggesting that prevalent conditions under water deficit modulate their conformation. We also show that the N-terminal region, conserved across all group 4 LEA proteins, is necessary and sufficient for conformational transitions and that their protective function is confined to this region, suggesting that folding into α -helix is required for chaperone-like activity under water limitation. We propose that these proteins can exist as different conformers, favoring functional diversity, a moonlighting property arising from their structural dynamics.

Low water availability caused by different environmental conditions such as drought, or low temperatures represents a vulnerable situation for many forms of life, particularly for plants. To contend with and to overcome these adverse environments, numerous complex response mechanisms have been selected in the different species of the plant kingdom. One of the most conserved responses is the accumulation of a group of proteins known as late embryogenesis abundant (LEA)³ pro-

teins (1). LEA proteins have been found in all the orthodox dry seeds (embryos) where they have been searched (1, 2), and they also accumulate in response to water limitation in all vegetative tissues (2, 3). Most LEA proteins show high hydrophilicity, high content of small amino acids, and absence or deficit of hydrophobic residues, properties that are extended to a larger set of proteins called hydrophilins, which have been found in species from the three domains of life and that also accumulate under water deficit (2, 4). The composition of these proteins is also characteristic of a group of proteins known as intrinsically disordered proteins (IDPs) (5, 6). Consistent with the predicted structural disorder for most LEA proteins, structural analyses have confirmed this property for some of them in aqueous solution (7–13). Based on their sequence similarity, LEA proteins have been classified in seven groups or families, each one characterized by the presence of specific sequence motifs (2). In *Arabidopsis thaliana* there are 51 genes encoding LEA proteins from six of the seven families (3). Group 4 LEA (LEA4) proteins are one of the smallest families of LEA proteins in *Arabidopsis* consisting of only three members: AtLEA4-1 (At1g32560), AtLEA4-2 (At2g35300), and AtLEA4-5 (At5g06760) (3, 14). Group 4 LEA proteins are enriched in charged and small amino acid residues, whereas they lack Cys, Phe, and Trp (2, 3, 14). This group is characterized by an N-terminal region ranging from 74 to 78 amino acid residues, containing conserved amino acid sequence motifs. *In silico* analysis predicts that this particular region is able to form an amphipathic α -helix structure. The C-terminal region in this protein family is more variable in sequence and length, and it is predicted to be structurally disordered (2, 14). A phylogenetic analysis of group 4 LEA proteins revealed two subclasses in this family (subgroups 4A and 4B) (14). In *Arabidopsis*, AtLEA4-1 and AtLEA4-2 proteins belong to subgroup 4A, whereas AtLEA4-5 protein fits into subgroup 4B (14).

From the 10 distinctive motifs found in this protein group, the high conservation of motif 2 at the N-terminal region constitutes a signature for this family. The same study also showed that both subgroups emerged from a very early duplication before branching of monocots and dicots, suggesting that this separation gave rise to a subfunctionalization of these subgroups (14). Group 4 LEA proteins and transcripts have been found in dry seeds but also in response to water deficit in vegetative and reproductive tissues (14, 15). Moreover, *Arabidopsis* mutants deficient in group 4 LEA proteins are sensitive to

* This work was partially supported by Grants IN208212 from the Dirección General de Apoyo al Personal Académico/Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT, Universidad Nacional Autónoma de México and Grants 132258 and 221448 from the Consejo Nacional de Ciencia y Tecnología. The authors declare that they have no conflicts of interest with the contents of this article.

The nucleotide sequence(s) reported in this paper has been submitted to the DDBJ/GenBank™/EBI Data Bank with accession number(s) AEE31503.1, AEC09091.1, and AED91062.1.

¹ Supported by a Consejo Nacional de Ciencia y Tecnología doctoral fellowship.

² To whom correspondence should be addressed. Tel.: 52-777-329-1643; Fax: 52-777-313-9988; E-mail: crobles@ibt.unam.mx.

³ The abbreviations used are: LEA, late embryogenesis abundant; IDP, intrinsically disordered proteins; LDH, lactate dehydrogenase; TFE, 2,2,2-trifluoroethanol.

LEA4 Proteins Fold in Response to Water Deficit

water deficit, indicating that these proteins participate in the tolerance to this stress condition (14).

Many different functions have been proposed for group 4 LEA proteins such as membrane protectors, sugar or metal binding, radical scavengers, and protein dehydro- and cryoprotectors (16–20). *A. thaliana* AtLEA4-5 protein was shown to prevent inactivation and conformational changes of reporter enzymes such as lactate dehydrogenase (LDH) and malate dehydrogenase after partial dehydration and freeze/thaw cycles from 1:1 molar ratios, indicating that group 4 LEA proteins have a chaperone-like function to protect other proteins from the effects of water deficit (19, 20).

Studies on animal and bacterial IDPs have shown that these proteins can gain structural order upon binding to a specific partner, interaction that leads to IDP folding (21–25); however, in some other examples, the function of globular chaperones is linked to an order to disorder structural transitions in response to environmental cues such as those imposed by changes in pH or redox state (26, 27). Even though it has been shown that severe dehydration can promote folding of some LEA proteins (8, 11, 13, 28–31), the possibility that the environmental effects caused by mild water deficit in the cell (such as those occurring in vegetative tissues) leads to higher structural order in those IDPs responsive to this stressful environment (*e.g.* LEA proteins and other hydrophilins) and whether this structural changes could be related to their function are still open questions.

In this work, we demonstrate that even though *Arabidopsis* members of subgroups 4A (AtLEA4-2) and 4B (AtLEA4-5) LEA proteins are structurally disordered in solution, low osmotic potentials and macromolecular crowding can induce significant levels of α -helix, particularly in the conserved AtLEA4-5 N-terminal region, whereas the C-terminal region displays high structural disorder. We also show that the AtLEA4-5 N-terminal region is necessary and sufficient for the protective effect of this protein on reporter enzyme activities after freeze-thaw cycles and partial dehydration at low molar ratios. Our data support the hypothesis that cellular environment modulates the structural organization of disordered proteins and that these structural changes are related to their functions.

Experimental Procedures

In Silico Analyses—AtLEA4 proteins were aligned using T-Coffee multiple sequence alignment. Secondary structure prediction was determined using AGADIR helical content predictor (32). Intrinsically disordered tendency was predicted using DISpro (33), PONDR (34), and DISOPRED3 (35).

Plasmid Constructions—ORFs of *AtLEA4-1*, *AtLEA4-2*, and *AtLEA4-5* genes were cloned using cDNA from RNA obtained from *Arabidopsis* dry seeds. AtLEA4-5 ORF was amplified by PCR using specific primers containing at their ends NcoI (5'-AAACCATGGAGTCGATGAAAGAAAC-3') and SalI (5'-GCGGTCGACCCGTTTATCCAGTATATCC-3') restriction sites. This cloning strategy led to a modification in the second codon, which in the recombinant version corresponds to glutamic acid (GAG) instead of glutamine (CAG). The amplicon was cloned into pJET1.2/blunt and subsequently digested with NcoI and SalI for its insertion into pTrc99A vector. To eliminate the Met³³ in AtLEA4-5 ORF used in bacterial cells as an

alternative translation initiation site and responsible of the production of an additional shorter AtLEA4-5 protein, directed mutagenesis of AtLEA4-5 ORF sequence was conducted using the following overlapping primers to exchange Met³³ for a Leu residue: sense (5'-GGAGGAAAAGCGGAGAAGCTGAA-GAC-3') and antisense (5'-GTCTTCAGCTTCTCCGCCTT-TTCTCC-3'). The modified DNA fragment was inserted into pJET1.2/blunt plasmid vector to produce pJET1.2:AtLEA4-5. For protein production, AtLEA4-5 ORF was transferred to pTrc99A plasmid vector by digesting with NcoI and SalI restriction enzymes. The DNA fragments encoding AtLEA4-5 N- and C-terminal regions were obtained from pJET1.2:AtLEA4-5 plasmid, using the following sense and antisense oligonucleotides: 5'-AAACCATGGAGTCGATGAAAGAAAC-3' and antisense 5'-CGCGTCCGACTCAGGTTCCGGCTCCAGC-CGC-3' and sense 5'-AAACCATGGCCGGTTTAGGTTTG-GGGAC-3' and antisense 5'-GCGGTCGACCCGTTTATCC-AGTATATCC-3', respectively, which were also inserted into pTrc99A to generate pTrc99:AtLEA4-5₁₋₇₇ and pTrc99:AtLEA4-5₇₈₋₁₅₈ plasmids, respectively. In all cases, nucleotide sequences were verified accordingly.

Because pTrc99A:AtLEA4-1, pTrc99A:AtLEA4-2, and pTrc99A:AtLEA4-5₇₈₋₁₅₈ did not lead to a successful protein expression in bacteria, instead corresponding ORFs were inserted into the pTYB11 vector to obtain them as intein fusion proteins (IMPACT-CN expression system; New England Biolabs Inc.). To this end, AtLEA4-1 and AtLEA4-2 coding sequences were amplified from pJET1.2 intermediary plasmids using specific oligonucleotides containing SapI and PstI restriction sites: 5'-GGTGGTTGCTCTTCCAACATGCAATCGGCCGAAAC-AGAAG-3' and 5'-GGTGGTCTGCAGTCATTAGTAGTGTGATGATTATGATGTCC-3' for AtLEA4-1 and, 5'-GGTGGTTGCTCTTCCAACATGCAGTCGCGGAAGG-3' and 5'-GGTGGTCTGCAGTCATTAGATCTGTCCC GGCG-3' for AtLEA4-2. To amplify the AtLEA4-5₇₈₋₁₅₈ coding sequence, the oligonucleotides used were 5'-GGTGGTTGCTCTTCCAACACCGGTTTAGGTTTGGGGAC-3' and 5'-GGTGGTCTGCAGTCATTATCCAGTATATCCCCCGC-3'.

Expression and Purification of Recombinant Proteins—Recombinant plasmids derived from pTrc99A or pTYB11 as described above were transformed into *Escherichia coli* BL21(DE3) pLysS competent cells (Promega). Single colonies were inoculated in fresh LB medium containing 100 μ g/ml ampicillin and grown overnight at 37 °C, from which 1 liter of fresh LB medium was inoculated to 0.01 A₆₀₀ and grown at 37 °C to 0.5–0.8 A₆₀₀. At this point, protein expression was induced with 1 mM isopropyl β -D-1-thiogalactopyranoside for 6 h at 25 °C. Cell cultures were harvested by centrifugation. For AtLEA4-5 and AtLEA4-5₁₋₇₇ purification, we used a straightforward method designed for nonacidic recombinant unstructured proteins as described by Campos *et al.* (36). After washing twice with acetone, the protein was resuspended in 10 mM sodium phosphate buffer, pH 7.5, and dialyzed extensively against the same buffer. For intein fused AtLEA4-1, AtLEA4-2, and AtLEA4-5₇₈₋₁₅₈, a different purification procedure was followed. Bacterial pellets were resuspended in lysis buffer (20 mM sodium phosphate, pH 8, 500 mM NaCl, 0.1% Triton X-100) containing one tablet of cOmplete protease inhibitor mix-

ture (Roche) per 50 ml of buffer. The cells were lysed by sonication on ice, and the extract was clarified by centrifugation at $20,000 \times g$ for 30 min at 4 °C. To obtain proteins lacking the intein tag, the clarified extract was loaded onto a chitin column following the procedure described by the manufacturer (IMPACTTM-CN kit). The eluted fractions were analyzed for the presence of the recombinant proteins by SDS-PAGE. In contrast to AtLEA4-2, this analysis showed that AtLEA4-1 and AtLEA4-5_{78–158} are most probably protease-susceptible proteins because we were unable to detect the complete corresponding polypeptides, despite using protease-deficient bacterial strains, as is the case for BL21 and RosettaTM 2(DE3) (Merck-Millipore), strains successfully used to purify a number of recombinant proteins from different organisms. Fractions containing AtLEA4-2 were pooled and extensively dialyzed against 10 mM sodium phosphate buffer, pH 7.5. The purity and identity of the different purified proteins were confirmed by SDS-PAGE and by LC-MS. LC-MS was performed by the Proteomic Facility of the Instituto de Biotecnología/Universidad Nacional Autónoma de México following standard methods using LQT-Orbitrap Velos (Thermo-Fisher) mass spectrometer with nanospray ionization system. Proteins were quantified using their molar extinction coefficient at 280 nm ($\epsilon = 2,980 \text{ M}^{-1} \text{ cm}^{-1}$). Because AtLEA4-5_{1–77} lacks aromatic amino acid residues, its concentration was determined by Bradford assay and verified by SDS-PAGE comparing with known concentrations of AtLEA4-5. Purified proteins were conserved by lyophilization or in aliquots at $-80 \text{ }^\circ\text{C}$ until used. Because it was not possible to obtain whole AtLEA4-5_{78–158}, the complete chemically synthesized polypeptide corresponding to this truncated protein was purchased from Biomatik (Cambridge, Canada) and extensively dialyzed before use. This polypeptide was verified by mass spectrometry and HPLC.

Far UV Circular Dichroism Spectroscopy—AtLEA4-2, AtLEA4-5, and AtLEA4-5_{1–77} recombinant proteins, as well as AtLEA4-5_{78–158} polypeptide were diluted to 0.3 mg/ml and far UV CD spectra were recorded using a Jasco J-715 CD spectropolarimeter (JASCO Analytical Instruments) on a 0.1-cm-path length cell from 190 to 250 nm. Desired temperature was regulated with a Peltier temperature-controlled cell holder (PTC-4235; JASCO). Three spectra were averaged and smoothed to reduce noise. Each spectrum was acquired every 1 nm with 2-s average time per point and 1-nm band pass. Secondary structure estimation was calculated using Dichroweb software (37, 38). The CDSSTR algorithm was used with 4, 7, and SP175 data sets. These assays were reproduced using proteins from at least three independent purification batches.

In Vitro Freeze-Thaw Assay—Freeze-thaw *in vitro* assays were carried out as previously described by Reyes *et al.* (19) with small modifications. Briefly, LDH from rabbit muscle (Roche) was diluted to a final concentration of 250 nM (monomer) with or without the corresponding test protein (AtLEA4-2, AtLEA4-5, AtLEA4-5_{1–77}, AtLEA4-5_{78–158}, or lysozyme) in buffer 25 mM Tris-HCl, pH 7.5. The different test proteins were set to the desired molar ratio from 0.5:1 to 20:1 (test protein: LDH), considering 250 nM as 1:1 molar ratio. Mixtures in a final volume of 100 μl were frozen for 30 s in liquid N₂ and subsequently thawed at 25 °C in a thermomixer (Eppendorf). This

procedure constituted one freeze-thaw cycle, which was repeated up to seven times. After the treatment, LDH activity was measured as reported (19). Enzyme activities for each sample were measured in at least three independent tests (each one with three technical replicates). These experiments were reproduced with proteins from at least three independent purification batches.

In Vitro Partial Dehydration Assay—Partial dehydration *in vitro* assays were performed as previously described by Reyes *et al.* (20) with small modifications. LDH from rabbit muscle (Roche) was diluted to 250 nM (monomer) as the final concentration, in the presence or absence of the corresponding test protein (AtLEA4-2, AtLEA4-5, AtLEA4-5_{1–77}, or AtLEA4-5_{78–158}) in buffer of 25 mM Tris-HCl, pH 7.5. The different test proteins were set to 5:1 molar ratio (1.25 μM). Mixtures in a final volume of 25 μl were placed in a SpeedVac concentrator (Savant Instruments), and water was evaporated to achieve $\geq 98\%$ water loss, keeping constant temperature. The percentage of partial water loss was defined as the amount of water evaporated from the samples by means of weight. Partially dehydrated samples were rehydrated to the initial weight with water, assuring that all solutes were completely resuspended. LDH activity was measured as described above. These assays were reproduced using protein samples from at least three independent purification batches.

Statistical Analyses—Statistical analyses were carried out using one-way analysis of variance test. Significant differences were calculated with Tukey's multiple comparison post-test ($p < 0.01$).

Results

A. thaliana Group 4 Late Embryogenesis Abundant Proteins Are Intrinsically Disordered Proteins in Aqueous Solution—In *A. thaliana*, there are three genes encoding group 4 LEA proteins: AtLEA4-1, AtLEA4-2, and AtLEA4-5 (3, 14). These proteins have low molecular masses between 10.5 and 16.2 kDa and pI values between 8.95 and 9.67. All of them present a 74–78-amino acid long N-terminal domain, highly conserved across all group 4 LEA proteins described so far in the plant kingdom, which contains distinctive motifs for this protein family (Fig. 1A). Because of the amino acid composition of this group of LEA proteins, it was proposed that they are IDPs (2, 12). *In silico* analyses using PONDR (VLXT) (34) (Fig. 1B), DISpro (33), and DISOPRED3 (35) (data not shown) indicated that, despite their sequence similarity, they possess different levels of disorder.

To characterize the LEA4 proteins structural features, we expressed and purified the three recombinant proteins in *E. coli*. AtLEA4-2 and AtLEA4-5 were obtained with more than 95% purity (data not shown). Because AtLEA4-1 was mostly degraded during different expression and purification procedures, the rest of the experiments were performed only with AtLEA4-2 and AtLEA4-5, representing the two LEA4 subgroups, 4A and 4B, respectively (14).

To determine the secondary structure of AtLEA4-2 and AtLEA4-5 in solution, the purified proteins were analyzed by far UV CD. The results obtained corroborated that both proteins are mostly disordered in solution over a wide range of temperatures (Fig. 2, A and B) and pH (data not shown), with

LEA4 Proteins Fold in Response to Water Deficit

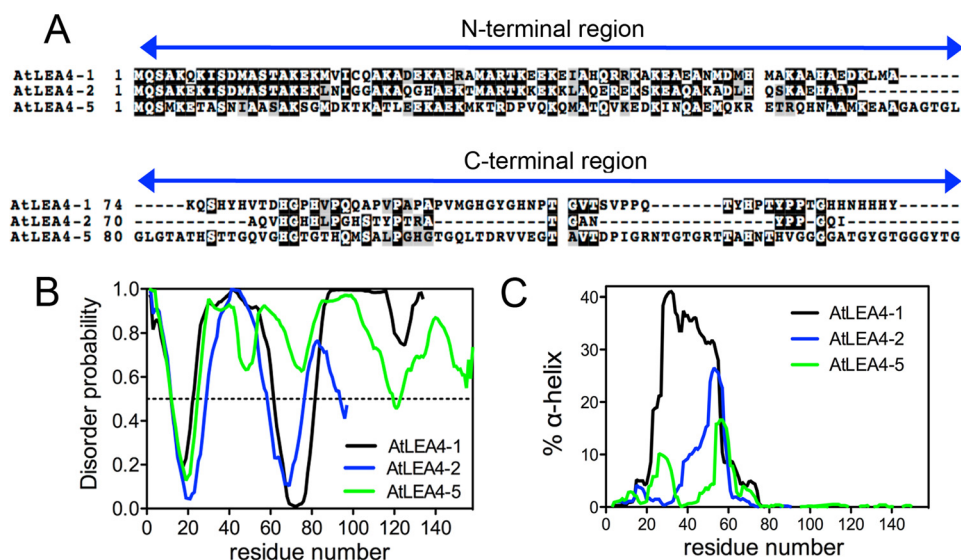


FIGURE 1. *In silico* structure prediction of *Arabidopsis* group 4 LEA proteins. *A*, sequence alignment of *Arabidopsis* group 4 LEA proteins showing N-terminal conserved and C-terminal variable regions. *B*, AtLEA4-1 (black), AtLEA4-2 (blue), and AtLEA4-5 (green) structural disorder levels using PONDR predictor. According to this algorithm, proteins with disorder probability values above 0.5 are considered highly disordered. *C*, percentage of α -helix predicted using AGADIR for AtLEA4-1 (black), AtLEA4-2 (blue), and AtLEA4-5 (green).

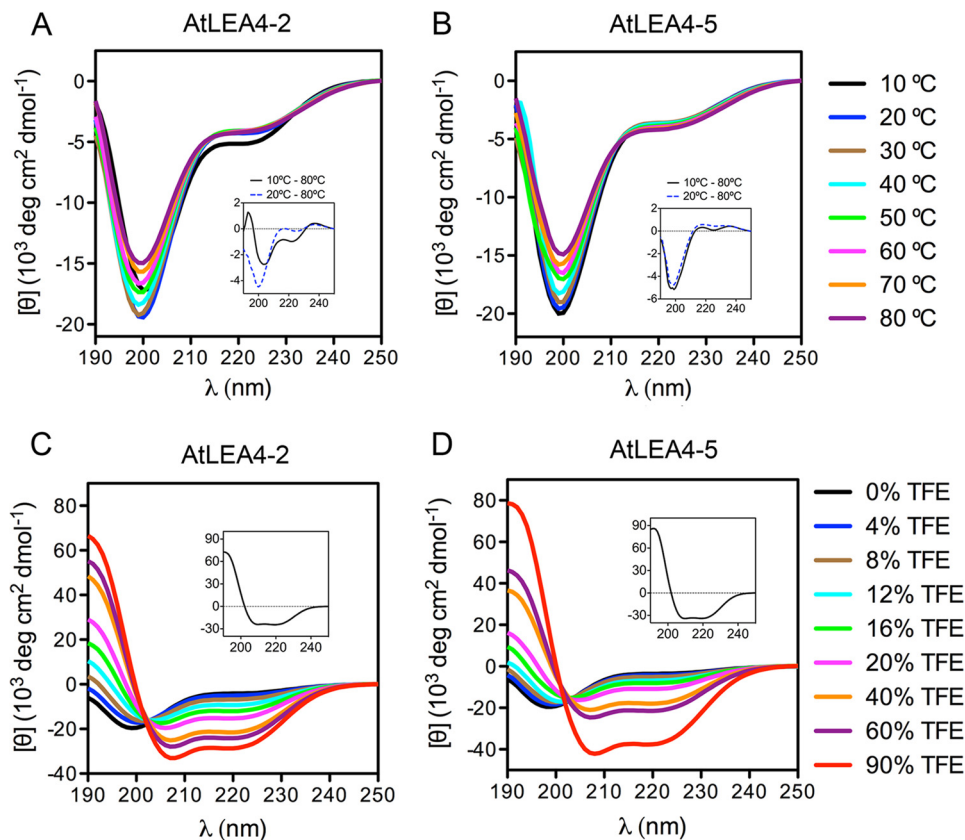


FIGURE 2. Far UV CD spectra of AtLEA4-2 and AtLEA4-5 in aqueous solution at various temperatures and in different TFE concentrations. *A* and *B*, far UV CD spectra of AtLEA4-2 (*A*) and AtLEA4-5 (*B*) at 10 °C (black), 20 °C (blue), 30 °C (brown), 40 °C (cyan), 50 °C (green), 60 °C (magenta), 70 °C (orange), and 80 °C (purple). Insets in *A* and *B* show $\Delta 10-80$ °C (continuous black line) and $\Delta 20-80$ °C (dashed blue line) difference spectra. *C* and *D*, AtLEA4-2 (*C*) and AtLEA4-5 (*D*) in TFE/water mixtures at 0% (black), 4% (blue), 8% (brown), 12% (cyan), 16% (green), 20% (magenta), 40% (orange), 60% (purple), and 90% (red) TFE. The difference spectra ($\Delta 90-0\%$ TFE) are shown as insets in *C* and *D*. These results were reproduced at least four times in autonomous experiments, using three independent purification batches of both proteins.

the characteristic negative band for random coil structures around 198 nm. Both spectra showed a significant negative signal around 222 nm (typical of α -helix), suggesting that these proteins possess residual α -helix structure. The comparative

analysis of the AtLEA4-2 CD difference spectra obtained at different temperatures ($\Delta 10-80$ °C and $\Delta 20-80$ °C) revealed that this protein is able to form α -helix structures at low temperatures (10 °C), whereas at higher temperatures the protein is

TABLE 1

Percentage of helix, strand, and unordered structures in AtLEA4-2 and AtLEA4-5

Secondary structure content in AtLEA4-2 and AtLEA4-5 proteins was obtained by far UV CD spectrometry and calculated with Dichroweb server.

Treatment	AtLEA4-2			AtLEA4-5		
	Helix	Strand	Unordered	Helix	Strand	Unordered
			%			%
Aqueous solution	5	33	61	5	34	61
4% TFE	7	32	61	6	33	61
8% TFE	16	25	58	6	32	61
12% TFE	21	23	56	17	23	59
16% TFE	35	15	51	18	25	55
20% TFE	42	15	44	32	17	51
40% TFE	57	9	33	49	13	39
60% TFE	62	9	30	57	8	35
90% TFE	80	3	16	88	2	10
10% glycerol	5	34	60	4	34	61
20% glycerol	8	32	61	7	36	57
30% glycerol	18	22	60	7	34	59
40% glycerol	21	23	57	18	24	57
50% glycerol	33	12	53	18	23	59
60% glycerol	43	14	43	34	15	51
70% glycerol	49	9	41	42	11	48
80% glycerol	54	12	33	46	9	45
45% PEG 5000	37	11	52	39	9	53

mostly disordered (Fig. 2A, inset). By contrast, a similar analysis indicated that AtLEA4-5 is mostly disordered under all temperatures tested (Fig. 2B, inset). We did not find evidence of extended helical conformations (e.g. poly-L-proline II) for any of these proteins, such as those found in LEA groups 1, 2, and 6 (10, 39). Altogether, these data demonstrate that AtLEA4 proteins are IDPs with residual α -helix structure in aqueous solution.

AtLEA4-2 and AtLEA4-5 Have the Potential to Acquire High Levels of Ordered Structure—To determine the intrinsic ability of AtLEA4-2 and AtLEA4-5 to attain helicity, CD analyses were performed in the presence of different concentrations of 2,2,2-trifluoroethanol (TFE), a well known α -helix inducer (40, 41). Increasing TFE concentrations promote α -helix formation in both proteins, as revealed by the progressive increase in $[\theta]_{198}$ toward positive values and the transition of the minimum at $[\theta]_{222}$ toward more negative values (Fig. 2, C and D). Difference spectra showed that both proteins are able to gain high helicity levels; however, $\Delta 90-0\%$ TFE indicates that AtLEA4-5 reaches higher α -helix percentage than AtLEA4-2 (Fig. 2, C and D, insets). Assessment of protein secondary structure using Dichroweb (37) indicated that AtLEA4-2 adopts 7% of α -helix in 4% TFE, and it reaches up to 80% α -helix in 90% TFE (Table 1), showing a decrease in its structural disorder from 61% in aqueous solution to 16% in 90% TFE. The α -helix content for AtLEA4-5 increased from 6% to 88% in 4% to 90% TFE, whereas its disordered structure decreased from 61% to 10% (Table 1). Both AtLEA4-2 and AtLEA4-5 CD spectra obtained with 4% to 90% TFE showed isodichroic points (Fig. 2, C and D), indicating that these proteins are able to adopt two structural conformations in equilibrium under these conditions: one mostly unstructured favored in aqueous solution (U_{4-2} and U_{4-5}) and a second one with higher helicity promoted by increasing TFE concentrations (F_{4-2} and F_{4-5}). The fitted straight lines obtained from transition diagrams for AtLEA4-2 and AtLEA4-5 support the formation of two conformers for both proteins

(Fig. 3, E and F), as was indicated by the presence of isodichroic points. These data demonstrate that AtLEA4-2 and AtLEA4-5 possess an intrinsic ability to form α -helical species, which under some conditions are in equilibrium with unfolded conformations ($U_{4-2} \leftrightarrow F_{4-2}$; $U_{4-5} \leftrightarrow F_{4-5}$).

AtLEA4-2 and AtLEA4-5 Fold to α -Helix in Response to Low Water Availability and Macromolecular Crowding Induced in Vitro—Previous studies have shown that recombinant LEA proteins from different groups acquire secondary structure, mostly α -helix, when subjected to complete dehydration (8, 11, 13, 28–31). Given the intrinsic potential of AtLEA4-2 and AtLEA4-5 to gain α -helix conformation and because these proteins accumulate even under mild water limitation, we hypothesized that *in vitro* conditions limiting water availability could induce changes in their secondary structure. Addition of increasing glycerol concentrations led to a notorious progressive gain in α -helix structure in both proteins, as shown by the $[\theta]_{198}$ change toward positive values and a deeper minimum at $[\theta]_{222}$ (Fig. 3, A and B). Analysis using Dichroweb estimated a small difference in α -helix content between these two proteins: 54% α -helix for AtLEA4-2 and 46% for AtLEA4-5 at the highest glycerol concentration (80%) (Table 1). We observed the presence of isodichroic points in both cases (Fig. 3, A and B), which was also supported by their corresponding transition diagrams (Fig. 3, E and F), that together with those obtained from TFE treatments showed that AtLEA4-2 and AtLEA4-5 seem to follow the same folding pathway to α -helix under both treatments (Fig. 3, E and F).

Inside living cells, macromolecules are present at very high concentrations (~ 400 g/liter) (42–44), a condition that is typically known as macromolecular crowding (45, 46). This state is further exacerbated in cells under water deficit, reaching macromolecular concentrations up to ~ 900 g/liter upon severe dehydration (42). PEG was used to simulate a crowded environment *in vitro*. The addition of 45% PEG 5000 to AtLEA4-2 or AtLEA4-5 solutions clearly induced changes in their structural conformations (Fig. 3, C and D). Dichroweb estimations indicate 37 and 39% α -helix gains for AtLEA4-2 and AtLEA4-5, respectively (Table 1). Together, these data indicate that AtLEA4-2 and AtLEA4-5 can acquire secondary structure under low water availability or macromolecular crowding *in vitro*, possibly reflecting what occurs in plant cells under water deficit.

The N-terminal Region of AtLEA4 Proteins Is Necessary and Sufficient for the Conformational Changes Induced by Water Deficit—Plant group 4 LEA proteins are characterized by the presence of conserved motifs at their N-terminal region (2, 14). *In silico* analysis predicts that this region of 70, 74, and 77 residues in AtLEA4-1, AtLEA4-2, and AtLEA4-5, respectively, has a higher propensity to adopt α -helical conformations than the C-terminal region (Fig. 1C). To test this prediction, we performed far UV CD experiments using AtLEA4-5 truncated versions: one containing the first 77 amino acids, named AtLEA4-5₁₋₇₇, and a second one corresponding to the 81 amino acid C-terminal region, from residues 78 to 158 (AtLEA4-5₇₈₋₁₅₈) (47).

In contrast to *in silico* predictions, AtLEA4-5₁₋₇₇ behaved as a disordered protein in aqueous solution under all tempera-

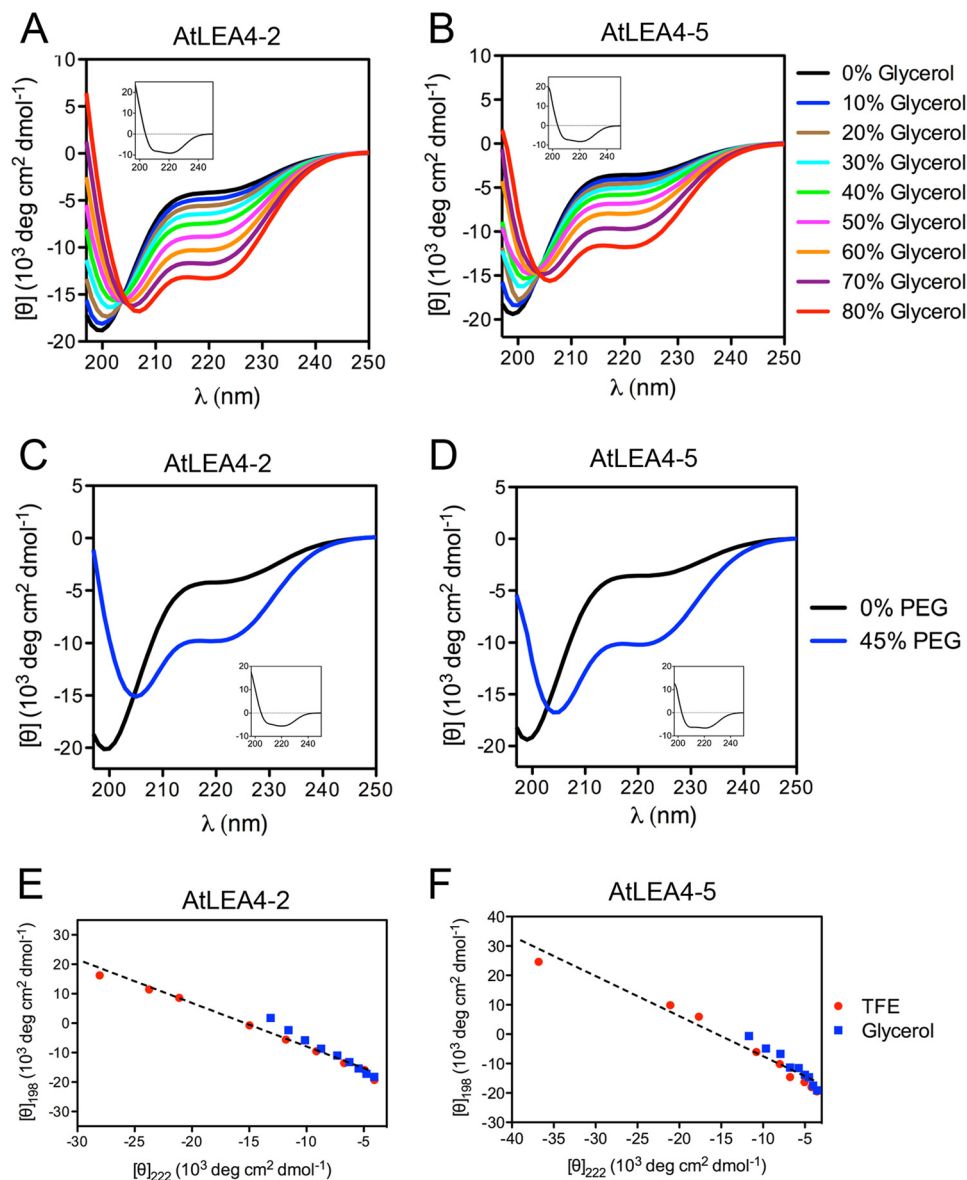


FIGURE 3. Far UV CD spectra of AtLEA4-2 and AtLEA4-5 in different concentrations of glycerol or PEG. A and B, far UV CD spectra of AtLEA4-2 (A) and AtLEA4-5 (B) in glycerol/water mixtures at 0% (black), 10% (blue), 20% (brown), 30% (cyan), 40% (green), 50% (magenta), 60% (orange), 70% (purple), and 80% (red) glycerol. Difference spectra ($\Delta 80-0\%$ glycerol) are shown as insets in A and B. C and D, AtLEA4-2 (C) and AtLEA4-5 (D) in aqueous solution (black line) and in PEG/water mixture at 45% PEG 5000 (blue line). Difference spectra ($\Delta 45-0\%$ PEG) are shown as insets in C and D. E and F, transition diagrams for AtLEA4-2 (E) and AtLEA4-5 (F) were obtained using the ellipticity values at 198 and 222 nm from TFE and glycerol titrations. Dashed lines represent the linear fits of the data. These results were reproduced at least four times in autonomous experiments, using three independent purification batches of both proteins.

tures tested (Fig. 4A), with a difference spectrum profile ($\Delta 10-80^\circ\text{C}$) similar to that obtained for the complete AtLEA4-5 (Figs. 2B and 4A, insets). However, the addition of increasing TFE concentrations progressively induced α -helix formation in the truncated protein containing the N-terminal region (Fig. 4B). Likewise, treatments with glycerol and PEG led to the same behavior in this protein (Fig. 4, C and D), reaching up to 55 and 42% α -helix at the highest glycerol and PEG concentrations, respectively (Table 2). Also complete AtLEA4-5 and AtLEA4-2, far UV CD spectra from AtLEA4-5₁₋₇₇ showed isodichroic points when treated with progressively increasing TFE or glycerol concentrations, indicating that this protein is in equilibrium between two states in these conditions: disordered and α -helix conformations (Fig. 4, B and C). The high linear corre-

lation of transition diagrams confirmed this observation (data not shown).

Using the chemically synthesized peptide corresponding to the AtLEA4-5C-terminal region (AtLEA4-5₇₈₋₁₅₈), far UV CD analysis confirmed that AtLEA4-5₇₈₋₁₅₈ is highly disordered in aqueous solution at different temperatures (Fig. 4E). Unlike AtLEA4-2, AtLEA4-5, or AtLEA4-5₁₋₇₇, the difference spectrum $\Delta 10-80^\circ\text{C}$ of AtLEA4-5₇₈₋₁₅₈ indicated the possible formation of poly-L-proline-like structures (Fig. 4E, inset). In contrast to the N-terminal region of these proteins, addition of up to 60% TFE had only minor effects on the structure of AtLEA4-5₇₈₋₁₅₈ (Fig. 4F), inducing 4% to 19% α -helix (Table 2). It was not until a high TFE concentration (90%) was reached that AtLEA4-5₇₈₋₁₅₈ became helical (Fig. 4F), indicating a low

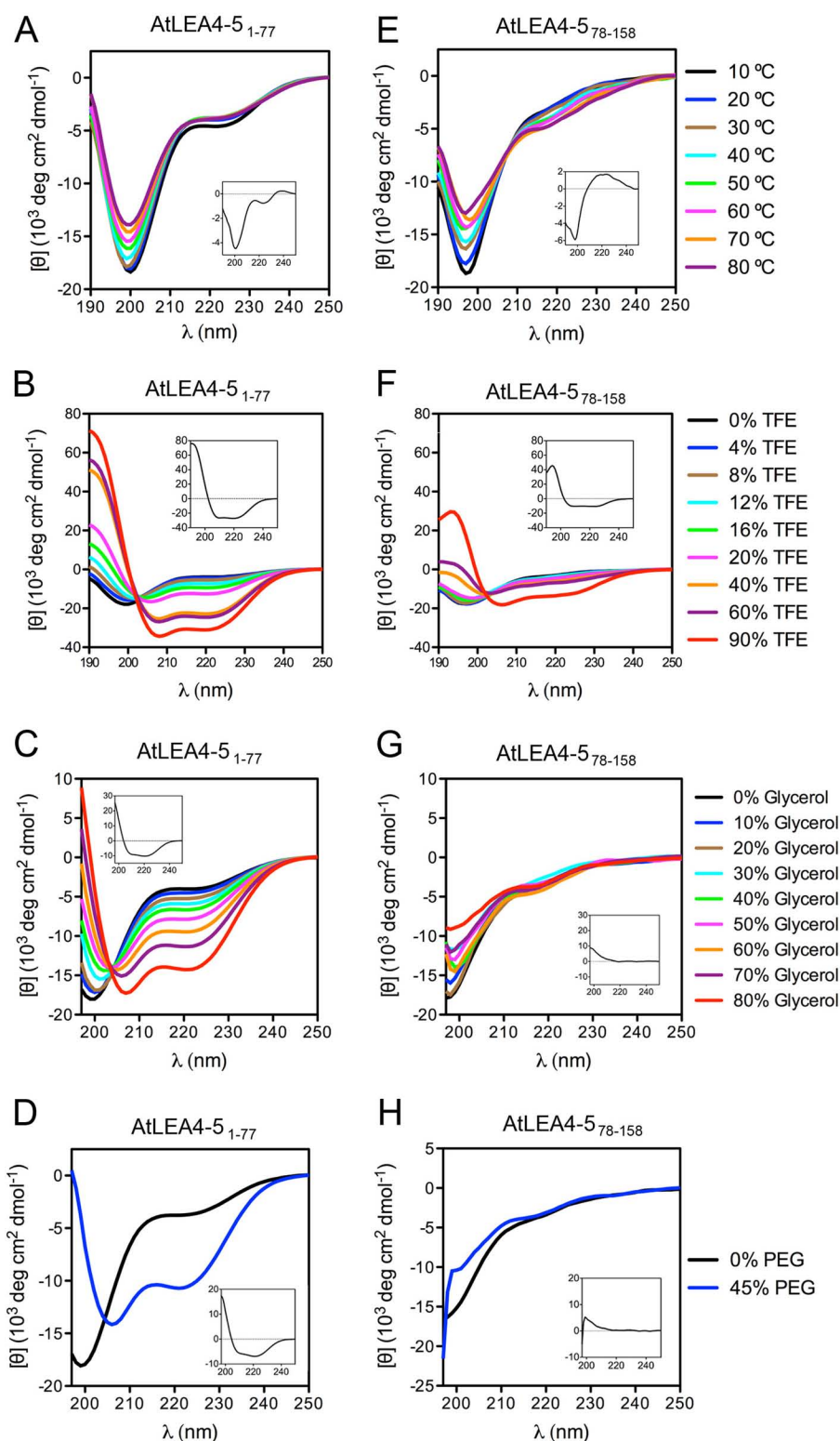


FIGURE 4. N-terminal conserved region of AtLEA4-5 is necessary and sufficient to induce folding to α -helix under the different conditions tested. Far UV CD spectra of AtLEA4-5₁₋₇₇ (A–D) and AtLEA4-5₇₈₋₁₅₈ (E–H) under different temperatures (A and E), different TFE (B and F) and glycerol concentrations (C and G), and under 45% PEG 5,000 (D and H). Difference spectra ($\Delta 10-80$ °C; $\Delta 90-0\%$ TFE; $\Delta 80-0\%$ glycerol; $\Delta 45-0\%$ PEG) are shown as insets in each graph. These results were reproduced at least four times in autonomous experiments, using three independent purification batches of AtLEA4-5₁₋₇₇ protein and one batch for AtLEA4-5₇₈₋₁₅₈ polypeptide.

intrinsic competence to acquire ordered structures. This finding was further supported by the results obtained from the addition of glycerol or PEG to AtLEA4-5₇₈₋₁₅₈ solutions, which showed no effect on its structure (Fig. 4, G and H), given that the

negative band at $[\theta]_{222}$ did not show any change and that only a slight increase in the $[\theta]_{198}$ signal was detected. Under the highest glycerol (80%) and PEG (45%) concentrations, AtLEA4-5₇₈₋₁₅₈ only showed 5% α -helix formation (Table 2). Comparison

LEA4 Proteins Fold in Response to Water Deficit

between the difference spectra for AtLEA4-5₁₋₇₇ and AtLEA4-5₇₈₋₁₅₈ in TFE, glycerol and PEG showed the magnitude of the AtLEA4-5₁₋₇₇ folding to α -helix, compared with AtLEA4-5₇₈₋₁₅₈ (Fig. 4, A–H, insets). In accordance, the transition diagram for glycerol titration of AtLEA4-5₇₈₋₁₅₈ showed no linear behavior (data not shown). Together, these data demonstrate that the AtLEA4-5 N-terminal region (AtLEA4-5₁₋₇₇) is necessary and sufficient to drive α -helix conformations in this protein under low water availability or macromolecular crowding conditions.

The N-terminal Region of AtLEA4 Proteins Is Necessary and Sufficient to Prevent Inactivation of Lactate Dehydrogenase Caused by Freeze-Thaw Cycles and Partial Dehydration—Because the motif conservation and ability to fold into α -helical conformations of the AtLEA4-5 are preferentially located at the N-terminal region as compared with its C-terminal region, we

TABLE 2

Percentage of helix, strand, and unordered structures in AtLEA4-5₁₋₇₇ and AtLEA4-5₇₈₋₁₅₈

Secondary structure content in AtLEA4-2 and AtLEA4-5 proteins was obtained by far UV CD spectrometry and calculated with Dichroweb server. NA, not available.

Treatment	AtLEA4-5 ₁₋₇₇			AtLEA4-5 ₇₈₋₁₅₈		
	Helix	Strand	Unordered	Helix	Strand	Unordered
		%			%	
Aqueous solution	6	32	61	3	38	58
4% TFE	6	33	59	4	37	58
8% TFE	7	32	60	4	37	58
12% TFE	18	25	57	5	36	59
16% TFE	22	24	53	5	35	59
20% TFE	36	16	48	5	36	59
40% TFE	59	9	31	8	33	58
60% TFE	65	7	27	19	28	54
90% TFE	80	3	16	49	15	35
10% glycerol	6	33	60	NA	NA	NA
20% glycerol	16	23	62	NA	NA	NA
30% glycerol	19	24	58	6	45	51
40% glycerol	19	23	57	0	53	45
50% glycerol	24	18	57	6	36	56
60% glycerol	39	12	48	4	41	55
70% glycerol	45	9	46	NA	NA	NA
80% glycerol	55	8	35	5	35	58
45% PEG 5000	42	9	50	NA	NA	NA

asked about the competence of these individual segments to protect the reporter enzyme LDH from deleterious effects caused by freeze-thaw cycles and partial dehydration, as was previously shown for the complete protein (19, 20). For this purpose, LDH in the presence or absence of AtLEA4-5, AtLEA4-2, AtLEA4-5₁₋₇₇, or AtLEA4-5₇₈₋₁₅₈ was subjected to *in vitro* freeze-thaw cycles and partial dehydration treatments. AtLEA4-2, the smallest protein of this group, mostly consisting of the N-terminal region (Fig. 1A), showed a similar protective effect on LDH activity as AtLEA4-5 after partial dehydration (Fig. 5A). Interestingly, a comparable protection was produced by AtLEA4-5₁₋₇₇, contrasting with the negligible protective levels showed by the AtLEA4-5 C-terminal region (AtLEA4-5₇₈₋₁₅₈), whose values were rather close to those shown by lysozyme, an unrelated globular protein (Fig. 5, A and B). A similar conclusion can be drawn when different molar ratios of these proteins were used in freeze-thaw *in vitro* assays, where it is evident that the various proteins containing the N-terminal region showed an equivalent protective trend, in opposition to the C-terminal region represented by AtLEA4-5₇₈₋₁₅₈ or to lysozyme (Fig. 5B). This analysis revealed the highest and lowest protective efficiencies for AtLEA4-5 and AtLEA4-5₇₈₋₁₅₈, respectively, among the different proteins tested (Fig. 5B). Because the hydrophilicity index and length are similar between the N-terminal and C-terminal regions, these data are consistent with the hypothesis that the protective activity of these proteins is rather dependent on the conserved motifs present in the N-terminal region and/or on its ability to fold into α -helical conformations.

Discussion

During the last nearly 20 years, research on the so-called IDPs has challenged the classical structure-function paradigm (48). IDPs that either completely lack a well defined three-dimensional structure or contain short regions of disorder, known as intrinsically disordered regions (IDRs) within folded

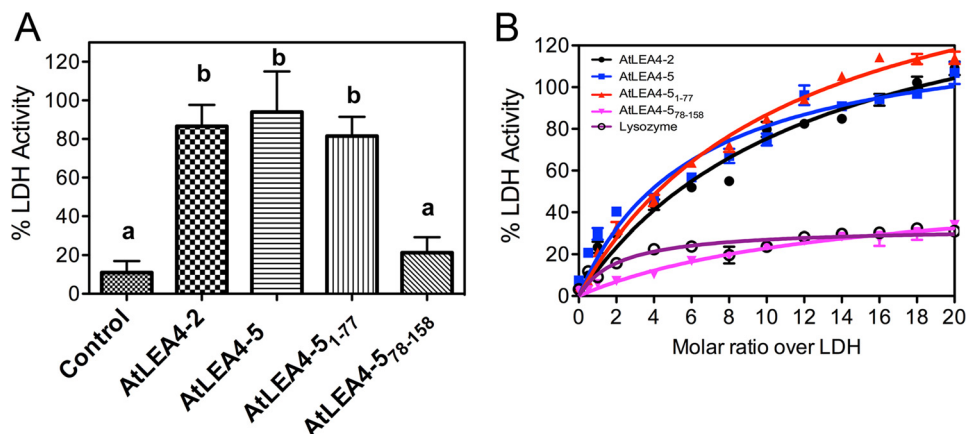


FIGURE 5. The N-terminal conserved region of AtLEA4-5 is necessary and sufficient to prevent inactivation of LDH after partial dehydration and freeze/thaw cycles *in vitro*. A, remaining LDH enzymatic activity after $\geq 98\%$ water loss induced *in vitro* of samples containing LDH without any additive (control) or with AtLEA4-2, AtLEA4-5, AtLEA4-5₁₋₇₇, and AtLEA4-5₇₈₋₁₅₈ in a 5:1 molar ratio (additive:LDH). B, remaining LDH enzymatic activity after seven freeze-thaw cycles of LDH without any additive or with increasing molar amounts (from 0:1 to 20:1 molar ratio additive:LDH) of AtLEA4-2 (black line with circles), AtLEA4-5 (blue line with squares), AtLEA4-5₁₋₇₇ (red line with up triangles), AtLEA4-5₇₈₋₁₅₈ (magenta line with inverted triangles), and lysozyme (purple line with open circles). Error bars indicate S.E. of three independent tests (with three internal repetitions). Letters indicate significant differences calculated with Tukey's multiple comparison post-test ($p < 0.01$). The data in B were fit to a hyperbola curve, and significant differences were subsequently calculated with Tukey's multiple comparison post-test ($p < 0.01$). These results were reproduced at least four times in autonomous experiments, using three independent purification batches of AtLEA4-2, AtLEA4-5, and AtLEA4-5₁₋₇₇ proteins, and one batch for AtLEA4-5₇₈₋₁₅₈ polypeptide.

domains are highly abundant in eukaryotic proteomes and perform important functions (6). In plants, IDPs participate in developmental control, light perception, transcriptional regulation, and response to abiotic stress (12). LEA proteins are the plant protein family with the highest number of proteins either predicted or characterized as IDPs (12). There is experimental evidence showing that LEA proteins from different groups acquire α -helix after complete drying (8, 11, 13, 28–31). These findings suggest that these proteins form α -helix in the dry seed, but little is known about LEA proteins structural behavior under less severe conditions, such as those present in vegetative tissues under drought. Mouillon *et al.* (49) showed that three *Arabidopsis* group 2 LEA proteins (Cor47, Lti29, and Lti30) are IDPs in solution and that they remain disordered when subjected to low water potential or macromolecular crowding simulated *in vitro* with glycerol and PEG, respectively. Even though Cor47 showed a slight folding into α -helix under high concentrations of glycerol and PEG, the authors propose that dehydrins have evolved to stay disordered under cellular conditions and that disorder might be required to properly fulfill their function (49). A similar effect was also observed experimentally for α -casein, MAP2c, and p21^{Cip1}, three different animal IDPs that show no conformational change when subjected to macromolecular crowding, supporting the idea that the physiological state of these IDPs is also disordered (50).

Even though, most LEA proteins can be considered as IDPs, the analysis of their amino acid sequences indicates variety in their potential to attain different levels of secondary structure (2). Because LEA proteins accumulate not only under severe dehydration, such as that occurring in the dry seed, but also in different tissues under a wide range of water limitation, we investigated the possibility that these proteins could adopt secondary structures under conditions prevailing upon water loss, not necessarily as extreme as those occurring in dry seeds. Group 4 LEA proteins in *Arabidopsis* represented a suitable set of proteins for this purpose, given that there are two protein subtypes (subgroups 4A and 4B): a long variant conformed by two distinctive N-terminal and C-terminal regions (AtLEA4-5, representative of subgroup 4B) and a short one with only the conserved N-terminal region (AtLEA4-2, representative of subgroup 4A) (14). Furthermore, *in silico* analysis showed that in contrast to the C-terminal region, the N-terminal region sequence has the potential to fold into α -helix conformations offering an appropriate system to experimentally compare their structural properties under different environments. In this work, we demonstrate not only that the N-terminal regions of AtLEA4-2 and AtLEA4-5 have an intrinsic capacity to fold into α -helix as shown by the far UV CD spectra in the presence of TFE but also that they gain significant helicity in low water potential and macromolecular crowded solutions. Interestingly, the AtLEA4-5 C-terminal region, which we show is disordered in aqueous solution as predicted, remains disordered under all conditions tested. These results indicate the presence of two functional domains in this LEA protein group. The N-terminal region can become ordered under conditions of water restriction, short of absolute dryness, whereas the persistent disorder of the C-terminal region could indicate a requirement to expose some amino acid residues for a more effective or

additional function. Such possibility could be related to the abundance of His residues in this region, which seems to be involved in their binding to metal ions (18), suggesting multifunctionality in some LEA protein families.

All LEA proteins studied to date from groups 1, 2, 3, 4, and 6 show structural disorder in aqueous solution, but most of them also have an intrinsic potential to acquire helical conformations in the presence of TFE (12). These results suggest that there are conditions where LEA proteins exhibit such structural transformation. For most of these LEA proteins, folding has been detected after extreme dehydration (8, 11, 13, 28–31), but extreme conditions may not be necessary in all cases to induce structural transformation. We show in this work that group 4 LEA proteins reach up to 54 and 39% α -helix under less severe environments regarding water availability and macromolecular crowding, respectively. The levels of α -helix formation for group 4 LEA proteins are significantly higher than those estimated for LEA proteins from other groups under similar treatments (12, 39), indicating that the ability to fold is not necessarily the same for different IDPs or intrinsically disordered regions. This is also supported by the wide range of α -helix formation observed among different LEA proteins upon comparable treatments (9, 12, 39, 49, 51, 52). Other proteins seem to be unable to gain ordered conformations such as Rab18, a group 2 LEA protein (9). This information indicates diversity in their structural plasticity, action mechanisms, and/or in their function.

The results from the protection activity assays demonstrate that the protective role of LEA 4 proteins on LDH under low water availability and/or molecular crowding is confined to the conserved N-terminal region of this protein family. There is no apparent participation of the C-terminal region, which completely lacks this safeguard function. These observations imply that the ability of a LEA4 N-terminal domain to gain helicity is related to this chaperone-like activity, particularly under conditions prevailing in water-deficit environments. Based on *in vitro* evidence, it has been proposed that one mechanism for this activity involves protein-protein interactions (20, 53–55). A comprehensive view brings into consideration the existence of different LEA4 conformers in equilibrium (partially folded or unfolded) under crowded or water-deficit environments, which supports the existence of preformed secondary structural elements, denominated as prestructured motifs that could be implicated in the recognition of different and specific binding partners (56). Our findings strongly suggest that water deficit leads to the stabilization of particular conformations in group 4 LEA proteins that may allow the exposure of different motifs necessary for the binding of their target molecules, hence supporting the idea that LEA proteins of this group function as a structural ensemble, whose dynamism can be modulated by environmental conditions (57–60). This proposed mode of action also exhibits possible binding promiscuity, in consonance with their role as chaperone-like molecules needed during water scarcity. At this point, we cannot discount the possibility that different conformers in these proteins could favor different functions such as protection of proteins and/or membranes, as well as metal binding, a moonlighting property arising from their structural dynamics.

LEA4 Proteins Fold in Response to Water Deficit

In conclusion, the findings reported here indicate that the *in vitro* chaperone-like function of the intrinsically disordered group 4 LEA proteins is closely associated to their ability to adopt ordered structural conformations under prevailing conditions in water-deficit environments. The high correlation found between accumulation under water deficit and intrinsic structural disorder, common features in typical LEA proteins and other hydrophilins (4), suggests a functional advantage for this attribute throughout evolution, not only to maintain structural plasticity that could avoid undesirable consequences under stressful environments, such as dramatic structural modifications leading to a functional breakdown, but also to gain functional and mechanistic diversity given their conformational freedom.

Author Contributions—C. L. C.-V. and A. A. C. designed the experimental strategy. C. L. C.-V. conducted cloning, protein expression, and protein purification. C. L. C.-V. and G. S.-R. performed circular dichroism spectroscopy experiments. C. L. C.-V. and J. L. R. carried out *in vitro* protection assays. A. A. C. supervised research. C. L. C.-V., G. S.-R., and A. A. C. analyzed data. C. L. C.-V. and A. A. C. wrote the article, which was read and approved by all authors.

Acknowledgments—We thank H. Jane Dyson for critical reviewing this manuscript, Rosa M. Solórzano for technical assistance in protein purification, and Francisco Campos for the gift of pTrc99A/AtLEA4-*S*_{1–77} plasmid. We are also grateful to the proteomics (Unidad de Proteómica) and DNA sequencing (Unidad de Síntesis y Secuenciación de ADN) core facilities at the Instituto de Biotecnología of the Universidad Nacional Autónoma de México.

References

1. Bray, E. A. (1997) Plant responses to water deficit. *Trends Plant Sci.* **2**, 48–54
2. Battaglia, M., Olvera-Carrillo, Y., Garcarrubio, A., Campos, F., and Covarrubias, A. A. (2008) The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* **148**, 6–24
3. Hundertmark, M., and Hinch, D. K. (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9**, 118
4. Garay-Arroyo, A., Colmenero-Flores, J. M., Garcarrubio, A., and Covarrubias, A. A. (2000) Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.* **275**, 5668–5674
5. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208
6. Tompa, P. (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **37**, 509–516
7. Hughes, S., and Graether, S. P. (2011) Cryoprotective mechanism of a small intrinsically disordered dehydrin protein. *Protein Sci.* **20**, 42–50
8. Hundertmark, M., Popova, A. V., Rausch, S., Seckler, R., and Hinch, D. K. (2012) Influence of drying on the secondary structure of intrinsically disordered and globular proteins. *Biochem. Biophys. Res. Commun.* **417**, 122–128
9. Mouillon, J. M., Gustafsson, P., and Harryson, P. (2006) Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiol.* **141**, 638–650
10. Rivera-Najera, L. Y., Saab-Rincón, G., Battaglia, M., Amero, C., Pulido, N. O., García-Hernández, E., Solórzano, R. M., Reyes, J. L., and Covarrubias, A. A. (2014) A group 6 late embryogenesis abundant protein from common bean is a disordered protein with extended helical structure and oligomer-forming properties. *J. Biol. Chem.* **289**, 31995–32009
11. Shih, M. D., Hsieh, T. Y., Lin, T. P., Hsing, Y. I., and Hoekstra, F. A. (2010) Characterization of two soybean (*Glycine max* L.) LEA IV proteins by circular dichroism and Fourier transform infrared spectrometry. *Plant Cell Physiol.* **51**, 395–407
12. Sun, X., Rikkerink, E. H., Jones, W. T., and Uversky, V. N. (2013) Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* **25**, 38–55
13. Tolleter, D., Jaquinod, M., Mangavel, C., Passirani, C., Saulnier, P., Manon, S., Teyssier, E., Payet, N., Avelange-Macherel, M. H., and Macherel, D. (2007) Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *Plant Cell* **19**, 1580–1589
14. Olvera-Carrillo, Y., Campos, F., Reyes, J. L., Garcarrubio, A., and Covarrubias, A. A. (2010) Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in *Arabidopsis*. *Plant Physiol.* **154**, 373–390
15. Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J. U. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**, 501–506
16. Dang, N. X., Popova, A. V., Hundertmark, M., and Hinch, D. K. (2014) Functional characterization of selected LEA proteins from *Arabidopsis thaliana* in yeast and *in vitro*. *Planta* **240**, 325–336
17. Hundertmark, M., Dimova, R., Lengefeld, J., Seckler, R., and Hinch, D. K. (2011) The intrinsically disordered late embryogenesis abundant protein LEA18 from *Arabidopsis thaliana* modulates membrane stability through binding and folding. *Biochim. Biophys. Acta* **1808**, 446–453
18. Liu, G., Xu, H., Zhang, L., and Zheng, Y. (2011) Fe binding properties of two soybean (*Glycine max* L.) LEA4 proteins associated with antioxidant activity. *Plant Cell Physiol.* **52**, 994–1002
19. Reyes, J. L., Campos, F., Wei, H., Arora, R., Yang, Y., Karlson, D. T., and Covarrubias, A. A. (2008) Functional dissection of hydrophilins during *in vitro* freeze protection. *Plant Cell Environ.* **31**, 1781–1790
20. Reyes, J. L., Rodrigo, M. J., Colmenero-Flores, J. M., Gil, J. V., Garay-Arroyo, A., Campos, F., Salamini, F., Bartels, D., and Covarrubias, A. A. (2005) Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects *in vitro*. *Plant Cell Environ.* **28**, 709–718
21. Borchers, W., Theillet, F. X., Kätzer, A., Finzel, A., Mishall, K. M., Powell, A. T., Wu, H., Manieri, W., Dieterich, C., Selenko, P., Loewer, A., and Daughdrill, G. W. (2014) Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat. Chem. Biol.* **10**, 1000–1002
22. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60
23. Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* **91**, 741–752
24. Rogers, J. M., Wong, C. T., and Clarke, J. (2014) Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *J. Am. Chem. Soc.* **136**, 5197–5200
25. Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025
26. Reichmann, D., Xu, Y., Cremers, C. M., Ilbert, M., Mittelman, R., Fitzgerald, M. C., and Jakob, U. (2012) Order out of disorder: working cycle of an intrinsically unfolded chaperone. *Cell* **148**, 947–957
27. Tapley, T. L., Körner, J. L., Barge, M. T., Hupfeld, J., Schauerer, J. A., Gafni, A., Jakob, U., and Bardwell, J. C. (2009) Structural plasticity of an acid-activated chaperone allows promiscuous substrate binding. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5557–5562
28. Goyal, K., Tisi, L., Basran, A., Browne, J., Burnell, A., Zurdo, J., and Tunnacliffe, A. (2003) Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *J. Biol. Chem.* **278**, 12977–12984
29. Popova, A. V., Hundertmark, M., Seckler, R., and Hinch, D. K. (2011) Structural transitions in the intrinsically disordered plant dehydration stress protein LEA7 upon drying are modulated by the presence of membranes. *Biochim. Biophys. Acta* **1808**, 1879–1887
30. Shih, M. D., Hsieh, T. Y., Jian, W. T., Wu, M. T., Yang, S. J., Hoekstra, F. A., and Hsing, Y. I. (2012) Functional studies of soybean (*Glycine max* L.) seed

- LEA proteins GmPM6, GmPM11, and GmPM30 by CD and FTIR spectroscopy. *Plant Sci.* **196**, 152–159
31. Shimizu, T., Kanamori, Y., Furuki, T., Kikawada, T., Okuda, T., Takahashi, T., Mihara, H., and Sakurai, M. (2010) Desiccation-induced structuralization and glass formation of group 3 late embryogenesis abundant protein model peptides. *Biochemistry* **49**, 1093–1104
 32. Muñoz, V., and Serrano, L. (1994) Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* **1**, 399–409
 33. Cheng, J. L., Sweredoski, M. J., and Baldi, P. (2005) Accurate prediction of protein disordered regions by mining protein structure data. *Data Min. Knowl. Disc.* **11**, 213–222
 34. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins* **42**, 38–48
 35. Jones, D. T., and Cozzetto, D. (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863
 36. Campos, F., Guillén, G., Reyes, J. L., and Covarrubias, A. A. (2011) A general method of protein purification for recombinant unstructured non-acidic proteins. *Protein Expr. Purif.* **80**, 47–51
 37. Lobley, A., Whitmore, L., and Wallace, B. A. (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **18**, 211–212
 38. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668–W673
 39. Soulages, J. L., Kim, K., Walters, C., and Cushman, J. C. (2002) Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. *Plant Physiol.* **128**, 822–832
 40. Buck, M. (1998) Trifluoroethanol and colleagues: cosolvents come of age: recent studies with peptides and proteins. *Q. Rev. Biophys.* **31**, 297–355
 41. Luo, P., and Baldwin, R. L. (1997) Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* **36**, 8413–8421
 42. Ellis, R. J. (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* **11**, 114–119
 43. Zimmerman, S. B., and Minton, A. P. (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomed. Struct.* **22**, 27–65
 44. Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **222**, 599–620
 45. Minton, A. P. (2000) Implications of macromolecular crowding for protein assembly. *Curr. Opin. Struct. Biol.* **10**, 34–39
 46. Minton, A. P. (1997) Influence of excluded volume upon macromolecular structure and associations in “crowded” media. *Curr. Opin. Biotech.* **8**, 65–69
 47. Campos, F., Zamudio, F., and Covarrubias, A. A. (2006) Two different late embryogenesis abundant proteins from *Arabidopsis thaliana* contain specific domains that inhibit *Escherichia coli* growth. *Biochem. Biophys. Res. Commun.* **342**, 406–413
 48. Wright, P. E., and Dyson, H. J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29
 49. Mouillon, J. M., Eriksson, S. K., and Harryson, P. (2008) Mimicking the plant cell interior under water stress by macromolecular crowding: disordered dehydrin proteins are highly resistant to structural collapse. *Plant Physiol.* **148**, 1925–1937
 50. Szasz, C. S., Alexa, A., Toth, K., Rakacs, M., Langowski, J., and Tompa, P. (2011) Protein disorder prevails under crowded conditions. *Biochemistry* **50**, 5834–5844
 51. Haaning, S., Radutoiu, S., Hoffmann, S. V., Dittmer, J., Giehm, L., Otzen, D. E., and Stougaard, J. (2008) An unusual intrinsically disordered protein from the model legume *Lotus japonicus* stabilizes proteins *in vitro*. *J. Biol. Chem.* **283**, 31142–31152
 52. McCubbin, W. D., Kay, C. M., and Lane, B. G. (1985) Hydrodynamic and optical-properties of the wheat-germ Em protein. *Can. J. Biochem. Cell Biol.* **63**, 803–811
 53. Cuevas-Velazquez, C. L., Rendón-Luna, D. F., and Covarrubias, A. A. (2014) Dissecting the cryoprotection mechanisms for dehydrins. *Front. Plant Sci.* **5**, 583
 54. Olvera-Carrillo, Y., Luis Reyes, J., and Covarrubias, A. A. (2011) Late embryogenesis abundant proteins: versatile players in the plant adaptation to water limiting environments. *Plant Signal. Behav.* **6**, 586–589
 55. Tompa, P., and Kovacs, D. (2010) Intrinsically disordered chaperones in plants and animals. *Biochem. Cell Biol.* **88**, 167–174
 56. Lee, S. H., Kim, D. H., Han, J. J., Cha, E. J., Lim, J. E., Cho, Y. J., Lee, C., and Han, K. H. (2012) Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. *Curr. Protein Pept. Sci.* **13**, 34–54
 57. Cino, E. A., Karttunen, M., and Choy, W. Y. (2012) Effects of molecular crowding on the dynamics of intrinsically disordered proteins. *PLoS One* **7**, e49876
 58. Mittag, T., Kay, L. E., and Forman-Kay, J. D. (2010) Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **23**, 105–116
 59. Mittag, T., Orlicky, S., Choy, W. Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M., and Forman-Kay, J. D. (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17772–17777
 60. Uversky, V. N. (2013) Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **1834**, 932–951

The Unstructured N-terminal Region of *Arabidopsis* Group 4 Late Embryogenesis Abundant (LEA) Proteins Is Required for Folding and for Chaperone-like Activity under Water Deficit

Cesar L. Cuevas-Velazquez, Gloria Saab-Rincón, José Luis Reyes and Alejandra A. Covarrubias

J. Biol. Chem. 2016, 291:10893-10903.

doi: 10.1074/jbc.M116.720318 originally published online March 22, 2016

Access the most updated version of this article at doi: [10.1074/jbc.M116.720318](https://doi.org/10.1074/jbc.M116.720318)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 60 references, 18 of which can be accessed free at <http://www.jbc.org/content/291/20/10893.full.html#ref-list-1>



Dissecting the cryoprotection mechanisms for dehydrins

Cesar L. Cuevas-Velazquez, David F. Rendón-Luna and Alejandra A. Covarrubias*

Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, México

Edited by:

Steffen P. Graether, University of Guelph, Canada

Reviewed by:

Vasileios Fotopoulos, Cyprus University of Technology, Cyprus
Saman Seneweera, The University of Melbourne, Australia
Steffen P. Graether, University of Guelph, Canada

*Correspondence:

Alejandra A. Covarrubias,
Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Colonia Chamilpa, 62210 Cuernavaca, Morelos, México
e-mail: crobles@ibt.unam.mx

One of the common responses of plants to water deficit is the accumulation of the so-called late embryogenesis abundant (LEA) proteins. *In vitro* studies suggest that these proteins can protect other macromolecules and cellular structural components from the impairments caused by water limitation. Their binding to phospholipids, nucleic acids and/or to divalent cations has suggested multi-functionality. Genetic analyses indicate that these proteins are required for an optimal adjustment of plants to this insult. This diverse information has conducted to propose different models for LEA proteins action mechanisms. Many of these properties are shared by group 2 LEA proteins or dehydrins (DHNs), one of the LEA protein families for which large amount of data is available. This manuscript focuses on the different mechanisms proposed for this LEA protein group by analyzing published data derived from *in vitro* cryoprotection assays. We compared the molar ratio of protectant:enzyme needed to preserve 50% of the initial activity per enzyme monomer to assess different mechanisms of action. Our results add evidence for protein-protein interaction as a protection mechanism but also indicate that some DHNs might protect by different means. The strength and weakness of the proposed protection mechanisms are discussed.

Keywords: dehydrins, late embryogenesis abundant proteins, cryoprotection, water deficit, abiotic stress, intrinsically disordered proteins

DEHYDRINS, A PLANT SPECIFIC GROUP OF LEA PROTEINS

Late embryogenesis abundant (LEA) proteins are a group of enigmatic proteins that have been strongly associated with plant responses to water deficit (Battaglia et al., 2008; Hinch and Thalhhammer, 2012). They accumulate mainly in dry seeds but also in vegetative tissues when plants experience water deficit such as drought, freezing and high salinity. Among the different LEA proteins, group 2 LEA proteins (D11) have been the most studied (Rorat, 2006; Hara, 2010; Hanin et al., 2011). Proteins in this group are also known as dehydrins (DHNs), and to date they have been found only in plants (Battaglia et al., 2008; Hara, 2010). Due to the vast majority of DHNs reports over other LEA proteins, people not familiar with these proteins assume that all LEA proteins are DHNs; however, this is not the case, LEA proteins represent a rather large group of diverse proteins. Depending on sequence similarity and the presence of particular motifs, LEA proteins have been classified in at least seven groups or families (Battaglia et al., 2008). Although there is sequence similarity within proteins in each group, different LEA protein groups exhibit virtually no sequence conservation with each other. Nevertheless, typical LEA proteins share distinctive physicochemical characteristics such as high hydrophilicity, high content of Gly, Ala and Ser, and lack or underrepresentation of Cys, Trp, and other hydrophobic amino acids (Dure, 1993; Garay-Arroyo et al., 2000; Hundertmark and Hinch, 2008). These characteristics suggest that LEA proteins from different groups might have similar functional properties.

Computational analyses have indicated that, as other LEA proteins, DHNs lack stable tridimensional structures, leading to be considered as intrinsically disordered proteins (IDPs; Close, 1996;

Battaglia et al., 2008). Some of them have been experimentally characterized as IDPs in solution (Ismail et al., 1999; Soulages et al., 2003; Bokor et al., 2005). In spite of this structural flexibility, it has been shown that some DHNs gain ordered structure in the presence and binding to other molecules such as lipids, nucleic acids, or metal ions (Koag et al., 2003; Hara et al., 2009; Eriksson et al., 2011; Rahman et al., 2011).

Dehydrins have been defined by the presence of a Lys-rich segment, or K-segment, which may be repeated several times. Moreover, they are further classified in sub-classes depending on the representation of some of their distinctive conserved motifs, such as the Y-segment (Tyr-rich), or the S-segment (Ser-track). Accordingly, five sub-classes can be distinguished: $K_{(n)}$, $SK_{(n)}$, $K_{(n)}S$, $Y_{(n)}K_{(n)}$, and $Y_{(n)}SK_{(n)}$ (n = number of repeats), for each of which different functions have been proposed (Hara, 2010; Hanin et al., 2011).

Several approaches have been followed to determine the function of these proteins, using *in vivo* and *in vitro* experimental systems. Even though for some LEA protein groups it has been shown their participation in the response of vascular plants to water limiting environments by reverse genetics (Manfre et al., 2006; Kim and Nam, 2010; Olvera-Carrillo et al., 2010), this has not been the case for DHNs, mostly due to the large number of members found within this family (*Arabidopsis* group 2 LEA proteins presents 10 members; Battaglia et al., 2008; Hundertmark and Hinch, 2008). However, a contribution to salt and osmotic stress tolerance was reported for the two DHN genes (*PpDHNA* and *PpDHNB*) of the moss *Physcomitrella patens*, for which targeted knockout mutants were characterized (Saavedra et al., 2006; Ruibal et al., 2012). Binding to other macromolecules such as

negatively charged lipids and DNA (Koag et al., 2003; Hara et al., 2009) have led to propose that DHNs can protect the integrity of biological membranes and nucleic acids from the effects caused by low water availability. Also, the binding to divalent cations such as Ca^{2+} , Zn^{2+} , Fe^{3+} , Co^{2+} , Ni^{2+} , and Cu^{2+} has suggested that DHNs might act as buffer for these metals under water deficit (Kruger et al., 2002; Alsheikh et al., 2005; Hara et al., 2005). The ability of DHNs to bind to such diverse set of ligands could be due to their structural flexibility.

An extensively explored possibility for DHNs function has been their competence to protect other proteins from the effects resulting from water scarcity in the cellular environment. This hypothesis has been addressed by *in vitro* assays, where water limitation is imposed by partial water dehydration or by freeze/thaw cycles. The results from these experiments have demonstrated that different DHNs can prevent the inactivation of reporter enzymes [lactate dehydrogenase (LDH); alcohol dehydrogenase (ADH); firefly luciferase; citrate synthase (CS); β -glucosidase G, (β glG); and glucose oxidase (GOD/POD)] under these different water deficit conditions indicating that some of them are cryo- and/or dehydro-protectors (Sanchez-Ballesta et al., 2004; Tantos et al., 2009; Brini et al., 2010; Drira et al., 2013).

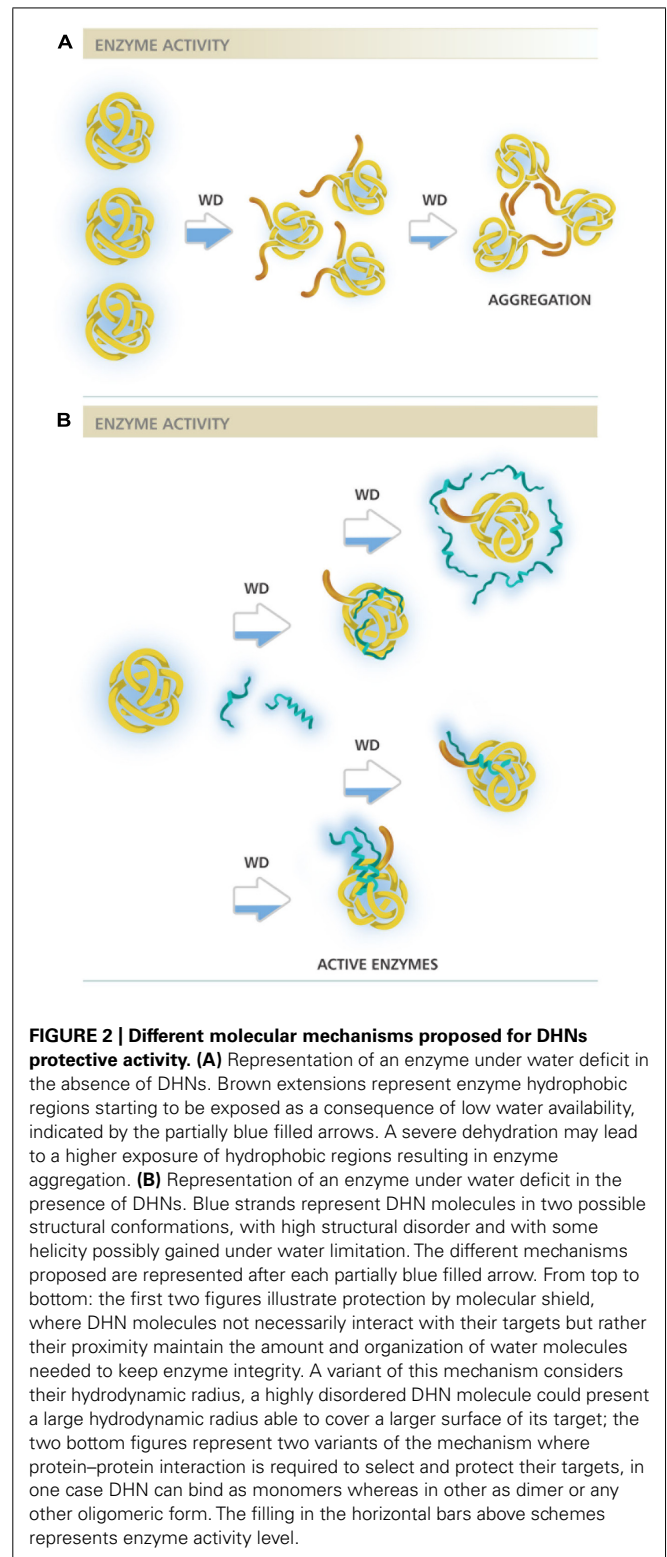
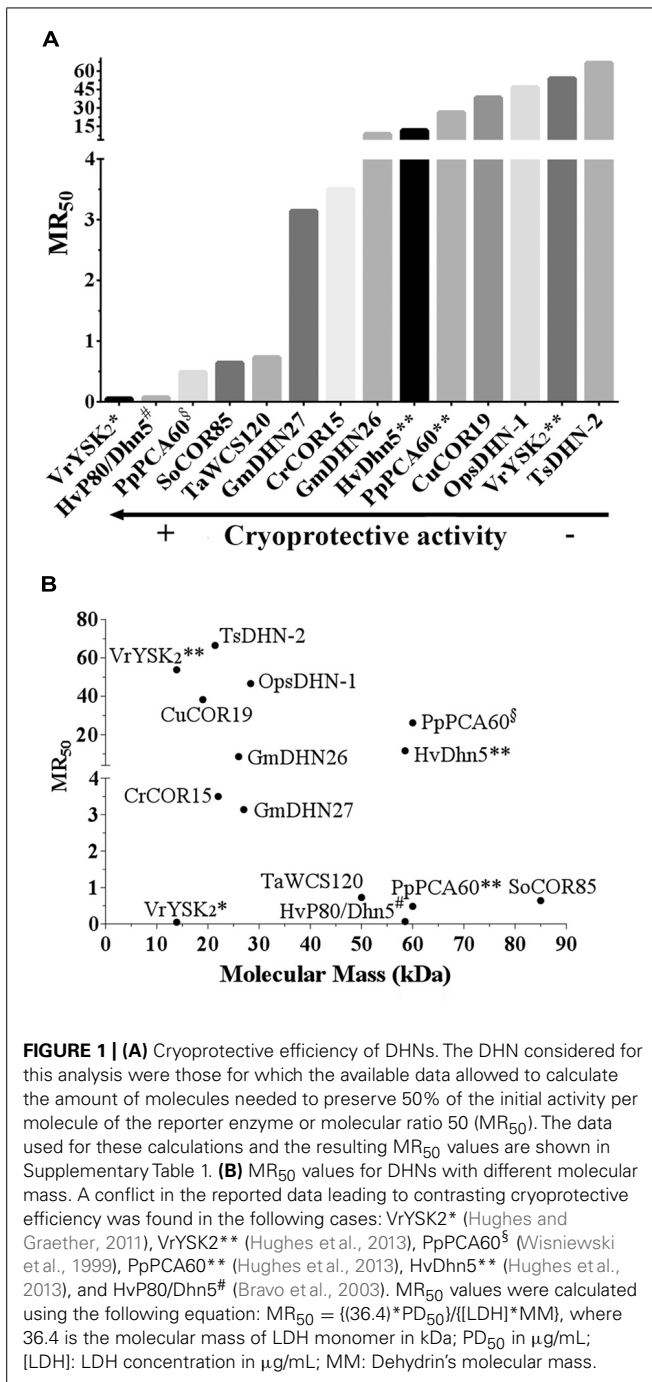
DISSECTING THE MOLECULAR MECHANISM OF DHNs PROTEIN PROTECTION

Several years ago, it was reported that addition of some LEA proteins, including DHNs, prevented the inactivation of reporter enzymes upon various freezing and thawing cycles (Lin and Thomashow, 1992; Kazuoka and Oeda, 1994). This *in vitro* assay is now a common approach to evaluate the protective activity of LEA proteins under these stress condition (Wisniewski et al., 1999; Hara et al., 2001; Bravo et al., 2003; Momma et al., 2003; Sanchez-Ballesta et al., 2004; Reyes et al., 2008; Hughes and Graether, 2011; Hughes et al., 2013). Kazuoka and Oeda (1994) reported that SoCOR85, an 85 KDa DHN from spinach, was able to keep LDH activity after freezing the sample for 24 h at -20°C and subsequent thawing at 4°C for 2 h. In that work, SoCOR85 showed a PD_{50} ("Protectant protein" concentration needed to preserve 50% of the reporter enzyme activity) smaller than that for a variety of proteins with no relation to DHNs, including BSA, a known cryoprotectant. These data indicated that SoCOR85 was an effective cryoprotectant, and suggested that this cryoprotective activity was specific (Kazuoka and Oeda, 1994). Subsequently, a similar activity was found for DHNs from different plant species such as TaWCS120 from wheat (Houde et al., 1995), PpPCA60 from peach (Wisniewski et al., 1999), CuCOR19 from *Citrus unshiu* (Hara et al., 2001), HvP-80/Dhn5 from barley (Bravo et al., 2003), GmDHN26 and GmDHN27 from soybean (Momma et al., 2003), CrCOR15 from fortune mandarin fruit (Sanchez-Ballesta et al., 2004), ERD10 from *Arabidopsis* (Reyes et al., 2008), RcDhn5 from *Rhododendron catawbiense* Michaux (Reyes et al., 2008), ERD14 from *Arabidopsis* (Tantos et al., 2009), TaDHN-5 from wheat (Drira et al., 2013), VrYSK₂ from *Vitis riparia* (Hughes and Graether, 2011; Hughes et al., 2013), TsDHN-2 from *Thellungiella salsuginea* (Hughes et al., 2013), and OpsDHN-1 from *Opuntia streptacantha* (Hughes et al., 2013).

Although there are some differences in the *in vitro* assays cited above (Supplemental Table 1), the data from this experimental system seem to be a good platform to compare the effectiveness of this activity among the different DHNs tested, and to address questions regarding the mechanism of action involved in such protecting effect. In order to be able to compare the data from these different reports, we have unified this information by considering not only the amount but also the molecular mass of the proteins assayed; hence, instead of comparing PD_{50} we compared the molar ratio of protectant:enzyme, which ponders the amount of molecules needed to preserve 50% of the initial activity per monomer of the reporter enzyme (Molar Ratio₅₀, MR_{50}). LDH was the common reporter enzyme in all cases analyzed here. The original and standardized data are shown in Supplemental Table 1. The comparison between the MR_{50} of the different DHNs showed a broad range of cryoprotection levels, from 0.05 to 66.5. Hv80/Dhn5 showed to be the most effective cryoprotectant with a MR_{50} of 0.07 indicating that only one molecule (or less) of DHN is required to protect one molecule of LDH monomer during freeze/thaw treatments (Figure 1A). Even considering TsDHN2, the less effective protectant with a MR_{50} of 66.5, DHNs cryoprotection effectiveness seems remarkable when it is compared with that attained by compatible osmolytes such as sucrose, for which 10^6 molecules are needed to protect one monomer of LDH (Houde et al., 1995).

This analysis allows envisaging at least two different molecular mechanisms of action to explain DHNs protective activity, both already proposed in different reports (Reyes et al., 2005; Tunnacliffe and Wise, 2007; Tompa and Kovacs, 2010; Olvera-Carrillo et al., 2011; Chakrabortee et al., 2012; Hughes et al., 2013). One, supported by the low DHN:enzyme molar ratios needed to get protection, in which protein-protein interaction is strongly suggested, and another where a higher amount of DHN molecules seem to be necessary to exert this protective effect (Figures 1A and 2). The first hypothesis is also supported by additional published data that was not considered in this study because the available data did not allow estimating MR_{50} values. This is the case for ERD10, an *Arabidopsis* DHN, which was able to keep 75% of LDH activity after five freeze/thaw cycles in a 1:1 (ERD10:LDH) molar ratio, suggesting a MR_{50} lower than 1 (Reyes et al., 2008). Similarly, ERD14 from *Arabidopsis* preserved 100% of the ADH activity when subjected to five freeze/thaw cycles in a 0.5:1 (ERD14:LDH) molar ratio (Tantos et al., 2009).

Those cases with a high MR_{50} , where many DHN molecules are required for cryoprotection, suggest a mechanism where DHN molecules would not necessarily need to contact the cryo-susceptible target protein but rather stay localized near the enzyme and, because their hydrophilic and highly disordered characteristics, they may provide an appropriate environment to stabilize a native and functional structure, a mechanism that has been referred as molecular shield (Tunnacliffe and Wise, 2007; Chakrabortee et al., 2012; Hughes et al., 2013). For this last mechanism, it has been proposed that the extended random coil structure reported for some of the characterized DHNs would result in proteins with a large hydrodynamic radius, which predicts that DHNs with this conformation would be able to align their hydrophilic amino acids around the surface of a target protein to circumvent



the loss of its bulk water and consequent changes in its native structure (Tunnacliffe and Wise, 2007; Hughes et al., 2013). This hypothesis is supported by a recent work, where different concatemers of the K₂ hypothetical protein from *V. riparia* were used (Hughes et al., 2013). In this report it was found that the level of protection conferred by these K-segment concatemers was directly proportional to their hydrodynamic radius. A similar correlation was observed by comparing the protection level of polyethylene glycol (PEG) polymers with different hydrodynamic radius. By contrast, no correlation was detected when globular proteins were

tested (Hughes et al., 2013). However, if we consider the MR₅₀ for these proteins (Supplemental Table 1), they showed values between 39.3 and 668 (up to 668 molecules of DHN are needed to protect one molecule of LDH). From these data, it is evident that

the cryoprotective efficiency of these concatemers is far low from that showed by other DHNs, where protection was detected even with one protein molecule (Figure 1A; Supplemental Table 1), suggesting that cryoprotection through molecular shield is less efficient than that obtained by protein–protein interaction. The need of a larger number of molecules that could be protecting through a molecular shield mechanism could also be interpreted as an unspecific effect produced just by the presence of many molecules close to a sensitive target, which also could occur by some other proteins apparently not devoted to such specific function. If only the hydrodynamic radius and the hydrophilicity of a protecting protein are relevant for this effect, then any polypeptide with these properties despite their amino acid sequence would be a cryoprotectant, which raises a question remaining to be addressed. On this regard, it has been reported that a large and highly hydrophilic amino acid-based polymer, poly L-Lysine was unable to protect LDH activity from freeze/thaw treatments (Reyes et al., 2008), implying the contribution of additional characteristics in a protein to be a good protectant. The relevance of the protein sequence also has to be considered for the data obtained using the K₂ concatemers because the proteins obtained not only are longer with larger hydrodynamic radius but they also contain an increasing number of the conserved K-segments, whose specific sequence could influence on this effect. The participation of K-segments in DHNs cryoprotection was suggested because their progressive deletion from *Arabidopsis thaliana* and *Rhododendron catawbiense* DHNs (ERD10 and RcDhn5, respectively) impaired their ability to protect LDH activity (Reyes et al., 2008). Similar findings were obtained for wheat DHN-5 (Drira et al., 2013).

The MR₅₀ from many of the DHN cryoprotection experiments reported also allowed us to look for a relation between the size of natural DHNs and their molecular protection effectiveness; however, no correlation was found between these two parameters (Figure 1B). There are low molecular mass DHNs with an MR₅₀ close to 70 and others whose MR₅₀ was lower than 10. Although some correlation could be detected for the K-segments analyzed by Hughes et al. (2013; see Supplementary Table 1), the data for any set of the natural DHNs, with low or high molecular mass (Figure 1B), are not consistent with a general mechanism in which the length of DHNs plays an important role in cryoprotection. Furthermore, we neither found any correlation by considering the levels of structural disorder determined using PONDR tools (Romero et al., 1997). HvP80/Dhn5 with the highest cryoprotection efficiency shows the same level of disorder (52.7%) than that obtained for TsDHN2 (52.43%) with the lowest cryoprotection efficiency, indicating that the protective effect is rather related to specific properties in each of the DHNs tested.

Results where direct binding between a reporter enzyme and a DHN was not detected favored the idea that the physicochemical properties of LEA proteins (including DHNs), such as the abundance of charged amino acid residues, promote electrostatic interactions to keep the two proteins closely enough to provide protection without binding (Hughes and Graether, 2011). However, data indicating that one or few LEA protein molecules are enough to protect reporter enzymes from the effects of water scarcity (Kazuoka and Oeda, 1994; Houde et al., 1995; Bravo

et al., 2003; Reyes et al., 2005; Nakayama et al., 2007) sustains the possibility of direct protein interaction, which is further supported by the analysis made in the present work. Moreover, evidence for a physical interaction between LEA proteins and target proteins has been obtained for some hydrophilins and LEA proteins (Alsheikh et al., 2005; Reyes et al., 2005; Nakayama et al., 2007; Kushwaha et al., 2012, 2013; Xie et al., 2012). This has also been the case for Y₂K₄-type DHN from *Medicago truncatula*, for which it was found *in vitro* and *in vivo* interaction with an ICE-type (inducer of CBF expression 1) transcription factor (Xie et al., 2012). The low occurrence of hydrophobic patches in the many DHNs and other LEA proteins predicts low affinity association with other proteins; however, attention should be given to their amino acid sequence as well as to possible structural modifications that could be promoted by changes in their environment (Mouillon et al., 2008; Olvera-Carrillo et al., 2011) and/or by their interaction with their clients, as it has been proposed for various IDPs (Dyson and Wright, 2005; Tompa and Fuxreiter, 2008; Pancsa and Tompa, 2012). Based on the physicochemical properties of IDPs and some experimental evidence (Pufall et al., 2005; Galea et al., 2008; Wang et al., 2011), it has also been hypothesized that the existence of a variety of structural states for a particular IDP or intrinsically disordered regions (IDRs) could lead to the formation of dynamic protein complexes, where this macromolecular ensembles may fluctuate between diverse structural organizations (Tompa and Fuxreiter, 2008). Hence, protein–protein interactions between LEA protein and their clients may be transient, making more challenging the finding of experimental conditions where these interactions could be stabilized.

FUTURE DIRECTIONS

Overall, considering the differences in the reported evidence and the analysis in this study, it cannot be discarded that different DHNs perform their protective function by different mechanisms or a combination of them (see Figure 2) depending on their particular sequences or even on the severity of the stress and/or cell type where they carry out their function. The possibility that DHNs perform multiple functions, a feature that seems to be common for IDPs (Jeffery, 2003; Sun et al., 2013), is something to be considered for their molecular mechanisms of action. Evidence for multi-functionality has been obtained *in vitro* for some DHNs, showing their ability to bind divalent cations, nucleic acids or some phospholipids, in addition to their cryoprotective effect (Koag et al., 2003, 2009; Hanin et al., 2011). It should be kept in mind the possible role of post-translational modifications in the modulation of different functions and interactions, considering a further structural and functional characterization of DHNs obtained from plant tissues. Their potential role in plant cells as protectant proteins or chaperones during water deficit needs additional evidence, identifying their natural targets or clients as well as regions or sequences in their structure relevant for their function.

The structural disordered nature of DHNs, their distinctive properties and sequences impose *ad hoc* experimental designs, some of them challenging but needed to get closer to the understanding of their function in the plant responses to abiotic and

biotic stress. It is imperative the analysis of a larger number of DHNs as well as the standardization of *in vitro* protection assays to be able to address various aspects of the proposed mechanisms. The emergence of new and high-resolution technologies represents a good opportunity to validate the proposed molecular mechanisms and to address their relevance in the plant cell.

ACKNOWLEDGMENTS

We are grateful to María Teresa Covarrubias for artwork design. This work was partially supported by Consejo Nacional de Ciencia y Tecnología-Mexico (CONACyT-132258) and Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT-DGAPA-UNAM; IN-208212). CLC-V and DFR-L were supported by CONACyT with a Ph.D. and MSci fellowships, respectively.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpls.2014.00583/abstract>

REFERENCES

- Alsheikh, M. K., Svensson, J. T., and Randall, S. K. (2005). Phosphorylation regulated ion-binding is a property shared by the acidic subclass dehydrins. *Plant Cell Environ.* 28, 1114–1122. doi: 10.1111/J.1365-3040.2005.01348.X
- Battaglia, M., Olvera-Carrillo, Y., Garcíarrubio, A., Campos, F., and Covarrubias, A. A. (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* 148, 6–24. doi: 10.1104/PP.108.120725
- Bokor, M., Csizmok, V., Kovacs, D., Banki, P., Friedrich, P., Tompa, P., et al. (2005). NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophys. J.* 88, 2030–2037. doi: 10.1529/Biophysj.104.051912
- Bravo, L. A., Gallardo, J., Navarrete, A., Olave, N., Martínez, J., Alberdi, M., et al. (2003). Cryoprotective activity of a cold-induced dehydrin purified from barley. *Physiol. Plant.* 118, 262–269. doi: 10.1034/J.1399-3054.2003.00060.X
- Brini, F., Saibi, W., Amara, I., Gargouri, A., Masmoudi, K., and Hanin, M. (2010). Wheat dehydrin DHN-5 exerts a heat-protective effect on beta-glucosidase and glucose oxidase activities. *Biosci. Biotechnol. Biochem.* 74, 1050–1054. doi: 10.1271/Bbb.90949
- Chakrabortee, S., Tripathi, R., Watson, M., Schierle, G. S. K., Kurniawan, D. P., Kaminski, C. F., et al. (2012). Intrinsically disordered proteins as molecular shields. *Mol. Biosyst.* 8, 210–219. doi: 10.1039/C1mb05263b
- Close, T. J. (1996). Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiol. Plant.* 97, 795–803. doi: 10.1111/J.1399-3054.1996.Tb00546.X
- Drira, M., Saibi, W., Brini, F., Gargouri, A., Masmoudi, K., and Hanin, M. (2013). The K-segments of the wheat dehydrin DHN-5 are essential for the protection of lactate dehydrogenase and β -glucosidase activities in vitro. *Mol. Biotechnol.* 54, 643–650. doi: 10.1007/S12033-012-9606-8
- Dure, L. (1993). “Structural motifs in lea proteins,” in *Plant Responses to Cellular Dehydration during Environmental Stress*, Vol. 10, eds T. J. Close and E. A. Bray (Rockville, MD: American Society of Plant Physiologists), 91–103.
- Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208. doi: 10.1038/Nrm1589
- Eriksson, S. K., Kutzer, M., Procek, J., Grobner, G., and Harryson, P. (2011). Tunable membrane binding of the intrinsically disordered dehydrin Lti30, a cold-induced plant stress protein. *Plant Cell* 23, 2391–2404. doi: 10.1105/TPC.111.085183
- Galea, C. A., Nourse, A., Wang, Y., Sivakolundu, S. G., Heller, W. T., and Kriwacki, R. W. (2008). Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27^{Kip1}. *J. Mol. Biol.* 376, 827–838. doi: 10.1016/J.jmb.2007.12.016
- Garay-Arroyo, A., Colmenero-Flores, J. M., Garcíarrubio, A., and Covarrubias, A. A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.* 275, 5668–5674. doi: 10.1074/jbc.275.8.5668
- Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., and Masmoudi, K. (2011). Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signal. Behav.* 6, 1503–1509. doi: 10.4161/psb.6.10.17088
- Hara, M. (2010). The multifunctionality of dehydrins: an overview. *Plant Signal. Behav.* 5, 503–508. doi: 10.4161/psb.11085
- Hara, M., Fujinaga, M., and Kuboi, T. (2005). Metal binding by citrus dehydrin with histidine-rich domains. *J. Exp. Bot.* 56, 2695–2703. doi: 10.1093/jxb/eri262
- Hara, M., Shinoda, Y., Tanaka, Y., and Kuboi, T. (2009). DNA binding of citrus dehydrin promoted by zinc ion. *Plant Cell Environ.* 32, 532–541. doi: 10.1111/J.1365-3040.2009.01947.X
- Hara, M., Terashima, S., and Kuboi, T. (2001). Characterization and cryoprotective activity of cold-responsive dehydrin from *Citrus unshiu*. *J. Plant Physiol.* 158, 1333–1339. doi: 10.1078/0176-1617-00600
- Hincha, D. K., and Thalhammer, A. (2012). LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem. Soc. Trans.* 40, 1000–1003. doi: 10.1042/Bst20120109
- Houde, M., Daniel, C., Lachapelle, M., Allard, F., Laliberte, S., and Sarhan, F. (1995). Immunolocalization of freezing-tolerance-associated proteins in the cytoplasm and nucleoplasm of wheat crown tissues. *Plant J.* 8, 583–593. doi: 10.1046/J.1365-313x.1995.8040583.X
- Hughes, S., and Graether, S. P. (2011). Cryoprotective mechanism of a small intrinsically disordered dehydrin protein. *Protein Sci.* 20, 42–50. doi: 10.1002/pro.534
- Hughes, S. L., Schart, V., Malcolmson, J., Hogarth, K. A., Martynowicz, D. M., Tralman-Baker, E., et al. (2013). The importance of size and disorder in the cryoprotective effects of dehydrins. *Plant Physiol.* 163, 1376–1386. doi: 10.1104/pp.113.226803
- Hundertmark, M., and Hincha, D. K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9:118. doi: 10.1186/1471-2164-9-118
- Ismail, A. M., Hall, A. E., and Close, T. J. (1999). Purification and partial characterization of a dehydrin involved in chilling tolerance during seedling emergence of cowpea. *Plant Physiol.* 120, 237–244. doi: 10.1104/pp.120.1.237
- Jeffery, C. J. (2003). Moonlighting proteins: old proteins learning new tricks. *Trends Genet.* 19, 415–417. doi: 10.1016/S0168-9525(03)00167-7
- Kazuoka, T., and Oeda, K. (1994). Purification and characterization of Cor85-oligomeric complex from cold-acclimated spinach. *Plant Cell Physiol.* 35, 601–611.
- Kim, S. Y., and Nam, K. H. (2010). Physiological roles of *ERD10* in abiotic stresses and seed germination of *Arabidopsis*. *Plant Cell Rep.* 29, 203–209. doi: 10.1007/S00299-009-0813-0
- Koag, M. C., Fenton, R. D., Wilkens, S., and Close, T. J. (2003). The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. *Plant Physiol.* 131, 309–316. doi: 10.1104/PP.011171
- Koag, M. C., Wilkens, S., Fenton, R. D., Resnik, J., Vo, E., and Close, T. J. (2009). The K-segment of maize DHN1 mediates binding to anionic phospholipid vesicles and concomitant structural changes. *Plant Physiol.* 150, 1503–1514. doi: 10.1104/pp.109.136697
- Kruger, C., Berkowitz, O., Stephan, U. W., and Hell, R. (2002). A metal-binding member of the late embryogenesis abundant protein family transports iron in the phloem of *Ricinus communis* L. *J. Biol. Chem.* 277, 25062–25069. doi: 10.1074/jbc.M201896200
- Kushwaha, R., Downie, A. B., and Payne, C. M. (2013). Uses of phage display in agriculture: sequence analysis and comparative modeling of late embryogenesis abundant client proteins suggest protein-nucleic acid binding functionality. *Comput. Math. Methods Med.* 2013, 470390. doi: 10.1155/2013/470390
- Kushwaha, R., Lloyd, T. D., Schafermeyer, K. R., Kumar, S., and Downie, A. B. (2012). Identification of late embryogenesis abundant (LEA) protein putative interactors using phage display. *Int. J. Mol. Sci.* 13, 6582–6603. doi: 10.3390/Ijms13066582
- Lin, C. T., and Thomashow, M. F. (1992). A cold-regulated *Arabidopsis* gene encodes a polypeptide having potent cryoprotective activity. *Biochem. Biophys. Res. Commun.* 183, 1103–1108. doi: 10.1016/S0006-291x(05)80304-3
- Manfre, A. J., Lanni, L. M., and Marcotte, W. R. (2006). The *Arabidopsis* group 1 LATE EMBRYOGENESIS ABUNDANT protein AtEM6 is required for normal seed development. *Plant Physiol.* 140, 140–149. doi: 10.1104/PP.105.072967
- Momma, M., Kaneko, S., Haraguchi, K., and Matsukura, U. (2003). Peptide mapping and assessment of cryoprotective activity of 26/27-kDa dehydrin from soybean seeds. *Biosci. Biotechnol. Biochem.* 67, 1832–1835. doi: 10.1271/bbb.67.1832

- Mouillon, J. M., Eriksson, S. K., and Harryson, P. (2008). Mimicking the plant cell interior under water stress by macromolecular crowding: disordered dehydrin proteins are highly resistant to structural collapse. *Plant Physiol.* 148, 1925–1937. doi: 10.1104/pp.108.124099
- Nakayama, K., Okawa, K., Kakizaki, T., Honma, T., Itoh, H., and Inaba, T. (2007). Arabidopsis Cor15am is a chloroplast stromal protein that has cryoprotective activity and forms oligomers. *Plant Physiol.* 144, 513–523. doi: 10.1104/pp.106.094581
- Olvera-Carrillo, Y., Campos, F., Reyes, J. L., Garcarrubio, A., and Covarrubias, A. A. (2010). Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in Arabidopsis. *Plant Physiol.* 154, 373–390. doi: 10.1104/pp.110.158964
- Olvera-Carrillo, Y., Reyes, J. L., and Covarrubias, A. A. (2011). Late embryogenesis abundant proteins: versatile players in the plant adaptation to water limiting environments. *Plant Signal. Behav.* 6, 586–589. doi: 10.4161/psb.6.4.15042
- Panca, R., and Tompa, P. (2012). Structural disorder in eukaryotes. *PLoS ONE* 7:e34687. doi: 10.1371/journal.pone.0034687
- Pufall, M. A., Lee, G. M., Nelson, M. L., Kang, H. S., Velyvis, A., Kay, L. E., et al. (2005). Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* 309, 142–145. doi: 10.1126/Science.1111915
- Rahman, L. N., Smith, G. S., Bamm, V. V., Voyer-Grant, J. A., Moffatt, B. A., Dutcher, J. R., et al. (2011). Phosphorylation of *Thellungiella salsuginea* dehydrins TsDHN-1 and TsDHN-2 facilitates cation-induced conformational changes and actin assembly. *Biochemistry* 50, 9587–9604. doi: 10.1021/bi201205m
- Reyes, J. L., Campos, F., Wei, H., Arora, R., Yang, Y. I., Karlson, D. T., et al. (2008). Functional dissection of hydrophilins during *in vitro* freeze protection. *Plant Cell Environ.* 31, 1781–1790. doi: 10.1111/j.1365-3040.2008.01879.X
- Reyes, J. L., Rodrigo, M. J., Colmenero-Flores, J. M., Gil, J. V., Garay-Arroyo, A., Campos, F., et al. (2005). Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects *in vitro*. *Plant Cell Environ.* 28, 709–718. doi: 10.1111/j.1365-3040.2005.01317.X
- Romero, P., Obradovic, Z., and Dunker, K. (1997). Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Inform. Ser. Workshop Genome Inform.* 8, 110–124.
- Rorat, T. (2006). Plant dehydrins — tissue location, structure and function. *Cell. Mol. Biol. Lett.* 11, 536–556. doi: 10.2478/s11658-006-0044-0
- Ruibal, C., Salamo, I. P., Carballo, V., Castro, A., Bentancor, M., Borsani, O., et al. (2012). Differential contribution of individual dehydrin genes from *Physcomitrella patens* to salt and osmotic stress tolerance. *Plant Sci.* 190, 89–102. doi: 10.1016/j.plantsci.2012.03.009
- Saavedra, L., Svensson, J., Carballo, V., Izmendi, D., Welin, B., and Vidal, S. (2006). A dehydrin gene in *Physcomitrella patens* is required for salt and osmotic stress tolerance. *Plant J.* 45, 237–249. doi: 10.1111/j.1365-313x.2005.02603.X
- Sanchez-Ballesta, M. T., Rodrigo, M. J., LaFuente, M. T., Granell, A., and Zacarias, L. (2004). Dehydrin from *Citrus*, which confers *in vitro* dehydration and freezing protection activity, is constitutive and highly expressed in the flavedo of fruit but responsive to cold and water stress in leaves. *J. Agric. Food Chem.* 52, 1950–1957. doi: 10.1021/jf035216
- Soulaiges, J. L., Kim, K., Arrese, E. L., Walters, C., and Cushman, J. C. (2003). Conformation of a group 2 late embryogenesis abundant protein from soybean. Evidence of poly(L-proline)-type II structure. *Plant Physiol.* 131, 963–975. doi: 10.1104/PP.015891
- Sun, X. L., Rikkerink, E. H. A., Jones, W. T., and Uversky, V. N. (2013). Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* 25, 38–55. doi: 10.1105/tpc.112.106062
- Tantos, A., Friedrich, P., and Tompa, P. (2009). Cold stability of intrinsically disordered proteins. *FEBS Lett.* 583, 465–469. doi: 10.1016/j.febslet.2008.12.054
- Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* 33, 2–8. doi: 10.1016/j.tibs.2007.10.003
- Tompa, P., and Kovacs, D. (2010). Intrinsically disordered chaperones in plants and animals. *Biochem. Cell Biol.* 88, 167–174. doi: 10.1139/O09-163
- Tunnacliffe, A., and Wise, M. J. (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94, 791–812. doi: 10.1007/S00114-007-0254-Y
- Wang, Y. F., Fisher, J. C., Mathew, R., Ou, L., Otieno, S., Sublet, J., et al. (2011). Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21. *Nat. Chem. Biol.* 7, 214–221. doi: 10.1038/Nchembio.536
- Wisniewski, M., Webb, R., Balsamo, R., Close, T. J., Yu, X. M., and Griffith, M. (1999). Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: a dehydrin from peach (*Prunus persica*). *Physiol. Plant.* 105, 600–608. doi: 10.1034/j.1399-3054.1999.105402.X
- Xie, C., Zhang, R. X., Qu, Y. T., Miao, Z. Y., Zhang, Y. Q., Shen, X. Y., et al. (2012). Overexpression of *MtCAS31* enhances drought tolerance in transgenic Arabidopsis by reducing stomatal density. *New Phytol.* 195, 124–135. doi: 10.1111/j.1469-8137.2012.04136.X

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 July 2014; accepted: 09 October 2014; published online: 29 October 2014.

Citation: Cuevas-Velazquez CL, Rendón-Luna DF and Covarrubias AA (2014) Dissecting the cryoprotection mechanisms for dehydrins. *Front. Plant Sci.* 5:583. doi: 10.3389/fpls.2014.00583

This article was submitted to Plant Physiology, a section of the journal *Frontiers in Plant Science*.

Copyright © 2014 Cuevas-Velazquez, Rendón-Luna and Covarrubias. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Group 1 LEA proteins, an ancestral plant protein group, are also present in other eukaryotes, and in the archaea and bacteria domains

F. Campos · C. Cuevas-Velazquez · M. A. Fares ·
J. L. Reyes · A. A. Covarrubias

Received: 7 January 2013 / Accepted: 26 June 2013 / Published online: 17 July 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Water is an essential element for living organisms, such that various responses have evolved to withstand water deficit in all living species. The study of these responses in plants has had particular relevance given the negative impact of water scarcity on agriculture. Among the molecules highly associated with plant responses to water limitation are the so-called late embryogenesis abundant (LEA) proteins. These proteins are ubiquitous in the plant kingdom and accumulate during the late phase of embryogenesis and in vegetative tissues in response to water deficit. To know about the evolution of these proteins, we have studied the distribution of group 1

LEA proteins, a set that has also been found beyond the plant kingdom, in *Bacillus subtilis* and *Artemia franciscana*. Here, we report the presence of group 1 LEA proteins in green algae (*Chlorophyta* and *Streptophyta*), suggesting that these group of proteins emerged before plant land colonization. By sequence analysis of public genomic databases, we also show that 34 prokaryote genomes encode group 1 LEA-like proteins; two of them belong to *Archaea* domain and 32 to bacterial phyla. Most of these microbes live in soil-associated habitats suggesting horizontal transfer from plants to bacteria; however, our phylogenetic analysis points to convergent evolution. Furthermore, we present data showing that bacterial group 1 LEA proteins are able to prevent enzyme inactivation upon freeze–thaw treatments in vitro, suggesting that they have analogous functions to plant LEA proteins. Overall, data in this work indicate that LEA1 proteins' properties might be relevant to cope with water deficit in different organisms.

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-013-0768-2) contains supplementary material, which is available to authorized users.

F. Campos (✉) · C. Cuevas-Velazquez ·

J. L. Reyes · A. A. Covarrubias

Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001 Col. Chamilpa, 62210 Cuernavaca, MOR, Mexico

e-mail: campos@ibt.unam.mx

C. Cuevas-Velazquez

e-mail: ccuevas@ibt.unam.mx

J. L. Reyes

e-mail: jlreyes@ibt.unam.mx

A. A. Covarrubias

e-mail: crobles@ibt.unam.mx

M. A. Fares

Instituto de Biología Molecular y Celular de Plantas, Universidad Politécnica de Valencia, Ingeniero Fausto Elio s/n, 46022 Valencia, Spain

e-mail: mfares@ibmcp.upv.es

Keywords LEA proteins · Em proteins · *GsiB* · *YciG* · Hydrophilins · Enzyme protection

Introduction

Water is an essential element for life. All cell functions depend on different levels of available water, so that changes in the optimal cell water status are cues leading to series of responses that allows the organisms to adjust to the new conditions. Plants have acquired along evolution a number of mechanisms to contend against low water availability, a condition that can disadvantageously affect their growth and development. As sessile organisms, plants possess a diverse set of responses to this stress condition

acting at different functional levels, during different developmental stages and upon a variety of stress severity. Remarkably, higher plants have developed specialized structures such as seeds and pollen that are able to retain their function upon extensive water removal. Several components involved in this trait have been identified, among them, non-reducing oligosaccharides such as sucrose, compatible solutes, small heat shock proteins (HSPs) and an heterogeneous group of proteins that accumulate in high levels during the late phase of embryogenesis, when seeds lose most of their water content, known as late embryogenesis abundant (LEA) proteins (Hoekstra et al. 2001). In addition, most LEA proteins accumulate in vegetative tissues in response to water deficit, cold, saline stress, or abscisic acid application (Ingram and Bartels 1996; Battaglia et al. 2008). Seven groups of LEA proteins have been described based on the similarities in their amino acid sequence and conserved motifs (Battaglia et al. 2008). With the exception of group 5, LEA proteins are highly hydrophilic, contain a high percentage of glycine and/or other small amino acid residues (alanine, serine and threonine), and lack or contain few tryptophan and cysteine residues. It is predicted from their amino acid sequences that they are intrinsically unstructured proteins and experimentally confirmed for a few of them in aqueous solution (Hundertmark et al. 2012). These general characteristics led to the identification of a wide group of proteins present in different taxonomic kingdoms, termed “hydrophilins”, in which typical LEA proteins are included (Garay-Arroyo et al. 2000).

LEA proteins of some groups have been detected in angiosperms, gymnosperms, non-vascular plants and algae, suggesting that they are ubiquitous in the *Viridiplantae* kingdom (Battaglia et al. 2008; Shih et al. 2008). Furthermore, some LEA-like proteins have been detected in other organisms different from plants (Hand et al. 2010), including bacteria (Stacy and Aalen 1998), rotifers (Tunnacliffe et al. 2005), crustaceans (Hand et al. 2007), insects (Kikawada et al. 2006), nematodes (Browne et al. 2002) and humans (Brandon et al. 2011).

Although LEA proteins were identified more than 30 years ago (Dure and Galau 1981), their precise function is still unknown. It has been reported that the ectopic over-expression of some LEA genes can improve stress tolerance of transgenic plants (Xu et al. 1996; Olvera-Carrillo et al. 2010). Furthermore, over-expression of LEA proteins in microorganisms, such as baker's yeast *Saccharomyces cerevisiae* or *Escherichia coli*, also lead to osmotic and salt tolerance (Imai et al. 1996; Liu and Zheng 2005). Regarding their function, some progress has been obtained in recent years using in vitro assays with purified LEA proteins from different groups. These studies have shown that LEA proteins might have a protective role, preventing

the inactivation of other proteins upon partial dehydration. This is the case of groups 2 and 4 LEA proteins from *Arabidopsis thaliana* that prevent the enzymatic inactivation of malate dehydrogenase and lactate dehydrogenase upon water limitation and freezing treatments (Reyes et al. 2005, 2008; Campos et al. 2011). Goyal et al. (2005) have found that group 1 LEA protein from wheat and a group 3 LEA-like protein from the nematode *Aphelenchus avenae* can prevent the aggregation of dehydration-labile proteins. Other activities reported for LEA proteins include radical scavenging (Hara et al. 2004), glass sugar stabilization (Wolkers et al. 2001), bacterial growth inhibition (Campos et al. 2006) and hydration buffer (Tompa et al. 2006). The apparent functional diversity found for the different LEA protein groups and their physicochemical characteristics have led to the hypothesis that these proteins may have a multi-functional role upon environmental changes during plant development (Olvera-Carrillo et al. 2011).

The amino acid composition of group 1 LEA (LEA1) proteins (also called Em proteins) is characterized by the relative high abundance of glutamic and glycine residues (11–22 %). As other hydrophilins, the lack of tryptophan and cysteine, and high content of glycine and charged residues, predicts structural disorder in aqueous solutions, which has been experimentally confirmed for some proteins of this group (Soulages et al. 2002). Group LEA1 proteins are highly conserved (Battaglia et al. 2008). Characteristics of this group are the presence of an internal 20-amino acid signature motif [(R/G)S(R/K)GGQTRKEQLGXEGT-XEM], which may be repeated up to four times, and two additional conserved regions, at the N- and C-termini, respectively (Espelund et al. 1992; Gaubier et al. 1993; Battaglia et al. 2008). All *LEA1* genes contain a single intron, separating the first exon, which encodes for a plant-specific N-terminus, from the second exon encoding the 20-mer motif and C-terminal region (Vicent et al. 2000).

Plant *LEA1* genes have been found in monocots, eudicots, gymnosperms and bryophytes (*Physcomitrella*) (Battaglia et al. 2008; Shih et al. 2008), organized as multigene families (Espelund et al. 1992; Vicent et al. 2000; Shih et al. 2010). The *A. thaliana* genome contains two *LEA1* genes, *AtEM1* and *AtEM6* (Gaubier et al. 1993). The *AtEM1* protein has four copies of the 20-amino acid motif, whereas *AtEM6* contains only one copy. *LEA1* genes are expressed in mature and developing embryos and young seedlings exposed to water deficit or ABA addition; however, unlike other *LEA* group genes, *LEA1* genes show low or none expression in mature vegetative tissues under ABA or stress conditions (Marcotte et al. 1989; Parcy et al. 1994; Bensmihen et al. 2004). The characterization by Manfre et al. (2006) of a T-DNA insertion mutant in the *AtEM6* gene (*atem6-1*) showed that the absence of this protein leads to the production of seeds, at the distal ends of the

siliques, that exhibit premature dehydration and maturation, a defect that does not alter their viability or germination (Manfre et al. 2006, 2009). A new variant of *LEA1* genes has been found in various grasses but not in dicotyledonous plants. These atypical *LEA1* genes encode for proteins with an N-terminal sequence that shows an incomplete version of the 20-mer motif, characteristic of this protein family (Fig. 1; Shih et al. 2010).

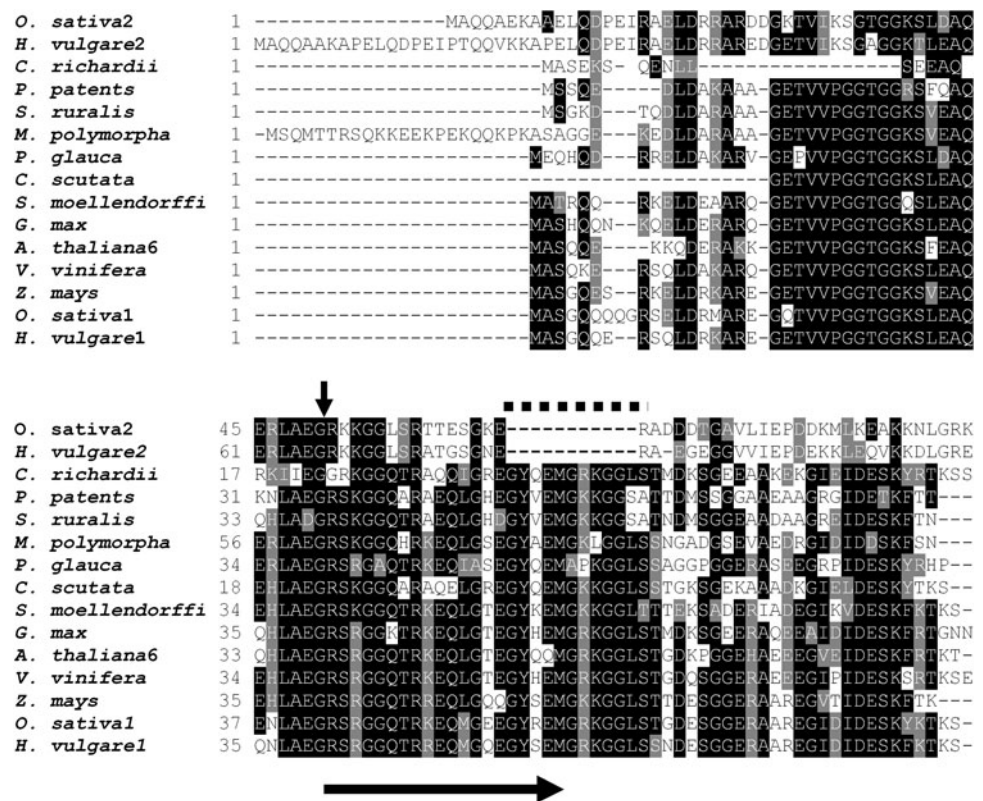
Analysis of the *cis*-elements in the promoters of those *LEA1* genes whose complete sequence is known showed that the ABA-responsive elements (ABRE or ABRE-like sequences) are the most conserved *cis*-elements (Shih et al. 2010). This observation is in agreement with their common expression at the onset of seed dehydration, when ABA shows high levels, and also consistent with their response to ABA treatments during vegetative growth (Delseny et al. 2001; Hattori et al. 1995; Bensmihen et al. 2004). The response to ABA via ABA-responsive elements in *LEA1* genes seems to be conserved throughout land plant evolution given the activation of this gene by exogenous ABA and the presence of ABREs in the moss *P. patens* *LEA1* gene promoter (Yotsui et al. 2013). No other *cis*-elements seem to be conserved in *LEA1* gene promoters.

Interestingly, *LEA1* proteins seem to be widely dispersed in life domains. A gene similar to group 1 *LEA*

genes, named *GsiB*, was described in *Bacillus subtilis* (Stacy and Aalen 1998). *GsiB* transcript and protein accumulate in stress situations, such as heat shock, salt stress, oxygen deprivation, glucose or phosphate starvation, and oxidative stress or during stationary phase (Maul et al. 1995). Furthermore, transcripts for similar proteins have also detected in cysts of the crustacean *Artemia franciscana* (Chen et al. 2009; Sharon et al. 2009). As for other *LEA* proteins, the function of *LEA1* proteins is unknown; however, structural and expression similarities among these proteins and their close relatives in bacteria (*GsiB*) and *A. franciscana* suggest that these proteins perform analogous functions.

To get insight into the evolution and function of *LEA1* proteins, in this work, we identified and analyzed algae, bacterial and *Archae* *LEA1* proteins, unearthing that these proteins are distributed throughout the three domains of life and that they have an ancient origin. A detailed analysis of the available sequences for *LEA1* proteins showed that 20-mer motifs are highly conserved among these proteins in the three different kingdoms. In addition, here we show that the *B. subtilis*, *Arabidopsis* and *Pseudomonas aeruginosa* *LEA1* protein relatives are capable of preventing enzyme inactivation upon freeze–thaw treatments in vitro, suggesting similar activities in very diverse organisms.

Fig. 1 Multiple-sequence alignment of group 1 *LEA* proteins from plant and algae species. Representative plant group 1 *LEA* sequences from eudicots, monocots and gymnosperms containing one 20-mer motif were selected. Black and grey shading represents identical and similar residues, respectively. The intron position in the genes is marked with a vertical arrowhead, and the 20-mer motif is indicated by a horizontal arrow. Discontinuous line indicates incomplete 20-mer motifs in rice and barley. Accession numbers for plant group 1 *LEA* proteins used in this alignment are described in Table 1



Materials and methods

Computational analysis

To identify plant and non-plant members of LEA1 group of proteins, several database searches were performed using the BLAST and TBLASTN algorithms. Searches in NCBI (<http://www.ncbi.nlm.nih.gov/>) and Phytozome (<http://www.phytozome.net/>) databases were conducted using *Arabidopsis* Em1 (NP_190749) and *Bacillus subtilis* GsiB (CAA40009) amino acid sequences as queries. A multiple alignment analysis was conducted using the MUSCLE and Boxshade online server (<http://www.ch.embnet.org>) with default parameters. Conserved motifs were analyzed using WebLogo (Crooks et al. 2004) programs (<http://weblogo.berkeley.edu/logo.cgi>).

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignments were built by MUSCLE using amino acid sequences. This alignment was then used to build an alignment of nucleotide sequences in which codons were concatenated according to the amino acid sequence alignment. RaxML Pthreads version 7.2.4 (Stamatakis 2006) was used to build a maximum likelihood (ML) phylogenetic tree after estimation of the appropriate amino acid substitution model by ProtTest version 2.4 (Abascal et al. 2005). The best ML tree was that based on LG substitution model (Le and Gascuel 2008) with GAMMA correction to account for rate heterogeneity among amino acid sites and empirical amino acid frequencies. Bootstrap analyses based on 100 pseudo-replicates were conducted on this phylogeny. The topology of the tree was drawn using the program FigTree v.1.1.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Cloning, expression and purification of selected group 1 LEA-like proteins

Open reading frames of *A. thaliana* *AtEM6*, *B. subtilis* *GsiB* and *P. aeruginosa* LEA1-like protein (referred to as PaEM protein) were obtained and cloned onto pTrc99A vector as described previously (Campos et al. 2006, 2011). The recombinant plasmids were transformed into *Escherichia coli* XL1 Blue (Stratagene). A selected colony was grown overnight in LB medium containing ampicillin at 37 °C. After a 1:100 dilution in fresh LB medium containing antibiotic, the culture was further grown at 37 °C to 1.0 OD₆₀₀ (approx. 2–3 h); at this point, protein expression was induced with 0.3 mM isopropyl β-D-thiogalactopyranoside (IPTG). At 2.0 OD₆₀₀ (approx. 2 h) the culture was harvested by centrifugation to proceed for recombinant protein purification as described by Campos et al. (2011).

The concentration of the purified proteins used in the in vitro freeze–thaw assays was determined considering their extinction coefficients.

Construction of pAtEM6del

The 20-mer amino acid motif (RSRGGQTRKEQLGTE-GYQQM) from *Arabidopsis* Em6 was deleted by the two-step recombinant-PCR method (Higuchi 1990), using as template the open reading frame of Em6 cloned into pTrc99A. To generate the deletion, two overlapping PCR fragments of AtEm6 were amplified, in the first step of PCR. In the second round of PCR, the deleted AtEm6 was generated using only the outer primers and the first PCR products as templates. Deleted open reading frame of AtEM6 (AtEM6del) was cloned into the pTrc99A vector as described elsewhere (Campos et al. 2006) and the identity of the insert was confirmed by DNA sequencing. The primer sequences were as follows: 99F, 5'-CTCCCGTTCTG GATAATG-3' (outer forward primer); 99R, 5'-ACT-TCTGAGTTCGGCATG-3 (outer reverse primer); EmR, 5'-GACCACCTTTGCGTCCCCCTTCAGCGAGATGC-3 (inner reverse primer), EmF, GCATCTCGCTGAAGGGG GACGCAAAGGTGGTC (inner forward primer). Recombinant plasmid (pAtEM6del) was transformed into *E. coli* XL1 Blue (Stratagene) and the pure recombinant AtEM6del protein was obtained as described (Campos et al. 2006, 2011).

In vitro freeze–thaw assays

To evaluate the protective ability of purified LEA1 proteins, we used a freeze–thaw in vitro assay previously developed by Reyes et al. (2008). LDH from rabbit muscle (Roche), an enzyme susceptible to freeze–thaw treatments, was used at 250 nM final concentration (monomer) in 25 mM Tris pH 7.5. AtEM6, GsiB and PaEM recombinant proteins were added at 1:1 ratio, at 250 nM final concentration; considering the LDH monomer molecular weight. Protein concentration was calculated using their molar extinction coefficient. The mixture in a final volume of 100 μL was frozen for 15 min in dry ice and thawed for 15 min in a water bath at 25 °C. This constituted one freeze–thaw cycle, which was repeated up to five times as indicated in the text. At the end of the treatment, samples were transferred to wet ice until enzyme activity was determined. LDH enzymatic activity was determined using aliquots of 15 μL in a final volume of 600 μL of the reaction assay buffer. The assay buffer was 25 mM Tris–HCl pH 7.5 containing 2 mM pyruvate (Sigma) and 0.15 mM NADH (Roche). LDH activities were monitored as the rate of decrease in absorbance at 340 nm for 1 min due to the conversion of NADH into NAD at 25 °C. The

rate determined for the untreated samples was considered as 100 % in all graphs. Enzyme activity for each sample shown was determined at least in three independent tests (with three internal repetitions), as indicated in the corresponding figure legend for each experiment.

Statistical analysis was carried out using one-way ANOVA test. Significant differences were calculated with Tukey's multiple comparisons post-test ($P < 0.05$).

Results

Group 1 LEA genes in algae genomes

LEA1 genes have been found in monocots, eudicots, gymnosperms, and in the *Physcomitrella* moss (Battaglia et al. 2008; Shih et al. 2008). To further investigate the extent of the phylogenetic distribution of LEA1 proteins,

we searched non-vascular and seedless plants, and algae sequence databases (NCBI and Phytozome) using as query the Arabidopsis LEA1 protein sequence to perform a BLAST analysis. Sequences similar to LEA1 proteins were identified in seedless vascular plants (*Selaginella moellendorffii* and *Ceratopteris richardii*), basal non-vascular plants (*Marchantia polymorpha*), streptophyte alga (*Coleochaete scutata*) and chlorophyta algae (*Chlamydomonas reinhardtii* and *Volvox carteri*) (Table 1). The alignment of the *Viridiplantae* LEA1 protein representative sequences including *Coleochaete scutata* alga showed a high similarity in the 20-mer motif and also in N-terminal and C-terminal regions, indicating a high conservation among these proteins (Fig. 1); however, no conservation was found in the N-terminus of the fern *Ceratopteris richardii* LEA1 protein sequences (Fig. 1). Although *Chlamydomonas* and *Volvox* LEA1 relatives show a high conservation in the 20-mer motif, this was

Table 1 Protein accessions of plant group 1 LEA proteins used in this work

Species	Accession number	Total length	Number of repeats
Eudicots			
<i>Arabidopsis thaliana</i>			
AtEm1	NP_190749	152	4
AtEM6	NP_181546	92	1
<i>Glycine max</i>	NP_001237186	101	1
<i>Vitis vinifera</i>	XP_002273618	95	1
Monocots			
<i>Oryza sativa</i>			
Osativa1	NP_001055273	95	1
Osativa2	NP_001042086	93	Incomplete
<i>Hordeum vulgare</i>			
Hvulgare1	BAK06033	108	Incomplete
Hvulgare2	Q05190	93	1
<i>Zea mays</i>	NP_001105429	91	1
Gymnosperms			
<i>Picea glauca</i>	AAB01568	91	1
Euphyllophyta			
<i>Ceratopteris richardii</i> ^a	CV735240	78	1
Lycophyta			
<i>Selaginella moellendorffii</i>	XP_002963571	106	1
Bryophyta			
<i>Physcomitrella patens</i>	XP_001752475	87	1
<i>Syntrichia ruralis</i> ^b	CN205670	89	1
Marchantiophyta			
<i>Marchantia polymorpha</i> ^c	BJ872620	112	1
Charophyta			
<i>Coleochaete scutata</i> ^d	JG444645	78 ^d	1
Chlorophyta			
<i>Chlamydomonas reinhardtii</i> ^e	Cre02.g116900	222	4
<i>Volvox carteri</i>	XP_002956564	1,124	3

^a Found 4 full ESTs

^b Found 5 full ESTs

^c Found 2 full ESTs

^d Only found one incomplete EST

^e Found in Phytozome

not the case for the amino- and carboxy-terminal motifs (data not shown).

All plant *LEA1* genes previously reported have one intron (Shih et al. 2010). The comparative analysis we carried out between the different plant *LEA1* genes showed that the intron position is conserved in eudicots, monocots, mosses (*Physcomitrella*) and spike mosses (*Selaginella*). Although this could be the case also for *Marchantia* and *Coleochaete* alga *LEA1* genes, the lack of information on their genomic sequences precludes the inclusion of these organisms in the conservation of the intron position. Contrasting with this result, the *Chlamydomonas* *LEA1* gene has six introns (Fig. 2); noticeably, each of four exons encodes only for the 20-mer EMGHKGGEARKEQLGHEGYQ (Fig. 2). An additional distinctive feature of the *Chlamydomonas* *LEA1* protein is that its amino-terminus is longer compared to plant *LEA1* proteins, even to those with a long amino terminal region due to the presence of a mitochondrial transit peptide (Fig. 2).

Identification of group 1 LEA-like proteins in non-plant organisms

To identify *LEA1* homologues in organisms other than plants, we have performed an exhaustive BLAST search in NCBI databases using the Arabidopsis AtEM1 and *Bacillus*

subtilis GsiB 20-mer amino acid repeat sequence (Espelund et al. 1992; Stacy and Aalen 1998) as query. Homologues of *LEA1* proteins were identified in the three taxonomic domains of life (Table 2). Starting with 1,515 bacterial and 81 archaeal fully sequenced genomes available from NCBI (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), we found 34 prokaryote genomes encoding polypeptides containing the *LEA1* motif, two archaeal and 32 bacterial (Table 2). Most of these microbes live in habitats commonly associated with soil or plant roots.

The detected *Archaea* genomes encoding *LEA1*-like proteins correspond to *Methanosaeta concilii* and *Methanocella paludicola*. The methanogen *M. paludicola* is a member of the Rice Cluster I lineage (RC-1) that usually inhabits roots and the surrounding rhizosphere of rice plants and is responsible for methane emissions in paddy rice fields (Erkel et al. 2006). *M. paludicola* encodes a large multigene *LEA1* family with 15 members (Table 2). *M. concilii* is also a methanogenic archaeobacteria isolated from anaerobic sludge from Ottawa sewage (Patel and Sprott 1990).

In the case of bacteria, we found *LEA1* protein relatives in only three of the 29 phyla that constitute this domain: these include 22 species of the *Firmicutes* phylum, 10 species in *Proteobacteria* phylum and one of the *Chloroflexi* phylum (*Ktedonobacter racemifer*) (Table 2). All 22 *Firmicutes* species correspond to *Bacilli* class. *Proteobacteria* containing *LEA1* protein motifs belong to δ - and γ -*Proteobacteria* classes. In the first group, one *LEA1*-like sequence corresponds to *Azotobacter* and four to *Pseudomonas* genera. γ -*Proteobacteria* *LEA1* relatives correspond to the *Myxococcales* bacterial order (*A. dehalogenans*, *M. fulvus*, *M. xanthus* and *S. aurantiaca*), bacteria able to form fruiting bodies upon starvation (Huntley et al. 2011).

The only non-plant eukaryotic *LEA1*-like protein containing the 20-mer repeat was detected in *Artemia franciscana*, a crustacean able to live in hyper-saline environments (Sharon et al. 2009; Table 2). The finding of multiple ESTs from RNA samples obtained from *A. franciscana* diapausic embryos or cysts indicates that, as in plants, *LEA1*-like proteins from this organism respond to water limitation, given that these specialized structures are able to resist severe stress conditions, including desiccation (Chen et al. 2009).

The number of *LEA1* genes varies widely among species. Most of the *Proteobacteria* and *Firmicutes* contain a single-copy of the *LEA1* gene (Table 2). By contrast, *LEA1* protein relatives are encoded by multigene families in *A. dehalogenans*, *A. vinelandii*, *B. megaterium*, *K. racemifer*, *P. polymixa*, *P. vortex* and archaea *M. paludicola*. In the crustacean *A. franciscana* seven *LEA1*-like

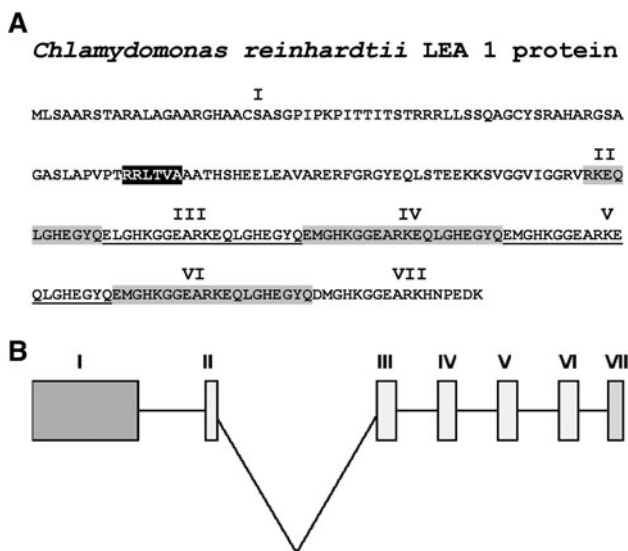


Fig. 2 *LEA1* gene and protein from *Chlamydomonas reinhardtii*. **a** Amino acid sequence of group 1 LEA protein. Roman numbers indicate the protein regions encoded by the different exons in the corresponding gene. The 20-mer motifs are shaded. Letters in the black box denote a possible mitochondrial targeting signal predicted by PSORT algorithm. **b** Schematic representation of introns and exons in the *LEA1* gene. Lines represent introns whereas exons are illustrated as grey boxes. Light grey boxes correspond to those exons encoding 20-mer motifs. Drawing is not to scale. The information on this protein was obtained from Phytozome (Cre02.g116900)

Table 2 Group 1 LEA proteins from non-plant organisms

Organism	Accession number	Total length	Number of repeats
Archaea			
<i>Methanocella paludicola</i> SANA E	YP_003357050	101	3
	YP_003356795	96	3
	YP_003355057	113	3
	YP_003357393	80	2
	YP_003357051	84	2
<i>Methanosaeta concilii</i> GP6	YP_003356143	83	2
	YP_003355656	93	2
	YP_003355090	77	2
	YP_003357373	61	1
	YP_003356866	54	1
	YP_003356572	57	1
	YP_003356555	55	1
	YP_003356164	59	1
	YP_003356148	66	1
	YP_003356147	61	1
	YP_004383121	63	1
Bacteria			
Firmicutes			
<i>Bacillus amyloliquefaciens</i> FZB42	YP_001420103	123	5
<i>Bacillus atrophaeus</i> 1942	YP_003976032	64	2
<i>Bacillus clausi</i> KSM-K16	BAD66040	133	5
<i>Bacillus coagulans</i> 36D1	YP_004859611	88	3
<i>Bacillus coahuilensis</i> m4-4	ZP_03227440	132	5
<i>Bacillus licheniformis</i> ATCC 14580	YP_077772	146	6
<i>Bacillus megaterium</i> QM B1551	YP_003562041	69	2
	YP_003566050	69	2
	YP_003565885	49	1
<i>Bacillus subtilis</i> str.168	CAA40009	123	5
<i>Carnobacterium</i> sp. 17-4	YP_004373749	54	1
<i>Lactobacillus sakei</i>	BAC99042	126	5
<i>Lysinibacillus fusiformis</i> ZC1	ZP_07049788	120	3
<i>Lysinibacillus sphaericus</i> C3-41	YP_001698192	52	1
<i>Melissococcus plutonius</i> ATCC 35311	YP_004457183	102	4
<i>Paenibacillus lactis</i> 154	ZP_09004132	152	5
<i>Paenibacillus mucilaginosus</i> KNP414	YP_004642050	148	5
	YP_003949390	115	3
	YP_003948259	82	3
<i>Paenibacillus vortex</i> V453	ZP_07902246	156	5
	ZP_07898955	81	3
<i>Pediococcus acidilactici</i> 7_4	ZP_06197568	125	5
<i>Pediococcus pentosaceus</i>	YP_003813015	128	5
<i>Staphylococcus carnosus</i> TM300	CAL28373	106	4
<i>Solibacillus silvestris</i> StLB046	BAK15770	94	1
<i>Sporosarcina newyorkensis</i> 2681	ZP_08678756	84	2
Proteobacteria			
<i>Anaeromyxobacter dehalogenans</i> strain 2CP-C	(1) YP_464900	131	4
	(2) YP_464899	64	1
	(3) YP_465692	61	1

Table 2 continued

Organism	Accession number	Total length	Number of repeats
<i>Azotobacter vinelandii</i> DJ	(1) YP_002798307	133	4
	(2) YP_002798308	55	1
	(3) YP_002801532	63	1
	(4) YP_002797466 ^a	95	1
<i>Myxococcus fulvus</i> HW-1	YP_004668452	150	4
<i>Myxococcus xanthus</i> DK 1622	YP_632032	153	4
<i>Pseudomonas aeruginosa</i> PAO1	NP_250880 ^a	114	1
<i>Pseudomonas fluorescens</i> WH6	ZP_07775951 ^a	116	1
<i>Pseudomonas mendocina</i> ymp	YP001187749 ^a	93	1
<i>Pseudomonas syringae</i> pv. japonica str. M301072PT	EGH33010 ^a	116	1
<i>Stigmatella aurantiaca</i> DW4/3-1	YP_003953991	148	4
Chloroflexi			
<i>Ktedonobacter racemifer</i> DMS 44963	ZP_06966625	93	3
	ZP_06969814	120	3
Eukarya			
<i>Artemia franciscana</i>	ABR67402	182	8
	ADE45145	142	6
	ADE45147	62	2
	ACX81197 ^b	97	2

^a These proteins contains YciG domains

^b This protein has putative a mitochondrial signal

proteins have been detected, containing 1–8 20-mer motif repeats (Warner et al. 2010; Table 2). Except for *LEA1* genes from plants (Battaglia et al. 2008; Shih et al. 2010), *Artemia franciscana* (Sharon et al. 2009), *Bacillus subtilis* (Maul et al. 1995) and *Bacillus clausii* (data not shown), it is unknown whether the rest of bacterial *LEA1* genes analyzed here are transcribed.

Comparison of group 1 LEA proteins between plant and non-plant relatives

The alignment of representative LEA1 protein amino acid sequences from different organisms shows that the regions with the highest similarity correspond to the 20-mer amino acid repeats (Fig. 3). Within the 20-mer sequence residues KGG and Y (denoted by arrows) are absolutely conserved among all kingdoms (Fig. 3). This alignment also shows that non-plant LEA1 protein relatives have a shorter amino-terminus as compared to the plant LEA1 proteins, and that the carboxyl-terminal region is variable between plant and non-plant LEA1 proteins (Fig. 3). The *A. franciscana* LEA1-like protein has been detected abundantly accumulated in mitochondria of encysted embryos, in agreement with the presence of a putative mitochondrial targeting sequence predicted in the N-terminal region of two *A. franciscana* LEA1-like proteins (ACX81198 and ACX81197) (Warner et al. 2010). Sequence analysis of

prokaryote and plant LEA1 protein relatives failed to identify the presence of a putative signal peptide or sub-cellular compartment targeting sequence, except in *Chlamydomonas*, as mentioned above. In addition, alignment analysis revealed the presence of a novel conserved motif, EAGR/QK, in the amino-terminal region of non-plant LEA1-like proteins (Fig. 3).

Analysis of amino acid motifs in group 1 LEA proteins

To define the consensus sequence of the 20-mer motif between plants and non-plants, multiple alignment analysis and the application of the WebLogo tool were carried out using the 20-mer motif. This analysis revealed that K-3, G-4, G-5, Y-17 and E-19 residues are the most conserved in microbial and plant relatives (Fig. 4). KGG motif is found in several bacterial stress-induced proteins (Astari et al. 2011); in plants this motif is best defined as XGG, where X is R or K (Fig. 4). When the alignment analysis was performed with microbial LEA1 protein relatives, it was evident that the G-1 and F-16 residues are well conserved in prokaryotes (Fig. 4). In the case of gram-positive bacteria orthologues, the 20-mer repeats are highly conserved, suggesting a recent emergence or the presence of evolutionary constraints specific to gram-positive LEA1-like proteins to maintain minimum changes in this repeat.

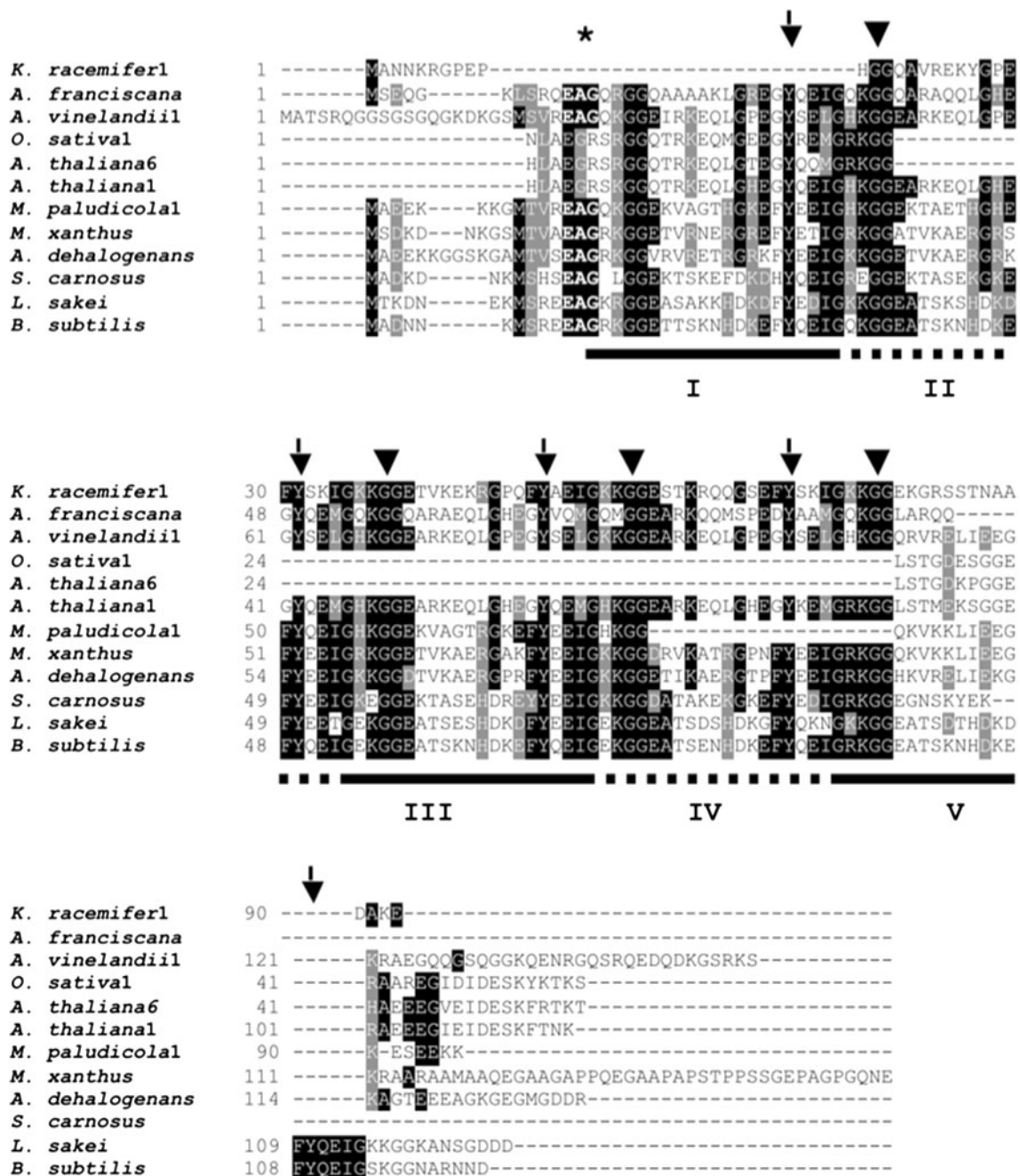


Fig. 3 Multiple-sequence alignment of plant and non-plant group LEA1 proteins. The 20-mer amino acid repeats are indicated in *continuous* and *discontinuous horizontal bars*. The Y and KGG residues are highly conserved among all domains and are indicated by

vertical arrows and *arrowheads*, respectively. The EAG prokaryotic motif is denoted with an *asterisk*. Accession numbers for plant and non-plant group 1 LEA proteins used in this alignment are described in Tables 1 and 2

Phylogenetic analysis

To examine the evolutionary relationship between plant group 1 LEA proteins and microbial relatives, a phylogenetic tree with bootstrap analysis was constructed (Fig. 5). This analysis only included the second exon of the plant LEA1 protein sequences used for the multiple-sequence

alignment (Fig. 1). Plant LEA1 and microbial GsiB proteins form two independent clades indicating that plant and microbial group 1 LEA proteins (including *A. franciscana*) are two independent monophyletic groups (Fig. 5).

A close phylogenetic relationship between the *A. franciscana* LEA1 and *Ktedonobacter racemifer* GsiB proteins was found, suggesting that it is the result of a horizontal

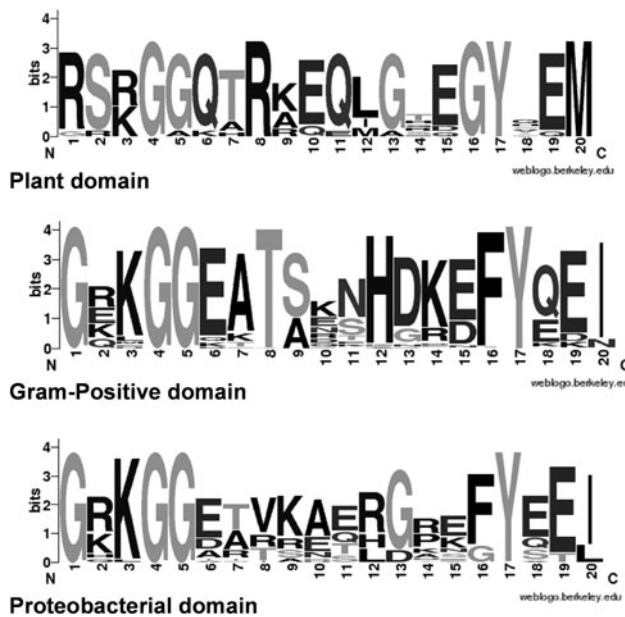


Fig. 4 Amino acid conservation of 20-mer repeats sequence in plants, gram-positive bacteria and proteobacteria. Consensus sequences were obtained by applying WebLogo program on a multiple alignment of the 20-mer motifs sequences

gene transfer event (HGT). Also, the phylogenetic analysis suggested dispersion from *Firmicutes* GsiB to γ -*Proteobacteria*. GsiB proteins from *Lactobacillus sakei* and *Pediococcus pentosaceus* are encoded within plasmids (Astari et al. 2011), which would allow horizontal gene transmission to *proteobacteria* and possibly to *Archaea*.

A subgroup of bacterial LEA1 protein relatives contains an additional conserved domain

According to the alignment analysis *P. aeruginosa*, *P. fluorescens*, *P. mendocina*, *P. syringae* and *A. vinelandii* LEA1-like protein sequences differ from their prokaryotic relatives in the longer length of their carboxy-terminus (Fig. 1 Supl.). These regions also show similarity with an intrinsically unstructured small protein, named YciG, initially found as an *Escherichia coli* hydrophilin, whose transcript accumulates in response to osmotic stress (Garay-Arroyo et al. 2000). Sequences similar to YciG were detected in NCBI databases not only in bacteria but also in some fungi, as in the case of *Neurospora crassa* CON-10 protein, whose transcript accumulates during conidiation and upon heat shock (Lee and Ebbole 1998). The analysis of the *E. coli* YciG protein sequence showed a possible 22-mer amino acid sequence with similarity to the 20-mer LEA1 protein motif that could be considered as an YciG distinctive motif (Fig. 6). *P. aeruginosa*, *P. fluorescens*, *P. mendocina*, *P. syringae* and *A. vinelandii* LEA1 protein relatives contain one LEA1 protein motif and three YciG

motifs (Fig. 6). The genomes of these microorganisms encode two types of LEA1-like proteins, one group containing hybrid LEA1-YciG motifs and another set of proteins with only the YciG motif. The *Firmicutes* and *Archaea* genomes that encode proteins with a LEA1 protein distinctive motif (Table 2) do not contain YciG-like proteins. From this analysis, can be inferred, that there is a close evolutionarily relationship between YciG and LEA1 proteins.

Protective activity of group 1 LEA-like plant and bacterial proteins

To test whether the bacterial group 1 LEA proteins are able to protect other proteins from the deleterious effects caused by freeze–thaw treatments, as it was previously shown for plant group 1 LEA proteins (Goyal et al. 2005; Zou et al. 2011), we carried out in vitro cryoprotection assays (see “Materials and methods” for details) using GsiB from *B. subtilis* and LEA1-YciG from *P. aeruginosa* (PaEM). As shown in Fig. 7, LDH is an enzyme susceptible to freeze–thaw treatments that is completely inactivated after five freeze–thaw cycles; however, the addition of a group 1 LEA protein from *A. thaliana* containing one copy of the 20-mer conserved motif (AtEM6) partially prevented this inactivation. A similar effect was detected when GsiB or PaEM proteins were added in these assays. In the three cases, this protective effect was detected at a 1:1 LEA:LDH molar ratio, and was concentration-dependent as a higher protection was observed when the LEA:LDH molar ratio was increased (20:1) in all three cases (Fig. 7). The most conserved sequence between plant and bacterial LEA1 proteins is that corresponding to the 20-mer motif. To determine the contribution of this sequence in the protective activity of these proteins, the 20-mer motif was deleted from AtEM6. The mutant recombinant protein, AtEM6del, was over-expressed and purified from *E. coli* to test its protective activity. The results showed that the removal of the 20-mer motif does not affect the ability of this LEA1 protein to prevent the inactivation of LDH upon freeze–thaw treatments when 1:1 ratio (LEA1:LDH) was used (Fig. 7); however, although a lower activity was detected for AtEM6del at 20:1 (LEA1:LDH) ratio, the difference between the complete and shortened proteins did not provide significant results, indicating that the deleted region does not play a relevant role in this activity under these particular conditions.

Discussion

Even though the physicochemical characteristics of LEA proteins are widespread in proteins of all taxonomic

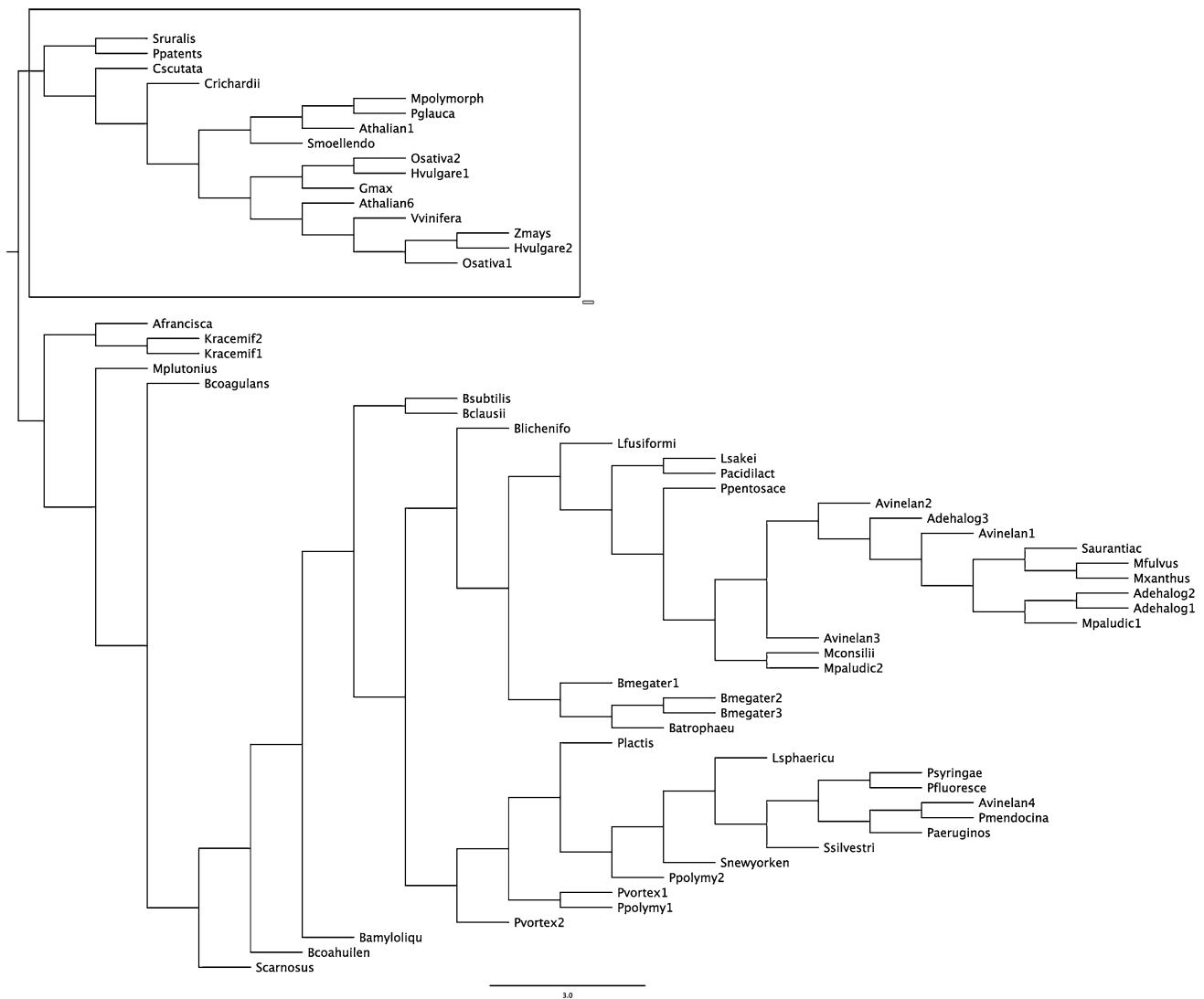


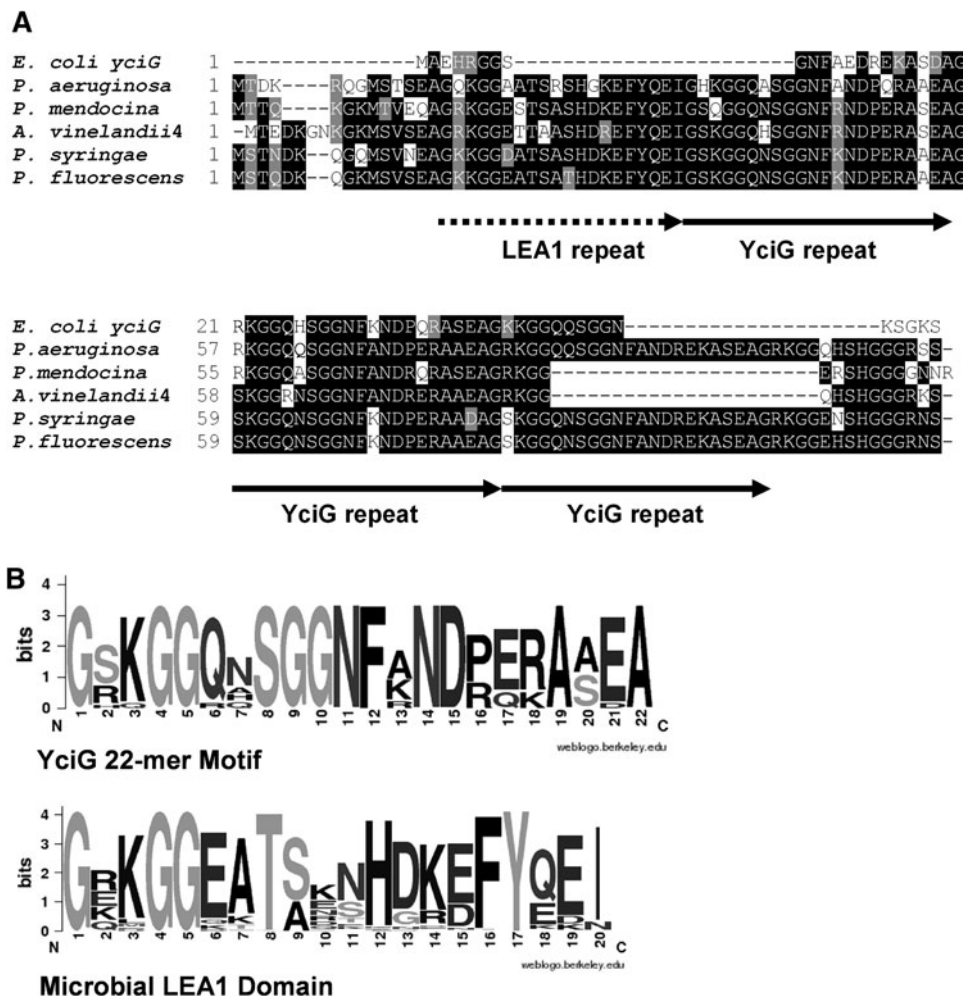
Fig. 5 Maximum-likelihood (RaxML) phylogenetic tree of group LEA1 proteins from plant and non-plant organisms. Plant clade is shown in a *box*. This analysis ignores the amino-terminal region

encoded by the first exon in plants. Accession numbers for all proteins in the tree are described in Tables 1 and 2. Only the topology of the maximum-likelihood tree is shown

domains, proteins with similar amino acid sequences are not common in the different kingdoms of life. Outside the plant kingdom, only proteins that present significant similarity to group 3 and group 1 LEA proteins or with some of their representative motifs have been found (Hand et al. 2010). Proteins similar to group 3 LEA proteins have been detected in insects, crustaceans, nematodes, rotifers and bacteria (Hand et al. 2010). For group 1 LEA proteins, similar polypeptides have been found in *B. subtilis* and crustacea (Chen et al. 2009; Sharon et al. 2009; Stacy and Aalen 1998). In this work, we explored the distribution of this protein group by searching for similarity in protein sequences obtained from the genome sequences of seed and seedless vascular and non-vascular plants, algae as well as of those non-*Viridiplantae* genome sequences

deposited in NCBI and Phytozome databases. This search allowed us to identify similar group 1 LEA proteins also in *Coleochate*, a streptophyte alga that lives in fresh water, and in the liverwort *Marchantia*, the most basal land plant described to date (Del Bem and Vincentz 2010). Among algae, streptophytes are considered the only phylum that successfully colonized the terrestrial habitats, in contrast to other algal groups that did not succeed in living on their substrate or those that effectively colonized marine habitats. The land conquest by streptophytes has favoured the hypothesis that land plants evolved from this algal group (Graham et al. 2012). The high similarity found between the *Coleochate* LEA1-like protein and that from *Marchantia*, not only between their 20-mer motifs but also between their C- and N-terminal regions, indicates that

Fig. 6 a, b Multiple-sequence alignment of group 1 LEA-YciG hybrid proteins with *Escherichia coli* YciG protein. **a** *Pseudomonas aeruginosa*, *Pseudomonas fluorescens*, *Pseudomonas mendocina*, *Azotobacter vinelandii* LEA1-YciG protein and *Escherichia coli* YciG were aligned using the MUSCLE program. Identical and similar amino acids are shaded in black and grey, respectively. The positions of the 22-mer repeat distinctive of the YciG protein is indicated by horizontal arrows, while LEA1 20-mer motif is indicated by a discontinuous line. **b** The YciG consensus sequence was obtained by applying the WebLogo program on all 22-mer repeats found in the aligned sequences



LEA1 proteins arose in streptophyte algae before land colonization, possibly selected as one of the factors that contribute to protect spores from dehydration. The presence of these proteins in bryophytes (liverworts and mosses), the first representatives of terrestrial plants, also suggests that they were inherited and maintained in these phyla, which confronted the crucial difficulty of moving from water to land.

The comparison of the amino acid sequences of LEA1-like proteins from *Coleochate* and *Marchantia* with that from *Chlamydomonas* showed a lower similarity than between the first two phyla; even though, there is a conserved 20-mer motif, the C- and N- regions did not show similarity. Because *Chlamydomonas* is a *Chlorophyta* alga, this observation is in agreement with the proposed phylogenetic distance between the *Chlorophyta* and *Streptophyta* lineages, and it is also consistent with the closer phylogenetic relationship between streptophytes and bryophytes. At this point, it is unclear whether *Chlorophyta* LEA1-like genes share a common ancestor with LEA1-like genes from

Streptophyta, or whether the similarities between the 20-mer motifs merely reflect a functional convergence.

An additional finding from this search was the presence of LEA1-like proteins in more than one group of Eubacteria. The previous description of a LEA1-like gene in *B. subtilis* suggested that this gene family was dispersed outside from the *Plantae* kingdom, a suggestion that is confirmed and extended with the data from this work, as we found that LEA1-like genes are present in the *Animalia* kingdom and also in various *Eubacteria* groups and in *Archaea*. However, noteworthy, the distribution of these genes among the different kingdoms and phyla seems somehow stochastic, maybe partially due to the limitation in genome sequence information. In the *Animalia* kingdom, we only found one sequence with similarity to LEA1 proteins in *A. franciscana* (*Arthropoda* phylum). Although the complete sequence of 2,019 bacterial and 106 archaeal genomes are now available (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), we only found LEA1-like proteins in two *Archeae* phyla. In the case of the Eubacterial

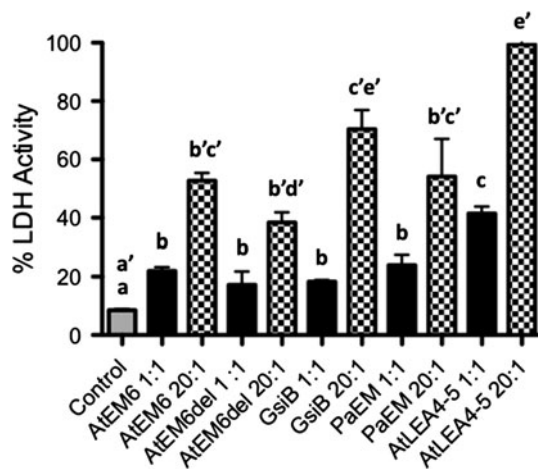


Fig. 7 Group1 LEA proteins from bacterial relatives prevent freeze–thaw effects on LDH activity. The freeze–thaw in vitro assay was carried out as described in “Materials and methods” using pure bacterial LEA1-like proteins. In these assays, the five freeze–thaw cycles applied led to a strong LDH inactivation. The addition previous to the freeze–thaw cycles of AtEM6, GsiB or PaEM recombinant proteins, in 1:1 or 20:1 ratios, prevented LDH inactivation in different extent. Letters indicate significant differences calculated with Tukey’s multiple comparisons post-test ($P < 0.05$). Letters with and without apostrophe represents comparisons among 20:1 and 1:1 (LEA:LDH molar ratio) samples, respectively

kingdom, 32 LEA1-like sequences were found; however, they appear to be distributed only in some species from the same order or family, even within the same phylum. One characteristic that is common to most bacteria, where this type of proteins were detected, is their association with soil habitats, which could suggest horizontal transfer from plants to bacteria; however, the phylogenetic analysis performed in this work indicates a different evolutionary path, maybe related to convergent evolution given the fact that eubacteria have had also to adapt to constant environmental changes in water availability, and these proteins could assist their survival under this condition. Further investigation will be needed to learn more about the evolutionary history of this protein family.

The presence of a sole LEA1-like protein in *A. franciscana*, the only member of the Animalia kingdom with a polypeptide similar to LEA1 proteins, could be explained by horizontal gene transfer with an ancestral bacterial endosymbiont already encoding a *LEA1* gene. This is supported by the finding of obligate α -proteobacterial endosymbionts such as bacteria from the *Wolbachia* genus that have been described in some arthropods (Dunning-Hotopp et al. 2007). A detailed genomic analysis of *A. franciscana* will be necessary to determine whether other bacterial sequences are also present in this arthropod genome.

As for other LEA protein families, little is known about the function of LEA1 proteins; however, it has been reported that proteins from this family are able to prevent

the inactivation of some reporter enzymes such as malate and lactate dehydrogenase and citrate synthase in vitro freeze–thaw assays (Goyal et al. 2005). The in vitro assay used in this study allowed us to detect protective activity for some bacterial LEA1-like proteins, comparable to that obtained for a LEA1 protein from *Arabidopsis*, suggesting that the bacterial proteins similar to plant LEA1 proteins have also retained analogous functions and hence might correspond to orthologous proteins. It should be mentioned that although the protective activity detected in this freeze–thaw in vitro assay for the proteins analyzed in this study was lower than that obtained for LEA proteins from other groups, and similar to that for other proteins such as RNase A, it was comparable to that reported for the wheat and soybean LEA1 proteins (Goyal et al. 2005; Zou et al. 2011). Even though, the ability of bacterial and plant LEA1-like proteins to prevent enzymatic inactivation upon freeze–thaw treatments supports the idea that the 20-mer motif is an important factor for this protective function, in agreement with previous evidence (Zou et al. 2011) and with its high conservation in different taxa, the non-significant effect observed in the freeze–thaw protection assays when the AtEM6 protein lacking the 20-mer motif (AtEM6del) was tested suggests that this sequence does not play a relevant role for the protective activity of LEA1 proteins, at least under the conditions tested. These observations confirm similar results obtained with wheat LEA1 protein (Gilles et al. 2007), and rather indicate that other regions in these proteins are important to prevent the inactivation of LDH upon freeze–thaw cycles. A more detailed analysis should be done to identify those residues or protein regions critical for LEA1 protein activity.

It is interesting to speculate that the acquisition of a LEA1-like gene during evolution may reflect the selective advantage of an increased capacity of the recipient organism to respond to adverse conditions. Future comparison of closely related species differing in the presence or absence of LEA1-like proteins will shed light into this interesting phenomenon.

Acknowledgments We thank Jaqueline Mazari for excellent technical assistance. We acknowledge to Paul Gaytán and Santiago Becerra from Oligonucleotide Synthesis and DNA Sequencing Facilities of the Instituto de Biotecnología-UNAM for providing us with the oligonucleotides and DNA sequences used in this work. This work was partially supported by CONACyT-Mexico to AAC (50485 and 132258). C.C.-V. was supported by a PhD fellowship from CONACyT.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105

- Asteri I-A, Boutou E, Anastasiou R, Pot B, Vorgias CE, Tsakalidou E, Papadimitriou K (2011) In silico evidence for the horizontal transfer of *gsiB*, a σ^B -regulated gene in gram-positive bacteria to lactic acid bacteria. *Appl Environ Microbiol* 77:3526–3531
- Battaglia M, Olvera-Carrillo Y, Garcarrubio A, Campos F, Covarrubias AA (2008) The enigmatic LEA proteins and other hydrophilins. *Plant Physiol* 148:6–24
- Bensmihen S, To A, Lambert G, Kroj T, Giraudat J, Parcy F (2004) Analysis of an activated *AB15* allele using a new selection method for transgenic *Arabidopsis* seeds. *FEBS Lett* 561:127–131
- Brandon M, Hall BM, Owsn KM, Singh KK (2011) Distinct functions of evolutionary conserved MSF1 and late embryogenesis abundant (LEA)-like domains in mitochondria. *J Biol Chem* 286:39141–39152
- Browne J, Tunnacliffe A, Burnell A (2002) Anhydrobiosis: plant desiccation gene found in a nematode. *Nature* 416:38
- Campos F, Zamudio F, Covarrubias AA (2006) Two different late embryogenesis abundant proteins from *Arabidopsis thaliana* contain specific domains that inhibit *Escherichia coli* growth. *Biochem Biophys Res Commun* 342:406–413
- Campos F, Guillen G, Reyes JL, Covarrubias AA (2011) A general method of protein purification for recombinant unstructured non-acidic proteins. *Protein Expr Purif* 80:47–51
- Chen WH, Ge X, Wang W, Yu J, Hu S (2009) A gene catalogue for post-diapause development of an anhydrobiotic arthropod *Artemia franciscana*. *BMC Genomics* 10:52
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190
- Del Bem LE, Vincentz MGA (2010) Evolution of xyloglucan-related genes in green plants. *BMC Evol Biol* 10:341
- Delseny M, Bies-Etheve N, Carles C, Hull G, Vicent C, Raynal M, Grellet F, Aspart L (2001) Late embryogenesis abundant (LEA) protein gene regulation during *Arabidopsis* seed maturation. *J Plant Physiol* 158:419–427
- Dunning-Hotopp J et al (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317:1753–1756
- Dure L, Galau GA (1981) Developmental biochemistry of cotton seed embryogenesis and germination. XIII. Regulation of biosynthesis of principal storage proteins. *Plant Physiol* 68:187–194
- Erkel C, Kube M, Reinhardt R, Liesack W (2006) Genome of rice cluster I Archaea—the key methane producers in the rice rhizosphere. *Science* 313:370–372
- Espelund M, Saebøe-Larssen S, Hughes DW, Galau GA, Larsen F, Jakobsen KS (1992) Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differentially by abscisic acid and osmotic stress. *Plant J* 2:241–252
- Garay-Arroyo A, Colmenero-Flores JM, Garcarrubio A, Covarrubias AA (2000) Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J Biol Chem* 275:5668–5674
- Gaubier P, Raynal M, Hull G, Huestis GM, Grellet F, Arenas C, Pagès M, Delseny M (1993) Two different Em-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. *Mol Gen Genet* 238:409–418
- Gilles GJ, Hines KM, Manfre AJ, Marcotte WR Jr (2007) A predicted N-terminal helical domain of a group 1 LEA protein is required for protection of enzyme activity from drying. *Plant Physiol Biochem* 45:389–399
- Goyal K, Walton LJ, Tunnacliffe A (2005) LEA proteins prevent protein aggregation due to water stress. *Biochem J* 388:151–157
- Graham LE, Arancibia-Avila P, Taylor WA, Strother PK, Cook ME (2012) Aeroterrestrial Coleochaete (*Streptophyta*, *Coleochaetales*) models early plant adaptation to land. *Am J Bot* 99:130–144
- Hand SC, Jones D, Menze MA, Witt TL (2007) Life without water: expression of plant LEA genes by an anhydrobiotic arthropod. *J Exp Zool* 307:62–66
- Hand SC, Menze MA, Toner M, Boswell L, Moore D (2010) LEA proteins during water stress: not just for plants anymore. *Annu Rev Physiol* 73:115–134
- Hara M, Fujinaga M, Kuboi T (2004) Radical scavenging activity and oxidative modification of citrus dehydrin. *Plant Physiol Biochem* 42:657–662
- Hattori T, Terada T, Hamasuna S (1995) Regulation of the *Osem* gene by abscisic acid and the transcriptional activator VPI: analysis of *cis*-acting promoter elements required for regulation by abscisic acid and VPI. *Plant J* 7:913–925
- Higuchi R (1990) Recombinant PCR. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp 177–183
- Hoekstra FA, Golosina EA, Buitink J (2001) Mechanisms of plant desiccation tolerance. *Trends Plant Sci* 6:431–438
- Hundertmark M, Popova AV, Rausch S, Seckler R, Hinch DK (2012) Influence of drying on the secondary structure of intrinsically disordered and globular proteins. *Biochem Biophys Res Commun* 417:122–128
- Huntley S, Hamann N, Wegener-Feldbrügge S, Treuner-Lange A, Kube M, Reinhardt R, Klages S, Müller R, Ronning CM, Nierman WC, Sogaard-Andersen L (2011) Comparative genomic analysis of fruiting body formation in *Myxococcales*. *Mol Biol Evol* 28:1083–1097
- Imai R, Chang L, Ohta A, Bray EA, Takagi M (1996) A *lea*-class gene of tomato confers salt and freezing tolerance when expressed in *Saccharomyces cerevisiae*. *Gene* 170:243–248
- Ingram J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. *Annu Rev Plant Physiol Plant Mol Biol* 47:377–403
- Kikawada T, Nakahara Y, Kanamori Y, Iwata K-I, Watanabe M, McGee B, Tunnacliffe A, Okuda T (2006) Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochem Biophys Res Commun* 348:56–61
- Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320
- Lee K, Ebbole DJ (1998) Analysis of two transcription activation elements in the promoter of the developmentally regulated *con-10* gene of *Neurospora crassa*. *Fungal Genet Biol* 23:259–268
- Liu Y, Zheng Y (2005) PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in *Escherichia coli*. *Biochem Biophys Res Commun* 331:325–332
- Manfre AJ, Lanni LM, Marcotte WR Jr (2006) The Arabidopsis group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. *Plant Physiol* 140:140–149
- Manfre AJ, LaHatte GA, Climer CR, Marcotte WR Jr (2009) Seed dehydration and the establishment of desiccation tolerance during seed maturation is altered in the *Arabidopsis thaliana* mutant *atem6-1*. *Plant Cell Physiol* 50:243–253
- Marcotte WR Jr, Russell SH, Quatrano RS (1989) Abscisic acid responsive sequences from the *Em* gene of wheat. *Plant Cell* 1:969–976
- Maul B, Volker U, Riethdorf S, Engelmann S, Hecker M (1995) Sigma B-dependent regulation of *gsiB* in response to multiple stimuli in *Bacillus subtilis*. *Mol Gen Genet* 248:114–120
- Olvera-Carrillo Y, Campos F, Reyes JL, Garcarrubio A, Covarrubias AA (2010) Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in *Arabidopsis thaliana*. *Plant Physiol* 154:373–390
- Olvera-Carrillo Y, Reyes JL, Covarrubias AA (2011) Late embryogenesis abundant proteins, versatile players in the plant

- adaptation to water limiting environments. *Plant Signal Behav* 6:586–589
- Parcy F, Valon C, Raynal M, Gaubier-Comella P, Delseny M, Giraudat J (1994) Regulation of gene expression programs during *Arabidopsis* seed development: roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell* 6:1567–1582
- Patel GB, Sprott GD (1990) *Methanosaeta concilii* gen. nov., sp. nov. (“*Methanothrix concilii*”) and *Methanosaeta thermoacetophila* nom. rev., comb. nov. *Int J Syst Bacteriol* 40:79–82
- Reyes JL, Rodrigo MJ, Colmenero-Flores JM, Gil JV, Garay-Arroyo A, Campos F, Salamini F, Bartels D, Covarrubias AA (2005) Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects in vitro. *Plant Cell Environ* 28:709–718
- Reyes JL, Campos F, Wei H, Arora R, Yang Y, Karlson DT, Covarrubias AA (2008) Functional dissection of hydrophilins during in vitro freeze protection. *Plant Cell Environ* 31:1781–1790
- Sharon MA, Kozarova A, Clegg JS, Vacratsis PO, Warner AH (2009) Characterization of a group I late embryogenesis abundant protein in encysted embryos of the brine shrimp *Artemia franciscana*. *Biochem Cell Biol* 87:415–430
- Shih MD, Hoekstra FA, Hsing YC (2008) Late embryogenesis abundant proteins. *Adv Bot Res* 48:211–255
- Shih MD, Huang LT, Wei FJ, Wu MT, Folkert A, Hoekstra FA, Hsing YC (2010) OsLEA1a, a new Em-like protein of cereal plants. *Plant Cell Physiol* 51:2132–2144
- Soulages JL, Kim K, Walters C, Cushman JC (2002) Temperature-induced extended helix/random coil transitions in a group I late embryogenesis-abundant protein from soybean. *Plant Physiol* 128:822–832
- Stacy RAP, Aalen RB (1998) Identification of sequence homology between the internal hydrophilic repeated motifs of Group I late-embryogenesis-abundant proteins in plants and hydrophilic repeats of the general stress protein GsiB of *Bacillus subtilis*. *Planta* 206:476–478
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690
- Tompa P, Bánki P, Bokor M, Kamasa P, Kovács D, Lasanda G, Tompa K (2006) Protein–water and protein–buffer interactions in the aqueous solution of an intrinsically unstructured plant dehydrin: NMR intensity and DSC aspects. *Biophysical J* 91:2243–2249
- Tunnacliffe A, Lapinski J, McGee B (2005) A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers. *Hydrobiologia* 181:315–321
- Vicient C, Hull G, Guillemot J, Devic M, Delseny M (2000) Differential expression of the *Arabidopsis* genes coding for Em-like proteins. *J Exp Bot* 51:1211–1220
- Warner AH, Miroshnychenko O, Kozarova A, Vacratsis PO, MacRae TH, Kim J, Clegg JS (2010) Evidence for multiple group I late embryogenesis abundant proteins in encysted embryos of *Artemia* and their organelles. *J Biochem* 148:581–592
- Wolkers WF, McCreedy S, Brandt WF, Lindsey GG, Hoekstra FA (2001) Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses in vitro. *Biochim Biophys Acta* 1544:196–206
- Xu D, Duan X, Wang B, Hong B, Ho T, Wu R (1996) Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol* 110:249–257
- Yotsui I, Saruhashi M, Kawato T, Taji T, Hayashi T, Quatrano RS, Sakata Y (2013) ABSCISIC ACID INSENSITIVE3 regulates abscisic acid-responsive gene expression with the nuclear factor Y complex through the ACTT-core element in *Physcomitrella patens*. *New Phytol*. doi:10.1111/nph.12251
- Zou Y, Hong R, He S, Liu G, Huang Z, Zheng Y (2011) Polyproline II structure is critical for the enzyme protective function of soybean Em (LEA1) conserved domains. *Biotechnol Lett* 33:1667–1673

LAS PROTEÍNAS DESORDENADAS Y SU FUNCIÓN: UNA NUEVA FORMA DE VER LA ESTRUCTURA DE LAS PROTEÍNAS Y LA RESPUESTA DE LAS PLANTAS AL ESTRÉS

César Luis Cuevas-Velázquez y Alejandra A. Covarrubias-Robles*

Departamento de Biología Molecular de Plantas, Instituto de Biotecnología,
Universidad Nacional Autónoma de México, Apdo. Postal 510-3, C.P. 62250
Cuernavaca, Mor., México. E-mail: *crobles@ibt.unam.mx

RESUMEN

El dogma que relaciona la función de una proteína con una estructura tridimensional definida ha sido desafiado durante los últimos años por el descubrimiento y caracterización de las proteínas conocidas como proteínas no estructuradas o desordenadas. Estas proteínas poseen una elevada flexibilidad estructural la cual les permite adoptar estructuras diferentes y, por tanto, reconocer ligandos diversos conservando la especificidad en el reconocimiento de los mismos. A las proteínas de este tipo, altamente hidrofílicas y que se acumulan ante condiciones de déficit hídrico (sequía, salinidad, congelamiento) se les ha denominado hidrofílicas. En plantas, las hidrofílicas mejor caracterizadas son las proteínas LEA (del inglés Late Embryogenesis Abundant) que se acumulan abundantemente en la semilla seca y en tejidos vegetativos cuando las plantas se exponen a condiciones de limitación de agua. Evidencia reciente ha demostrado que las proteínas LEA se requieren para que las plantas toleren y se adapten a condiciones de baja disponibilidad de agua. Esta revisión describe los datos más relevantes que asocian las características fisicoquímicas de estas proteínas con su flexibilidad estructural y cómo se afecta ésta por las condiciones ambientales; así como, aquéllos relacionados con sus posibles funciones en la célula vegetal ante situaciones de limitación de agua.

Palabras Clave: Déficit hídrico, estrés ambiental, hidrofílicas, proteínas intrínsecamente no estructuradas (PINES), proteínas LEA.

ABSTRACT

The dogma that relates the function of a protein with a defined three-dimensional structure has been challenged in recent years by the discovery and characterization of a set of proteins known as unstructured or disordered proteins. These proteins have a high structural flexibility which allows them to adopt different structures and therefore recognize different ligands while retaining their specificity. Proteins of this type, which are highly hydrophilic and accumulate under water deficit (drought, salinity, freezing) have been recently characterized and named hydrophilins. In plants, the best characterized hydrophilins are the LEA proteins (for Late Embryogenesis Abundant), which accumulate abundantly in the dry seed and in vegetative tissues when plants are exposed to water-limited environments. Recent evidence has shown that LEA proteins are required for plants to tolerate and adapt to conditions of low-water availability. This review describes the most relevant data regarding their structural flexibility and how this is affected by environmental conditions. Also, it addresses information related to their possible functions in plant cells that are exposed to water deficit.

Key Words: Water deficit, environmental stress, hydrophilins, intrinsically unstructured proteins (IUP), LEA proteins.

INTRODUCCIÓN

Las proteínas son las biomoléculas más versátiles y diversas de la célula. Están presentes en todos los procesos biológicos, indicativo de la capacidad que tienen para desarrollar un amplio número de funciones. ¿Qué hay detrás de la versatilidad funcional de estas moléculas?, la respuesta es: su estructura tridimensional. La estructura de una proteína es la forma espacial más estable que adopta en cierto ambiente, la cual depende directamente de su secuencia de aminoácidos. A partir del descubrimiento de las primeras estructuras cristalográficas comenzó a establecerse la idea de que la función de cada proteína depende en gran medida de su estructura tridimensional, a tal grado que se postuló que *“para que una proteína sea funcional, debe poseer una estructura tridimensional bien definida”*¹, idea que por muchos años ha prevalecido en muchas áreas de la ciencia. El paradigma de estructura-función fue totalmente aceptado por la comunidad científica durante muchos años. Los conceptos de estructuras secundarias como hélices alfa y láminas beta fueron ampliamente utilizados y se aceptaron como las unidades estructurales fundamentales de las proteínas. Sin embargo, a lo largo del tiempo se han encontrado proteínas a las cuales no se les ha podido asignar, con las metodologías especializadas, alguna estructura secundaria conocida, lo que generó la noción de que existen proteínas que carecen de una estructura estable; es decir, que poseen una “estructura o plegamiento azaroso” (en inglés “random coil”)². Inicialmente, este tipo de plegamiento flexible, azaroso o sin estructura se encontraba solamente en ciertas regiones de las proteínas que sólo se consideraban como conectores entre regiones con estructuras definidas o dominios

funcionales. A medida que ha aumentado el número de estructuras cristalográficas descritas, también ha aumentado la cantidad de proteínas con regiones con plegamientos azarosos, que varían en longitud y en número en una misma proteína. Por otro lado, el descubrimiento de nuevas proteínas y la caracterización de algunas de las ya descritas ha revelado la existencia de múltiples proteínas que, en la mayor parte de su extensión, presentan una estructura flexible o un plegamiento azaroso. Esto ha sido un reto para el concepto clásico de estructura-función, ya que de acuerdo a este concepto, estas proteínas no tendrían función alguna; sin embargo, ya se había demostrado que proteínas en las que abunda este tipo de plegamiento están involucradas en diferentes vías de señalización, algunas otras funcionan como factores transcripcionales, o bien son proteínas abundantes en algunos tejidos de diferentes organismos. Este cúmulo de evidencias, junto con el hecho de que son proteínas comunes en todas las especies, ha llevado a abrir una nueva sección en el capítulo de estructura-función de las proteínas, en el que se incluyen a este nuevo grupo de proteínas a las que se les ha denominado como “proteínas intrínsecamente desordenadas” (PIDs o IDPs, del inglés Intrinsically Disordered Proteins) o “proteínas no estructuradas” (PINES o IUPs, del inglés Intrinsically Unstructured Proteins)².

¿SIN ESTRUCTURA NO HAY FUNCIÓN?

Las PIDs o PINES son proteínas en las que abundan las regiones intrínsecamente desordenadas (RID o IDR, del inglés Intrinsically Disordered Regions) y que, por tanto, carecen total o parcialmente de una estructura tridimensional bien definida (Fig. 1). Como se

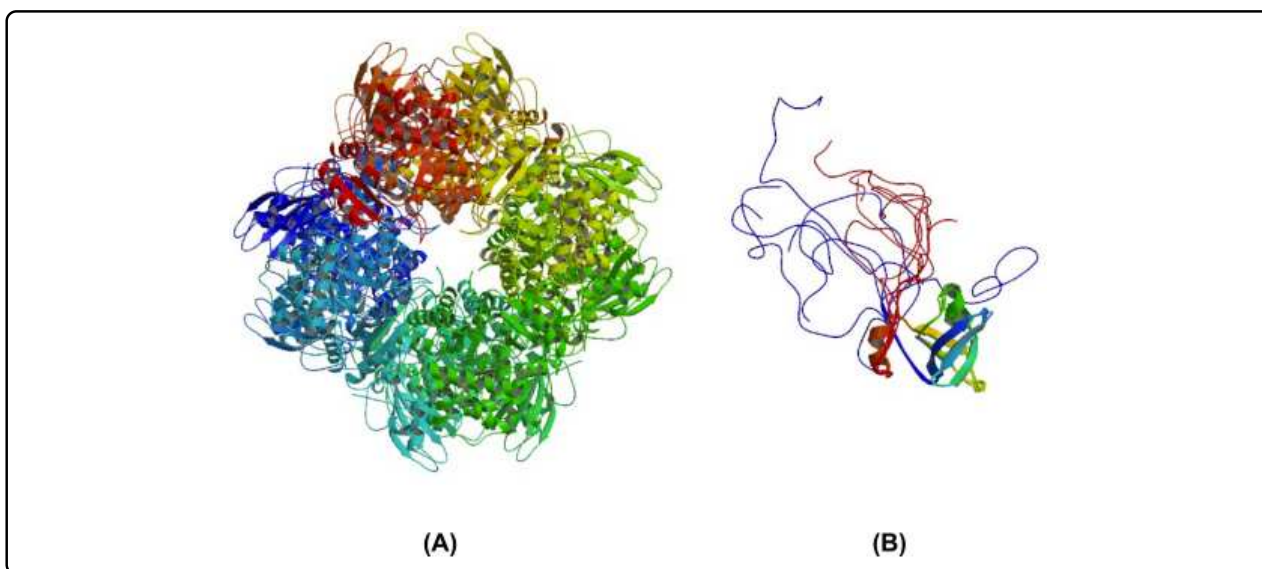


Figura 1. Estructuras de dos proteínas distintas: (A) la ribulosa 1,5-bisfosfato-carboxilasa-oxigenasa (rubisco) de espinaca (PDB:1RCX) es una proteína globular; (B) el factor de inicio de la traducción (eIF1A) de humano (PDB:1D7Q), en el que la mayor parte de su estructura es desordenada. Las líneas en (B) representan las posibles conformaciones que adquiere dicha zona de la proteína.

menciona arriba, la literatura especializada contiene cada vez más evidencia que demuestra que, contrario a lo establecido en el paradigma de estructura-función, las PINEs son capaces de llevar a cabo una o más funciones. De hecho, son proteínas que desempeñan funciones muy importantes en la célula, por ejemplo, entre las PINEs y/o proteínas con RIDs podemos encontrar receptores membranales, proteínas de andamiaje, proteínas del citoesqueleto, factores transcripcionales y receptores nucleares de hormonas, entre otras³. Su importancia resalta aún más ya que mutaciones en algunas de estas proteínas, las cuales comprometen su función cambiando su ambiente estructural, están asociadas a enfermedades tan serias como el cáncer y desórdenes neurodegenerativos⁴. El advenimiento de los métodos masivos para obtener información genómica y la aplicación de la informática predictiva ha permitido en los últimos años obtener información que indica que las PINEs comprenden una alta proporción de las proteínas codificadas en un genoma eucarionte, alrededor del 30% se predicen como proteínas parcial o totalmente desordenadas⁵. Esto ha llevado a múltiples preguntas de relevancia biológica referentes a la relación estructura-función en una proteína y al significado del desorden o flexibilidad estructural en estas moléculas. Preguntas como ¿por qué se ha

favorecido el desorden estructural en las proteínas a lo largo de la evolución? ¿representa alguna ventaja evolutiva para una proteína el poseer diferentes niveles de desorden? ¿qué tipo de funciones se encuentran asociadas a estas características fisicoquímicas?

En general, las PIDs o PINEs son proteínas que presentan una composición de aminoácidos particular, son proteínas de baja complejidad que tienen una baja abundancia o carecen de aminoácidos hidrofóbicos y/o voluminosos (Val, Ile, Met Phe, Trp, Tyr) que son la base de las proteínas globulares, pues promueven un plegamiento espontáneo y, por otro lado, poseen una alta proporción de aminoácidos cargados o polares (Gln, Ser, Pro, Glu, Lys, Arg) y en algunos casos tienen una alta abundancia de ciertos aminoácidos pequeños (Gly, Ala)⁶. Muchas de las regiones ricas en estos aminoácidos han sido reconocidas desde hace varios años como fundamentales para ciertas funciones en proteínas regulatorias, como lo son los factores transcripcionales, en los cuales estas regiones son la base para la interacción con los ácidos nucleicos y, por tanto, para su función como activadores o represores de la expresión genética⁷. Estudios recientes muestran que frecuentemente las PINEs contienen

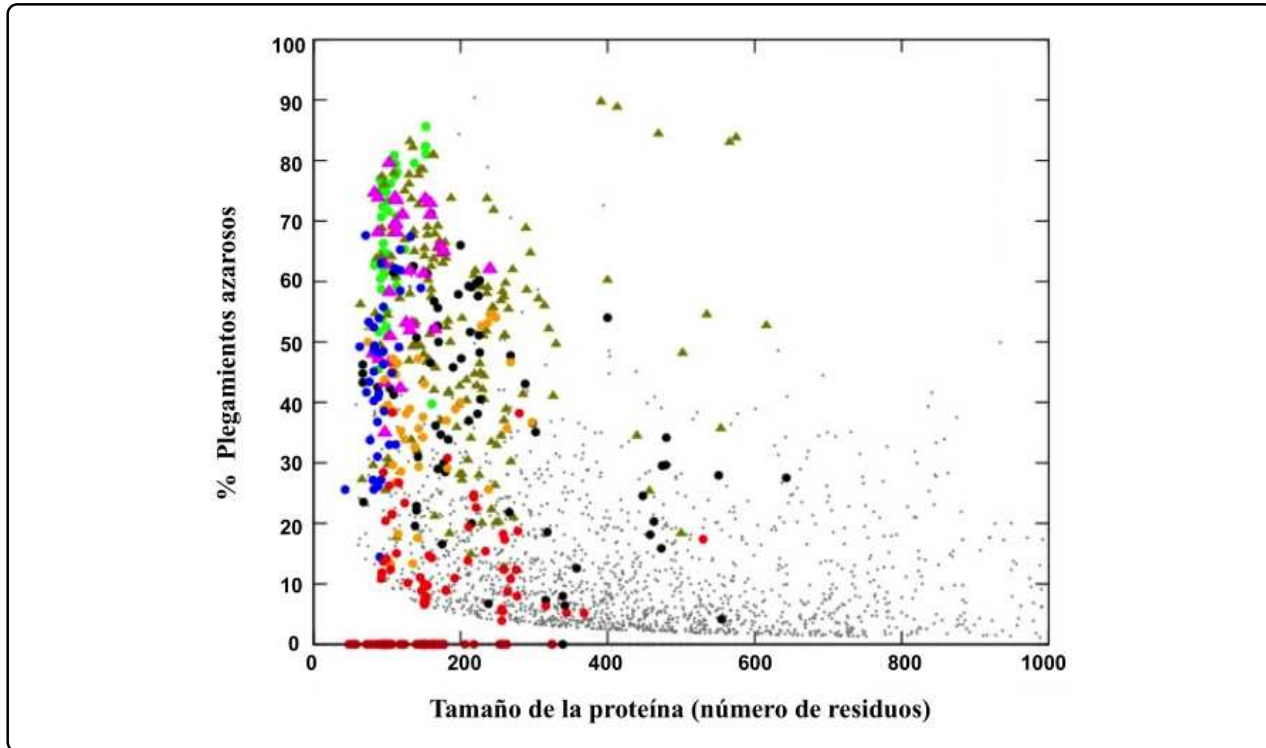


Figura 2. Representación esquemática del rango de plegamientos azarosos (*random coil*) existente en las proteínas LEA. Los porcentajes de plegamientos azarosos y el número de residuos en las proteínas se muestra en los ejes de las ordenadas y de las abscisas, respectivamente. Las proteínas LEA están representadas por círculos y triángulos llenos, en tanto que los puntos más pequeños representan proteínas diferentes a las proteínas LEA descritas para las células vegetales. La mayoría de las proteínas en las células vegetales presenta porcentajes de plegamiento azaroso iguales o menores a 0 y no se muestran en esta figura (Modificada de Battaglia *et al.*⁸).

regiones repetidas, las cuales se ha propuesto que pudieron haber evolucionado por expansión⁹. Es su composición de aminoácidos y características como éstas últimas las responsables de que la proteína sea incapaz de obtener una estructura estable en solución acuosa. Tenemos que considerar que todo tipo de estructura en las macromoléculas se ha caracterizado primordialmente en solución acuosa, de tal forma que la descripción del orden o flexibilidad estructural de una proteína corresponde generalmente a la estructura más estable que llega a adquirir en solución acuosa. Como se mencionó antes, hay que recordar que entre las PINEs las hay con diferentes grados de desorden, de tal forma que entre ellas existen proteínas que tienen múltiples dominios los cuales están conectados por regiones desordenadas o flexibles y que en cierto punto les permiten obtener una estructura compacta y más ordenada; otras que poseen amplias regiones intrínsecamente desordenadas (glóbulos fundidos o en inglés 'molten globules') y otras ordenadas, y las que son mayoritariamente desplegadas, heterogéneas o desordenadas (Fig. 2). Es probable que esta flexibilidad estructural que se reconoce en solución acuosa rara vez se dé bajo condiciones fisiológicas en la célula, ya que las condiciones del microambiente podrían promover la interacción entre ciertos aminoácidos que en solución acuosa no ocurre, de tal forma que las proteínas podrían tender a adquirir cierto grado de estructura secundaria en determinadas regiones.

Con esta información ahora regresemos a la pregunta de ¿por qué se ha favorecido el desorden estructural en las proteínas a lo largo de la evolución? Reflexiones sobre esta pregunta ha llevado a proponer varias hipótesis; quizás una de las más favorecidas sea la que considera que las PINEs son proteínas más flexibles o maleables, de tal forma que pueden adoptar diferentes estructuras funcionales que les permitirían reconocer ligandos diversos. Es posible incluso pensar que este tipo de proteínas puede presentar un mayor número de interfases intermoleculares en un momento determinado que, en el caso de las proteínas ordenadas o globulares, implicaría que éstas fuesen mucho más grandes. Esto probablemente representó una restricción evolutiva pues el aumentar la capacidad funcional de las proteínas ordenadas o globulares por un aumento en su tamaño llevaría a un mayor amontonamiento molecular en la célula, probablemente desventajoso para algunas funciones pues disminuiría el agua disponible, entre otros efectos. Por otro lado, cabe considerar que la mayor flexibilidad estructural de las PINEs no sólo abre la posibilidad de presentar más de una interfase intermolecular, sino también de exhibir más de una estructura funcional dependiendo de las condiciones microambientales y de la presencia de posibles ligandos a través de los cuales se reflejaría su función. Por ejemplo, de acuerdo a la situación metabólica de la célula es posible que se modifique el pH intracelular, el contenido de ciertos iones, el estado oxidante, el agua disponible, etc.; lo cual podría favorecer una u otra estructura y, por tanto, una u otra función. Dependiendo del estado de desarrollo de un organismo podrían estar presentes

ciertas moléculas y otras no, lo que favorecería la interacción con ciertas moléculas en un estadio determinado y con otras en uno diferente, y, por tanto, en cada caso realizaría una función distinta. Sin duda un escenario como éste sería ventajoso para muchos organismos pues no sería necesario incrementar su capacidad de almacenaje genético (genomas más grandes) y resultaría en la adquisición de respuestas adaptativas más rápidas y, quizás, más exitosas. Hipótesis como éstas se ven favorecidas por observaciones como las que resultan del análisis de la secuencia de genomas de organismos complejos. Por ejemplo, la secuencia del genoma humano ha mostrado que nuestro material genético codifica para aproximadamente 25,000 proteínas, el mismo número aproximado de proteínas codificadas por el genoma de la pequeña hierba *Arabidopsis thaliana*; esto ha llevado a una paradoja aún no resuelta relacionada con la falta de correlación entre la complejidad de un organismo y el número de genes que codifican para proteínas. Una explicación parcial a esta paradoja pudiera ser que la complejidad funcional de un organismo no necesariamente se resuelve con tener más genes que codifiquen proteínas diferentes, sino algunos genes que codifiquen proteínas multifuncionales (también llamadas 'moonlighting' en inglés) como lo serían las PINEs.

Algunas de las ideas que se describen arriba han sido comprobadas experimentalmente de tal manera que ahora sabemos que las PINEs tienen la capacidad de fluctuar entre diferentes estados conformacionales en una escala de tiempo muy corta (del orden de nano a micro-segundos)¹⁰. Esta alta flexibilidad por unidad de tiempo les da la habilidad de interactuar con un abanico de moléculas distintas con las que se unen con baja afinidad pero conservando una alta especificidad¹⁰, lo cual representa una gran ventaja desde el punto de vista biológico y evolutivo. Ante esta posibilidad, las PINEs serían capaces incluso de interactuar específicamente con la misma molécula pero de formas distintas (Fig. 3). Así pues, algunas PINEs tienen la capacidad de funcionar como interruptores (*switches*); es decir, la interacción de una PINE, en un estado conformacional dado, con una molécula blanco determinada promueve que una vía de señalización se encuentre activa; sin embargo, si el estado conformacional de la PINE cambia (modificando la dinámica de la interacción con su blanco), la vía se inactiva (Fig. 3). Éste es el caso de las p21 y p27, las cuales pueden tener tanto un efecto inhibidor, de hecho conocidas como proteínas inhibidoras de las quinasa dependientes de ciclinas (CDKs), como activador sobre las mismas CDKs, lo cual se reportó posteriormente¹¹. Otros ejemplos similares se han descrito en la literatura con proteínas como la DHPR (receptor de la dihidropiridina, detector de voltaje)¹², la CFTR (regulador de la conductancia transmembranal en la fibrosis quística)¹³; o bien, el caso de la calpastatina, un inhibidor de la proteasa activada por calcio, calpaina. La calpastatina favorece la activación de la calpaina facilitando su unión a calcio; sin embargo, también la inhibe a través de unirse por un dominio específico¹⁴ (para mayor información ver Dyson y Wright⁵).

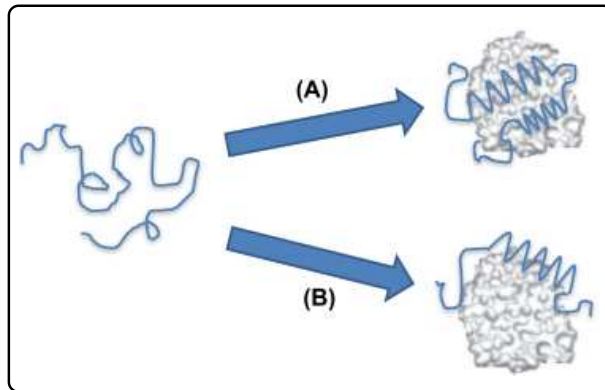


Figura 3. Las PINEs son capaces de realizar más de una función, gracias a su flexibilidad. Una PINE puede unir un ligando con una conformación dada y realizar una función específica (A); sin embargo, la misma PINE puede interactuar con el mismo ligando (o con otro distinto) a través de una conformación diferente y, por tanto, realizar una función completamente distinta (B). Las PINEs, por lo tanto, son capaces de realizar más de una función, por lo que se constituyen como un grupo de proteínas multifuncionales (que en inglés se refieren como “moonlighting”).

La multifuncionalidad que se ha observado en algunas de las PINEs es el resultado de modificaciones postraduccionales que favorecen ciertos cambios conformacionales, lo que les permite presentar una interfase diferente y con ello modular el reconocimiento a sus moléculas blanco (Fig. 4). Éste es el caso de modificaciones por fosforilación en diferentes PINEs, como sucede con CFTR (regulador de la conductancia transmembranaral en la fibrosis quística), el cual ante una señal es fosforilado, lo que induce un cambio en su conformación y en el modo de unión de esta proteína a sus blancos¹⁵. Otro ejemplo bien caracterizado ocurre en CBP (CREB Binding Protein), cuyo dominio KIX interactúa con el dominio KID de CREB. Esta interacción requiere de la fosforilación del dominio KID, la cual habilita la alta afinidad por el dominio KIX. En este caso, tanto la forma fosforilada como la no fosforilada del dominio KID se encuentran no estructuradas en solución; sin embargo, la KID fosforilada se pliega al unirse al dominio KIX, de esta manera la flexibilidad intrínseca del dominio KID de CBP, ante la modificación por fosforilación, facilita la interacción con el dominio KIX de CREB⁶.

La flexibilidad estructural de las PINEs les permite unirse de forma específica a ligandos diversos, a la fecha se han descrito como moléculas interactoras de las PINEs a otras proteínas, ARN, ADN, membranas e incluso moléculas pequeñas como iones o azúcares¹⁶. Así pues, la función de las PINEs es también diversa, sus funciones incluyen regulación transcripcional y traduccional, transducción de señales celulares, fosforilación de proteínas, almacenamiento de moléculas pequeñas, chaperonas de proteínas y de ARN, regulación del ensamblaje de grandes complejos multiproteicos y del ribosoma, entre otras¹⁶.

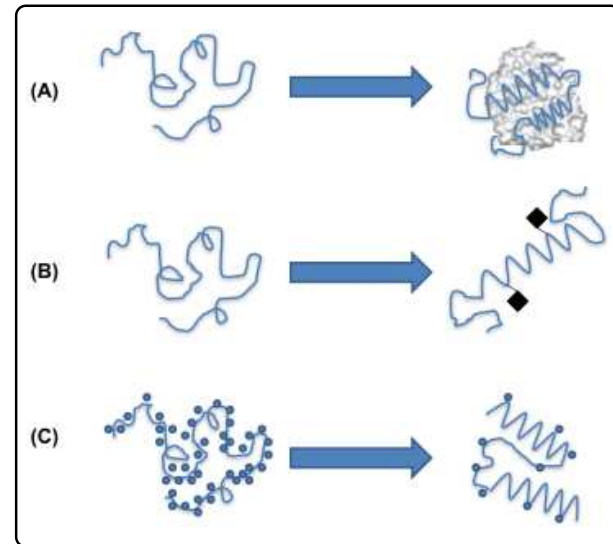


Figura 4. El plegamiento de las PINEs puede ser inducido por diferentes moléculas o condiciones ambientales en la célula. La interacción con los aminoácidos de un ligando (A) puede inducir en la PINE una conformación más estable. (B) Las modificaciones postraduccionales son otro factor que influye en el plegamiento de una PINE, el cual puede acompañarse además de la interacción con su(s) ligando(s). (C) La disponibilidad de agua también puede ser capaz de modificar la estructura de las PINEs; al cambiar la cantidad de agua disponible en el ambiente, la interacción de las PINEs con el agua se modificaría, lo que tiene como consecuencia un cambio en su conformación. Los rombos representan alguna modificación postraduccional (fosforilación, glicosilación, etc); los círculos representan moléculas de agua.

La plasticidad intrínseca de las PINEs no significa necesariamente que sean incapaces de adquirir una estructura más estable. Como se mencionaba antes, es posible que los cambios en su microambiente favorezcan una mayor estabilidad o un mayor orden estructural. Además, es posible que la interacción o unión con un ligando no sólo facilite un cambio conformacional sino que también lo estabilice (Fig. 4). Esta posibilidad ha sido demostrada para algunas PINEs y, en algunos casos, se ha estudiado con detalle. Algunos reportes recientes indican que los cambios estructurales que llevan a una conformación ordenada en una PINE se dan de manera acoplada, concomitantemente, con la unión a su ligando⁶. El costo de la entropía necesaria para plegar a una proteína desplegada lo paga la entalpía generada por la unión a su ligando. Evidencias como ésta han llevado a considerar que las PINEs adquieren la estructura necesaria para llevar a cabo su función al unirse a sus moléculas blanco. Sin embargo, aún queda abierta la pregunta de cómo es que las PINEs son capaces de reconocer a sus ligandos, ¿acaso se da en ausencia de un orden estructural determinado? o ¿es necesario cierto orden para presentar las interfaces adecuadas en el momento apropiado? ¿es el ambiente el

responsable de proporcionar las condiciones fisicoquímicas necesarias para facilitar o inducir tales cambios conformacionales o todo se reduce al azar? Éstas son preguntas que aún quedan por dilucidar.

LAS PINEs Y LA RESPUESTA AL ESTRÉS

Hasta el momento ha sido demostrado que las PINEs son proteínas importantes en la célula y que son determinantes para el adecuado cumplimiento de las funciones celulares basales. Más recientemente, se ha encontrado que las PINEs también participan en la respuesta de los organismos a distintos estímulos o condiciones desfavorables para la célula. Un ejemplo clásico de las proteínas no estructuradas o flexibles que participan en la respuesta a condiciones adversas son las "hidrofilinas". Las hidrofilinas se definen como aquellas proteínas que tienen un índice de hidrofilicidad mayor a 1 y un contenido de glicina y de otros aminoácidos pequeños de más del 6%¹⁷. Además de estar presentes en todos los dominios taxonómicos, se ha encontrado que su acumulación está asociada notablemente con la exposición de las células o de los organismos a condiciones de limitación de agua. Llama la atención que en los diferentes *phyla* en donde se han encontrado, están acumuladas en respuesta a una baja disponibilidad de agua en el ambiente, dígase arqueas, bacterias, hongos, plantas, insectos, artrópodos y otros animales¹⁷. También se acumulan en aquellas células que transitan por etapas de desarrollo que involucran un déficit hídrico y en estructuras u órganos que pasan por estados de deshidratación severa¹⁷.

El hecho de que las hidrofilinas compartan la gran mayoría de sus características fisicoquímicas con las descritas para las PINEs y, de que se haya predicho por análisis *in silico* un alto porcentaje de desorden en todas ellas, permite incluirlas, al menos hipotéticamente, en el grupo de las PINEs o PIDs; sin embargo, a la fecha sólo de muy pocas hidrofilinas conocemos datos experimentales sobre sus características estructurales.

La hipótesis de que la presencia de las hidrofilinas confiera una ventaja a los diferentes organismos para contender con la limitación de agua parece evidente; sin embargo, aún quedan por responder preguntas como: ¿cuál es el mecanismo de acción de las hidrofilinas a nivel molecular? ¿cuál es la importancia de poseer una estructura flexible en este tipo de proteínas? y ¿por qué las características de alta hidrofilicidad y desorden estructural se han mantenido en estas proteínas a lo largo de la evolución en los diferentes organismos y se han relacionado íntimamente con situaciones de déficit hídrico? Éstas son interrogantes que han sido abordadas recientemente y que aún no se entienden del todo.

LA RESPUESTA DE LAS PLANTAS A UN DÉFICIT HÍDRICO REQUIERE DE CIERTO NIVEL DE DESORDEN

Un caso muy interesante que surge del estudio de las hidrofilinas en las plantas es el de un grupo de proteínas muy amplio y

abundante, conocido como proteínas LEA (por sus siglas en inglés Late Embryogenesis Abundant). Estas proteínas se encontraron por primera vez, hace ya casi 30 años, por ser proteínas abundantes en la etapa tardía de la embriogénesis de la semilla, etapa en la cual ocurre una desecación natural muy alta¹⁸. Posteriormente se encontró que, además de acumularse en la semilla, se acumulan abundantemente en tejidos vegetativos sometidos a déficit hídrico (sequía, salinidad y congelamiento)¹⁹. La correlación encontrada entre la acumulación de las proteínas LEA y condiciones adversas de limitación de agua sugirió que estas proteínas juegan un papel fundamental en la respuesta y en la tolerancia de las plantas a estas condiciones. Esta hipótesis se ha visto reforzada por el hecho de que la gran mayoría de las proteínas LEA se acumulan en respuesta a tratamientos con ácido abscísico (ABA)²⁰, conocido como la hormona del estrés, por ser una de las hormonas responsables de desencadenar una de las cascadas de señalización que dan lugar a la expresión de proteínas de respuesta que conllevan a la adaptación y/o tolerancia al estrés.

Desde el descubrimiento de las proteínas LEA, numerosos esfuerzos se han enfocado en determinar si son realmente una estrategia adaptativa utilizada por las plantas. En especial, estudios de genética funcional utilizando la planta modelo *Arabidopsis thaliana* han demostrado que al anular la expresión de 1, 2 ó 3 diferentes proteínas LEA de un mismo grupo, causa una deficiencia en la resistencia a sequía o a estrés osmótico, en comparación con las plantas silvestres²¹. Esto se ha visto también para proteínas LEA de otros grupos y, se ha encontrado que al complementar las mutantes con las proteínas correspondientes, el fenotipo se restablece. Experimentos como éstos han demostrado que las proteínas LEA son fundamentales para que las plantas puedan contender con situaciones de déficit hídrico.

Las proteínas LEA se han clasificado en al menos siete grupos, con base en la similitud en su secuencia de aminoácidos, de tal forma que cada grupo se distingue por ciertos motivos consenso distintivos⁸. En estudios previos utilizando difracción circular se ha demostrado que algunas proteínas LEA presentan estructuras con un alto nivel de desorden²². Sin embargo, algunos de estos estudios también demuestran que al someterlas a un tratamiento con compuestos que compiten por el agua cercana a las proteínas, como lo es el trifluoroetanol (TFE), estas proteínas tienden a adquirir cierto nivel de estructura, en su mayoría estructuras tipo alfa hélices u hojas beta. Estos datos indican que las condiciones de disponibilidad de agua son capaces de inducir cambios conformacionales en estas proteínas que las lleven a adquirir un mayor orden estructural. Consistente con estos datos, se encontró que el amontonamiento molecular inducido por la combinación de PEG (poli-etilen-glicol) y glicerol también es capaz de promover cambios conformacionales en estas proteínas. Estos resultados sugieren que las condiciones ambientales son capaces de modificar la organización estructural de las PINEs y,

en particular, de las proteínas LEA, lo cual adiciona un elemento más a las hipótesis propuestas dirigidas a entender el funcionamiento de este tipo de proteínas. Así, nosotros proponemos que las proteínas LEA presentan diferentes organizaciones estructurales dependiendo de la cantidad de agua disponible y/o del amontonamiento molecular en la célula, lo que a su vez podría modular el reconocimiento de diferentes ligandos, en concordancia con las hipótesis previamente propuestas para las PINEs. Estas proteínas son altamente flexibles en solución acuosa, condición poco frecuente en la célula; ante una disminución en la disponibilidad de agua y, ante el consecuente incremento en la concentración intracelular de las macromoléculas (amontonamiento molecular), las proteínas LEA tenderán a adquirir una conformación más definida que les permitiría reconocer algún o algunos ligandos y de esta forma estabilizarse (Fig. 5).

Esta hipótesis se ha combinado con el modelo de trabajo al que hemos llegado, de acuerdo con los datos obtenidos a lo largo de varios años, los cuales muestran que las proteínas LEA son capaces de prevenir los cambios estructurales que se promueven por los efectos de la limitación de agua en proteínas/enzimas reporteras (como la lactato deshidrogenasa, LDH), que pudieran llegar a su desnaturalización y consecuente agregación, y que tienen como consecuencia una desactivación de éstas. Este efecto protector es evidente en ensayos *in vitro* en los que al deshidratar de manera progresiva a la enzima reportera, ésta

pierde paulatinamente su actividad; sin embargo, si antes de iniciar la deshidratación se le agrega una proteína LEA, en una relación molar 1:1, la actividad de la enzima reportera se mantiene en un mayor nivel que cuando no es protegida por la adición de la proteína LEA²³. El hecho de que estas proteínas sean capaces de proteger a la enzima reportera en una relación molar 1:1 es indicativo de que uno de los posibles mecanismos de protección sea a través de una interacción directa proteína LEA-proteína blanco. Adicionalmente, datos provenientes de este tipo de ensayos nos han indicado que la capacidad protectora de las proteínas LEA de los diferentes grupos es distinta, algunas presentan una protección de casi 100% aún bajo condiciones de pérdida de agua mayores al 98%, aún mejor que la misma trehalosa, un disacárido con propiedades protectoras demostradas bajo diferentes condiciones²³ (Fig. 5). Sin embargo, otras proteínas LEA no muestran protección en estos ensayos. Es probable que las limitaciones de estos ensayos *in vitro*; entre otros, el uso restringido de enzimas reporteras, susceptibles a estos tratamientos, que no son los ligandos naturales, no permitan detectar su capacidad protectora. También es posible que su función esté dirigida a proteger otro tipo de moléculas, como ácidos nucleicos (ADN y/o ARN); o bien, algunas estructuras celulares como podrían ser membranas o citoesqueleto. Esto nos obligará a buscar alternativas experimentales que nos permitan explorar otros posibles blancos o funciones. Cabe mencionar que también hemos observado protección de enzimas reporteras por proteínas LEA en ensayos

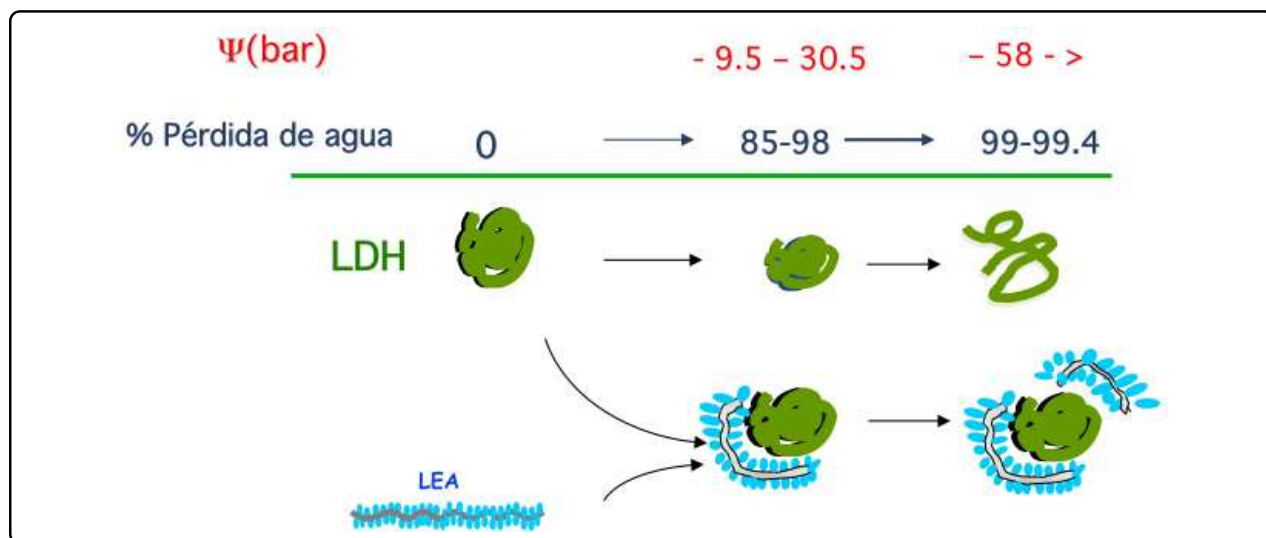


Figura 5. El modelo de acción de algunas de las proteínas LEA surge a partir de los ensayos *in vitro* descritos por Reyes et al.^{23,24} (ver texto). Cuando las células sufren pérdida de agua, estas condiciones pueden promover cambios conformacionales en las proteínas de mantenimiento, provocando una disminución en sus actividades catalíticas o una menor eficiencia en sus funciones; condiciones más severas, pueden llegar a inducir una pérdida total de su estructura o desnaturalización y, con ello una pérdida también total de su actividad. En condiciones en las que existe una cantidad óptima de agua, las proteínas LEA se encontrarían mayoritariamente desplegadas. Cuando la cantidad de agua disponible disminuye, se podría modificar la conformación de las proteínas LEA hacia un estado "activo", en el que ahora serían capaces de interactuar con las proteínas de mantenimiento (u otros ligandos o blancos), previniendo así la pérdida de su estructura y, por lo tanto, de su función.

de congelamiento-descongelamiento²⁴, otra condición que impone un déficit hídrico. Por otro lado, la aplicación de diferentes ensayos *in vitro* nos han mostrado que aunque estas proteínas pueden prevenir la desactivación de enzimas reporteras causada por los efectos de la deshidratación parcial, las proteínas LEA no son capaces de proveer tal protección ante temperaturas elevadas, lo cual indica que su función está más bien dirigida a contender con situaciones de baja disponibilidad de agua. Interesantemente, chaperonas moleculares como los son las proteínas de choque térmico de bajo peso molecular (sHSPs, del inglés small Heat Shock Proteins) no son capaces de prevenir la pérdida de actividad de la enzima reportera, ni su posterior desnaturalización, ante situaciones de deshidratación severa, lo cual indica que las sHSPs son incapaces de ejercer su función renaturalizadora en estas condiciones de estrés²³ y, por otro lado, refuerza la importancia funcional de las características fisicoquímicas de las proteínas LEA ante condiciones de déficit hídrico.

A pesar de las limitaciones que presentan estos ensayos *in vitro*, nos han permitido detectar una función para algunas de estas proteínas, y se presentan como una herramienta para explorar algunas otras preguntas en relación a las características estructurales de las proteínas LEA. Entre otras, utilizando proteínas mutantes en estos motivos, generadas *in vitro*, podríamos contestar cuál es el significado funcional de los motivos conservados en alguna de las familias o grupos de estas proteínas, cómo contribuyen estos motivos en la función protectora de estas proteínas, en estos ensayos, y cuáles son los aminoácidos relevantes para ello. La información que obtengamos será de utilidad para predecir la relevancia de posibles cambios conformacionales en estas condiciones. Aún más importante, será la posibilidad de probar estas mutantes *in vivo*, transformando plantas con estas variantes y evaluando los fenotipos que se obtengan, lo cual nos permitirá extrapolar nuestros resultados *in vitro* a su función en la planta.

La falta de similitud con otras proteínas de la célula, junto con su desorden o flexibilidad estructural, han sido obstáculos que han frenado un mayor avance en el conocimiento de la función de estas proteínas. Sin embargo, estas mismas características hacen de estas proteínas un modelo no sólo enigmático, sino también atractivo para su estudio. Las proteínas LEA se podrían considerar como un buen modelo para estudiar la función de las PINES en plantas y, sobre todo, para entender el por qué proteínas como éstas han sido seleccionadas a lo largo de la evolución como una solución general a un problema común en los diferentes organismos como lo es la limitación de agua.

AGRADECIMIENTOS

Este trabajo fue apoyado parcialmente por DGAPA (IN222309). César Luis Cuevas es apoyado con una beca de doctorado por el CONACyT.

REFERENCIAS

1. Stryer, L. Biochemistry (Freeman and Co, New York, 1995). 1064 págs.
2. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527-533 (2002).
3. Ward, J.J., *et al.* Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635-645 (2004).
4. Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z. & Dunker, A.K. Intrinsic disorder in cell-signalling and cancer-associated proteins. *J. Mol. Biol.* **323**, 573-584 (2002).
5. Schad, E., Tompa, P. & Hegyi, H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol.* **12** (2011).
6. Dyson, H.J. & Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197-208 (2005).
7. Mitchell, P.J. & Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378 (1989).
8. Battaglia, M., *et al.* The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* **148**, 6-24 (2008).
9. Tompa, P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays* **25**, 847-855 (2003).
10. Tsvetkov, P., Reuven, N. & Shaul, Y. The nanny model for IDPs. *Nat. Chem. Biol.* **5**, 778-781 (2009).
11. Cheng, M., *et al.* The p21(Cip) and p27(Kip) CDK "inhibitors" are essential activators of cyclin D-dependent kinases in murine fibroblasts. *EMBO J.* **18**, 1571-1583 (1999).
12. Haarmann, C.S., *et al.* The random-coil 'C' fragment of the dihydropyridine receptor II-III loop can activate or inhibit native skeletal ryanodine receptors. *Biochem. J.* **372**, 305-316 (2003).
13. Ma, J. Stimulatory and inhibitory functions of the R domain on CFTR chloride channel. *News Physiol. Sci.* **15**, 154-158 (2000).
14. Tompa, P., Mucsi, Z., Orosz, G. & Friedrich, P. Calpastinin subdomains A and C are activators of calpain. *J. Biol. Chem.* **277**, 9022-9026 (2002).
15. Schwiebert, E.M., *et al.* CFTR is a conductance regulator as well as a chloride channel. *Physiol. Rev.* **79**, S145-S166 (1999).
16. Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **18**, 1169-1175 (2004).
17. Garay-Arroyo, A., Colmenero-Flores, J.M., Garcarrubio, A. & Covarrubias, A.A. Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.* **275**, 5668-5674 (2000).
18. Dure, L. & Chlan, C. Developmental biochemistry of cotton seed embryogenesis and germination. XII. Purification and properties of principal storage proteins. *Plant Physiol.* **68**, 180-186 (1981).
19. Bies-Ethève, N., *et al.* Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* **67**, 107-124 (2008).
20. Zimmermann, P., Hirsch-Hoffmann, M., Henning, L. & Gruissem, W. GENEVESTIGATOR: *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* **136**, 2621-2632 (2004).
21. Olvera-Carrillo, Y., Campos, F., Reyes, J.L., Garcarrubio, A. & Covarrubias, A.A. Functional analysis of the group 4 late

-
- embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in *Arabidopsis*. *Plant Physiol.* **154**, 373-390 (2010).
22. Soulages, J.L., Kim, K., Walters, C. & Cushman, J.C. Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. *Plant Physiol.* **128**, 822-832 (2002).
23. Reyes, J.L., *et al.* Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects *in vitro*. *Plant Cell Environ.* **28**, 709-718 (2005).
24. Reyes, J.L., *et al.* Functional dissection of hydrophilins during *in vitro* freeze protection. *Plant Cell Environ.* **31**, 1781-1790 (2007).
-