



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Elementos estructurales relacionados con la capacidad protectora de una proteína LEA del grupo 4 de *Arabidopsis thaliana*

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M. en C. David Felipe Rendón Luna

TUTOR PRINCIPAL

Dra. Alejandra Alicia Covarrubias Robles
Instituto de Biotecnología, UNAM

MIEMBROS DEL COMITÉ TUTOR

Dra. Rosario Muñoz Clares
Facultad de Química, UNAM

Dr. Lorenzo Patrick Segovia Forcella
Instituto de Biotecnología, UNAM

Cuernavaca, Morelos. Junio, 2024



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



**PROTESTA UNIVERSITARIA DE INTEGRIDAD
Y HONESTIDAD ACADÉMICA Y
PROFESIONAL
(Graduación con trabajo escrito)**

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la Institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado:

**Elementos estructurales relacionados con la capacidad protectora de una proteína LEA
del grupo 4 de *Arabidopsis thaliana***

que presenté para obtener el grado de Doctorado es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi programa de posgrado, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de graduación.

Atentamente

**M. en C. David Felipe Rendón Luna
Número de Cuenta UNAM 513016532**

*«Notte e giorno faticar,
Per chi nulla sa gradir,
Piova e vento sopportar,
Mangiar male e mal dormir.»*

Aria de Leporello, de la ópera Don Giovanni, de Wolfgang Amadeus Mozart

Agradecimientos Oficiales

A la Dra. Alejandra Covarrubias Robles, por dirigir mi preparación doctoral y la elaboración de la presente tesis, desarrollada en el laboratorio 22 del Departamento de Biología Molecular de Plantas, en el Instituto de Biotecnología de la Universidad Nacional Autónoma de México.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por otorgarme una beca para mis estudios doctorales, beca con número 429013, asociada a mi CVU, número 488918. También por otorgarme apoyo del proyecto 09900 P-9900 CONACYT Fronteras 1615. Al “Programa de Apoyo a los Estudios de Posgrado” (PAEP) de la UNAM por apoyarme con financiamiento del para la asistencia a congresos.

A mi Comité Tutorial, por la dirección crítica del proyecto, por sus consejos y opiniones, por las discusiones constructivas, por las risas y los regaños. El comité estuvo constituido por:

Dra. Alejandra Covarrubias Robles Instituto de Biotecnología - UNAM

Dra. Rosario Muñoz Clares Facultad de Química - UNAM

Dr. Lorenzo Patrick Segovia Forcella Instituto de Biotecnología - UNAM

A los miembros del jurado evaluador de la tesis de doctorado, por el tiempo invertido en revisar el presente trabajo, comentarlo, engrandecerlo, pulirlo y criticarlo. El jurado estuvo constituido por:

Dr. Enrique Rudiño Piñera Instituto de Biotecnología - UNAM

Dra. Adriana Garay Arroyo Instituto de Ecología – UNAM

Dr. Enrique García Hernández Instituto de Química - UNAM

Dra. Lilian González Segura Facultad de Química – UNAM

Dr. Carlos Amero Tello Centro de Investigaciones Químicas - UAEM

Agradecimientos Generales

Carl Sagan decía: *«La ciencia es una empresa colaborativa que se extiende por generaciones. Cuando nos permite ver nuevos horizontes, recordamos a aquellos que prepararon el camino, viendo también por ellos»*. En ese sentido, la ciencia es un arte vivo a lo largo del tiempo y del espacio, construido en comunidad en beneficio de la humanidad. Por lo tanto, en ningún momento es un constructo aislado, sino que es el producto de muchas interacciones.

La presente tesis no habría sido posible sin la asesoría de la Dra. Alejandra Covarrubias Robles, quien decidió apostar por mí. Gracias por tu confianza; por tu guía; por el intercambio de ideas; por tus sugerencias experimentales; por las discusiones constructivas; por las pláticas amenas; por los regaños necesarios; por darme la oportunidad de crecer como ser humano y alcanzar un potencial que no pensaba tener.

A los miembros actuales del Consorcio Covarrubias-Reyes. A José Luis Reyes (Pepe), por sus críticas constructivas, por permitirme desahogar mis problemas, por las pláticas junto a la máquina de café. A Inti Alberto Arroyo, por tu amistad sincera; por tus valores y tu nobleza; por compartir tus conocimientos; por tu ayuda en experimentos, por «jugar con ideas» juntos como buenos científicos; por tu apoyo moral. A María Beatriz Pérez (Betty), por saber escuchar mis problemas; por preocuparte por los demás; por convertirte en una mano derecha para el funcionamiento del laboratorio. A Coral Martínez, por las risas a base de fotografías, memes y stickers; por desahogar nuestros problemas imaginando formas de destruir el sistema, por ser la «Sra. De Los Plásmidos». A Gabriela Pérez, simplemente por todas las risas; por mentarnos la madre como forma de cariño; por permitirme compartir mis conocimientos contigo. A Francisco Campos, por sus comentarios a mi proyecto en seminarios; por apoyarme con algunas de las proteínas mutantes. A Alfonso Sierra, Yuniet Hernández, David Garcias, Alberto De la Cruz, Abisai Jeréz, Luis Reyes, Teresa Nava, Dayana Martínez, Rosa Brianda, Laura Martínez, Jesús Moreno, entre otros. Todos aportaron un granito de arena en mi formación.

A personas que han pasado por el laboratorio y que han sido importantes en mi formación. A Cesar Cuevas, por ser mi mentor de laboratorio de quien aprendí muchas técnicas; por los consejos, pláticas y discusiones; por tu ayuda para preparar seminarios, conferencias y posters. A Carlos de la Rosa, por brindarme su amistad desde que llegué al laboratorio; por tantos consejos de vida; por todos los momentos divertidos rodeados de amigos y cervezas. A Omar Bringas, por permitirme compartir mi conocimiento contigo y considerarte mi primer alumnito no oficial; por las risas a la hora de la comida. A Esmeralda Ramírez, por las pláticas que tuvimos; por escucharme; por tus consejos.

A personas del IBt y otros institutos de la UNAM, campus Morelos. A Edmundo Calva (QEPD), por ser de las pocas personas con las que pude discutir acerca de arte, de la ciencia como un arte, de música, de lo que es bello, de filosofía. Nos faltó tocar juntos el Dúo de Mazas No. 2. A Luis Lozano, por las pláticas varias. A Shirley Alquicira, por incluirme en los talleres de bioinformática. A Irma Martínez, por siempre escucharme. A Filiberto Sánchez, por los whiskies. A Oswaldo López, por los mezcales y los vinos. A Verónica Jiménez y Leticia Vega, por arrastrarme al camino de la programación. A Ramsés García, por ser como un hermano mayor. A la Rucabanda, por tantas risas y por mostrar el fruto del trabajo en equipo. A los vecinos, el grupo Adrián Ochoa, por traer un poco de ser nortño al laboratorio. A Melany, a Luigi, a Juan Pablo, a Luis Morado, por su amistad.

La salud mental es esencial en la vida del ser humano. Un agradecimiento especial a mi psicóloga Lúa Castañeda, por haberme rescatado del agujero existencial en el que me encontraba cuando toqué fondo a mitad del doctorado; por guiarme a descubrir quién realmente soy; por ayudarme a controlar mis emociones en diversas situaciones; por ser alguien que me escuchó y me orientó de forma imparcial; por permitirme desahogarme cuando lo necesitaba. Si no hubieras estado cuando lo necesité, quizá yo no estaría vivo el día de hoy.

La vida te pone en el camino a grandes personas, provenientes de diversas esferas. Al Teniente Raúl Ezequiel Reyes Calderón, por ser un ejemplo de disciplina para mí, por enseñarme a no tirar la toalla, aunque la vida nos ponga pruebas complicadas. A Tonatiuh Valdez, por compartir tu conocimiento de negocios del cual yo no tengo ni la menor idea. A Abril Osorio, por tener un gran corazón para con sus alumnos y para con sus amigos. A Daniel Morales, por su amistad incondicional; por las caminatas, por los mangos.

Hay amigos, hay familia, y luego hay amigos que se convierten en familia. Quiero agradecer a dos personas que han formado parte especial de mi vida:

A José Antonio Flores Carrillo, por la amistad tan incondicional y sincera que surgió entre nosotros; por las risas y las lágrimas; por tantas aventuras que hemos vivido; por ser alguien a quien admiro por su conocimiento; por las experiencias de vida que hemos compartido; por la forma en como tu familia se convirtió en mi segunda familia; por siempre estar para una plática, unas cervezas, un regaño, una corrección, una ayuda, una compañía; por todas las enseñanzas recíprocas; por todos los buenos y malos momentos que al final construyen una verdadera y sólida amistad; por ser un gran ser humano con buenos valores, aunque no lo quieras reconocer; porque, quieras o no, nos tenemos para apoyarnos. En definitiva y simplemente, por ser el amigo más grande ypreciado que he tenido en toda mi vida. ¡Muchas gracias por todo y por lo que falta!

A Juan Ángel Olayo Miguez, en quien encontré una amistad totalmente inesperada, rodeada de coctelería, libros, videojuegos, e historias de vida; por entender lo que significa hacer un doctorado; por siempre estar ahí para desahogar mis problemas; por las pláticas que tuvimos de tantos temas; simplemente, por ser alguien que me supo querer como a un hermano.

La familia es un salvavidas en el mar de la vida. Quiero agradecer muy especialmente a mis padres. A mi mamá, María del Consuelo, por ser un pilar esencial en mi crecimiento; por nunca dejarme solo; por todo el apoyo que recibí de ti, que me permitió llegar a ser lo que soy. A mi papá, Jaime Roberto (QEPD), por las enseñanzas que me hacen ser la persona que

soy, por inculcarme el amor a los libros, por siempre querer ser alguien racional y la persona que procura llevar todo en paz, por ser el mejor maestro que he tenido. Aunque pocos lo sepan, fuiste alguien de quien aprendí muchísimo. Papá, te envidio, porque después de tantas pláticas, tú ya sabes lo que hay del otro lado del velo de esta vida. Nos vemos en el otro lado. A mi hermana Penélope, mi cuñado Luis y mis sobrinos Sebastián, Sarah y Jorge, por estar junto a mi mamá estando yo lejos; por apoyarme en mis proyectos.

A mi tío Toño y mi tía Norma, a mi tío Emilio y a mi tía Judith, a mi tía Lulu y mi tío Víctor, a mi tía Blanquita. A mi tío Álvaro (QEPD) de quien aprendí el amor por los libros. A mi tío Cesar y mi tía Nérida. A mi tía Marta. A mis primos, Judith, Luis Hugo, Vanesa, Álvaro César, Saúl, Herbey, Liz, Álvaro, Laura, Ricardo, Plácido, etc. A toda mi familia. Me doy cuenta de que la familia es lo más importante que tenemos.

A todos mis amigos de licenciatura, particularmente a Joel Arce, a Diego Estrella, a Edgardo Espinoza, a Luis Adrián Cabral, por ser personas que me apoyaron en mi crecimiento; por ayudarme a darme cuenta de quien soy; por ser amigos a quienes aprecio por siempre estar conmigo; por todas las historias que hemos vivido.

Al Daniel Rodríguez Leal por «torearme» a hacer un doctorado; por ser la primera persona que me acercó a la biología molecular y a la ciencia en general. Su no hubiera sido por ti, quizá yo no hubiera buscado una carrera científica. Eres un modelo a quien seguir.

A maestros que se han convertido en amigos. A Pedro Hernández porque siempre me permitió involucrarme en la transmisión del conocimiento. A Jorge Cid por formarme como un biólogo con aptitudes, valores y sabiduría necesaria para mi formación profesional. A Flor Millán por dejarme explotar a mi científico interior; por escucharme cuando lo necesitaba.

A todas las personas que he conocido en esta aventura llamada vida, que me han ayudado, de una u otra manera, a no darme por vencido. Yo no estaría aquí si no fuera por todos ustedes.
¡GRACIAS!

INDICE GENERAL

RESUMEN	1
ABSTRACT	3
INTRODUCCIÓN	5
El estrés hídrico en relación con el ambiente.....	5
Mecanismos de las plantas para combatir el déficit hídrico.....	5
El descubrimiento de las proteínas LEA.....	6
La importancia del desorden estructural en las proteínas.....	7
ANTECEDENTES	10
Las proteínas LEA del grupo 4 como modelos de estudio.....	10
Técnicas para el estudio de IDPs.....	12
HIPÓTESIS	14
OBJETIVOS	14
MATERIALES Y MÉTODOS.....	15
Obtención de secuencias de proteínas LEA4 y alineamiento múltiple de secuencias.	15
Construcción del árbol filogenético a partir de las secuencias de las proteínas LEA4.....	15
Selección de los residuos para realizar la mutagénesis o de las regiones a eliminar.	16
Obtención de fragmentos de DNA modificados mediante la técnica de PCR de traslape.....	16
Integración del producto de PCR modificado en el vector de expresión pTrc99a.....	17
Producción y purificación de las proteínas mutantes mediante precipitación con ácido tricloroacético (TCA).....	18
Producción y purificación de proteína mutante mediante sistema IMPACT-CN	19
Inmunodetección mediante ensayos tipo Western blot.....	20
Cuantificación de la concentración de las proteínas.....	20
Experimentos de dicroísmo circular	21

Ensayos de protección ante tratamientos de congelamiento-descongelamiento	21
Medición de la actividad enzimática de lactato deshidrogenasa (LDH)	22
Análisis de datos	22
RESULTADOS	24
Distribución filogenética de diversas proteínas LEA4.....	24
Análisis de proteínas LEA4 mutantes en residuos conservados en el N-terminal.....	27
Análisis de proteínas LEA4 truncadas.....	31
Análisis de proteínas con mutaciones que desestabilizan la formación de hélices alfa en la región N-terminal.....	35
Análisis de crioprotección utilizando diferentes relaciones molares.....	39
DISCUSIÓN	42
Las proteínas LEA4 como modelo para estudiar IDPs en plantas	42
El análisis filogenético de las proteínas LEA4 muestra una separación en los A y B después de su aparición en las briofitas.....	43
Las mutaciones en residuos conservados no afectan la estructura ni la función de la proteína AtLEA4-5	44
La eliminación de algunas zonas de la proteína AtLEA4-5 afectan negativamente su actividad protectora	45
Las mutaciones por prolinas eliminan la formación de hélices alfa pero no así la función protectora de la proteína completa.....	46
Posibles mecanismos de acción	48
LIMITACIONES DEL PROYECTO	51
Uso de LDH como enzima reportera.....	51
Los tratamientos de congelamiento-descongelamiento como una herramienta para el análisis funcional <i>in vitro</i>	52
CONCLUSIONES	54
PERSPECTIVAS	55
LITERATURA CITADA.....	57

ANEXOS64

INDICE DE FIGURAS

- Figura 1. La región N-terminal de la proteína LEA4-5 es suficiente para prevenir la inactivación de la LDH sometida a condiciones de congelamiento-descongelamiento.** Se muestra la actividad remanente de LDH después de siete ciclos de congelamiento-descongelamiento. Tomado de Cuevas-Velazquez *et al.* (2016). 11
- Figura 2. Árbol filogenético de las proteínas LEA4.** Los nombres de las proteínas sombreadas en amarillo corresponden a proteínas del subgrupo A, y las sombreadas en azul corresponden a las proteínas del subgrupo B. Las secuencias que no están sombreadas corresponden a proteínas identificadas en las briofitas. 26
- Figura 3. Alineamiento múltiple de secuencias más cercanas filogenéticamente a la proteína LEA4-5 de *A. thaliana* (AtLEA4-5).** Se muestra la ubicación de motivos conservados en el MSA de acuerdo con Battaglia *et al.* (2008). Los residuos son coloreados de acuerdo con el código de colores Clustal en Jalview, donde: rojo: residuos cargados negativamente; azul: residuos cargados positivamente; verde: residuos polares sin carga; rosa: residuos con cadena lateral hidrofóbica; magenta: glicina y prolina; naranja: tirosina y fenilalanina; amarillo; cisteína. Los asteriscos de color muestran la ubicación de los residuos que se modificaron en la proteína AtLEA4-5. Negro: LEA4-5-NT1; azul: LEA4-5-NT2; verde: LEA4-5-NT3; amarillo: LEA4-5-NT4; rojo: LEA4-5-NT9P. 27
- Figura 4. Ubicación de las mutaciones en residuos conservados.** Se indica además la identidad del residuo mutado y el residuo por el que fue mutado. 28
- Figura 5. Espectros de DC de mutantes en residuos conservados.** Se muestran los espectros obtenidos utilizando glicerol al 80% como tratamiento simulador de falta de agua. Gráfica interna: espectros diferencia obtenidos al restar los espectros a 0% de glicerol de los espectros a 80% de glicerol..... 29

Figura 6. Proporción de estructuras secundaria obtenidos mediante DichroWeb de mutantes en residuos conservados . Para cada proteína se muestra la proporción de estructura secundaria calculada por el algoritmo DichroWeb en muestras con 0 % y 80 % (v/v) de glicerol..... 30

Figura 7. Experimento de protección de LDH ante congelamiento-descongelamiento usando mutantes en residuos conservados. Se utilizaron siete ciclos de congelamiento-descongelamiento y una relación molar de 10:1 (Protector:LDH). Cada punto indica el promedio de un experimento con tres réplicas. Los datos se normalizaron considerando 100% la actividad de cada mezcla protector-LDH sin someterse a congelamiento-descongelamiento. ns: sin diferencia significativa comparado con LEA4-5. No LEAP: *No LEA protein*. ns: sin diferencia significativa comparado con LEA4-5 31

Figura 8. Ubicación de las mutaciones con eliminaciones de secciones de la región N-terminal. Se indica además la identidad del residuo mutado y el residuo por el que fue mutado..... 32

Figura 9. Espectros de DC de mutantes en residuos conservados. Se muestran los espectros obtenidos utilizando etilenglicol al 80% como tratamiento simulador de falta de agua. Gráfica interna: espectros diferencia obtenidos al restar los espectros obtenidos en presencia de 0 % de etilenglicol de los espectros colectados en presencia de 80 % (v/v) de etilenglicol. 33

Figura 10. Proporción de estructuras secundarias obtenida mediante DichroWeb de las truncadas. Para cada proteína se muestra la proporción de estructura calculada por el algoritmo DichroWeb a 0% y 80% de etilenglicol..... 34

Figura 11. Protección de LDH ante congelamiento-descongelamiento usando las mutantes truncadas. Se utilizaron siete ciclos de congelamiento-descongelamiento y una relación molar de 10:1 (Protector:LDH). Cada punto indica el promedio de un experimento con tres réplicas. Los datos se normalizaron considerando 100% la actividad de cada mezcla protector-LDH sin someterse a congelamiento-descongelamiento. No LEAP: *No LEA protein*. ns: sin diferencia significativa comparado con LEA4-5. **: Diferente significativamente con $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$ 35

Figura 12. Ubicación de las mutaciones a prolinas en la región N-terminal y mutantes truncadas en el extremo C-terminal. Se indica además la identidad del residuo mutado y el residuo por el que fue mutado.	36
Figura 13. Espectros de DC de mutantes de prolinas y mutantes truncada en el extremo C-terminal. Se muestran los espectros obtenidos utilizando etilenglicol al 80 % (v/v) como tratamiento simulador de falta de agua. Gráfica interna: espectros diferencia obtenidos al restar los espectros colectados en 0 % de etilenglicol de los espectros colectados en presencia de 80 % (v/v) de etilenglicol.....	37
Figura 14. Proporción de estructuras obtenidos mediante DichroWeb de mutantes con prolinas y mutantes truncadas en el extremo C-terminal. Para cada proteína se muestra la proporción de estructura calculada por el algoritmo DichroWeb en presencia de 0 % y 80 % (v/v) de etilenglicol.	37
Figura 15. Experimento de protección de LDH ante congelamiento-descongelamiento de mutantes con prolinas y mutantes truncadas en el extremo C-terminal Se utilizaron siete ciclos de congelamiento-descongelamiento y una relación molar de 10:1 (Protector:LDH). Cada punto indica el promedio de un experimento con tres réplicas. Los datos se normalizaron considerando 100% la actividad de cada mezcla protector-LDH sin someterse a congelamiento-descongelamiento. No LEAP: <i>No LEA protein</i> . ns: sin diferencia significativa comparado con LEA4-5. ****: Diferente significativamente con $p < 0.0001$	38
Figura 16. Experimento de protección de LDH ante congelamiento-descongelamiento de mutantes por prolinas y mutantes truncadas en el extremo C-terminal. Se incluyó a la lisozima como control negativo de protección. Las medias de protección a diferentes relaciones molares se usaron para realizar un ajuste de los datos a una curva sigmoideal.....	40
Figura 17. Experimento de protección de LDH ante congelamiento-descongelamiento de mutantes a prolinas y mutantes truncadas en el extremo C-terminal. El valor de MR_{50} se obtuvo a partir de los datos de ajuste a curva sigmoideal de la Figura.	41

ABREVIATURAS y ACRÓNIMOS

ABA: *Abscic Acid* (Acido abscísico)

CD: *Circular Dichroism* (Dicroísmo Circular)

EG: *Ethylene Glycol* (Etilenglicol)

IDP: *Intrinsically Disordered Protein* (Proteína Intrínsecamente Desordenada)

LEA: *Late Embryogenesis Abundant* (Abundante en la Embriogénesis Tardía)

PFAM: *Protein Family* (Familia de proteínas)

FELLS: *Fast Estimator of Latent Local Structure* (Estimador Rápido de Estructura Local Latente)

MUSCLE: *Multiple Sequence Comparison by Log-Expectation* (Comparación Múltiple de Secuencias por Expectativa Logarítmica)

PCR: *Polymerase Chain Reaction* (Reacción en cadena de la Polimerasa)

ORF: *Open Reading Frame* (Marco Abierto de Lectura)

MWCO: *Molecular Weight CutOff* (Límite de peso molecular)

IMPACT-CN: *Intein Mediated Purification with an Affinity Chitin-binding Tag at the C- or N-terminal* (Purificación mediada por inteína con una etiqueta de afinidad de unión a quitina en los C- o N-terminales)

BLAST: *Basic Local Alignment Search Tool* (Herramienta básica de búsqueda de alineamientos locales).

FRET: *Förster Resonance Energy Transfer* (Transferencia de energía de resonancia de Förster)

MSA: *Multiple Sequence Alignment* (Alineamiento múltiple de secuencias)

BSA: *Bovine Serum Albumin* (Albumina de suero bovino)

CD-HIT: *Cluster Database at High Identity with Tolerance* (Base de datos de grupos con alta identidad con tolerancia.)

MR50: *Molar Ratio for 50% protection* (Relación molar para el 50% de protección)

RESUMEN

Las proteínas Abundantes en la Embriogénesis Tardía (LEA, por las siglas del inglés *Late Embryogenesis Abundant*) son hidrofílicas y proteínas intrínsecamente desordenadas (IDPs, por las siglas del inglés *Intrinsically Disordered Proteins*) que se acumulan durante la desecación de la semilla y en tejidos vegetativos bajo déficit hídrico. Se clasifican en siete grupos de acuerdo con los motivos conservados en su secuencia. El grupo 4 (LEA4) presenta una región N-terminal conservada y una región C-terminal variable en composición y longitud. En *Arabidopsis thaliana*, las proteínas LEA4 son abundantes en semillas secas, y en tejidos vegetativos y botones florales en respuesta a tratamiento con sal y manitol. Estas respuestas parecen estar moduladas por ácido abscísico (ABA), ya que tratamientos con esta fitohormona también inducen el aumento en la abundancia de su transcrito y proteína. La falta de estas proteínas disminuye la capacidad de germinación en condiciones de estrés iónico y osmótico, así como la recuperación de plantas sometidas a sequía, demostrando estas proteínas se requieren para una respuesta de ajuste óptima ante de las plantas ante condiciones adversas. Estas proteínas muestran actividad protectora *in vitro* sobre enzimas sometidas a congelamiento-descongelamiento y deshidratación gradual. El que estas proteínas adquieren estructuras con hélices alfa en su N-terminal durante los tratamientos *in vitro* indica el impacto del ambiente que las rodea sobre la selección de sus ensamblajes conformacionales. A la fecha, hay poca evidencia directa de que la aparición de este tipo de estructura secundaria esté relacionada con su capacidad protectora *in vitro*. En el presente estudio, se utilizó a la proteína LEA4-5 de *A. thaliana* (AtLEA4-5) para determinar si la adquisición de estructura secundaria es indispensable para su función protectora *in vitro*. Con el fin de determinar la participación de ciertos residuos y de diferentes regiones de la proteína con respecto a la aparición de hélices alfa, se generaron mutantes de la proteína AtLEA4-5 para modificar residuos conservados, para eliminar algunas regiones de la proteína, o para bloquear la formación de hélices alfa. Asimismo, este enfoque también permitió evaluar su posible correlación con la actividad protectora causada por el déficit hídrico y por cambios en la temperatura. Los resultados

mostraron que la eliminación de residuos conservados o de regiones de la proteína no afectaron la función de ésta. De interés fue el hecho de que la eliminación del extremo C-terminal desordenado provocó una disminución en la actividad protectora. Sorprendentemente, el bloqueo en la formación de la estructura secundaria no afectó la función protectora en la proteína completa. Sin embargo, esta incapacidad de adquirir hélices alfa impactó negativamente la actividad protectora de una versión corta de la proteína AtLEA4-5, carente del extremo C-terminal. Este trabajo demostró que se requiere la formación de estructura secundaria y la presencia del C-terminal para que la proteína pueda ejercer una función protectora óptima. Además, se demostró que esta proteína puede presentar conformaciones alternativas con función protectora, evidenciando la plasticidad funcional y conformacional de esta proteína LEA. Esta tesis también contribuye con información que indica que ciertos cambios conformacionales en una proteína pueden estar asociados a una misma función, evidenciando la plasticidad de la relación estructura-función en las proteínas intrínsecamente desordenadas.

ABSTRACT

Late Embryogenesis Abundant (LEA) proteins are hydrophilins and intrinsically disordered proteins (IDPs) that accumulate during seed desiccation and in vegetative tissue under conditions of water deficit. These proteins are classified into seven groups according to conserved motifs in their sequence. Group 4 (LEA4) proteins present a conserved N-terminal region and a C-terminal region variable in composition and length. In *Arabidopsis thaliana*, LEA4 proteins are abundant in dry seeds, vegetative tissue and flower buds during salt and mannitol treatments. These responses appear to be modulated by abscisic acid (ABA), because treatments with this phytohormone also induced the increase in the abundance of its transcript and protein. Furthermore, the lack of these proteins decreases germination under conditions of ionic and osmotic stress, as well as the recovery of plants subjected to drought, showing that these proteins are required for an optimal adjustment response of the plant to adverse conditions. LEA4 proteins show protective activity *in vitro* on enzymes subjected to freezing and dehydration. The fact that these proteins acquire structures with alpha helices at their N-terminus during *in vitro* treatments indicates the impact of the environment that surrounds them on the selection of their conformational assemblies. To this date, there is little direct evidence that the appearance of this type of secondary structure is related to its protective capacity *in vitro*. In the present study, the *A. thaliana* LEA4-5 protein (AtLEA4-5) was used to determine whether the acquisition of secondary structure is essential for its protective function *in vitro*. To determine the participation of certain residues and different regions of the protein with respect to the appearance of alpha helices, mutants of the AtLEA4-5 protein were generated to modify conserved residues, to eliminate some regions of the protein, or to block the formation of alpha helices. Likewise, this approach also allowed us to evaluate its correlation with the protective activity caused by water deficit and changes in temperature. The results showed that the removal of conserved residues or regions of the protein did not affect its function. Of interest was the fact that removal of the disordered C-terminus caused a decrease in protective activity. Surprisingly, blocking secondary structure formation did not affect protective function in the full-length protein. However, this inability to acquire alpha helices negatively impacted the protective

activity of a short version of the AtLEA4-5 protein, lacking the C-terminus. This work demonstrated that the formation of secondary structure and the presence of the C-terminus are required for the protein to exert an optimal protective function. Furthermore, it was demonstrated that this protein can present alternative conformations with a protective function, evidencing the functional and conformational plasticity of this LEA protein. This thesis also contributes with information that indicates that certain conformational changes in a protein can be associated with the same function, evidencing the plasticity of the structure-function relationship in intrinsically disordered proteins.

.

INTRODUCCIÓN

El estrés hídrico en relación con el ambiente

Las plantas son organismos sésiles, que viven anclados al suelo y, por ende, no pueden escapar de condiciones ambientales o factores biológicos que les sean perjudiciales (Wu *et al.*, 2007). Uno de los factores ambientales más perjudiciales para el desarrollo de las plantas es la disminución en la disponibilidad de agua o déficit hídrico. Dicho estrés puede deberse a diversos factores: a la disminución en el potencial hídrico del ambiente debido a la reducción de la disponibilidad de las moléculas de agua por evaporación, o por su retención debido a una elevada concentración de sales u otros solutos, fenómeno que ocurre principalmente en el suelo (Carillo *et al.*, 2011). La disponibilidad del agua también dependerá del estado físico en el que se encuentren las moléculas de agua, donde una mayor viscosidad o la solidificación del agua provocarán una menor disponibilidad de esta molécula en el medio. Este fenómeno se presenta cuando ocurre una disminución considerable en la temperatura (Augé, 2001). Por lo tanto, las propiedades intrínsecas del agua pueden variar en función de las condiciones ambientales, donde, por ejemplo, las bajas temperaturas se relacionan al congelamiento y las altas temperaturas se asocian a la sequía.

Mecanismos de las plantas para combatir el déficit hídrico

Para hacer frente a este tipo de estrés, las plantas han desarrollado diversos mecanismos que les permiten resistir estas condiciones. Puesto que la sequía impone una restricción de agua en las plantas, las respuestas de las plantas relacionadas con esta condición de estrés se han clasificado en: evasión, tolerancia, y recuperación de la sequía (Yordanov *et al.*, 2000; Gaspar *et al.*, 2002; Fang *et al.*, 2015). La tolerancia a la sequía requiere que las plantas establezcan una serie de respuestas que les permitan ajustar sus procesos fisiológicos fundamentales, como son el crecimiento, el desarrollo o el metabolismo, a las condiciones adversas en las que se encuentren (Buchanan *et al.*, 2015). Por tal motivo, el conocer los mecanismos bioquímicos y moleculares involucrados en estas respuestas resultan de relevancia, no sólo para entenderlos desde el punto de vista biológico,

ecológico y evolutivo, sino, de ser posible, para mejorarlos. Asimismo, resulta factible generar herramientas que permitan diseñar estrategias de selección en programas de mejoramiento de cultivos y de condiciones de siembra, para mantener o incrementar el rendimiento de cultivos en condiciones poco favorables.

La información relacionada con diferentes aspectos de las respuestas de las plantas a condiciones ambientales adversas ha crecido enormemente en los últimos años. Ahora conocemos diferentes aspectos de estos procesos a nivel genético, celular, fisiológico, metabólico y molecular. En general, los cambios en el ambiente provocan modificaciones en la expresión de diversos genes en las plantas. Dichos cambios alteran los niveles de transcritos de diversas enzimas, tales como enzimas desintoxicantes, cinasas, fosfatasas, catalizadoras de la síntesis de osmolitos y de hormonas, acuaporinas, transportadores de azúcares, factores transcripcionales, entre muchas otras. Cabe resaltar también el aumento en la abundancia de proteínas que realizan un papel protector sobre otras proteínas o enzimas como son las chaperonas (Vashisht *et al.*, 2006; Jones *et al.*, 2013).

El descubrimiento de las proteínas LEA

Durante el estudio de proteínas de almacenamiento de semillas de *Gossypium hirsutum* L., se identificaron proteínas que se acumulan al final de la embriogénesis y que disminuyen durante la germinación (L. Dure *et al.*, 1981). Posteriormente, se determinó que dichas proteínas también se acumulan en ciertos tejidos expuestos a condiciones de deshidratación, por lo que se les atribuyó un papel para confrontar condiciones de limitación de agua (L. Dure, 3rd *et al.*, 1981; Baker *et al.*, 1988). A dichas proteínas se les denominó como *Late Embryogenesis Abundant (LEA) proteins* (proteínas Abundantes en la Embriogénesis Tardía) (Galau *et al.*, 1981).

La definición de LEA *stricto sensu* implica aquel conjunto de proteínas acumuladas en la embriogénesis tardía y en diferentes tejidos vegetativos de las plantas al someterse a condiciones de

déficit hídrico. Aún con esta definición, no existe, a nivel de estructura primaria, una secuencia consenso que permita identificar a una proteína como una proteína LEA. Sin embargo, diversos esfuerzos se han realizado para poder clasificarlas de acuerdo con diversos motivos que se pueden detectar en sus secuencias (Dure *et al.*, 1989; Bray, 1993; Battaglia *et al.*, 2008; Bies-Etheve *et al.*, 2008; Finn *et al.*, 2016). Para esta tesis se utilizará la clasificación propuesta por Battaglia *et al.* (2008).

Aunque no existe un motivo común en su secuencia primaria que identifique a las proteínas LEA en general, diversos estudios han permitido identificar algunas características peculiares de la mayoría de las proteínas LEA: son relativamente pequeñas, se acumulan durante la formación de las semillas y en diversos tejidos somáticos sometidos a estrés hídrico, y presentan una composición sesgada de aminoácidos, siendo abundantes en residuos polares cargados y no cargados, es decir residuos hidrofílicos, son escasas en residuos hidrofóbicos, y en general ausentes de residuos de fenilalanina, triptófano y cisteína (Battaglia *et al.*, 2008). Además, derivado de estas características, se pueden considerar a la mayoría de las proteínas LEA como proteínas intrínsecamente desordenadas (IDPs); es decir, proteínas que no tienen una estructura tridimensional estable y aun así son funcionales (Dunker *et al.*, 2008; Tompa *et al.*, 2009; Hinch *et al.*, 2012). Más aún, la mayoría de las proteínas LEA no solamente pueden clasificarse como IDPs, sino también como hidrofilinas, las cuales fueron descritas como un subconjunto de proteínas que se caracterizan, *lato sensu*, por tener un alto contenido de aminoácidos hidrofílicos, dando lugar a valores de hidropatía mayores de uno, así como y un alto porcentaje de residuos de glicina (superior al 6% de la composición total de aminoácidos) (Garay-Arroyo *et al.*, 2000)

La importancia del desorden estructural en las proteínas

La comprensión de la estructura de una proteína, y más específicamente, de su estructura tridimensional, ha sido un pilar fundamental de la biología moderna. Desde los primeros estudios

de estructuras tridimensionales de proteínas utilizando cristalografía de rayos X, se pensó que una función proteica particular debería relacionarse con una estructura tridimensional única que permitiera realizarla, lo cual, si bien es cierto para muchos ejemplos, no es una generalización absoluta. Aquellas regiones cuya estructura tridimensional no podía resolverse se consideraron «no estructuradas» y por tanto «no funcionales» (Uversky, 2013a). Con el paso del tiempo, la evidencia de la presencia de dichas regiones flexibles en diversas proteínas se incrementó y surgió la idea de que dichas regiones sin estructura podrían tener algún rol funcional (Dunker *et al.*, 2001). No fue sino hasta inicio del siglo XXI que fue inevitable reconocer la importancia del estudio de las regiones desordenadas, lo cual trajo una oleada de estudios respecto al desorden estructural y su relevancia para el funcionamiento de diversas proteínas (Chouard, 2011; Uversky, 2011).

El desorden estructural en proteínas puede entenderse como la capacidad de proteínas completas, o de regiones de estas, de permanecer en un estado flexible al no contar con un mínimo particular en el paisaje energético que determina la estructura de la proteína, en una conformación dada. Así entendido, las proteínas desordenadas cambian constantemente de estructura entre conformaciones poco estables y energéticamente equivalentes. Dicho paisaje energético puede modificarse debido a condiciones ambientales, a la presencia de proteínas blanco, o a modificaciones postraduccionales. Estas condiciones pueden provocar la aparición de embudos energéticos donde la proteína desordenada puede adquirir una estructura transitoria estable, la cual puede ser perturbada y regresar a estados desordenados cuando las condiciones ambientales vuelvan a cambiar. Esta plasticidad conformacional se ha relacionado a la capacidad de realizar más de una función debido a esta plasticidad estructural, lo cual se ha denominado como actividad “*moonlight*” (Jeffery, 2017).

Hay algunos ejemplos clásicos de proteínas donde se ha demostrado que su desorden estructural le permite realizar una o varias funciones. Tal es el caso de la fosfo proteína nuclear p53 de humanos,

a la cual se le ha denominada la “guardiana del genoma”. La proteína p53 funciona como un homo-tetrámero, donde cada monómero tiene seis dominios. De ellos, el dominio ubicado en la región N-terminal de p53 es intrínsecamente desordenada y puede interactuar con diferentes proteínas blanco para ejercer diferentes efectos, como son la regulación transcripcional o la detección y reparación del DNA dañado por diversos eventos (Burger *et al.*, 2014; Uversky, 2016).

Para el caso de las proteínas de interés con un origen vegetal, las proteínas LEA han demostrado ser buenos candidatos para comprender la importancia que el desorden estructural o la plasticidad conformacional tienen en plantas sometidas a condiciones ambientales adversas (Olvera-Carrillo *et al.*, 2011).

ANTECEDENTES

Las proteínas LEA del grupo 4 como modelos de estudio.

En *Arabidopsis thaliana* (de ahora en adelante *A. thaliana*), existen 51 genes que codifican a proteínas LEA (Hundertmark *et al.*, 2008), las cuales pueden agruparse en 7 grupos (Battaglia *et al.*, 2008). Las proteínas del grupo 4 se caracterizan por una región N-terminal altamente conservada, la cual se predice que forma una estructura tipo hélice alfa, y una región C-terminal desordenada, de composición y longitud variable (Battaglia *et al.*, 2008; Cuevas-Velazquez *et al.*, 2016). En *A. thaliana*, el grupo LEA4 cuenta con tres miembros, denominados AtLEA4-1, AtLEA4-2 y AtLEA4-5, cuyo sufijo denota el cromosoma en el que se encuentran los respectivos locus. Al tener pocos miembros, el grupo 4 de proteínas LEA en *A. thaliana* resulta útil para realizar estudios genéticos y comprender la función de estas proteínas en las plantas.

Algunos análisis fenotípicos de mutantes afectadas en la producción de estas proteínas mostraron sensibilidad durante la germinación en condiciones de estrés osmótico (350 mM manitol) e iónico (250 mM NaCl), en comparación con las semillas de plantas silvestres. Además, plantas adultas de estas mutantes de pérdida de función, mostraron ser más sensibles a condiciones de deshidratación, así como tener un menor peso seco y menos botones florales (Olvera-Carrillo *et al.*, 2010).

Por otro lado, en el estudio realizado por Cuevas-Velazquez *et al.* (2016) se determinó que la proteína AtLEA4-5 muestra una actividad protectora sobre la enzima lactato deshidrogenasa (LDH) sometida a tratamientos de congelamiento-descongelamiento y de deshidratación. Más aún, se determinó que la región N-terminal de dicha proteína (LEA4-5₁₋₇₇) es suficiente para ejercer la protección sobre esta enzima sometida a tratamientos de congelamiento-descongelamiento. Además, mediante el uso de dicroísmo circular (DC), se observó que esta región adquiere una estructura secundaria tipo hélice alfa en condiciones que simulan

amontonamiento molecular o falta de agua. Por el contrario, se determinó que la región C-terminal de la proteína LEA4-5 (LEA4-5₇₈₋₁₅₈) no adquiere estructura secundaria en condiciones de amontonamiento molecular o falta de agua y no protege la actividad de la enzima LDH en solución sometida a ciclos de congelamiento-descongelamiento (Figura 1).

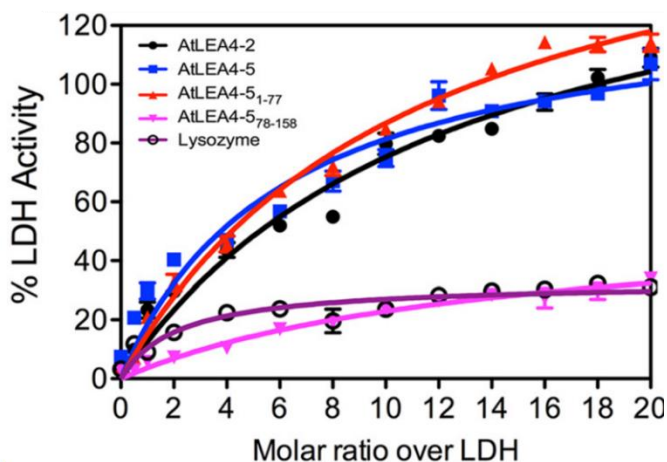


Figura 1. La región N-terminal de la proteína LEA4-5 es suficiente para prevenir la inactivación de la LDH sometida a condiciones de congelamiento-descongelamiento. Se muestra la actividad remanente de LDH después de siete ciclos de congelamiento-descongelamiento. Tomado de Cuevas-Velazquez *et al.* (2016).

A partir de los resultados anteriores, se concluyó que la función protectora de la proteína AtLEA4-5 era provocada por la adquisición de cierto tipo de estructura en la región N-terminal en condiciones de déficit hídrico, región que posee los motivos más conservados de esta familia de proteínas. Sin embargo, debido a la naturaleza misma de los experimentos de DC y de protección enzimática *in vitro* a ciclos de congelamiento-descongelamiento y deshidratación, los datos no demuestran de forma precisa el papel que podrían desempeñar las hélices alfa y/o los motivos conservados de la región N-terminal en la actividad protectora de la proteína AtLEA4-5.

Para establecer una relación más estrecha entre la capacidad de las proteínas LEA del grupo 4 de adquirir estructura secundaria en condiciones de déficit hídrico y su funcionalidad como protectoras de otras proteínas en esas condiciones, en la presente tesis se buscó mutar residuos o regiones específicas de la proteína modelo LEA4-5 de *A. thaliana*. Esta idea se basó en que la

modificación de la proteína modelo tendrá un impacto detectable en la capacidad de estructurarse y de proteger en condiciones de déficit hídrico *in vitro*, el cual permitirá tener una idea más clara del mecanismo de acción de estas proteínas.

Técnicas para el estudio de IDPs

La estrategia anteriormente planteada tiene como fundamento el hecho de que es posible identificar, o al menos proponer, la función de una proteína modificando residuos involucrados en las interacciones que lleva a cabo la proteína con su ligando (Shortle, 1992). Si bien la naturaleza desordenada de las IDPs dificulta identificar residuos particulares responsables de dichas interacciones, la conservación de regiones apoya la noción de que existen características que la evolución ha conservado, muy probablemente por ser funcionales. Para el caso de proteínas LEA4, esa región de interés corresponde principalmente la región N-terminal.

Una vez identificada una región de interés, es común utilizar una serie de técnicas que permiten estudiar la naturaleza conformacional de las IDPs, técnicas que incluyen: resonancia magnética nuclear (NMR), resonancia paramagnética del electrón (EPR), dicroísmo circular (DC) y fluorescencia por transferencia de energía de resonancia (FRET) en una sola molécula (Na *et al.*, 2018). También es posible utilizar aproximaciones computacionales como cálculos de dinámica molecular (Duong *et al.*, 2018), análisis filogénicos (Riley *et al.*, 2023), identificación de características de reconocimiento molecular (Sharma *et al.*, 2018), y predicción de desorden estructural (Y. Liu *et al.*, 2019).

De estas técnicas, el dicroísmo circular (DC) es una de las más utilizadas para el estudio de IDPs (Woody, 2010; Chemes *et al.*, 2012) y, específicamente, para el estudio de proteínas LEA (Shih *et al.*, 2010; Navarro-Retamal *et al.*, 2016; Koubaa *et al.*, 2019; LeBlanc *et al.*, 2021). Dicha técnica se basa en la absorción diferencial de luz polarizada circularmente por diferentes compuestos

químicos. Dicha absorción, en el caso de proteínas, esta mediada por diferentes elementos estructurales que actúan como cromóforos, como pueden ser el enlace peptídico, las cadenas aromáticas, los puentes disulfuro, la presencia de flavinas o protoporfirinas, etc. (Kelly *et al.*, 2005). También los diferentes tipos de estructura secundaria presente en proteínas pueden actuar como cromóforos, generando espectros particulares que pueden ser utilizados para determinar los tipos y proporciones de estructuras secundarias presentes en una proteína (Greenfield, 2006). En el presente trabajo se optó por el uso de esta técnica debido a su facilidad de aplicación y al tipo de información que se puede obtener de la misma.

HIPÓTESIS

Las mutaciones en la región N-terminal de la proteína LEA4-5 de *A. thaliana* que modifiquen su estructura secundaria y/o propiedades fisicoquímicas afectarán la capacidad protectora que esta proteína ejerce sobre enzimas reporteras sometidas a condiciones de baja disponibilidad de agua *in vitro*.

OBJETIVOS

- **GENERAL**
 - Identificar los aminoácidos y/o las regiones de la proteína LEA4-5 de *A. thaliana* necesarios para que se lleve a cabo su función protectora sobre enzimas reporteras sometidas a condiciones de baja disponibilidad de agua *in vitro*.
- **ESPECÍFICOS**
 - Identificar aminoácidos candidatos para mutagénesis usando alineamientos múltiples de secuencias de las proteínas LEA del grupo 4 de diversos grupos de plantas.
 - Generar versiones mutantes del gen *AtLEA4-5* donde se modifiquen aminoácidos conservados de la proteína.
 - Generar versiones mutantes del gen *AtLEA4-5* donde se modifiquen aminoácidos poco conservados pero cercanos a regiones propensas a formar hélices alfa.
 - Purificar, verificar la integridad y cuantificar las proteínas mutantes obtenidas.
 - Evaluar la capacidad de las proteínas mutantes generadas para proteger a enzimas reporteras en ensayos de congelamiento–descongelamiento *in vitro*.
 - Evaluar la capacidad de las proteínas mutantes generadas para adquirir estructura secundaria en condiciones que simulen falta de agua mediante estudios de dicroísmo circular.

MATERIALES Y MÉTODOS

Obtención de secuencias de proteínas LEA4 y alineamiento múltiple de secuencias.

Se realizó un BLAST con la secuencia de la proteína AtLEA4-5 contra la base de datos no redundantes de NCBI, acotando el análisis al taxón *Viridiplatae*. BLAST detectó el superdominio de PFAM LEA_1 en el extremo N-terminal de la secuencia de las proteínas encontradas. La secuencia consenso de dicho superdominio se utilizó como anzuelo para detectar proteínas reportadas en la base de datos Phytozome v11 (Goodstein *et al.*, 2012). Se limpiaron manualmente las secuencias, eliminando aquellas con caracteres que no representan algunos de los 20 residuos de aminoácidos típicos presentes en proteínas, así como proteínas que tuvieran un sitio alternativo de inicio de traducción y que estuvieran extendidas hacia el extremo N-terminal. Posteriormente, para facilitar el procesamiento de las secuencias, se eliminaron secuencias redundantes utilizando el programa CD-HIT (Li *et al.*, 2006; Fu *et al.*, 2012), el cual agrupa las secuencias con una similitud mayor al 95% y selecciona una de ellas como la representante del grupo. Las secuencias obtenidas se alinearon utilizando el programa MUSCLE v3.8.31 (Edgar, 2004).

Construcción del árbol filogenético a partir de las secuencias de las proteínas LEA4.

Se utilizó el programa Modeltest Next Generation v.0.1.3 (Darriba *et al.*, 2020) con el archivo de alineamiento múltiple de secuencias obtenido por MUSCLE para determinar el mejor modelo evolutivo de las secuencias. Una vez determinado el mejor modelo evolutivo, se realizó un análisis filogenético utilizando el programa RAxML v.8.2.10 (Stamatakis, 2006), aplicando una metodología para máxima verosimilitud, repitiendo el análisis utilizando 1000 semillas aleatorias, y llevando a cabo un análisis de bootstrap con 200 remuestreos. El mejor árbol fue visualizado y anotado mediante la paquetería “ggtree” de R.

Selección de los residuos para realizar la mutagénesis o de las regiones a eliminar.

El alineamiento múltiple obtenido de MUSCLE se visualizó utilizando el programa JalView. Los aminoácidos candidatos para ser mutados en la proteína AtLEA4-5 se seleccionaron considerando tres características: 1) su alto grado conservación, por identidad o similitud, los cuales se cambiarían por alaninas, evitando alterar en la medida de lo posible las estructuras tipo hélice alfa, 2) las regiones en los extremos de la proteína, así como algunas regiones internas propensas a la adquisición de hélices alfa, las cuales se eliminaron, o 3) su cercanía a regiones con tendencia de formar hélice alfa para cambiarlos por prolinas, interrumpiendo la formación de este tipo de estructura secundaria. Para predecir la tendencia de las proteínas mutantes para formar hélices alfa, se utilizó el programa FIELDS (Piovesan *et al.*, 2017) (Anexo VI), mientras que la presencia de regiones desordenadas se calculó mediante Metapredict v2.2 (Emenecker *et al.*, 2021) (Anexo VII). Los datos obtenidos se graficaron con GraphPad Prism v8.0.2 para Windows (GraphPad Software, Boston, Massachusetts, USA, www.graphpad.com).

Obtención de fragmentos de DNA modificados mediante la técnica de PCR de traslape

Después de seleccionar los aminoácidos a mutar, se diseñaron oligonucleótidos con los codones modificados correspondientes a los aminoácidos respectivos y se extendieron entre 10-15 bases a partir del codón cambiado, tanto hacia el 5' y como hacia el 3' del gen. Se procuró realizar las mutaciones considerando los codones más utilizados por *Escherichia coli* K12. El oligonucleótido «reverse» correspondió al reverso complementario exacto del oligonucleótido «forward» mutado. Los oligonucleótidos utilizados se muestran en Anexo I.

Posteriormente, se siguió la técnica de PCR de traslape (Horton *et al.*, 1990). Brevemente, se utilizaron dos pares de oligonucleótidos; el primero, que amplificó desde el ATG correspondiente al inicio de la traducción, hasta la zona donde se encuentran las mutaciones; el segundo, desde la zona de las mutaciones hasta el final del gen. Cabe señalar que a los oligonucleótidos del inicio y

final del gen se les agregó la secuencia de reconocimiento de las enzimas de restricción *NcoI* (5'-CCATGG-3') y *SaII* (5'-GTCGAC-3'), respectivamente, para que los amplicones se pudieran integrar posteriormente al vector de expresión pTrc99a. Como templado de DNA se usó el marco abierto de lectura (ORF) de la proteína AtLEA4-5 con la mutación M33L, que elimina una metionina interna, que puede ser utilizada por *E. coli* como sitio alternativo para iniciar la traducción. Primero, se realizaron las PCRs necesarias para obtener los dos fragmentos del gen, los cuales se separaron por electroforesis en agarosa al 1%. Se purificaron las bandas obtenidas mediante el uso del kit *GeneJET Gel Extraction Kit* (# cat. K0691, Thermo Scientific). Posteriormente, se realizó una PCR utilizando los oligonucleótidos «Forward» y «Reverse» del gen completo, agregando 2 µL de cada uno de los segmentos previamente purificados. De esta manera, durante la primera ronda de termociclaje, los segmentos se unieron por complementariedad y la polimerasa reconstituyó el gen, permitiendo posteriores amplificaciones del gen completo mutante. Los pares de oligonucleótidos para la obtención de los fragmentos usados en la PCR de traslape se muestran en Anexo II.

Para la producción de los amplicones de genes truncados se realizó una PCR de punto final con los oligonucleótidos apropiados para eliminar los extremos deseados. Para el caso de las mutantes con eliminación de secuencias internas, se siguió una estrategia similar a la PCR de traslape, pero con oligonucleótidos que compartieran regiones en común y que fueran dichas regiones las que se traslaparan. La combinación de oligonucleótidos utilizados para la generación de estas mutantes se muestra en el Anexo III.

Integración del producto de PCR modificado en el vector de expresión pTrc99a

Una vez realizado la PCR de traslape, se purificó la banda correspondiente al amplicón deseado y se subclonó en el vector pJET1.2/blunt (*Fermentas CloneJET PCR Cloning Kit* #K1231). Se transformaron células *E. coli* DH5α con la construcción y se seleccionaron colonias positivas en medio LB adicionado con ampicilina (100 µg/mL). Se purificó el plásmido mediante el uso del kit

GeneJET Plasmid Miniprep (# cat. K0502, Thermo Fischer Scientific) y se mandó a secuenciar el gen de interés. Se realizaron los alineamientos correspondientes para verificar la presencia de las mutaciones en las secuencias de los genes. Posteriormente, se realizó una reacción de restricción con enzimas *NcoI* y *SalI* para liberar el gen de interés del vector pJET. Se corrió la restricción en gel de agarosa al 1% y se purificó la banda correspondiente al gen. El producto obtenido y cuantificado mediante el uso de un Nanodrop 1000 (Thermo Fischer Scientific) se ligó usando el protocolo de T4 ligasa (#cat. EL0014, Thermo Fischer Scientific) al vector pTrc99a, previamente digerido con las mismas enzimas de restricción y purificado. La reacción de ligación se utilizó para transformar células *E. coli* DH5α y se seleccionaron colonias positivas en medio LB adicionado con ampicilina (100 µg/mL). Se purificó el plásmido de algunas colonias, se verificó su correcta integración mediante secuenciación, y se utilizaron los plásmidos correctamente construidos para transformar células *E. coli* XL1-Blue.

Producción y purificación de las proteínas mutantes mediante precipitación con ácido tricloroacético (TCA)

Se tomaron como referencia los protocolos de purificación de proteínas básicas y desordenadas de Campos *et al.* (2011) y Romero-Perez *et al.* (2023), a los cuales se les hicieron algunas modificaciones. Brevemente, se preparó un inóculo de 100 mL de medio líquido LB, adicionado con ampicilina (100 µg/mL), con la cepa de *E. coli* XL1-Blue transformada con la construcción que tiene la proteína ser expresada. Se incubó durante 16 horas a 37°C con agitación constante. Posteriormente, se inoculó un matraz Fernbach con 1 L de medio líquido LB modificado (10 g de triptona, 10 g de extracto de levadura y 10 g de NaCl, por litro). Se incubó a 37 °C por dos horas con agitación constante (200 rpm) y se indujo la producción de la proteína de interés mediante la adición de IPTG 1 mM, concentración final, por otras dos horas. Se centrifugó el cultivo a 7,000 RCF a 4 °C por 10 minutos. La pastilla se resuspendió en solución amortiguadora de Tris 20 mM pH 8.0 con NaCl 10 mM. Se lisaron las células en baño María por 10 minutos y el lisado se incubó en hielo por 10 minutos. Se centrifugó el lisado a 12,100 RCF a 4 °C por 10 minutos. El

sobrenadante se distribuyó en tubos Eppendorf de 2 mL (aproximadamente 1.8 mL cada uno) y se añadió TCA al 100% hasta una concentración final de 2 %. Se agitó con vórtex y se incubó en hielo por 15 minutos. Se centrifugó de nuevo a 12,100 RCF a 4 °C por 10 minutos, y el sobrenadante se distribuyó en tubos Eppendorf nuevos donde se agregó TCA al 100 % hasta una concentración final de 8%. Se agitó con ayuda de un vórtex y se incubó en hielo por 15 minutos. En esta ocasión, la pastilla del precipitado se lavó dos veces con acetona fría al 100%, se secó al aire libre y se resuspendió en aproximadamente 100-500 µL de solución amortiguadora de elección, aunque en la mayoría de las veces se utilizó agua destilada.

Producción y purificación de proteína mutante mediante sistema IMPACT-CN

Las proteínas denominadas LEA4-5₇₈₋₁₅₈ y LEA4-5_{1-77(9P)} tuvieron rendimientos de purificación muy bajos al ser purificadas por TCA. Para ellas, se optó por una estrategia distinta que fue la purificación por el sistema IMPACT-CN (New England Biolabs), que se basa en la fusión de una proteína de interés a una inteína y a un dominio de afinidad a quitina para una purificación mediante cromatografía a través de una matriz de quitina. Brevemente, los genes para LEA4-5₇₈₋₁₅₈ y LEA4-5_{1-77(9P)} se introdujeron en el vector pTYB11 mediante restricción enzimática, en esta ocasión cortando con *SapI* y *SmaI* en los extremos 5' y 3'. Se realizó una ligación con T4 ligasa, se verificó la inserción por secuenciación y se transformaron células de *E. coli* XL1-Blue. Se inoculó un matraz de 100 mL con medio líquido LB adicionado con ampicilina (100 µg/mL) y se incubó toda la noche. Dicho cultivo se utilizó para inocular 1L de medio líquido LB modificado como previamente se describió. Después de incubar a 37 °C por dos horas, se realizó la inducción de las proteínas de interés adicionando IPTG 1 mM, concentración final, por tres horas, y las células se colectaron por centrifugación a 7,000 RCF, 4 °C por 10 minutos. Posteriormente, se siguió el protocolo de acuerdo con las especificaciones del fabricante. Se sabe que la proteína LEA4-5₇₈₋₁₅₈ no se puede detectar por tinción con Azul de Coomassie, por lo que se realizó un experimento tipo *Western blot* para detectar las fracciones que contuvieran dicha proteína (Anexo IX). Las fracciones apropiadas se mezclaron y se dializaron contra agua Mili-Q usando una membrana de

diálisis de 3.5 kDa MWCO. Para determinar la pureza e integridad de las proteínas LEA4-5_{1-77(9P)} y LEA4-5₇₈₋₁₅₈ después de la diálisis, se realizó un SDS-PAGE, seguida de una tinción con Azul de Coomassie e inmunodetección, para las proteínas respectivas.

Inmunodetección mediante ensayos tipo Western blot

Para llevar a cabo la inmunodetección por ensayos tipo Western blot, se utilizaron 15 µL de las diferentes fracciones colectadas durante la purificación de la proteína LEA4-5₇₈₋₁₅₈. Se utilizó un anticuerpo primario producido contra la región carboxilo de la proteína LEA4-5 (Agrisera AS22 4831), en una dilución de 1:1,000, y un anticuerpo secundario (anti-conejo acoplado a HRP, Zymed), en una dilución de 1:10,000. Las señales se produjeron mediante la reacción de la HRP con luminol (Supersignal West Pico, Thermo Fisher Scientific). Las membranas se expusieron a películas fotográficas azules para rayos X (Kodak), y se revelaron. Las películas reveladas fueron fotografiadas utilizando un fotodocumentador ImageQuant 300 (GE Healthcare).

Cuantificación de la concentración de las proteínas

La composición de residuos sesgada de las IDPs dificulta la cuantificación de la concentración de estas proteínas por los métodos convencionales, como Bradford, Lowry o la medición de absorbancia a 280 nm (Contreras-Martos *et al.*, 2018). Debido a lo anterior, se optó por el uso del ensayo de cuantificación de proteína por Qubit. Este ensayo se basa en el uso de moléculas fluorescentes específicas para proteínas. Dichas moléculas tienen poca fluorescencia en solución, pero aumentan su fluorescencia al interactuar con moléculas proteicas. Brevemente, se hicieron diluciones seriales de las proteínas purificadas y se usaron para calcular los valores de fluorescencia que mejor se ajustaran a la curva estándar que provee el sistema de detección del equipo (Qubit 4, Thermo Scientific), lo que permitió calcular la concentración final de la proteína en cuestión.

Experimentos de dicroísmo circular

Para obtener los espectros de dicroísmo circular (DC), se diluyeron las muestras hasta una concentración final de 0.3 mg/mL en solución amortiguadora de fosfatos 10 mM, en ausencia o presencia de cantidades crecientes (0%, 40% y 80%) de glicerol o etilenglicol (EG), dos compuestos que son utilizados para eliminar la interacción de los residuos de las proteínas con las moléculas de agua de sus proximidades, simulando una falta de agua en el ambiente. Se utilizó EG debido a que es menos viscoso que el glicerol, facilitando el manejo de las muestras y la reproducibilidad de los experimentos (Anexo XII). La efectividad del EG para disminuir la cantidad de agua disponible se verificó calculando la osmolalidad de las soluciones en comparación con las obtenidas para el glicerol. Las soluciones de glicerol al 40% y 80% (v/v) alcanzan osmolalidades de 5.47 and 10.95 Osm; en tanto que de EG a las mismas concentraciones generan un potencial osmótico de 7.17 and 14.34 Osm, respectivamente. Las mezclas se analizaron utilizando un espectropolarímetro JASCO J-715, en una celda de 1 mm de ancho, usando longitudes de onda de 190 a 260 nm. La temperatura se mantuvo a 25 °C con un controlador Peltier (PTC-4235, JASCO). Se ajustó la velocidad de lectura del espectropolarímetro a 5 segundos por nm. Se obtuvieron tres espectros, se promediaron y se suavizaron las curvas para reducir el ruido, de acuerdo con los algoritmos del programa para obtención de datos del espectropolarímetro. Los espectros fueron analizados utilizando el programa en línea DichroWeb (Whitmore *et al.*, 2004) para obtener el porcentaje total de estructura secundaria de cada proteína en cada tratamiento, usando el algoritmo CDSSTR y la base de datos SP175t.

Ensayos de protección ante tratamientos de congelamiento-descongelamiento

Se utilizó el protocolo seguido por Reyes *et al.* (2008) y Cuevas-Velazquez *et al.* (2016), y detallado por Rendon-Luna *et al.* (2020). Brevemente, se preparó una mezcla maestra (volumen final de 600 μ L) con solución amortiguadora de Tris-HCl 25 mM pH 7.5 y LDH 0.25 μ M (lactato deshidrogenasa de músculo de conejo, Sigma-Aldrich) y el protector a ser evaluado en diversas

relaciones molares, considerando siempre la molaridad de los monómeros. Se distribuyó la mezcla en seis tubos (100 μ L por tubo). Tres de los tubos permanecieron en hielo y el resto se sometieron a siete ciclos de congelamiento-descongelamiento, consistentes en un baño en nitrógeno líquido por un minuto y un posterior descongelamiento a 37°C por cinco minutos.

Medición de la actividad enzimática de lactato deshidrogenasa (LDH)

Para la medición de la actividad enzimática remanente de LDH, se preparó la mezcla de reacción para LDH, que consistió en Tris-HCl 100 mM pH 7.5, KCl 25 mM, NADH 148 μ M y ácido pirúvico 2.5 mM. Se colocaron 600 μ L de dicha mezcla en una celda de cuarzo para espectrofotómetro y se colocó la celda dentro del espectrofotómetro, previamente ajustado para leer a una absorbancia de 340 nm (UV) durante 40 segundos cada 0.5 segundos. Se tomaron 15 μ L de uno de los tubos con la mezcla a analizar, se agregaron a los 600 μ L de la mezcla de reacción y se mezclaron rápidamente. Se comenzó a leer inmediatamente y se registraron todos los valores obtenidos para su análisis posterior.

Análisis de datos

Los datos obtenidos para cada cinética se procesaron mediante un guion informático en R, diseñado para graficar individualmente cada una de las curvas obtenidas. Después, se realizó una regresión lineal de los puntos obtenidos en los primeros 20 segundos de la reacción y se obtuvo la pendiente de la recta. Los datos se normalizaron considerando como 100 % de actividad el valor de la pendiente obtenida del análisis de cada mezcla de LDH no sometida al tratamiento de congelamiento-descongelamiento, con o sin crioprotector. Finalmente, se graficaron los valores de actividad normalizados de todas las muestras. Para determinar diferencias estadísticas entre las medias del porcentaje de actividad remanente de LDH en los ensayos de protección a congelamiento-descongelamiento, se realizó una ANOVA de una vía, seguida por una prueba *post hoc* de comparación múltiple de Dunnett ($\alpha = 0.05$) considerando una comparación de muchos a

uno, donde la media de protección de cada mutante se comparó con la media de protección de la proteína nativa AtLEA4-5.

RESULTADOS

Distribución filogenética de diversas proteínas LEA4

Usualmente, se recurre a técnicas como la cristalografía de rayos X o a la de NMR para determinar la estructura de una proteína y los residuos importantes de la misma, los cuales pueden estar involucrados en su función. Desafortunadamente, estos métodos no son tan fácilmente aplicables a las características conformacionales de las IDPs debido a su alta flexibilidad estructural. Para identificar residuos o regiones conservados evolutivamente, y quizá relevantes para el funcionamiento de las proteínas LEA4, se decidió realizar un alineamiento múltiple de secuencias (MSA, por las siglas del *Multiple Sequence Alignment*). Primeramente, se realizó un BLAST en la base no redundante de secuencias de proteínas reportadas en NCBI, filtrando por el taxón *Viridiplantae*. Dicho análisis arrojó secuencias que no se anotaban como LEA4. Además, este análisis detectó, en las proteínas LEA4, el dominio de una superfamilia reportado en la base de datos de PFAM, localizado en la región N-terminal de estas proteínas, denominado en PFAM como dominio LEA_1. Para disminuir la ambigüedad de los datos, se decidió utilizar la secuencia consenso del dominio de PFAM LEA_1 para extraer las secuencias de proteínas reportadas con ese dominio de la base de datos de Phytozome (versión 11), la cual cuenta con registros mejor anotados para diversos organismos vegetales. Las secuencias obtenidas se depuraron eliminando secuencias con residuos mal anotados o con sitios de inicio de la traducción alternos. Del resultado de este análisis se obtuvo un archivo en formato fasta con 275 secuencias de proteínas, las cuales se alinearon mediante el software MUSCLE. Posteriormente, se realizó un análisis para identificar el mejor modelo evolutivo que explica las variaciones en secuencias (ModelTest) y un posterior análisis filogenético (RaxML) (Ver Materiales y Métodos). Este análisis derivó en un árbol que muestra las relaciones de parentesco o evolutivas entre las proteínas incluidas. Lo más evidente de este análisis es que las 275 secuencias de proteínas LEA4 se separan en dos subgrupos, el subgrupo A (136 secuencias) y subgrupo B (134 secuencias) (Figura 2). Los subgrupos muestran regiones conservadas en el N-terminal, así como una región C-terminal más corta en el subgrupo A que en el B. Es interesante mencionar que, cinco secuencias no se clasificaron en ninguno de los

subgrupos. Curiosamente, estas secuencias inclasificables, por el método empleado en esta tesis, pertenecen a plantas denominadas briofitas, las cuales se consideran como representantes actuales de las primeras plantas que poblaron el medio terrestre. Esto podría indicar que, si bien el origen de las proteínas LEA4, o proteínas con secuencias similares, se remonta al momento en que las plantas acuáticas tuvieron que enfrentarse a un ambiente con menos agua disponible en su entorno, y la subdivisión que dio origen a los subgrupos A y B ocurrió posteriormente en la evolución, probablemente con la aparición de plantas con flores y frutos (angiospermas). Debido a que la base de datos de Phytozome no cuenta con registros de gimnospermas ni helechos, al menos hasta la versión 11, y con el fin de no incorporar bases de datos con diferentes procesos de curación y anotación, no fue posible indagar respecto al momento evolutivo de aparición de los subgrupos A y B. Aun así, la conservación de secuencias de ambos subgrupos en los mismos organismos revela que evolutivamente surgieron temprano en la evolución de las angiospermas. Además, es probable que las proteínas que se agrupan en estos subgrupos realicen funciones distintas o complementarias, lo que explicaría que se hayan conservado a lo largo de la evolución en los mismos organismos.

Puesto que el análisis funcional previamente reportado se realizó con proteínas LEA4 de *A. thaliana* (Cuevas-Velazquez *et al.*, 2016), se decidió adoptar como modelo para el análisis propuesto en esta tesis a la proteína AtLEA4-5, ya que es la única representante del subgrupo B en *A. thaliana* y que tiene la región C-terminal más larga. Considerando el MSA, se ubicaron los residuos más conservados como la base para la selección de los residuos a modificar. En esta secuencia también se localizaron los tres motivos que identifican a las proteínas de esta familia, localizados en la región N-terminal y reportados previamente (Battaglia *et al.*, 2008) (Figura 3).

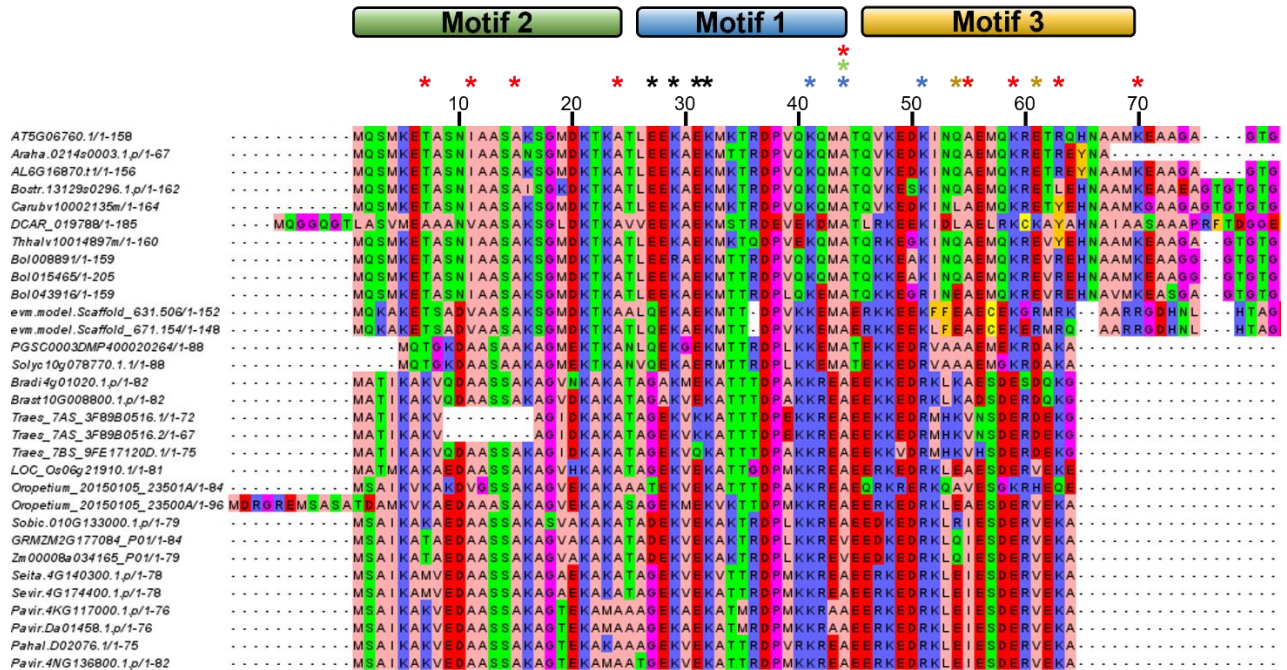


Figura 3. Alineamiento múltiple de secuencias más cercanas filogenéticamente a la proteína LEA4-5 de *A. thaliana* (AtLEA4-5). Se muestra la ubicación de motivos conservados en el MSA de acuerdo con Battaglia et al. (2008). Los residuos son coloreados de acuerdo con el código de colores Clustal en Jalview, donde: rojo: residuos cargados negativamente; azul: residuos cargados positivamente; verde: residuos polares sin carga; rosa: residuos con cadena lateral hidrofóbica; magenta: glicina y prolina; naranja: tirosina y fenilalanina; amarillo: cisteína. Los asteriscos de color muestran la ubicación de los residuos que se modificaron en la proteína AtLEA4-5. Negro: LEA4-5-NT1; azul: LEA4-5-NT2; verde: LEA4-5-NT3; amarillo: LEA4-5-NT4; rojo: LEA4-5-NT9P.

Análisis de proteínas LEA4 mutantes en residuos conservados en el N-terminal.

Los análisis de las secuencias mostraron que existen residuos conservados en el N-terminal de las proteínas LEA4 independientemente del subgrupo, por lo que algunos de esos residuos se seleccionaron en la proteína AtLEA4-5 para su posterior modificación. Las modificaciones puntuales por realizar fueron: E27A, K29A, E31A, K32A, K41A, A44K (el único residuo conservado incluso en briofitas y que se decidió cambiar por lisina), K51A, Q54A y E61A. La mayoría de las mutaciones se hicieron hacia residuos de alanina, el aminoácido quiral sin carga más simple cuya cadena lateral después del carbono β no introduce flexibilidad conformacional. Tales modificaciones se agruparon en 4 diferentes grupos, lo cual produjo cuatro proteínas

mutantes, denominadas LEA4-5-NT1 a -NT4. (Figura 4, Anexo V). Dichas proteínas se purificaron por el método de TCA (ver Materiales y Métodos) y se verificó su integridad y pureza (Anexo VIII).

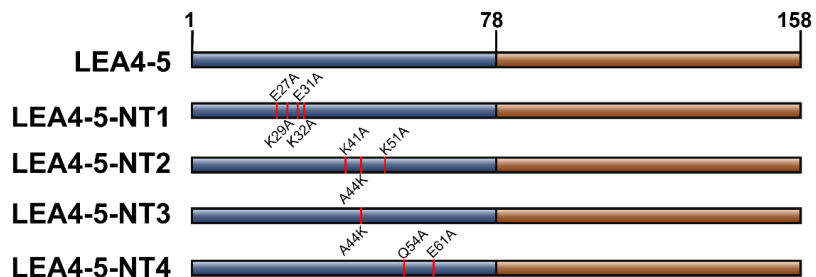


Figura 4. Ubicación de las mutaciones en residuos conservados. Se indica además la identidad del residuo mutado y el residuo por el que fue mutado.

Se realizaron experimentos de DC para determinar los niveles de estructura secundaria de estas proteínas bajo condiciones de deficiencia de agua, utilizando glicerol al 40 % y 80 % (v/v), y sin glicerol como referencia (Anexo X). Los espectros al 80% de glicerol mostraron que las proteínas LEA4-5-NT1 a -NT3 se comportaron como la proteína LEA4-5 silvestre (Figura 5). Sin embargo, se detectó un comportamiento diferente en el espectro de DC diferente de la proteína LEA4-5-NT4. Para realizar un análisis comparativo entre los espectros, se recurrió al análisis del espectro diferencia de los espectros a 80% del tratamiento contra los espectros obtenidos a 0% de tratamiento. En esta ocasión, se observó que el comportamiento de todos los espectros diferencia es prácticamente idéntico. (Figura 5, gráfico interno).

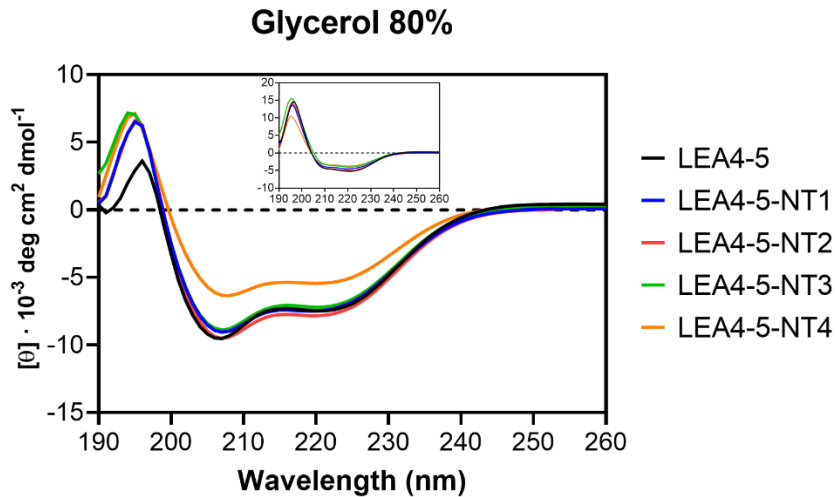


Figura 5. Espectros de DC de mutantes en residuos conservados. Se muestran los espectros obtenidos utilizando glicerol al 80% como tratamiento simulador de falta de agua. Gráfica interna: espectros diferencia obtenidos al restar los espectros a 0% de glicerol de los espectros a 80% de glicerol.

Para obtener un cálculo aproximado de la proporción de estructura secundaria a partir de los datos por DC, se utilizó el algoritmo DichroWeb (Figura 6). Los resultados mostraron que, prácticamente, la capacidad de adquisición de los diferentes tipos de estructuras que considera DichroWeb es similar en las diferentes proteínas analizadas, excepto en el caso de la proteína LEA4-5-NT4, que muestra una menor proporción de hélices alfa en presencia de 80% glicerol, comparado con la proteína silvestre. Si bien este tipo de análisis puede ayudarnos a tener una mejor idea de la proporción de estructura de la proteína, hay que recordar que el algoritmo DichroWeb predice la proporción de estructura secundaria utilizando como base de datos espectros de DC de proteínas globulares caracterizadas mayoritariamente con estudios cristalográficos, por lo que no resultaría extraño observar algún comportamiento anómalo en datos provenientes de IDPs.

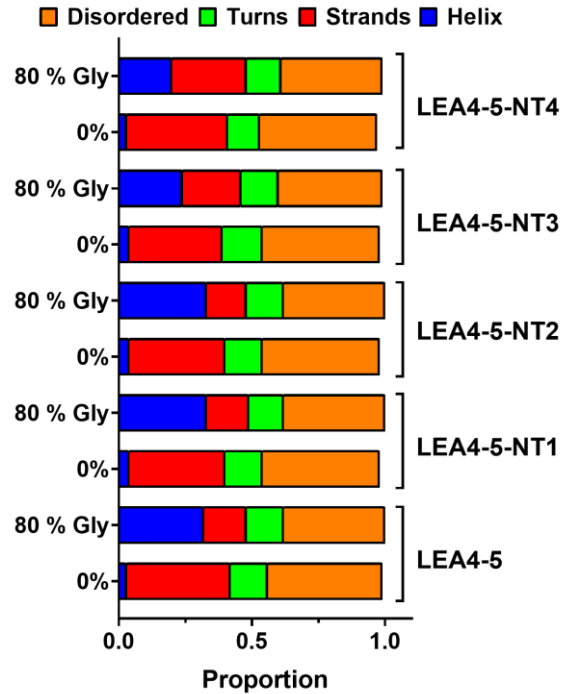


Figura 6. Proporción de estructuras secundaria obtenidos mediante DichroWeb de mutantes en residuos conservados . Para cada proteína se muestra la proporción de estructura secundaria calculada por el algoritmo DichroWeb en muestras con 0 % y 80 % (v/v) de glicerol.

Para determinar si los cambios que se introdujeron en las proteínas mutantes producían algún efecto en la actividad protectora de esta proteína, se realizaron experimentos de protección frente a congelamiento-descongelamiento. Se realizaron mezclas de protector:LDH (relación molar 10:1, considerando los monómeros) y se aplicaron siete ciclos de congelamiento-descongelamiento acorde con datos previos en los que se demostró que este número de ciclos anulaba totalmente la actividad de la LDH (Cuevas-Velazquez *et al.*, 2016). Los resultados no mostraron diferencias significativas entre la protección alcanzada por la proteína LEA4-5 silvestre y las proteínas mutantes LEA4-5-NT1 a -NT4 (Figura 7).

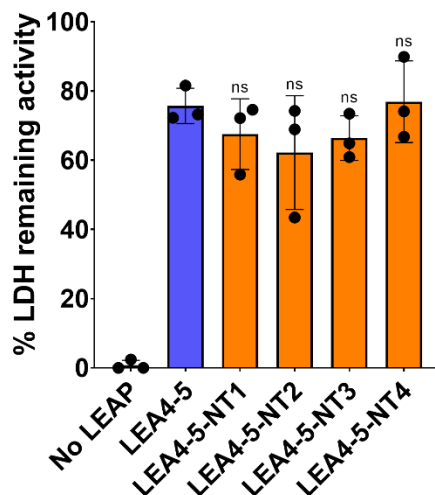


Figura 7. Experimento de protección de LDH ante congelamiento-descongelamiento usando mutantes en residuos conservados. Se utilizaron siete ciclos de congelamiento-descongelamiento y una relación molar de 10:1 (Protector:LDH). Cada punto indica el promedio de un experimento con tres réplicas. Los datos se normalizaron considerando 100% la actividad de cada mezcla protector-LDH sin someterse a congelamiento-descongelamiento. ns: sin diferencia significativa comparado con LEA4-5. No LEAP: *No LEA protein*. ns: sin diferencia significativa comparado con LEA4-5

Estos datos indican que los residuos conservados en la región N-terminal de la proteína AtLEA4-5 no tienen una participación significativa en la capacidad de adquisición de estructura secundaria ante la deficiencia de agua, ni en la actividad protectora de esta proteína *in vitro*.

Análisis de proteínas LEA4 truncadas.

Debido a que las mutaciones en residuos conservados no mostraron efectos en la capacidad de adquirir estructura secundaria ni en la capacidad protectora, se decidió eliminar de algunos segmentos a partir del extremo N-terminal. Se obtuvieron mutantes en donde se eliminaron los primeros 25, 42, 56 y 77 residuos del extremo N-terminal, en función de la localización *in silico* de zonas de formación de hélices alfa (Anexo VI). También se removieron fragmentos en una zona de la región N-terminal, en donde se mostraba mayor propensión a formar hélices alfa (residuos 61-78 y 44-78), conformando las proteínas LEA4-5 $_{\Delta 61-78}$ y LEA4-5 $_{\Delta 44-78}$, respectivamente (Figura 8).

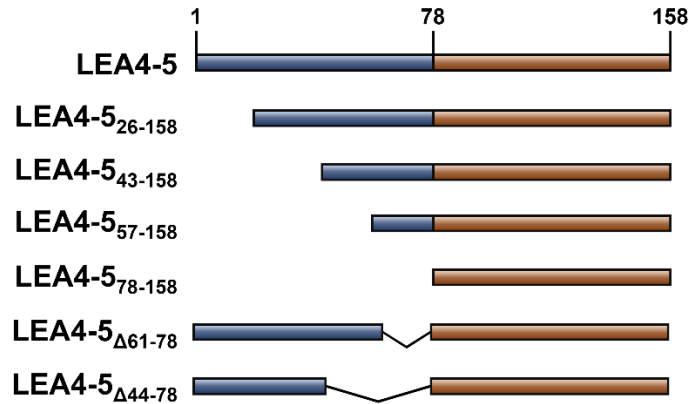


Figura 8. Ubicación de las mutaciones con eliminaciones de secciones de la región N-terminal. Se indica además la identidad del residuo mutado y el residuo por el que fue mutado.

Utilizando las proteínas mutantes purificadas, se realizaron los experimentos de DC correspondientes. En esta ocasión, para estos experimentos se utilizó etilenglicol (EG) en lugar de glicerol, debido a que el etilenglicol es un compuesto menos viscoso que el glicerol a 25 °C, lo cual facilita su manipulación en las celdas utilizadas en el espectropolarímetro (ver Materiales y Métodos), favorece una mejor homogenización y, por tanto, una mayor reproducibilidad de los datos. Las soluciones de ambos compuestos generan potenciales osmóticos similares a los que se obtienen con glicerol por lo que aseguran condiciones similares en cuanto a la disponibilidad de agua (Anexo XII).

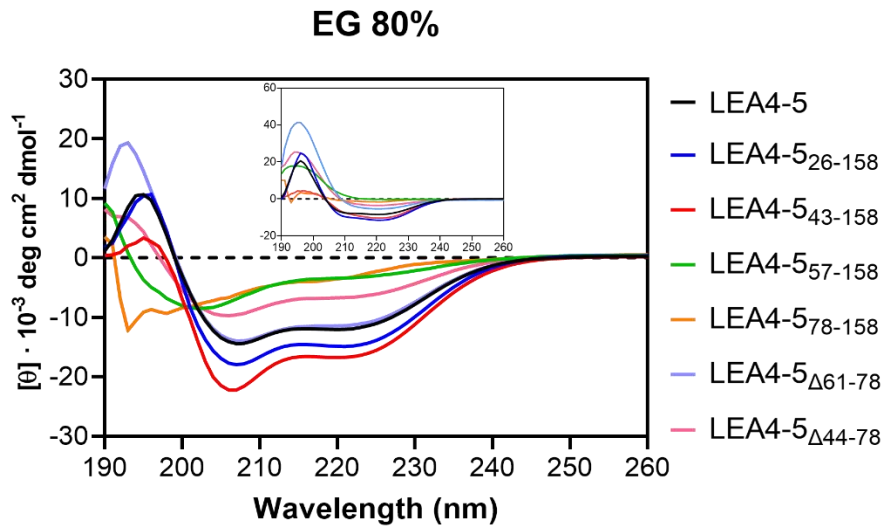


Figura 9. Espectros de DC de mutantes en residuos conservados. Se muestran los espectros obtenidos utilizando etilenglicol al 80% como tratamiento simulador de falta de agua. Gráfica interna: espectros diferencia obtenidos al restar los espectros obtenidos en presencia de 0 % de etilenglicol de los espectros colectados en presencia de 80 % (v/v) de etilenglicol.

Los resultados del análisis de DC, utilizando una concentración de EG del 80 % mostraron que las mutantes LEA4-5₂₆₋₁₅₈ y LEA4-5₄₃₋₁₅₈ presentan una mayor formación de hélices alfa, evidente por el incremento de la señal negativa a 208 y 222 nm de sus espectros de DC (Figura 9). Aun cuando las proteínas mutantes son más cortas que la proteína silvestre, y debido a que la región C-terminal es incapaz de ganar estructura bajo estas condiciones (Cuevas-Velazquez *et al.*, 2016), estos datos sugieren que el segmento remanente de la región N-terminal adquirió mayor proporción de hélices alfa que la región silvestre en estas condiciones (Figura10, Anexo X). En contraste, las proteínas LEA4-5₅₇₋₁₅₈ y LEA4-5₇₈₋₁₅₈, mostraron una menor ganancia de estructura secundaria en hélice alfa, lo cual se esperaba debido a que en estas proteínas se eliminaron zonas con mayor propensión a formar estructura secundaria (Figura 9, Figura 10). Acorde con datos previos (Cuevas-Velazquez *et al.*, 2016), la eliminación completa de la región N-terminal mostró un espectro típico de una proteína desordenada; es decir, sin la aparente ganancia de estructura bajo estas condiciones. Por su parte, el espectro de DC de la proteína LEA4-5_{Δ61-78} en 80% de EG se comportó como la proteína

silvestre, mientras que el correspondiente a la proteína LEA4-5 Δ 44-78 en la misma condición mostró menor capacidad de formación de estructura secundaria hélice alfa (Figura 9, Figura 10). Los resultados de estas dos últimas mutantes sugieren que la región comprendida entre los residuos 44 a 78 tiene una capacidad de formar hélices alfa, predicha *in silico*, que se elimina efectivamente en estas mutantes (Anexo VI).

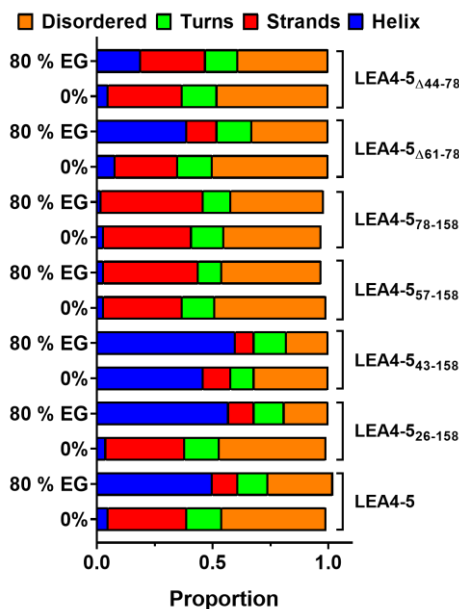


Figura 10. Proporción de estructuras secundarias obtenida mediante DichroWeb de las truncadas. Para cada proteína se muestra la proporción de estructura calculada por el algoritmo DichroWeb a 0% y 80% de etilenglicol.

Para determinar el efecto de la eliminación de distintas regiones en la capacidad protectora de estas mutantes, se realizaron los ensayos de congelamiento-descongelamiento como en los casos anteriores. Los resultados de tales experimentos se muestran en la Figura 11. Se observó que las proteínas LEA4-5 Δ 26-158 y LEA4-5 Δ 43-158 tienen una capacidad protectora comparable a la proteína silvestre, mientras que la proteína LEA4-5 Δ 57-158 protege significativamente menos que la proteína silvestre, siendo aún menor en el caso de la proteína LEA4-5 Δ 78-158. Los resultados con esta última mutante indican que el extremo C-terminal puede tener cierta actividad protectora, lo cual resultó curioso. Inesperadamente, la proteína LEA4-5 Δ 61-78 mostró valores de protección mayores que la

proteína silvestre, mientras que la proteína LEA4-5 Δ_{44-78} no mostró diferencias en su capacidad protectora en comparación con la proteína silvestre.

En conjunto, estos datos indican que la eliminación progresiva de secciones extremas de la región N-terminal genera una pérdida progresiva del efecto protector, mientras que la eliminación de las regiones internas conlleva a una capacidad similar a la de la proteína silvestre.

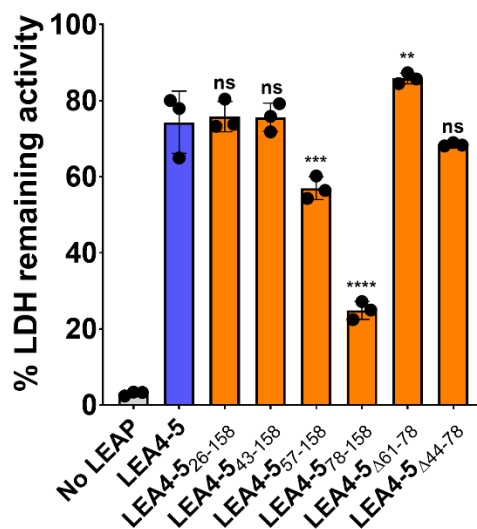


Figura 11. Protección de LDH ante congelamiento-descongelamiento usando las mutantes truncadas. Se utilizaron siete ciclos de congelamiento-descongelamiento y una relación molar de 10:1 (Protector:LDH). Cada punto indica el promedio de un experimento con tres réplicas. Los datos se normalizaron considerando 100% la actividad de cada mezcla protector-LDH sin someterse a congelamiento-descongelamiento. No LEAP: *No LEA protein*. ns: sin diferencia significativa comparado con LEA4-5. **: Diferente significativamente con $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

Análisis de proteínas con mutaciones que desestabilizan la formación de hélices alfa en la región N-terminal.

Los resultados de la sección anterior apuntaron a una posible relevancia de la formación de hélices alfa en la región N-terminal para la capacidad protectora de la proteína LEA4-5, en concordancia con los datos reportados por Cuevas-Velazquez *et al.* (2016), donde la región N-terminal es

suficiente para ejercer un efecto protector en los ensayos *in vitro*. Para probar esta hipótesis se generaron las mutantes LEA4-5-NT2P, LEA4-5-NT4P y LEA4-5-NT9P con mutantes en diferentes residuos, que incluyeron las mutaciones: T7P, I11P, A15P, A24P, A44P, A55P, K59P, R63P y K70P (Figura 12).

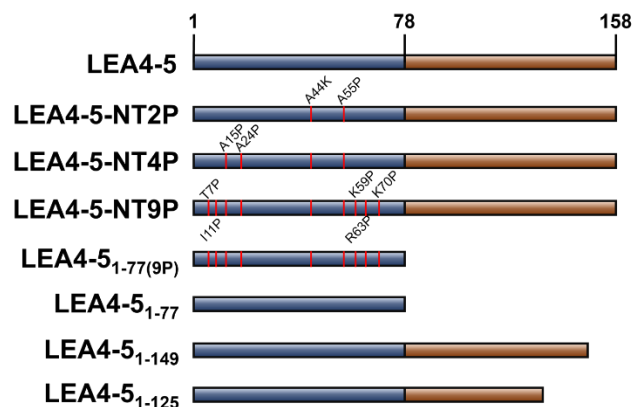


Figura 12. Ubicación de las mutaciones a prolinas en la región N-terminal y mutantes truncadas en el extremo C-terminal. Se indica además la identidad del residuo mutado y el residuo por el que fue mutado.

Para el caso de estas mutaciones, los espectros de DC y su deconvolución por DichroWeb mostraron que, a mayor cantidad de mutaciones por prolina, la proteína tuvo menos capacidad de adquirir estructura secundaria (Figura 13, Figura 14). Sin embargo, no se observó un efecto en la capacidad protectora de estas proteínas, lo cual resultó inesperado (Figura 15).

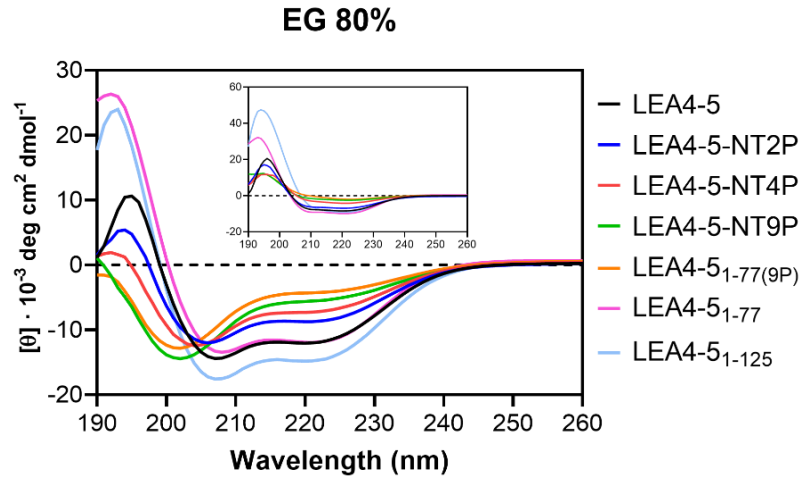


Figura 13. Espectros de DC de mutantes de prolina y mutantes truncada en el extremo C-terminal. Se muestran los espectros obtenidos utilizando etilenglicol al 80 % (v/v) como tratamiento simulador de falta de agua. Gráfica interna: espectros diferencia obtenidos al restar los espectros colectados en 0 % de etilenglicol de los espectros colectados en presencia de 80 % (v/v) de etilenglicol.

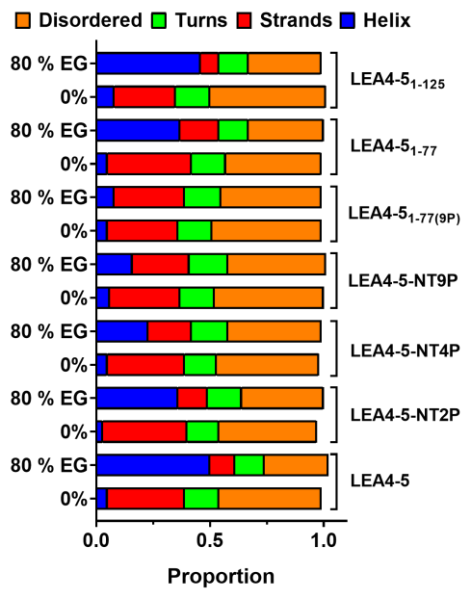


Figura 14. Proporción de estructuras obtenidos mediante DichroWeb de mutantes con prolina y mutantes truncadas en el extremo C-terminal. Para cada proteína se muestra la proporción de estructura calculada por el algoritmo DichroWeb en presencia de 0 % y 80 % (v/v) de etilenglicol.

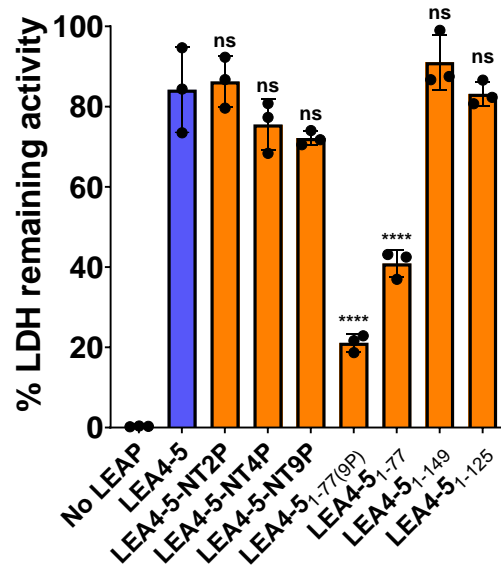


Figura 15. Experimento de protección de LDH ante congelamiento-descongelamiento de mutantes con prolinas y mutantes truncadas en el extremo C-terminal Se utilizaron siete ciclos de congelamiento-descongelamiento y una relación molar de 10:1 (Protector:LDH). Cada punto indica el promedio de un experimento con tres réplicas. Los datos se normalizaron considerando 100% la actividad de cada mezcla protector-LDH sin someterse a congelamiento-descongelamiento. No LEAP: *No LEA protein*. ns: sin diferencia significativa comparado con LEA4-5. ****: Diferente significativamente con $p < 0.0001$.

Para eliminar una posible influencia del C-terminal en el fenotipo de protección y considerando que se ha reportado que el N-terminal tienen la capacidad suficiente de proteger como la proteína silvestre (Cuevas-Velazquez *et al.*, 2016), se optó por producir una proteína mutante con las nueve inserciones de prolina, pero donde se eliminara el extremo C-terminal LEA4-5_{1-77(9P)}. Para contar con un buen control para esta proteína, también se produjo la proteína LEA4-5₁₋₇₇ (Figura 12). La introducción de los nueve residuos de prolina disminuyó la hidrofobicidad media y la carga neta de la proteína LEA4-5_{1-77(9P)} en comparación con LEA4-5₁₋₇₇ (Anexo XIV)

Como se esperaba, los resultados de los experimentos de DC con o sin EG (40 % y 80 %) mostraron que la proteína LEA4-5_{1-77(9P)} adquiere una menor proporción de hélices alfa al estar en presencia de 80 % de EG en comparación con el LEA4-5₁₋₇₇ silvestre (Figura 13, Figura 14, Anexo

VI). Sorprendentemente, los ensayos de protección ante congelamiento-descongelamiento mostraron la proteína LEA4-5₁₋₇₇ tiene una capacidad protectora menor que la proteína completa. Aún con esto, al analizar la proteína LEA4-5_{1-77(9P)} se observó una disminución de la capacidad protectora aún mayor (Figura 15).

Por otro lado, derivado de los resultados obtenidos de la proteína LEA4-5₇₈₋₁₅₈, y considerando la presencia de dos tirosinas en el C-terminal, se decidió producir dos mutantes que eliminaran una sección del C-terminal, y se denominaron LEA4-5₁₋₁₄₉ and LEA4-5₁₋₁₂₅. (Figura 12). Los experimentos de DC utilizando la proteína mutante LEA4-5₁₋₁₂₅ mostraron que la proteína se estructura (Figura 13, Figura 14), y los experimentos de protección ante congelamiento-descongelamiento que se comportan como la proteína silvestre (Figura 15).

Análisis de crioprotección utilizando diferentes relaciones molares

Dados los resultados anteriores, se infirió que las mutaciones a prolina afectaron negativamente la estructuración y funcionalidad de la proteína LEA4-5_{1-77(9P)}, mientras que las mismas mutaciones en la versión completa (LEA4-5-NT9P) afectaron negativamente capacidad de estructuración, pero no la funcionalidad. Para indagar más respecto a dicho fenotipo, se realizaron experimentos de protección ante congelamiento-descongelamiento a diferentes relaciones molares para observar el comportamiento de la capacidad protectora de algunas mutantes. La gráfica de protección en función de la relación molar de algunas proteínas mutantes se muestra en la Figura 16.

Estos experimentos mostraron que la actividad protectora de las proteínas LEA4-5, LEA4-5-NT4P y LEA4-5-NT9P es similar en diferentes relaciones molares, lo cual indica que la disminución de la proporción de hélices alfa en efecto no impactó en la actividad de estas proteínas. Por otro lado, la proteína LEA4-5₁₋₇₇ mostró un nivel de protección menor que proteína LEA4-5 en relaciones molares entre 8:1 y 20:1. Sin embargo, la capacidad protectora de la mutante LEA4-5₁₋₇₇ alcanza los mismos niveles de protección que la proteína LEA4-5 en relaciones molares mayores a 20:1. Por su

parte, se observó una diferencia entre la capacidad protectora de la proteína LEA4-5₁₋₇₇ y su versión mutante de prolina, LEA4-5_{1-77(9P)}. Esta última mostró una capacidad protectora menor en las diferentes relaciones molares analizadas, lo que confirma que la capacidad protectora de esta proteína mutante si fue negativamente afectada por la incapacidad de formar estructura hélice alfa. Por su parte, la región C-terminal de la proteína LEA4-5, representada por la mutante truncada LEA4-5₇₈₋₁₅₈, no mostró una capacidad protectora significativa en relaciones molares bajas, menores a 10:1, pero se observó una creciente capacidad protectora conforme se aumentó la relación molar.

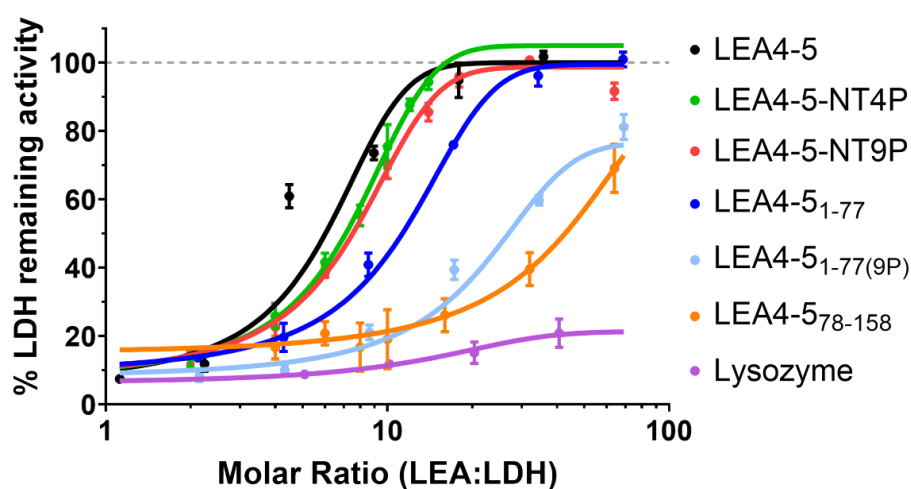


Figura 16. Experimento de protección de LDH ante congelamiento-descongelamiento de mutantes por prolina y mutantes truncados en el extremo C-terminal. Se incluyó a la lisozima como control negativo de protección. Las medias de protección a diferentes relaciones molares se usaron para realizar un ajuste de los datos a una curva sigmoideal.

Para analizar con mayor detalle las diferencias entre la capacidad protectora de estas proteínas, se utilizaron los datos de ajuste a las curvas sigmoideas calculadas (Figura 16) para obtener la relación molar necesaria para alcanzar el 50% de protección (MR_{50}) de esas mismas proteínas (Cuevas-Velazquez *et al.*, 2014). Los valores calculados de MR_{50} se muestran en la Figura 17. Se determinó que la MR_{50} de las mutantes LEA4-5-NT4P y LEA4-5-NT9P son prácticamente idénticas a la proteína silvestre, mientras que la MR_{50} de la proteína mutante LEA4-5₁₋₇₇ fue mayor

que la silvestre. Por su parte, la mutante LEA4-5_{1-77(9P)} mostró una MR₅₀ considerablemente mayor que su control silvestre, la proteína LEA4-5₁₋₇₇, lo cual evidencia el impacto de la falta de estructura en esta proteína. Por último, el extremo C-terminal de la LEA4-5, la mutante LEA4-5₇₈₋₁₅₈, mostró una MR₅₀ aún mayor, por lo que, aunque su actividad protectora no es nula, es menor que la LEA4-5₁₋₇₇ y que la LEA4-5_{1-77(9P)}.

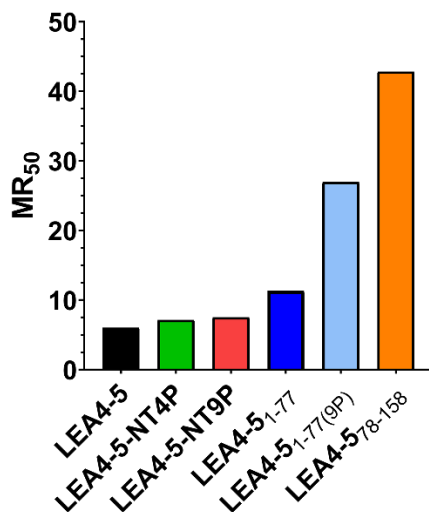


Figura 17. Experimento de protección de LDH ante congelamiento-descongelamiento de mutantes a prolina y mutantes truncadas en el extremo C-terminal. El valor de MR₅₀ se obtuvo a partir de los datos de ajuste a curva sigmoideal de la Figura.

DISCUSIÓN

Las proteínas LEA4 como modelo para estudiar IDPs en plantas

El papel que desempeñan las proteínas LEA en la capacidad de las plantas de sobrevivir ante condiciones de falta de agua es un fenómeno descrito por varios autores (Tunnacliffe *et al.*, 2007; Olvera-Carrillo *et al.*, 2011; Hinch *et al.*, 2012; Hernández-Sánchez *et al.*, 2022). De los diferentes grupos de proteínas LEA (Battaglia *et al.*, 2008), el grupo 4 resulta atractivo para llevar a cabo estudios genéticos y funcionales debido a que la planta modelo *A. thaliana* sólo cuenta con tres miembros (Hundertmark *et al.*, 2008; Cuevas-Velazquez *et al.*, 2017). Además, existen fenotipos reportados de plantas mutantes en algunos de los miembros de proteínas LEA4 que señalan la participación de estas proteínas en la contención del daño provocado por la limitación de agua (Olvera-Carrillo *et al.*, 2010).

Las propiedades fisicoquímicas de las proteínas LEA4 permite clasificarlas como IDPs (Cuevas-Velazquez *et al.*, 2017), y se sabe que, aunque son proteínas con una estructura flexible en solución acuosa, pueden adquirir elementos de estructura secundaria tipo hélice alfa en condiciones de falta de agua, y que la estructura adquirida es importante para ejercer una función protectora *in vitro* (Cuevas-Velazquez *et al.*, 2016). Con el fin de conocer mejor las características de la proteína relacionadas con su función, nos preguntamos si el efecto protector observado *in vitro* es producto de la adquisición de estructura tipo hélice alfa o por la presencia de ciertos residuos de aminoácidos en esa región que pudieran participar en la interacción entre la proteína LEA y su blanco. Sin embargo, la elevada flexibilidad de las IDPs impone un reto debido a que estas proteínas no permanecen en una conformación estable. Por ese motivo, el estudio de IDPs ha implicado aproximaciones metodológicas diferentes.

En esta tesis, abordamos esta problemática a través de la generación de mutantes y su análisis por medio de evaluaciones funcionales y estructurales *in vitro*. Se modificaron ciertos residuos o

regiones de una proteína representante de la familia LEA4 (tLEA4-5), seguido de evaluar su capacidad de estructurarse en baja disponibilidad de agua y de prevenir la pérdida de actividad de enzimas reporteras en estas condiciones estresantes.

El análisis filogenético de las proteínas LEA4 muestra una separación en los A y B después de su aparición en las briofitas.

Es común que una misma enzima con la misma actividad catalítica presente estructuras secundarias similares asociadas a dicha función. Los residuos de aminoácidos claves asociados a estas estructuras se pueden detectar al realizar alineamientos múltiples de secuencias de proteínas. Siguiendo esta lógica, para detectar residuos importantes para la función de las proteínas LEA4, se obtuvieron las secuencias de dichas proteínas reportadas en la base de datos Phytozome y se realizaron alineamientos múltiples de secuencias, con los que se identificaron los residuos conservados (Figura 3). Además, se realizó un análisis filogenético para determinar la historia evolutiva de estas proteínas (Figura 2). De forma general, se observó que, en todas las proteínas anotadas como LEA4 en Phytozome, el N-terminal está altamente conservado y presenta una considerable cantidad de residuos cargados, mientras que el C-terminal es de longitud variable y con menor cantidad de residuos con carga. Es importante notar que, aún con más secuencias utilizadas para la elaboración de este análisis, la división en subgrupos A y B se mantuvo (Figura 2), lo cual fortalece lo reportado previamente (Olvera-Carrillo *et al.*, 2010). Cabe mencionar que esta subdivisión se conserva si el análisis filogenético se realiza sólo con las secuencias de la región N-terminal, sugiriendo que en este segmento existe información relacionada con una diversidad funcional entre estos dos subgrupos. Asimismo, la diferencia más notable entre estos dos subgrupos es la extensión de la región C-terminal; más larga en el subgrupo B que en el subgrupo A. Esta observación es indicativa de que la región C-terminal en el subgrupo B es relevante para la función de las proteínas LEA4 de este subgrupo.

Las proteínas LEA4 de grupos basales de plantas no pertenecen a ningún subgrupo, lo que muestra que, aunque estas proteínas evolucionaron como respuesta de las primeras plantas que colonizaron ambientes terrestres a un ambiente con menor disponibilidad de agua, la especialización que permitiría originar los dos subgrupos ocurrió después de la evolución de las plantas terrestres, y antes de la aparición de las primeras plantas con flor.

Las mutaciones en residuos conservados no afectan la estructura ni la función de la proteína AtLEA4-5

Con la idea de que los residuos conservados en una proteína son relevantes para su función, se seleccionaron dichos residuos en la región N-terminal de la proteína AtLEA4-5 para modificarse (Figura 4). Los residuos elegidos se caracterizan por ser, en la mayoría de los casos, residuos cargados y por estar en las zonas con mayor propensión a formar hélice alfa (Anexo VI). Al analizar las proteínas mutantes (LEA4-5-NT1 a -NT4) mediante DC, no se observaron diferencias notables en sus espectros. La proteína LEA4-5-NT4 mostró un espectro con una elipticidad molar menor que las demás mutantes, indicativo de una menor adquisición de estructura (Figura 5, Figura 6). Sin embargo, al analizar los espectros de diferencia, todos los espectros son similares. Contrario a lo esperado, ninguna de estas proteínas mutantes mostró diferencia en su capacidad protectora en comparación con la proteína silvestre (Figura 7). Esto indica que los residuos conservados no parecen ser importantes para la función protectora *in vitro* de la LDH, la enzima reportera utilizada. Es posible que los residuos conservados estén relacionados a su función hacia sus blancos naturales, de tal manera que mientras los ensayos se realicen con un sustrato artificial, difícilmente se podría determinar la relevancia funcional de los mismos *in vivo*. Sin embargo, esto no excluye el hecho de que los ensayos *in vitro* constituyan una herramienta para conocer posibles mecanismos de acción de estas proteínas de forma general. Aun así, es importante considerar que la interacción de la proteína LEA con su blanco, y por tanto su efecto protector, puede estar mediada por un conjunto de eventos, como puede ser la atracción electrostática, el efecto hidrofóbico o la detección de MoRFs (*Molecular Recognition Features*, características de

reconocimiento molecular). Por lo tanto, la mutación de algunos residuos, aunque conservados, se tolera en las condiciones utilizadas en este trabajo.

La eliminación de algunas zonas de la proteína AtLEA4-5 afectan negativamente su actividad protectora .

Adicionalmente, se optó por eliminar algunas regiones de la proteína AtLEA4-5, iniciando con el extremo N-terminal, debido a su elevada conservación. Se generaron mutantes en donde se eliminaron segmentos progresivos de esta región y algunas zonas internas (Figura 8). Las proteínas LEA4-5₂₆₋₁₅₈ y LEA4-5₄₃₋₁₅₈ mostraron espectros de DC con una mayor señal a 208 y 222 nm que la proteína silvestre, evidenciando una mayor formación de estructura hélice alfa, en coincidencia con la mayor propensión predicha para la región que permanece. Esta observación es consistente con los espectros obtenidos para las proteínas LEA4-5₅₇₋₁₅₈ y LEA4-5₇₈₋₁₅₈, típicos para una proteína desordenada (Figura 9). Por último, la proteína LEA4-5_{Δ61-78} mostró un espectro de DC como la proteína silvestre, mientras que la proteína LEA4-5_{Δ44-78} mostró una disminución de las señales a 208 y 222 nm (Figura 8). En conjunto, esto podría indicar que la región comprendida entre los residuos 44 y 61 posee una mayor propensión a formar una estructura hélice alfa bajo déficit hídrico.

Por otro lado, los experimentos de protección ante congelamiento-descongelamiento *in vitro* mostraron que las proteínas LEA4-5₂₆₋₁₅₈, LEA4-5₄₃₋₁₅₈ y LEA4-5_{Δ44-78} tienen el mismo nivel de protección que la proteína silvestre, mientras que las proteínas LEA4-5₅₇₋₁₅₈ y LEA4-5₇₈₋₁₅₈ mostraron una reducción de la capacidad protectora (Figura 10). Cabe hacer notar que en este trabajo se observó cierta capacidad protectora por parte de la proteína LEA4-5₇₈₋₁₅₈, siendo que en estudios previos se determinó que dicha región de la proteína LEA4-5 no mostraba protección (Cuevas-Velazquez *et al.*, 2016). Una explicación para tal observación es que, a diferencia del trabajo anterior, donde la proteína LEA4-5₇₈₋₁₅₈ se sintetizó químicamente, en el presente trabajo se logró purificar dicha proteína mediante la metodología de IMPACT-NC (ver Materiales y

Métodos). Sumado a esto, en este trabajo decidimos determinar la concentración de las proteínas utilizando un método más confiable para proteínas desordenadas (Qubit), dado que éstas carecen o son deficientes en aminoácidos aromáticos, lo cual mostró diferencias importantes con los métodos convencionales usados comúnmente. Esto da solidez a los resultados obtenidos aquí y permite asegurar que el extremo C-terminal de la proteína LEA4-5 ejerce actividad protectora.

Por el contrario, el nivel de protección dado por la proteína LEA4-5 $_{\Delta 61-78}$ fue mayor. Este resultado fue un tanto imprevisto puesto que al eliminar parte del N-terminal se esperaba que la actividad protectora disminuyera como sucedió con las proteínas LEA4-5 $_{57-158}$ y LEA4-5 $_{78-158}$. Sin embargo, podría hipotetizarse que se requiere de una interacción entre el N-terminal y el C-terminal para ejercer una protección efectiva. Dicha interacción se favorecería al eliminar una región central de la proteína silvestre, lo cual explicaría el fenotipo observado.

Considerando la presencia de residuos de tirosina en la región más extrema del C-terminal que pudieran ser importantes, se realizaron un par de mutantes truncadas adicionales, llamadas LEA4-5 $_{1-149}$ y LEA4-5 $_{1-125}$ (Figura 12). Se observó un incremento en la formación de estructura secundaria en la proteína LEA4-5 $_{1-125}$ (Figura 13, Figura 14) pero esto pudo deberse a la eliminación de regiones intrínsecamente desordenadas de la proteína. En cuanto a la protección, ninguna de estas dos proteínas se comportó diferente a la proteína silvestre (Figura 15). Por lo tanto, la eliminación de hasta 33 residuos del extremo C-terminal no afectó negativamente el funcionamiento de la proteína. Esta evidencia no descarta la posibilidad de que esta región y, en particular los residuos de tirosina participen en el reconocimiento de los blancos naturales de la proteína AtLEA4-5.

Las mutaciones por prolinas eliminan la formación de hélices alfa pero no así la función protectora de la proteína completa

La idea de eliminar regiones de una proteína para evaluar su contribución a la función de ésta tiene una desventaja intrínseca, que es la eliminación de residuos que pudieran servir para interactuar

con otras proteínas. Esto impide determinar con certeza la relación entre la adquisición de estructura con una función. Otra aproximación metodológica es posible para solventar esta situación, la cual implica la sustitución de residuos cercanos a regiones propensas a formar hélices alfa por residuos de prolina. Dicho residuo introduce rigidez en el enlace peptídico e impide la formación de estructura hélice alfa en las cercanías. De esta manera, es posible disminuir efectivamente la formación de hélices sin alterar el tamaño de la proteína.

Con esta idea en mente, se diseñaron proteínas en donde se sustituyeron residuos no conservados y cercanos a zonas con propensión a formar hélices alfa por residuos de prolina a lo largo de la región N-terminal de la proteína LEA4-5. Dichas mutaciones se realizaron tanto en la proteína completa (LEA4-5-NT2P, LEA4-5-NT4P y LEA4-5-NT9P) como en la proteína LEA4-5₁₋₇₇ (Figura 12) (LEA4-5_{1-77(9P)}), considerando que se ha reportado al extremo N-terminal como la región funcional de la proteína (Cuevas-Velazquez *et al.*, 2016). Como se esperaba, las sustituciones por prolina disminuyeron progresivamente la cantidad de estructura secundaria de las diferentes proteínas mutantes (Figura 13, Figura 14, Anexo VI, Anexo VII). Asimismo, la actividad protectora de la proteína LEA4-5_{1-77(9P)} disminuyó respecto a la región N-terminal silvestre, aunque no se eliminó totalmente, posiblemente debido a cierto nivel de estructura secundaria remanente. Cabe mencionar que el efecto protector de esta región detectado en este trabajo fue menor que el reportado previamente Cuevas-Velazquez *et al.* (2016), lo cual se explica por las diferencias encontradas en la determinación de las concentraciones de proteínas utilizando la metodología de Qubit en donde se consideran los enlaces peptídicos y no la abundancia de aminoácidos aromáticos. La comparación entre los métodos convencionales y Qubit mostró que los primeros sobrestiman la concentración de la proteína, por lo que en los ensayos anteriores se utilizó una mayor cantidad de éstas.

Sorprendentemente, los experimentos de protección utilizando la proteína completa con las sustituciones de prolina (LEA4-5-NT9P) mostraron un nivel de protección similar a la proteína

silvestre (Figura 15), indicando que en presencia de la región C-terminal, la formación de hélices alfa no es necesaria para ejercer un efecto protector en estas condiciones. Un escenario que explicaría este resultado es un efecto sinérgico entre las regiones N-terminal y C-terminal, que podría darse al tener más residuos expuestos capaces de asociarse por interacciones electrostáticas a su blanco. Esta hipótesis se soporta por datos de experimentos de FRET, en los que sólo se obtuvo señal de FRET, bajo condiciones de deficiencia de agua, con la proteína silvestre y no así con la mutante, indicando que esta última se mantiene en una conformación menos compacta (T. Nava, datos no publicados).

Con el fin de contar con una descripción comparativa más rigurosa de la actividad protectora de la proteína LEA4-5 y de algunas de las mutantes obtenidas, se realizaron experimentos de protección utilizando diferentes relaciones molares de las proteínas LEA4, con los cuales se generaron curvas de protección y se obtuvieron las relaciones molares con las cuales estas se alcanzaron el 50% de protección (MR_{50}) (Figura 16, Figura 17). Así, al comparar el efecto protector de las proteínas LEA4-5-NT4P y LEA4-5-NT9P con la proteína silvestre se obtuvieron MR_{50} similares entre ellas, reforzando que estas mutaciones no afectan esta actividad bajo ninguna concentración. La proteína LEA4-5₁₋₇₇ mostró una MR_{50} mayor (11.33), apoyando la conclusión de que es necesaria una mayor cantidad de proteína para ejercer un efecto protector similar a la proteína silvestre (MR_{50} 6.03), aun cuando es posible alcanzar el 100% de protección. Por su parte, la proteína LEA4-5_{1-77(9P)} mostró un MR_{50} considerablemente mayor (26.97) al de la silvestre (6.03) y no alcanzó el 100% de protección. De acuerdo con los datos mencionados anteriormente, la región C-terminal (LEA4-5₇₈₋₁₅₈) mostró el mayor MR_{50} (42.81), lo que habla de que esta región aun cuando puede ejercer protección, esta es muy ineficiente.

Posibles mecanismos de acción

La distribución de cargas a lo largo de una proteína pueda tener un papel esencial para su funcionamiento, puesto que muchas interacciones entre proteínas se dan precisamente por

interacciones electrostáticas (Kumar *et al.*, 2002; Furuki *et al.*, 2016; Yang *et al.*, 2020). Para el caso de la proteína LEA4-5, existen residuos cargados positiva y negativamente en el N-terminal, mientras que en el C-terminal existen algunos residuos positivos (Figura 3, Anexo XIII). Además, evolutivamente se ha conservado el extremo N-terminal, región capaz de formar hélices alfa, por lo que es posible pensar que la naturaleza de los residuos y/o la adquisición de estructura en dicha región resulta relevante para la funcionalidad de la proteína. Es razonable sugerir que la presencia de estas características favorece la interacción de las proteínas LEA4 con su blanco y que esto redundaría en la función que las proteínas LEA4 desempeñan.

Se sabe por estudios con fragmentos de IDPs que la interacción de este tipo de proteínas con sus blancos puede ocurrir ya sea mediante un plegamiento inducido por la interacción *per se* con el blanco o mediante un plegamiento inducido por el ambiente, donde la estructura producida podría favorecer la interacción proteica (Uversky, 2013b; Arai *et al.*, 2015). Dada la dificultad de los experimentos con las proteínas LEA4 para obtener estructuras con resolución atómica por su naturaleza desordenada, es complicado establecer si la interacción de la LEA4-5 con la LDH induce un plegamiento en la LEA. Por otro lado, los experimentos de DC donde se simula la falta de moléculas de agua alrededor de la proteína con ayuda de compuestos químicos de diferente naturaleza (glicerol o etilenglicol) (Anexo X) demuestran que la proteína LEA4-5 y algunas versiones mutantes pueden llevar a cabo un plegamiento inducido por el ambiente.

La capacidad del N-terminal de plegarse en estructuras tipo hélices alfa en condiciones de falta de agua, y lo observado con la proteína LEA4-5_{1-77(9P)} comparada con la LEA4-5₁₋₇₇, sugiere la importancia de la formación de hélices alfa en el N-terminal. Sin embargo, lo observado con la proteína LEA4-5-NT9P apunta a lo contrario, ya que en dicha proteína no es necesaria la adquisición de estructura para ejercer la función protectora. Aunque los datos parecen contradictorios, en realidad resaltan un papel fundamental previamente no considerado del extremo C-terminal en la actividad protectora de la proteína.

Es posible pensar que la formación de estructura hélices alfa en el N-terminal de la proteína LEA4-5 sí está relacionada con la capacidad protectora de dicha proteína, pero que dicha relación se encuentra enmascarada por la presencia del extremo C-terminal, resultando innecesarias las estructuras hélices alfa siempre que el C-terminal esté presente. Este tipo de relación resulta muy peculiar y no se tiene conocimiento de algún reporte que mencione a alguna proteína LEA con un comportamiento similar. Además, la idea de una región N-terminal desordenada con capacidad de adquirir estructura y una región C-terminal intrínsecamente desordenada, apoya la idea de que la presencia del desorden puede correlacionarse con la plasticidad de adaptación de estructura y función en diversas condiciones ambientales estresantes además del congelamiento-descongelamiento explorado en esta tesis.

Por lo anterior, es posible proponer un modelo teórico del funcionamiento de estas proteínas. El extremo C-terminal, desordenado, poco cargado e hidrofóbico (Anexo XIV), podría reconocer zonas de proteínas en vías de desnaturalización, donde suelen exponerse algunas regiones hidrofóbicas del blanco. Dicho reconocimiento inicial, colocaría al extremo N-terminal en las cercanías del blanco, favoreciendo interacciones electrostáticas. Las interacciones dadas entre el C-terminal y el N-terminal con el blanco, evitarían la posterior desnaturalización del blanco y/o interrumpiría la formación de agregados moleculares del blanco. Al término del estrés, la proteína LEA perdería su estructura, disminuirían las interacciones con su blanco, se liberaría de éste y quedaría suspendida en el ambiente acuoso en espera de una nueva condición estresante. En ausencia del C-terminal, podría actuar un mecanismo distinto, donde, al no existir un C-terminal que funcione como ancla, el extremo N-terminal interactuaría electrostáticamente con la proteína blanca de forma menos eficiente, pero aun relativamente funcional, lo que se traduciría en observar cierta protección.

LIMITACIONES DEL PROYECTO

Uso de LDH como enzima reportera

Los resultados antes discutidos evidencian un mecanismo de acción propio de las condiciones experimentales *in vitro* utilizadas, en condiciones de laboratorio que reflejan lo que ocurre en la naturaleza. Entre alguna de las consideraciones está la naturaleza de la enzima reportera modelo utilizada, en este caso lactato deshidrogenasa, la cual proviene de músculo de corazón de conejo, que evidentemente no es un blanco real *in planta*. No es descartable la idea de que las proteínas LEA puedan tener distintos mecanismos de acción dependiendo de la proteína con la que interactúen. Se han utilizado otras enzimas modelo para comprender el funcionamiento de las proteínas LEA, como alcohol deshidrogenasa (Lv *et al.*, 2018), citrato sintasa (Goyal *et al.*, 2005), malato deshidrogenasa (Reyes *et al.*, 2005), luciferasa, lisozima (Kovacs *et al.*, 2008), rodanasa, fumarasa (Grelet *et al.*, 2005), o catalasa (Hara *et al.*, 2001). Aun así, los ensayos con LDH son los más comunes, pero en ningún caso se ha confirmado a alguna de estas enzimas como blancos reales de proteínas LEA.

Considerando lo anterior, y teniendo en cuenta que la alteración de residuos cargados o la disminución de la capacidad de formar hélices alfa no afectaron la capacidad protectora de las proteínas mutantes aquí utilizadas, es posible hipotetizar que dichas mutaciones sí podrían afectar la interacción de la proteína LEA con blancos reales, provocando fenotipos de disminución de protección. De ser así, la conservación evolutiva de residuos cargados como probables residuos responsables de la interacción entre proteínas directamente relacionados con la función de las proteínas LEA tiene sentido.

Los tratamientos de congelamiento-descongelamiento como una herramienta para el análisis funcional *in vitro*.

Otra de las consideraciones es la naturaleza misma del tratamiento utilizado. Diferentes aproximaciones *in vitro* intentan simular algún estado celular o condición ambiental que permita comprender el impacto de dichos estados en la estructura y/o función de una proteína. Para el caso de proteínas LEA, dos de las metodologías más usadas para simular estrés hídrico *in vitro* son los experimentos de protección enzimática ante deshidratación y ante ciclos de congelamiento-descongelamiento (Goyal *et al.*, 2005; Reyes *et al.*, 2005; Reyes *et al.*, 2008; Cuevas-Velazquez *et al.*, 2016; Rendon-Luna *et al.*, 2020). Particularmente, el congelamiento-descongelamiento aquí utilizado implicó siete ciclos de congelamiento en nitrógeno líquido y descongelamiento a 25 °C. Se considera que este tratamiento simula estrés hídrico ya que el agua disminuye las interacciones con las proteínas presentes en la solución al congelarse. Estos experimentos no replican lo que puede ocurrir *in vivo*. En primer lugar, en los experimentos *in vitro* no están presentes todos los componentes y compuestos celulares con los que podría estar interaccionando la proteína analizada y que pudiera ayudar a ejercer su función. En segundo, un congelamiento súbito seguido de descongelamiento es algo que no ocurriría de forma natural durante el ciclo de vida de una planta.

Aun así, es innegable que existen proteínas que pueden proteger a enzimas reporteras en estas condiciones artificiales, por lo que resultan una herramienta para poder aproximar la función de dichas proteínas. Estos experimentos han sido utilizados para observar la capacidad protectora de proteínas como: AtLEA4-5, AtLEA4-2 (Reyes *et al.*, 2005; Reyes *et al.*, 2008; Cuevas-Velazquez *et al.*, 2016), ZmLEA3 (Yang Liu *et al.*, 2016), WCI16 (Sasaki *et al.*, 2014), TdLEA3 (Koubaa *et al.*, 2019), CuCOR19 (Hara *et al.*, 2001), DHN-5 (Drira *et al.*, 2013), AtHIRD11 (Yokoyama *et al.*, 2020), OsLEA5 (He *et al.*, 2012), P-80, Dhn5 (Bravo *et al.*, 2003), K₂ (S. Hughes *et al.*, 2011; S. L. Hughes *et al.*, 2013), 23 kDa (Momma, 2019), PCA69 (Wisniewski *et al.*, 1999), AtHIRD11

(Yokoyama *et al.*, 2020), DHN1a (Rosales *et al.*, 2014), ERD10 y RcDHN5 (Reyes *et al.*, 2008), por mencionar algunos. Puede observarse que es una estrategia aplicada comúnmente para comprender el funcionamiento de IDPs, particularmente LEAs, de plantas *in vitro*.

CONCLUSIONES

En el presente trabajo se observó que los residuos conservados de la proteína AtLEA4-5 son de poca importancia para la actividad protectora *in vitro* de la proteína, y que la presencia del extremo C-terminal sí es importante para que la proteína desempeñe completamente su función.

Más aún, la estructura tipo hélice alfa que puede formarse en el extremo N-terminal de la proteína también resultó importante para que la actividad protectora fuera óptima, si bien dicho efecto solamente fue detectable cuando se eliminó el extremo C-terminal desordenado.

Finalmente, la eliminación de pequeñas porciones de los extremos N- o C-terminales de la proteína no afectaron la actividad protectora, mientras que la eliminación de toda la región C-terminal tuvo un impacto en dicha función. Inesperadamente, la eliminación de regiones internas del N-terminal con la mayor propensión a formar hélices alfa, no afectó negativamente la capacidad protectora, contrario a la suposición de una disminución de la actividad protectora.

En conjunto, se determinó que el N-terminal de la proteína LEA4-5, conservado y con capacidad de estructurarse, y la región C-terminal, desordenada, actúan de forma sinérgica para lograr que la proteína LEA4-5 en su conjunto tenga una actividad protectora óptima *in vitro*.

PERSPECTIVAS

Aunque las estrategias seguidas en la presente investigación aportan evidencias más contundentes respecto al funcionamiento de la proteína AtLEA4-5, aún quedan algunas cuestiones de importancia que resolver para establecer un modelo de acción más concreto para estas proteínas.

Si bien se observó la importancia del extremo C-terminal, no es del todo claro si ambas partes de la proteína actúan de forma conjunta o independiente para el reconocimiento del blanco y posterior protección. Para ello, se podrían realizar experimentos de protección en congelamiento-descongelamiento, evaluando a los extremos N-terminal y C-terminal separados, pero en la misma mezcla. Además, podrían existir interacciones intramoleculares entre ambas regiones necesarias para la actividad protectora. Dichas interacciones podrían provocar una compactación general de la proteína, quizá estabilizando la formación de las estructuras hélices alfa del extremo N-terminal. La compactación podría corroborarse por ensayos de FRET en ambientes que simulen déficit hídrico. Por otro lado, el C-terminal podría ser necesario para la interacción con la proteína blanco, lo cual podría corroborarse mediante la mutación de residuos del C-terminal por residuos cargados que volvieran dicha región más hidrofílica y menos propensa a interactuar con proteínas en vías de desarrollo.

Además, debido a que se observó en trabajos previos que algunas proteínas LEA pueden estar fosforiladas *in vivo* (Rivera-Najera *et al.*, 2014), aunado a la presencia considerable de residuos cargados en el extremo N-terminal de proteínas LEA4, se podría evaluar el impacto de las fosforilaciones *in vitro* en la actividad protectora de la proteína.

*Heil sei euch Geweihten!
Ihr dranget durch Nacht.*

...

*Es siegte die Stärke
Und krönet zum Lohn
Die Schönheit und Weisheit
Mit ewiger Kron'.*

Coro final, de la ópera Die Zauberflöte, de Wolfgang Amadeus Mozart

LITERATURA CITADA

- Arai, M., Sugase, K., Dyson, H. J., y Wright, P. E. (2015). "Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding". *Proc Natl Acad Sci U S A*, 112(31), 9614-9619. doi: 10.1073/pnas.1512799112
- Augé, R. M. (2001). "Water relations, drought and vesicular-arbuscular mycorrhizal symbiosis". *Mycorrhiza*, 11(1), 3-42. doi: 10.1007/s005720100097
- Baker, J., Van Dennsteele, C., y Dure, L., 3rd. (1988). "Sequence and characterization of 6 Lea proteins and their genes from cotton". *Plant Mol Biol*, 11(3), 277-291. doi: 10.1007/bf00027385
- Battaglia, M., Olvera-Carrillo, Y., Garcarrubio, A., Campos, F., y Covarrubias, A. A. (2008). "The enigmatic LEA proteins and other hydrophilins". *Plant Physiol.*, 148(1), 6-24. doi: 10.1104/pp.108.120725
- Bies-Etheve, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., Cooke, R., y Delseny, M. (2008). "Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in Arabidopsis thaliana". *Plant Mol. Biol.*, 67(1-2), 107-124. doi: 10.1007/s11103-008-9304-x
- Bravo, L. A., Gallardo, J., Navarrete, A., Olave, N., Martinez, J., Alberdi, M., Close, T. J., y Corcuera, L. J. (2003). "Cryoprotective activity of a cold-induced dehydrin purified from barley". *Physiol Plant*, 118(2), 262-269. doi: 10.1034/j.1399-3054.2003.00060.x
- Bray, E. A. (1993). "Molecular Responses to Water Deficit". *Plant Physiol*, 103(4), 1035-1040. doi: 10.1104/pp.103.4.1035
- Buchanan, B. B., Gruissem, W., y Jones, R. L. (2015). *Biochemistry and molecular biology of plants*: John Wiley & sons.
- Burger, V. M., Gurry, T., y Stultz, C. M. (2014). Intrinsically Disordered Proteins: Where Computation Meets Experiment. *Polymers*, 6(10), 2684-2719. doi:10.3390/polym6102684
- Campos, F., Guillén, G., Reyes, J. L., y Covarrubias, A. A. (2011). "A general method of protein purification for recombinant unstructured non-acidic proteins". *Protein Expr. Purif.*, 80(1), 47-51. doi: 10.1016/j.pep.2011.06.007
- Carillo, P., Annunziata, M. G., Pontecorvo, G., Fuggi, A., y Woodrow, P. (2011). "Salinity stress and salt tolerance". *Abiotic stress in plants-Mechanisms and adaptations*, 1, 21-38. doi: 10.5772/22331
- Chemes, L. B., Alonso, L. G., Noval, M. G., y de Prat-Gay, G. (2012). "Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains". *Methods Mol Biol*, 895, 387-404. doi: 10.1007/978-1-61779-927-3_22
- Chouard, T. (2011). "Structural biology: Breaking the protein rules". *Nature*, 471(7337), 151-153. doi: 10.1038/471151a
- Contreras-Martos, S., Nguyen, H. H., Nguyen, P. N., Hristozova, N., Macossay-Castillo, M., Kovacs, D., Bekesi, A., Oemig, J. S., Maes, D., Pauwels, K., Tompa, P., y Lebrun, P. (2018).

- "Quantification of intrinsically disordered proteins: A problem not fully appreciated". *Front. Mol. Biosci.*, 5, 83. doi: 10.3389/fmolb.2018.00083
- Cuevas-Velazquez, C. L., Rendón-Luna, D. F., y Covarrubias, A. A. (2014). "Dissecting the cryoprotection mechanisms for dehydrins". *Front. Plant Sci.*, 5, 583. doi: 10.3389/fpls.2014.00583
- Cuevas-Velazquez, C. L., Reyes, J. L., y Covarrubias, A. A. (2017). "Group 4 late embryogenesis abundant proteins as a model to study intrinsically disordered proteins in plants". *Plant Signal. Behav.*, 12(7), 10893-10903. doi: 10.1080/15592324.2017.1343777
- Cuevas-Velazquez, C. L., Saab-Rincón, G., Reyes, J. L., y Covarrubias, A. A. (2016). "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(20), 10893-10903. doi: 10.1074/jbc.M116.720318
- Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., y Flouri, T. (2020). "ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models". *Mol Biol Evol*, 37(1), 291-294. doi: 10.1093/molbev/msz189
- Drira, M., Saibi, W., Brini, F., Gargouri, A., Masmoudi, K., y Hanin, M. (2013). "The K-segments of the wheat dehydrin DHN-5 are essential for the protection of lactate dehydrogenase and beta-glucosidase activities in vitro". *Mol Biotechnol*, 54(2), 643-650. doi: 10.1007/s12033-012-9606-8
- Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., *et al.* (2001). "Intrinsically disordered protein". *J Mol Graph Model*, 19(1), 26-59. doi: 10.1016/s1093-3263(00)00138-8
- Dunker, A. K., Silman, I., Uversky, V. N., y Sussman, J. L. (2008). "Function and structure of inherently disordered proteins". *Curr Opin Struct Biol*, 18(6), 756-764. doi: 10.1016/j.sbi.2008.10.002
- Duong, V. T., Chen, Z., Thapa, M. T., y Luo, R. (2018). "Computational Studies of Intrinsically Disordered Proteins". *J Phys Chem B*, 122(46), 10455-10469. doi: 10.1021/acs.jpcc.8b09029
- Dure, L., 3rd, Greenway, S. C., y Galau, G. A. (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. doi: 10.1021/bi00517a033
- Dure, L., y Chlan, C. (1981). "Developmental biochemistry of cottonseed embryogenesis and germination : XII. Purification and properties of principal storage proteins". *Plant Physiol.*, 68(1), 180-186. doi: 10.1104/pp.68.1.180
- Dure, L., III, Crouch, M., Harada, J., Ho, T.-H., Mundy, J., Quatrano, R., Thomas, T., y Sung, Z. R. (1989). "Common amino acid sequence domains among the LEA proteins of higher plants". *Plant Mol Biol*, 12(5), 475-486. doi: 10.1007/BF00036962
- Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput". *Nucleic Acids Res.*, 32(5), 1792-1797. doi: 10.1093/nar/gkh340

- Emenecker, R. J., Griffith, D., y Holehouse, A. S. (2021). "Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure". *Biophys. J.*, 120(20), 4312-4319. doi: 10.1016/j.bpj.2021.08.039
- Fang, Y., y Xiong, L. (2015). "General mechanisms of drought response and their application in drought resistance improvement in plants". *Cell Mol Life Sci*, 72(4), 673-689. doi: 10.1007/s00018-014-1767-0
- Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., y Bateman, A. (2016). "The Pfam protein families database: towards a more sustainable future". *Nucleic Acids Res.*, 44(D1), D279-D285. doi: 10.1093/nar/gkv1344
- Fu, L., Niu, B., Zhu, Z., Wu, S., y Li, W. (2012). "CD-HIT: accelerated for clustering the next-generation sequencing data". *Bioinformatics*, 28(23), 3150-3152. doi: 10.1093/bioinformatics/bts565
- Furuki, T., y Sakurai, M. (2016). "Group 3 LEA protein model peptides protect enzymes against desiccation stress". *Biochim Biophys Acta*. doi: 10.1016/j.bbapap.2016.04.012
- Galau, G. A., y Dure, L., 3rd. (1981). "Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by reciprocal heterologous complementary deoxyribonucleic acid--messenger ribonucleic acid hybridization". *Biochemistry*, 20(14), 4169-4178.
- Garay-Arroyo, A., Colmenero-Flores, J. M., Garcarrubio, A., y Covarrubias, A. A. (2000). "Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit". *Journal of Biological Chemistry*, 275(8), 5668-5674. doi: 10.1074/jbc.275.8.5668
- Gaspar, T., Franck, T., Bisbis, B., Kevers, C., Jouve, L., Hausman, J. F., y Dommes, J. (2002). "Concepts in plant stress physiology. Application to plant tissue cultures". *Plant Growth Regulation*, 37(3), 263-285. doi: 10.1023/A:1020835304842
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., y Rokhsar, D. S. (2012). "Phytozome: a comparative platform for green plant genomics". *Nucleic Acids Res.*, 40(Database issue), D1178-1186. doi: 10.1093/nar/gkr944
- Goyal, K., Walton, L. J., y Tunnacliffe, A. (2005). "LEA proteins prevent protein aggregation due to water stress". *Biochem. J.*, 388(1), 151-157. doi: 10.1042/bj20041931
- Greenfield, N. J. (2006). "Using circular dichroism spectra to estimate protein secondary structure". *Nature Protocols*, 1(6), 2876-2890. doi: 10.1038/nprot.2006.202
- Grelet, J., Benamar, A., Teyssier, E., Avelange-Macherel, M.-H., Grunwald, D., y Macherel, D. (2005). "Identification in Pea Seed Mitochondria of a Late-Embryogenesis Abundant Protein Able to Protect Enzymes from Drying". *Plant Physiol*, 137(1), 157-167. doi: 10.1104/pp.104.052480

- Hara, M., Terashima, S., y Kuboi, T. (2001). "Characterization and cryoprotective activity of cold-responsive dehydrin from Citrus unshiu". *Journal of Plant Physiology*, 158(10), 1333-1339. doi: Doi 10.1078/0176-1617-00600
- He, S., Tan, L., Hu, Z., Chen, G., Wang, G., y Hu, T. (2012). "Molecular characterization and functional analysis by heterologous expression in E. coli under diverse abiotic stresses for OsLEA5, the atypical hydrophobic LEA protein from Oryza sativa L". *Mol Genet Genomics*, 287(1), 39-54. doi: 10.1007/s00438-011-0660-x
- Hernández-Sánchez, I. E., Maruri-López, I., Martínez-Martínez, C., Janis, B., Jiménez-Bremont, J. F., Covarrubias, A. A., Menze, M. A., Graether, S. P., y Thalhammer, A. (2022). "LEAing through literature: late embryogenesis abundant proteins coming of age—achievements and perspectives". *J. Exp. Bot.*, 73(19), 6525-6546. doi: 10.1093/jxb/erac293
- Hincha, D. K., y Thalhammer, A. (2012). "LEA proteins: IDPs with versatile functions in cellular dehydration tolerance". *Biochem. Soc. Trans.*, 40(5), 1000-1003. doi: 10.1042/bst20120109
- Horton, R. M., Cai, Z. L., Ho, S. N., y Pease, L. R. (1990). "Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction". *Biotechniques*, 8(5), 528-535. doi: 10.2144/000114017
- Hughes, S., y Graether, S. P. (2011). "Cryoprotective mechanism of a small intrinsically disordered dehydrin protein". *Protein Sci*, 20(1), 42-50. doi: 10.1002/pro.534
- Hughes, S. L., Schart, V., Malcolmson, J., Hogarth, K. A., Martynowicz, D. M., Tralman-Baker, E., Patel, S. N., y Graether, S. P. (2013). "The importance of size and disorder in the cryoprotective effects of dehydrins". *Plant Physiol*, 163(3), 1376-1386. doi: 10.1104/pp.113.226803
- Hundertmark, M., y Hincha, D. K. (2008). "LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana". *BMC Genom.*, 9, 118. doi: 10.1186/1471-2164-9-118
- Jeffery, C. J. (2017). "Protein moonlighting: what is it, and why is it important?". *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1738), 20160523. doi: 10.1098/rstb.2016.0523
- Jones, R., Ougham, H., Thomas, H., y Waaland, S. (2013). "The molecular life of plants".
- Kelly, S. M., Jess, T. J., y Price, N. C. (2005). "How to study proteins by circular dichroism". *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1751(2), 119-139. doi: 10.1016/j.bbapap.2005.06.005
- Koubaa, S., Bremer, A., Hincha, D. K., y Brini, F. (2019). "Structural properties and enzyme stabilization function of the intrinsically disordered LEA_4 protein TdLEA3 from wheat". *Sci Rep*, 9(1), 3720. doi: 10.1038/s41598-019-39823-w
- Kovacs, D., Kalmar, E., Torok, Z., y Tompa, P. (2008). "Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins". *Plant Physiol*, 147(1), 381-390. doi: 10.1104/pp.108.118208

- Kumar, S., y Nussinov, R. (2002). "Close-range electrostatic interactions in proteins". *ChemBioChem*, 3(7), 604-617. doi: 10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X
- LeBlanc, B. M., y Hand, S. C. (2021). "Target enzymes are stabilized by AfrLEA6 and a gain of alpha-helix coincides with protection by a group 3 LEA protein during incremental drying". *Biochim Biophys Acta Proteins Proteom*, 1869(6), 140642. doi: 10.1016/j.bbapap.2021.140642
- Li, W., y Godzik, A. (2006). "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences". *Bioinformatics*, 22(13), 1658-1659. doi: 10.1093/bioinformatics/btl158
- Liu, Y., Liang, J., Sun, L., Yang, X., y Li, D. (2016). "Group 3 LEA Protein, ZmLEA3, Is Involved in Protection from Low Temperature Stress". 7. doi: 10.3389/fpls.2016.01011
- Liu, Y., Wang, X., y Liu, B. (2019). "A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction". *Brief Bioinform*, 20(1), 330-346. doi: 10.1093/bib/bbx126
- Lv, A., Su, L., Liu, X., Xing, Q., Huang, B., An, Y., y Zhou, P. (2018). "Characterization of Dehydrin protein, CdDHN4-L and CdDHN4-S, and their differential protective roles against abiotic stress in vitro". *BMC Plant Biol*, 18(1), 299. doi: 10.1186/s12870-018-1511-2
- Momma, M. (2019). "Detection and Cryoprotective Activity of Dehydrin Proteins from Rice Bran and Soybean Whey". *Journal of Nutritional Science and Vitaminology*, 65(Supplement), S129-S133. doi: 10.3177/jnsv.65.S129
- Na, J.-H., Lee, W.-K., y Yu, Y. G. (2018). How Do We Study the Dynamic Structure of Unstructured Proteins: A Case Study on Nopp140 as an Example of a Large, Intrinsically Disordered Protein. *Int J Mol Sci*, 19(2). doi:10.3390/ijms19020381
- Navarro-Retamal, C., Bremer, A., Alzate-Morales, J., Caballero, J., Hinch, D. K., Gonzalez, W., y Thalhammer, A. (2016). "Molecular dynamics simulations and CD spectroscopy reveal hydration-induced unfolding of the intrinsically disordered LEA proteins COR15A and COR15B from Arabidopsis thaliana". *Phys Chem Chem Phys*, 18(37), 25806-25816. doi: 10.1039/c6cp02272c
- Olvera-Carrillo, Y., Campos, F., Reyes, J. L., Garcarrubio, A., y 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(1), 373-390. doi: 10.1104/pp.110.158964
- Olvera-Carrillo, Y., Luis Reyes, J., y Covarrubias, A. A. (2011). "Late embryogenesis abundant proteins: versatile players in the plant adaptation to water limiting environments". *Plant Signaling & Behavior*, 6(4), 586-589. doi: 10.4161/psb.6.4.15042
- Piovesan, D., Walsh, I., Minervini, G., y Tosatto, S. C. E. (2017). "FELLS: fast estimator of latent local structure". *Bioinformatics*, 33(12), 1889-1891. doi: 10.1093/bioinformatics/btx085
- Rendon-Luna, D. F., Romero-Perez, P. S., Cuevas-Velazquez, C. L., Reyes, J. L., y Covarrubias, A. A. (2020). "Determining the protective activity of IDPs under partial dehydration and

- freeze-thaw conditions". *Methods Mol. Biol.*, 2141, 519-528. doi: 10.1007/978-1-0716-0524-0_26
- Reyes, J. L., Campos, F., Wei, H., Arora, R., Yang, Y., Karlson, D. T., y Covarrubias, A. A. (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, J. L., Rodrigo, M.-J., Colmenero-Flores, J. M., Gil, J.-V., Garay-Arroyo, A., Campos, F., Salamini, F., Bartels, D., y Covarrubias, A. A. (2005). "Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects in vitro". *Plant Cell Environ.*, 28(6), 709-718. doi: 10.1111/j.1365-3040.2005.01317.x
- Riley, A. C., Ashlock, D. A., y Graether, S. P. (2023). "The difficulty of aligning intrinsically disordered protein sequences as assessed by conservation and phylogeny". *PLoS One*, 18(7), e0288388. doi: 10.1371/journal.pone.0288388
- Rivera-Najera, L. Y., Saab-Rincon, G., Battaglia, M., Amero, C., Pulido, N. O., Garcia-Hernandez, E., Solorzano, R. M., Reyes, J. L., y 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* *Iskdhgñasdlgk*, 289(46), 31995-32009. doi: 10.1074/jbc.M114.583369
- Romero-Perez, S. P., Covarrubias, A. A., y Campos, F. (2023). "A simple method to purify intrinsically disordered proteins by adjusting trichloroacetic acid concentration". *Protein Expr. Purif.*, 202, 106183. doi: 10.1016/j.pep.2022.106183
- Rosales, R., Romero, I., Escribano, M. I., Merodio, C., y Sanchez-Ballesta, M. T. (2014). "The crucial role of Phi- and K-segments in the in vitro functionality of Vitis vinifera dehydrin DHN1a". *Phytochemistry*, 108, 17-25. doi: 10.1016/j.phytochem.2014.10.006
- Sasaki, K., Christov, N. K., Tsuda, S., y Imai, R. (2014). "Identification of a Novel LEA Protein Involved in Freezing Tolerance in Wheat". *Plant & Cell Physiology*, 55(1), 136-147. doi: 10.1093/pcp/pct164
- Sharma, R., Raicar, G., Tsunoda, T., Patil, A., y Sharma, A. (2018). "OPAL: prediction of MoRF regions in intrinsically disordered protein sequences". *Bioinformatics*, 34(11), 1850-1858. doi: 10.1093/bioinformatics/bty032
- Shih, M. D., Hsieh, T. Y., Lin, T. P., Hsing, Y. I., y 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(3), 395-407. doi: 10.1093/pcp/pcq005
- Shortle, D. (1992). "Mutational studies of protein structures and their stabilities". *Quarterly Reviews of Biophysics*, 25(2), 205-250. doi: 10.1017/S0033583500004674
- Stamatakis, A. (2006). "RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models". *Bioinformatics*, 22(21), 2688-2690. doi: 10.1093/bioinformatics/btl446
- Tompa, P., y Fersht, A. (2009). *Structure and function of intrinsically disordered proteins*: CRC press.

- Tunnacliffe, A., y Wise, M. J. (2007). "The continuing conundrum of the LEA proteins". *Naturwissenschaften*, 94(10), 791-812. doi: 10.1007/s00114-007-0254-y
- Uversky, V. N. (2011). "Intrinsically disordered proteins from A to Z". *The International Journal of Biochemistry & Cell Biology*, 43(8), 1090-1103. doi: 10.1016/j.biocel.2011.04.001
- Uversky, V. N. (2013a). "A decade and a half of protein intrinsic disorder: biology still waits for physics". *Protein Sci*, 22(6), 693-724. doi: 10.1002/pro.2261
- Uversky, V. N. (2013b). "Unusual biophysics of intrinsically disordered proteins". *Biochim. Biophys. Acta.*, 1834(5), 932-951. doi: 10.1016/j.bbapap.2012.12.008
- Uversky, V. N. (2016). "p53 Proteoforms and Intrinsic Disorder: An Illustration of the Protein Structure-Function Continuum Concept". *Int J Mol Sci*, 17(11). doi: 10.3390/ijms17111874
- Vashisht, A. A., y Tuteja, N. (2006). "Stress responsive DEAD-box helicases: a new pathway to engineer plant stress tolerance". *J Photochem Photobiol B*, 84(2), 150-160. doi: 10.1016/j.jphotobiol.2006.02.010
- Whitmore, L., y Wallace, B. A. (2004). "DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data". *Nucleic Acids Res.*, 32(Web Server issue), W668-673. doi: 10.1093/nar/gkh371
- Wisniewski, M., Webb, R., Balsamo, R., Close, T. J., Yu, X.-M., y Griffith, M. (1999). "Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: A dehydrin from peach (*Prunus persica*)". *Physiol Plant*, 105(4), 600-608. doi: 10.1034/j.1399-3054.1999.105402.x
- Woody, R. W. (2010). "Circular dichroism of intrinsically disordered proteins". *Instrumental analysis of intrinsically disordered proteins: Assessing structure and conformation*, 303-321. doi: 10.1002/9780470602614.ch10
- Wu, G., Zhang, C., Chu, L.-Y., y Shao, H.-B. (2007). "Responses of higher plants to abiotic stresses and agricultural sustainable development". *Journal of Plant Interactions*, 2(3), 135-147. doi: 10.1080/17429140701586357
- Yang, J., Zeng, Y., Liu, Y., Gao, M., Liu, S., Su, Z., y Huang, Y. (2020). "Electrostatic interactions in molecular recognition of intrinsically disordered proteins". *J Biomol Struct Dyn*, 38(16), 4883-4894. doi: 10.1080/07391102.2019.1692073
- Yokoyama, T., Ohkubo, T., Kamiya, K., y Hara, M. (2020). "Cryoprotective activity of Arabidopsis KS-type dehydrin depends on the hydrophobic amino acids of two active segments". *Archives of Biochemistry and Biophysics*, 691, 108510. doi: 10.1016/j.abb.2020.108510
- Yordanov, I., Velikova, V., y Tsonev, T. (2000). "Plant Responses to Drought, Acclimation, and Stress Tolerance". *Photosynthetica*, 38(2), 171-186. doi: 10.1023/A:1007201411474

ANEXOS

Anexo I. Secuencia de oligonucleótidos usados para obtener los genes que codifican para las proteínas mutantes. Las secuencias en cursivas y subrayadas corresponden a sitios de restricción.

Las secuencias en negrita y subrayadas corresponden a codones modificados

Nombre	Secuencia (5' → 3')
TRC45F	AAA <u>CCA TGG</u> AGT CGA TGA AAG AAA C
TRC45R	GCG <u>GTC GAC</u> CCG TTT ATC CAG TAT ATC C
45F-27a32	CAA AAC CAA AGC TAC CTT <u>GGC GGA AGC GGC GGC GAT</u> GAA GAC ACG AGA CCC
45R-27a32	GGG TCT CGT GTC TTC AGC <u>GCC GCC GCC GCT TCC GCC</u> AAG GTA GCT TTG GTT TTG
45F-41a51	CGA GAC CCT GTT CAG <u>GCG</u> CAG ATG <u>AAA</u> ACA CAG GTT AAA GAA GAT <u>GCG</u> ATC AAT CAG GCT GAG
45R-41a51	CTC AGC CTG ATT GAT <u>CGC</u> ATC TTC <u>TTT</u> AAC CTG TGT TTT CAT CTG <u>CGC</u> CTG AAC AGG GTC TCG
45F-A44K	CCT GTT CAG AAA CAG ATG <u>AAA</u> ACA CAG GTT AAA GAA GAT AAG
45R-A44K	CTT ATC TTC TTT AAC CTG TGT <u>TTT</u> CAT CTG TTT CTG AAC AGG
45F-54y61	GAA GAT AAG ATC AAT <u>GCG</u> GCT GAG ATG CAG AAG AGA <u>GCG</u> ACG CGT CAG CAC AAC
45R-54y61	GTT GTG CTG ACG CGT <u>CGC</u> TCT CTT CTG CAT CTC AGC <u>CGC</u> ATT GAT CTT ATC TTC
MtL45P12Fw	CGA ATA TTG CAG CTT <u>CTC CAA</u> AAT CTG GCA TGG ACA AAA CCA <u>AAC CTA</u> CCT TGG AGG AAA AG
MtL45P12Rv	CTT TTC CTC CAA GGT <u>AGG</u> TTT GGT TTT GTC CAT GCC AGA TTT <u>TGG</u> AGA AGC TGC AAT ATT CG
MtL45P34Fw	CTG TTC AGA AAC AGA <u>TGC CTA</u> CAC AGG TTA AAG AAG ATA AGA TCA ATC <u>AAC CTG</u> AGA TGC AGA AGA GAG
MtL45P34Rv	CTC TCT TCT GCA TCT <u>CAG GTT</u> GAT TGA TCT TAT CTT CTT TAA CCT GTG <u>TAG GCA</u> TCT GTT TCT GAA CAG
M33LAPNonu-F	AAA CCA TGG AGT CGA TGA AAG AAC CAG CTT CGA ATC CTG CAG CTT CTC C
M33LAPNonu-R	GGC GTC GAC TCA GGT TCC GGC TCC AGC CGC TTC TGG CAT GGC CGC GTT GTG CTG AGG CGT TTC TCT TGG CTG CAT CTC
M33LAPNonu-R-no-ter	GGC GTC GAC <u>ACC</u> GGT TCC GGC TCC AGC CGC TTC TGG CAT GGC CGC GTT GTG CTG AGG CGT TTC TCT TGG CTG CAT CTC
SD26.Fw	TAC <u>CAT GGA</u> GGA AAA GGC GGA G
M430.Fw	AAA <u>CCA TGG</u> CTA CAC AGG TTA AAG
LEA4M57.Fw	AAA <u>CCA TGG</u> AGA AGA GAG AAA CGC G
H61RV	GGT GGT GTC GAC TCA TTC TCT CTT CTG CAT CTC AG
Moxo1.Rv	GCG <u>GTC GAC</u> TTA CCC GGT GGC ACC ACC
Moxo2.Rv	GCG <u>GTC GAC</u> TTA CGG GTC GGT CAC AGC CG
FwGap	GAT GCA GAA GAG ATT AGG TTT GGG GAC GGC C

GapRv	GCT GAG ATG CAG AAG AGA TTA GGT TTG GGG
F43Gap	CAG AAA CAG ATG TTA GGT TTG GGG ACG GCC
Gap43R	CCC TGT TCA GAA ACA GAT GTT AGG TTT GGG G
TYB11NonuF	GGT GGT GGT <u>TGC TCT TCC</u> AAC ATG GAG TCG ATG AAA GAA CCA GC
TYB11NonuR	GGT GGT GGT <u>CCC GGG</u> TCA GGT TCC GGC TCC AGC CGC TTC TGG C
TRC4RC-F	AAA CCA TGG CCG GTT TAG GTT TGG GGA C
CLEA4H-R	TTA TCC GGC TCC AGC CGC TCC TTT CAT GGC
pTRC99a_R	GCG TTC TGA TTT AAT CTG TAT CAG G

Anexo II. Pares de oligonucleótidos utilizados para generar los fragmentos de genes necesarios para el PCR de traslape.

Proteína mutante	Oligonucleótidos para el segmento A		Oligonucleótidos para el segmento B	
	Forward	Reverse	Forward	Reverse
LEA4-5-NT1	TRC45F	45R-27a32	45F-27a32	TRC45R
LEA4-5-NT2	TRC45F	45R-41a51	45F-41a51	TRC45R
LEA4-5-NT3	TRC45F	45R-A44K	45F-A44K	TRC45R
LEA4-5-NT4	TRC45F	45R-54y61	45F-54y61	TRC45R
LEA4-5-NT2P	TRC45F	MtL45P34Rv	MtL45P34Fw	TRC45R
LEA4-5-NT4P^a	TRC45F	MtL45P12Rv	MtL45P12Fw	TRC45R
LEA4-5-NT9P^b	TRC45F	M33LAPNonu-R-no-ter	M33LAPNonu-F-no-ter	TRC45R
LEA4-5_{Δ61-78}	TRC45F	GapRv	FwGap	TRC45R
LEA4-5_{Δ44-78}	TRC45F	Gap43R	F43Gap	TRC45R

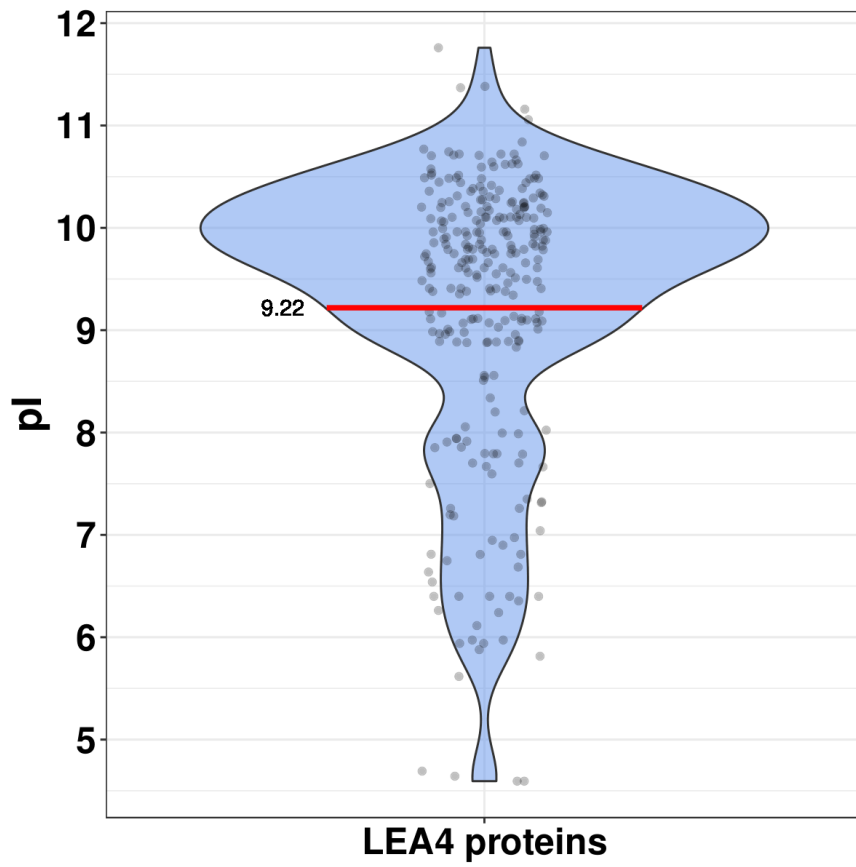
^a El DNA templado para esta mutante fue el gen de la mutante LEA4-5-NT2P.

^b El DNA templado para esta mutante fue el gen de la mutante LEA4-5-NT4P.

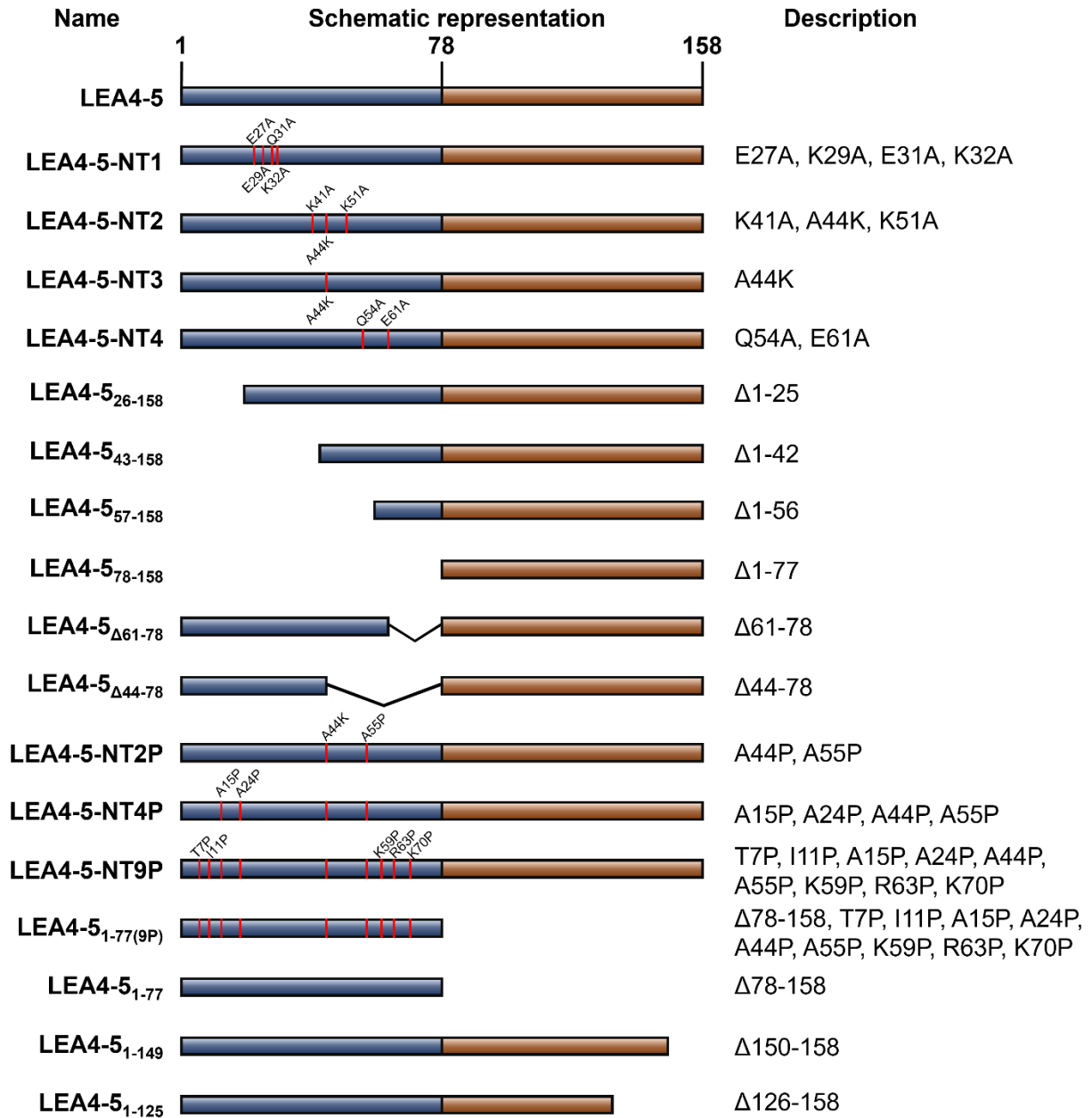
Anexo III. Pares de oligonucleótidos utilizados para generar los amplicones truncados por PCR de punto final.

Proteína mutante	Oligonucleótido Forward	Oligonucleótido Reverse
LEA4-5₂₆₋₁₅₈	SD26.Fw	TRC45R
LEA4-5₄₃₋₁₅₈	M430.Fw	TRC45R
LEA4-5₅₇₋₁₅₈	LEA4M57.Fw	TRC45R
LEA4-5₇₈₋₁₅₈	TRC4RC-F	TRC45R
LEA4-5₁₋₁₄₉	TRC45F	Moxo1.Rv
LEA4-5₁₋₁₂₅	TRC45F	Moxo2.Rv
LEA4-5₁₋₇₇	TRC45F	CLEA4H-R
LEA4-5_{1-77(9P)}^a	M33LAPNonu-F TYB11NonuF	M33LAPNonu-R TYB11NonuR

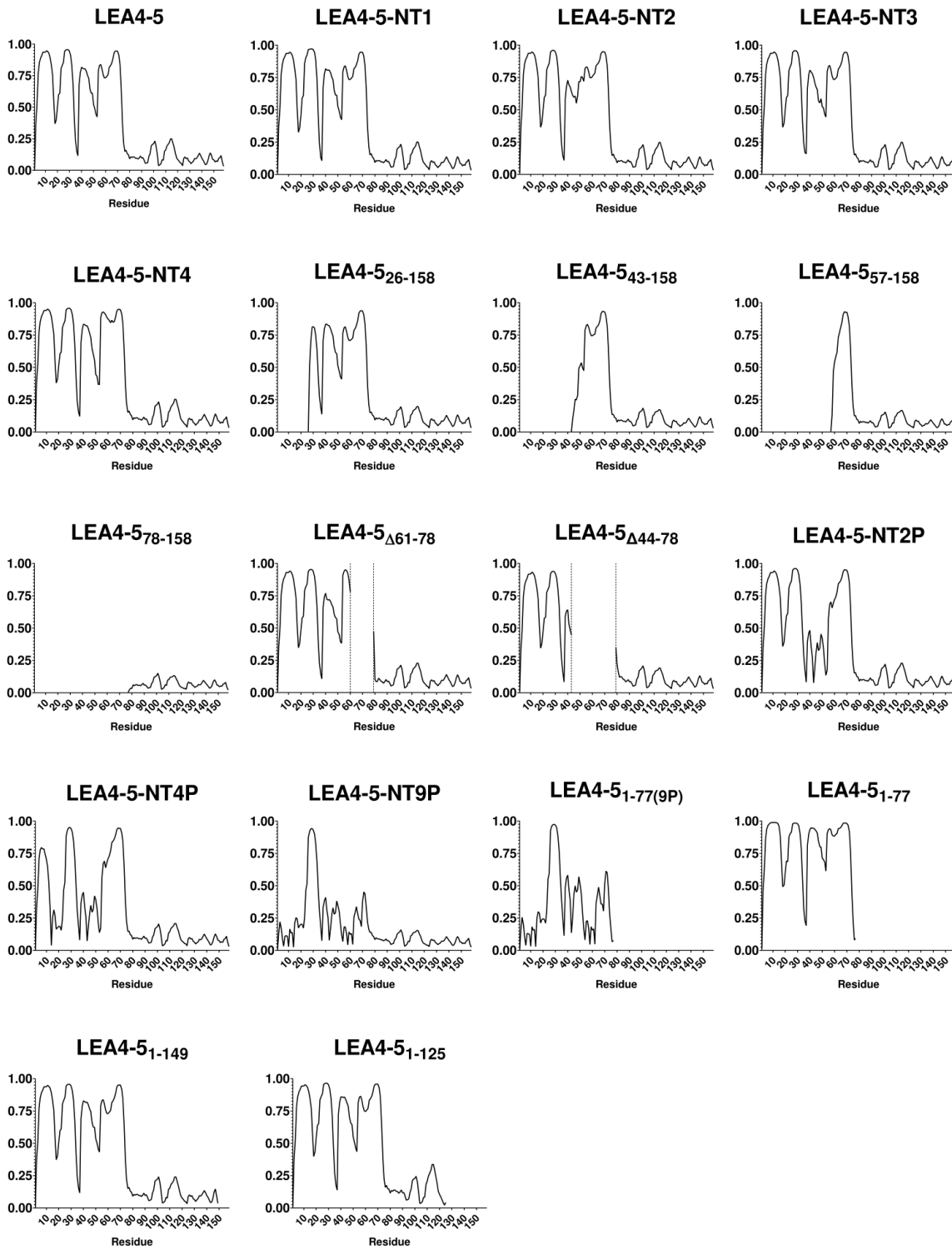
^aEl templado de DNA para esta mutante fue el gen LEA4-5-NT9P. Esta mutante se obtuvo en dos rondas de PCR. Primero, usando el primer par de oligonucleótidos, reemplazamos el codón 78 (GGT) por un codón de paro (TGA). Luego, el segundo par de oligonucleótidos se usaron para añadir sitios de restricción para poder clonar los amplicones en el vector de expresión pTYB11.



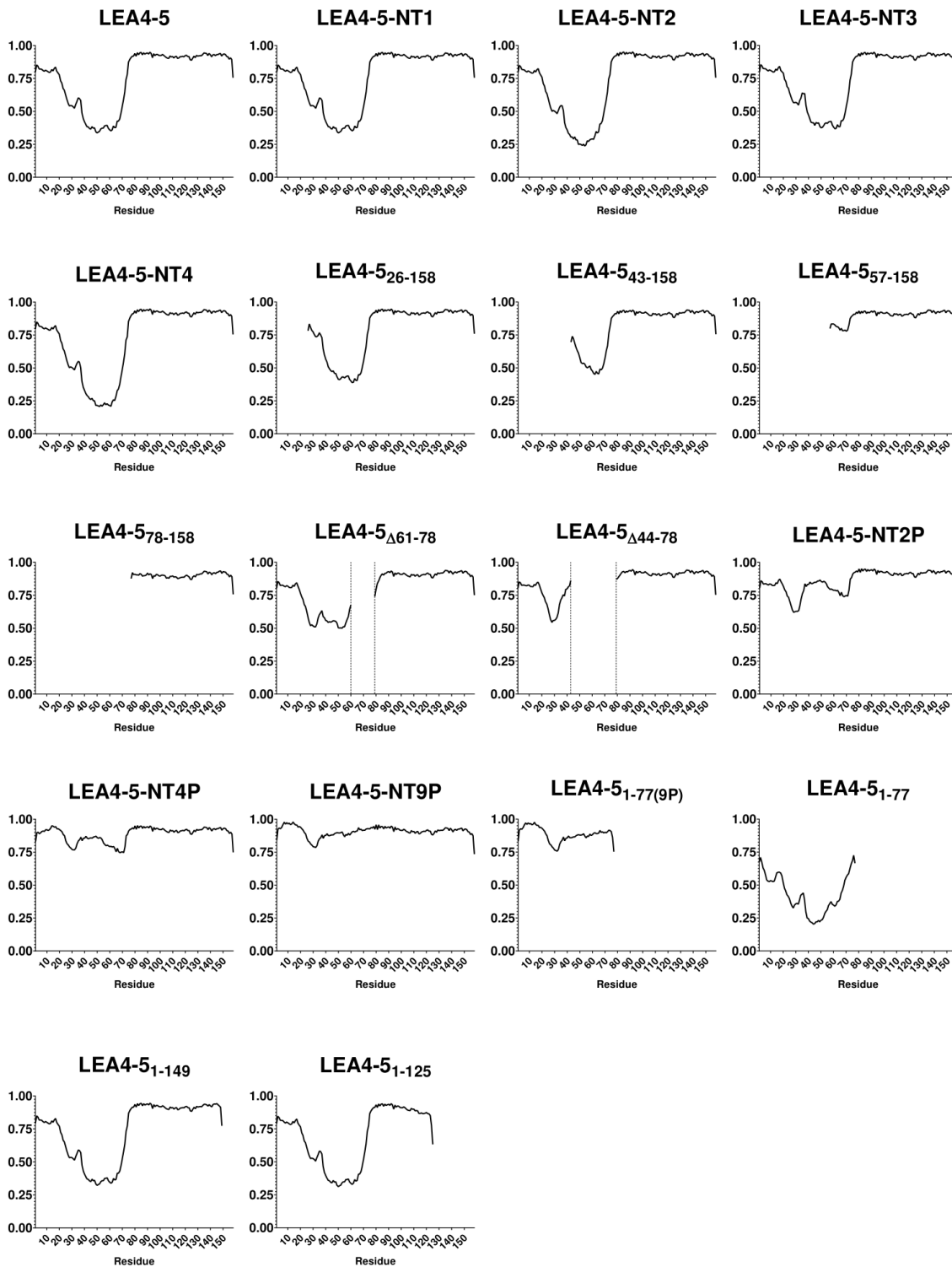
Anexo IV. Distribución de los pI de las diferentes secuencias de proteínas LEA4. Se marca en rojo el pI promedio.



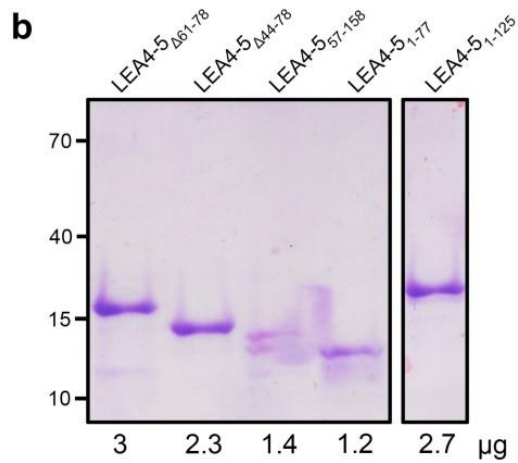
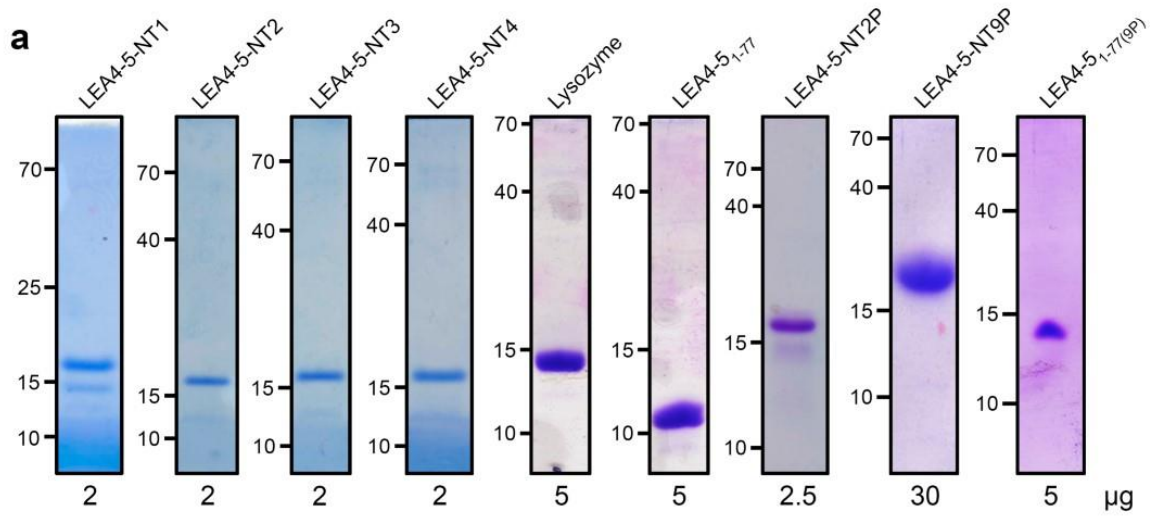
Anexo V. Esquema general de todas las mutaciones producidas en este estudio. Se muestra una descripción de las mutaciones para una mejor identificación de las modificaciones.



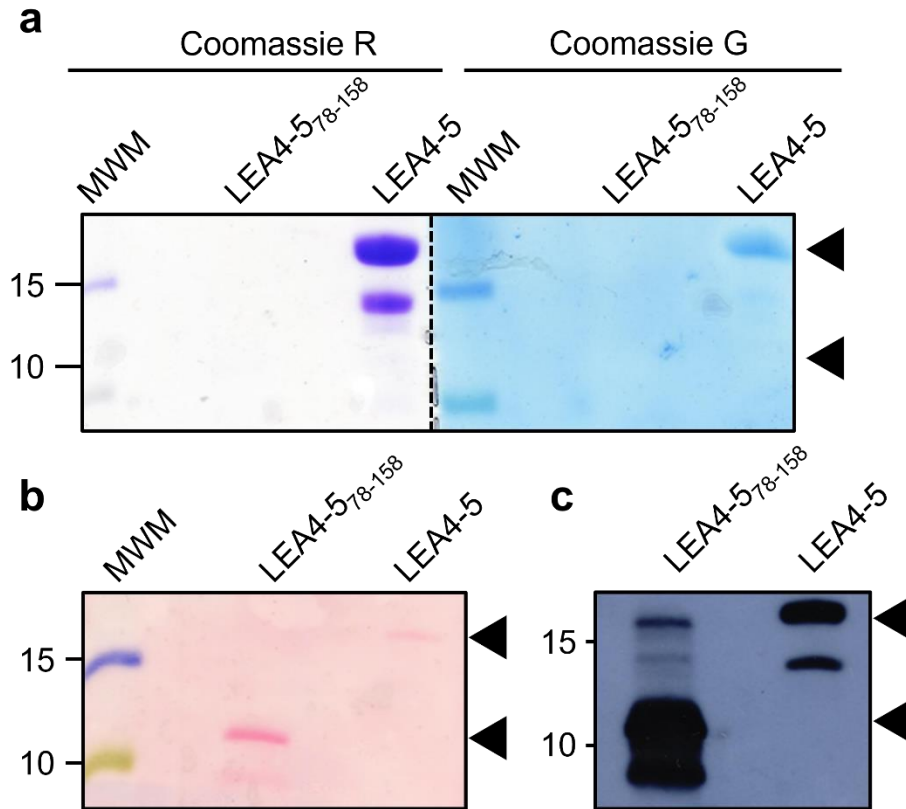
Anexo VI. Predicción bioinformática de la propensión de formación de hélices alfa, usando el algoritmo FELLs. El eje Y muestra el valor de propensión asignado por el algoritmo.



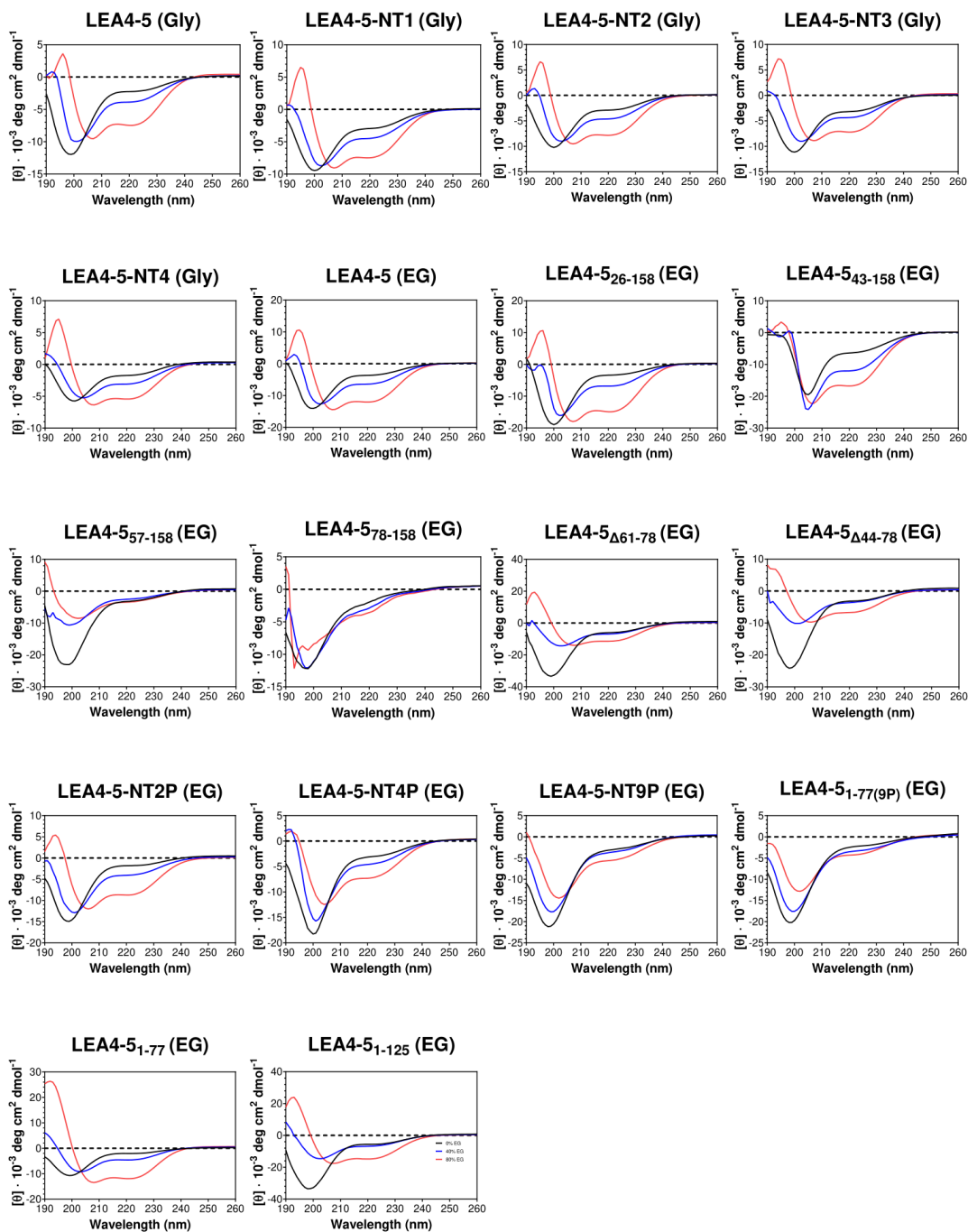
Anexo VII. Predicción bioinformática de desorden estructural, usando el algoritmo **Metapredict**. El eje Y muestra el valor de desorden asignado por el algoritmo.



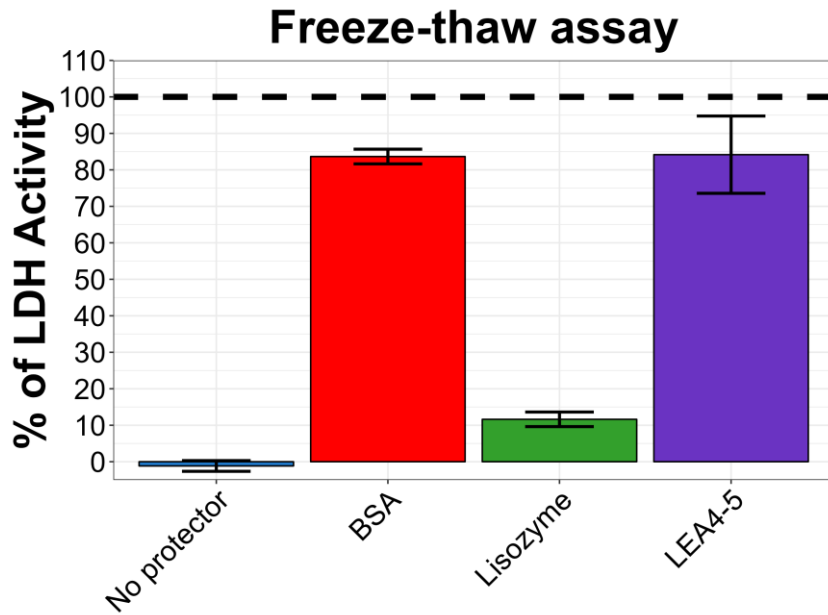
Anexo VIII. SDS-PAGE de proteínas mutantes. Se verificó la pureza de las proteínas mutantes por SDS-PAGE y tinción por Coomassie (G o R). Las proteínas se cuantificaron por el método de Qubit. La cantidad de proteína cargada en cada carril se indica debajo de cada panel. Las imágenes de muestras que fueron separadas en diferentes geles se muestran de forma individual. El carril completo se muestra en todos los casos. Los números a la izquierda corresponden a la migración del marcador de peso molecular en kDa (Prestained Protein Ladder, Thermo Fischer Scientific) para cada gel.



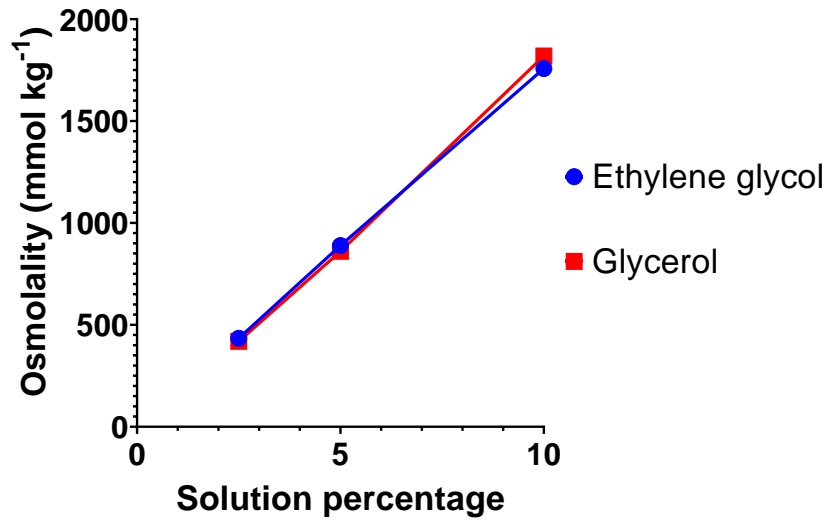
Anexo IX. Verificación de pureza de la proteína LEA4-5₇₈₋₁₅₈. a) La tinción por azul de Coomassie R o Coomassie G no tiñe a la proteína LEA4-5₇₈₋₁₅₈, lo cual es característico de la región C-terminal de la LEA4-5. La imagen muestra la tinción del marcador de peso molecular y de la proteína LEA4-5 (5 µg) pero no de la LEA4-5₇₈₋₁₅₈ (5.5 µg). b) Durante el experimento Western Blot, se observó que la LEA4-5₇₈₋₁₅₈ se tiñó con Rojo Ponceau después de la transferencia a membrana de nitrocelulosa (tamaño de poro de 0.2 µm); la banda correspondiente a la proteína LEA4-5 también se detectó débilmente durante la tinción. c) Experimento de Western Blot realizado con anticuerpo primario específico para la región C-terminal (1:2,500) y anticuerpo secundario anti-conejo acoplado a HRP. Las cabezas de flechas negras apuntan a la migración de las proteínas LEA4-5 y LEA4-5₇₈₋₁₅₈. MWM: *Molecular weight marker* (marcador de peso molecular) en kDa.



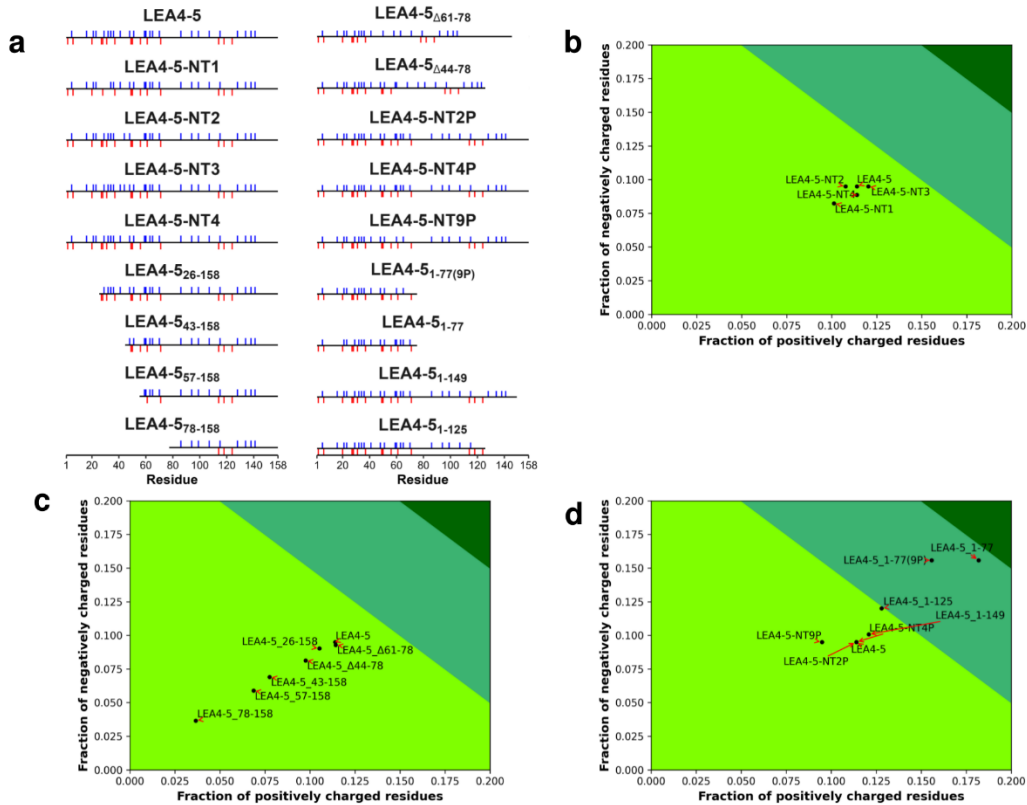
Anexo X. Espectros de dicroísmos circulares obtenidos en este estudio. Gly y EG se refieren a glicerol y etilenglicol, respectivamente. Los espectros negros, azules y rojos representan concentraciones de 0%, 40% y 80% del inductor de déficit hídrico. $[\Phi]$ = Elipticidad molar.



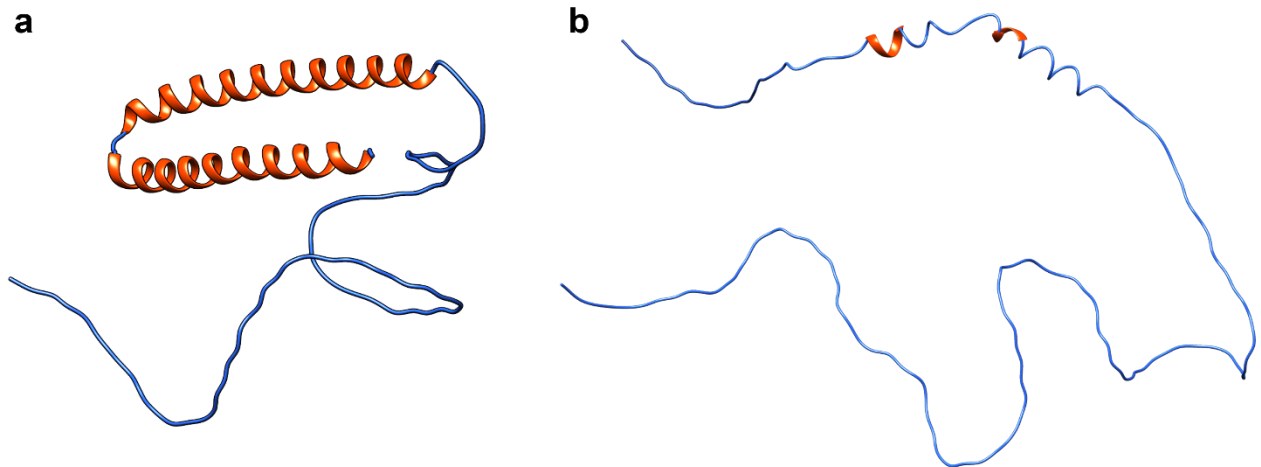
Anexo XI. Experimento control de los ensayos de congelamiento-descongelamiento. Se realizó un experimento control para detectar el nivel de inactivación de la LDH sola o en presencia de algunas proteínas (BSA, lisozima o AtLEA4-5). Se utilizaron siete ciclos de congelamiento-descongelamiento. Los resultados corresponden a tres réplicas.



Anexo XII. Determinación experimental de la osmolalidad de soluciones porcentuales de sustancias que simulan falta de agua *in vitro*. Se utilizaron porcentajes bajos en comparación con los usados para DC debido a que el osmómetro Wescor Vapro 5600 no pudo trabajar con soluciones porcentuales más concentradas. Se infirió que, a concentraciones mayores, la osmolalidad de estas sustancias se comportaría de la misma manera.



Anexo XIII. Distribución de residuos cargados en todas las proteínas mutantes. a) Distribución lineal de residuos cargados en la secuencia de las diferentes proteínas caracterizadas en este estudio. La posición de los residuos se indica debajo de cada diagrama. Azul: residuos cargados positivamente. Rojo: residuos cargados negativamente. Distribución de la AtLEA4-5 nativa y mutantes en el diagrama de estado de Das-Pappu. Este diagrama ofrece una idea del comportamiento conformacional de ensamblajes no estructurados en proteínas desordenadas. b) LEA4-5, LEA4-5 NT1 - NT4. c) LEA4-5, LEA4-5₂₆₋₁₅₈, LEA4-5₄₃₋₁₅₈, LEA4-5₅₇₋₁₅₈, LEA4-5₇₇₋₁₅₈, LEA4-5_{Δ44-78}, y LEA4-5_{Δ61-78}. d) LEA4-5, LEA4-5₁₋₇₇, LEA4-5_{1-77(9P)}, LEA4-5-NT2P, LEA4-5-NT4P, y LEA4-5-NT9P.



Anexo XIV. Modelos tridimensionales de la estructura de proteínas producidos por AlphaFold. En azul se muestra las zonas en estado desordenado y en naranja las regiones que conforman hélices alfa. A) AtLEA4-5 nativa. B) Mutante LEA4-5-NT9P.

Anexo XIV. Parámetros fisicoquímicos de la proteína AtLEA4-5 y sus versiones mutantes.

Name	# Residues	M.W. (Da)	Theoretical pI	Neg. Residues (Asp + Glu)	Pos. Residues (Arg + Lys)	Aliphatic Index	Grand Average of Hydrophobicity (GRAVY)	Net Charge Per Residue (NCPR)	Kappa	% of alpha helix residues (FELLS)	% of disordered residues (MetaPredict)	Pos. Residues (Arg + Lys + Hys)
LEA4-5	158	16161.84	9.22	15	18	46.46	-0.803	0.018987342	0.06	39.9	79.7	25
LEA4-5-NT1	158	15931.58	9.3	13	16	48.99	-0.663	0.018987342	0.071	39.2	69.6	23
LEA4-5-NT2	158	16104.74	8.93	15	17	47.09	-0.766	0.012658228	0.073	41.1	78.5	24
LEA4-5-NT3	158	16218.94	9.4	15	19	45.82	-0.839	0.025316456	0.07	40.5	80.4	26
LEA4-5-NT4	158	16046.75	9.43	14	18	47.72	-0.735	0.025316456	0.087	38.6	77.8	25
LEA4-5 ₂₆₋₁₅₈	133	13608.94	9.05	12	14	45.56	-0.847	0.015037594	0.06	30.1	80.5	21
LEA4-5 ₄₃₋₁₅₈	116	11568.58	8.22	8	9	45.52	-0.697	0.00862069	0.07	21.6	91.4	16
LEA4-5 ₅₇₋₁₅₈	102	9982.78	8.31	6	7	43.14	-0.643	0.009803922	0.081	14.7	100	14
LEA4-5 ₇₇₋₁₅₈	82	7844.44	7.02	3	3	47.56	-0.449	0	0.216	0	100	9
LEA4-5 _{Δ61-78}	140	14379.92	9.26	13	16	48.86	-0.782	0.021428571	0.078	34.3	99.3	22
LEA4-5 _{Δ44-78}	123	12380.65	9.05	10	12	48.46	-0.669	0.016260163	0.077	27.6	100	18
LEA4-5-NT2P	158	16213.92	9.22	15	18	45.19	-0.846	0.018987342	0.06	32.3	100	25
LEA4-5-NT4P	158	16265.99	9.22	15	18	43.92	-0.889	0.018987342	0.06	24.7	100	25
LEA4-5-NT9P	158	16124.77	7.07	15	15	41.46	-0.885	0	0.069	7	100	22
LEA4-5 _{1-77(9P)}	77	8429.55	6.58	12	12	34.42	-1.314	0	0.061	29.9	100	13
LEA4-5 ₁₋₇₇	77	8377.52	9	12	14	48.57	-1.003	0.025974026	0.047	87	37.7	15
LEA4-5 ₁₋₁₄₉	149	15348.02	9.3	15	18	49.26	-0.811	0.020134228	0.06	42.3	78.5	25
LEA4-5 ₁₋₁₂₅	125	13115.66	8.04	15	16	51.68	-0.826	0.008	0.051	51.2	74.4	21

Name	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
LEA4-5	18	0	5	10	0	27	7	3	12	6	7	5	3	9	6	6	25	7	0	2
LEA4-5-NT1	22	0	5	8	0	27	7	3	10	6	7	5	3	9	6	6	25	7	0	2
LEA4-5-NT2	19	0	5	10	0	27	7	3	11	6	7	5	3	9	6	6	25	7	0	2
LEA4-5-NT3	17	0	5	10	0	27	7	3	13	6	7	5	3	9	6	6	25	7	0	2
LEA4-5-NT4	20	0	5	9	0	27	7	3	12	6	7	5	3	8	6	6	25	7	0	2
LEA4-5 ₂₆₋₁₅₈	13	0	4	8	0	26	7	2	8	5	5	4	3	9	6	2	22	7	0	2
LEA4-5 ₄₃₋₁₅₈	12	0	3	5	0	26	7	2	4	4	4	4	2	7	5	2	21	6	0	2
LEA4-5 ₅₇₋₁₅₈	10	0	2	4	0	26	7	1	2	4	3	3	2	4	5	2	20	5	0	2
LEA4-5 ₇₇₋₁₅₈	5	0	2	1	0	24	6	1	0	4	2	2	2	3	3	2	18	5	0	2
LEA4-5 _{Δ61-78}	13	0	5	8	0	24	6	3	11	6	6	4	3	8	5	6	23	7	0	2
LEA4-5 _{Δ44-78}	11	0	4	6	0	24	6	2	8	6	5	3	3	5	4	6	22	6	0	2
LEA4-5-NT2P	16	0	5	10	0	27	7	3	12	6	7	5	5	9	6	6	25	7	0	2
LEA4-5-NT4P	14	0	5	10	0	27	7	3	12	6	7	5	7	9	6	6	25	7	0	2
LEA4-5-NT9P	14	0	5	10	0	27	7	2	10	6	7	5	12	9	5	6	24	7	0	2
LEA4-5 _{1-77(9P)}	9	0	3	9	0	3	1	1	10	2	6	3	10	6	2	4	6	2	0	0
LEA4-5 ₁₋₇₇	16	0	3	9	0	3	1	2	11	2	6	3	1	6	3	2	7	2	0	0
LEA4-5 ₁₋₁₄₉	18	0	5	10	0	22	7	3	12	6	7	5	3	9	6	6	23	7	0	0
LEA4-5 ₁₋₁₂₅	16	0	5	10	0	14	5	2	12	6	7	3	3	9	4	6	17	6	0	0