



**Universidad Nacional Autónoma de México**  
**Facultad de Estudios Superiores Iztacala**

Factores de transcripción con regiones intrínsecamente desordenadas como sensores ambientales en *Arabidopsis thaliana*

**TESIS**

que para obtener el título de bióloga

**PRESENTA:**

Alcántara González María del Rocío

**DIRECTOR DE TESIS**

Dr. César Luis Cuevas Velázquez



Los Reyes Iztacala, Tlalnepantla, Edo. Mex. 2023



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

*“Aprende del ayer, vive para el hoy, ten esperanza para el mañana.  
Lo importante es no dejar de hacerse preguntas.”*  
Albert Einstein

## **Agradecimientos**

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM IA209920. Agradezco a la DGAPA-UNAM por la beca recibida.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el donativo de Ciencia de Frontera CONACYT con clave CF-2019/252952 al Dr. César Luis Cuevas Velázquez.

Este material está basado en trabajo realizado gracias al financiamiento del University of California Institute for Mexico and the Unites States (UC MEXUS) y el Consejo Nacional de Ciencia y Tecnología de México (CONACYT). Proyecto CN-20-113 titulado “Sensibilidad al ambiente de proteínas intrínsecamente desordenadas de plantas inducidas por deshidratación”.

Al Programa de Apoyo a la Investigación y el Posgrado (PAIP) de la Facultad de Química con la clave 5000-9182 al Dr. César Luis Cuevas Velázquez.

## Índice

RESUMEN.....	1
INTRODUCCIÓN.....	2
ANTECEDENTES.....	10
HIPÓTESIS.....	12
OBJETIVOS.....	12
MATERIALES Y MÉTODOS.....	13
RESULTADOS.....	15
DISCUSIÓN.....	31
CONCLUSIONES.....	35
REFERENCIAS.....	36

## RESUMEN

Las plantas al ser organismos sésiles están constantemente expuestas a los cambios ambientales. Sin embargo, poseen vías genéticas que les permiten tolerar períodos de estrés. Estas vías están reguladas por factores de transcripción. Se ha demostrado que un alto número de factores de transcripción poseen regiones intrínsecamente desordenadas (IDR). Se realizó un análisis de 1986 factores de transcripción (FT) de *Arabidopsis thaliana* para determinar los porcentajes de desorden, la predicción de formar alfa-hélice; y sus niveles de expresión frente a estrés osmótico. Se encontró que los FT son ricos en regiones desordenadas con una media de 7 regiones. Además, se determinó que el tamaño promedio de la IDR más larga es de 29 aminoácidos y esta representa un 13 % de FT completo. Finalmente correlacionando los niveles de desorden y la predicción de alfa-hélice se encontraron algunos FT que presentan niveles significativos, dentro de estos se encuentran MBF1c, GBF3 y H8SFA6A, que además están relacionados con la respuesta y tolerancia al estrés hídrico lo cual relaciona directamente a las regiones desordenadas con una respuesta al ambiente.

## INTRODUCCIÓN

### **Ambiente y factores abióticos**

El ambiente es determinante para la supervivencia de las especies y de los individuos. Interesantemente, las plantas son un grupo que se destaca por la fina regulación que tiene el ambiente en muchos de sus procesos, por ejemplo, el crecimiento de las hojas y raíces de una planta depende de la cantidad de luz percibida, donde participan principalmente hormonas y algunas proteínas. Durante la transición a la floración, las hojas detectan cambios en la cantidad de luz y envían una señal (una molécula de mRNA) al meristemo apical que le indica el momento adecuado para florecer (Wu & Gallagher, 2012). Las semillas germinan sólo si la temperatura y la cantidad de agua son las adecuadas, las hojas captan los fotones y los utilizan como energía para la fotosíntesis, y como estos, existen muchos otros ejemplos que nos permiten resaltar lo importante que es el ambiente para las plantas (Brodsky et al., 2020).

El ambiente guía grandes procesos relacionados con la evolución, y para entender cómo se distribuyen las especies, es necesario considerar la importante relación que tienen con este, el cual, las ha acompañado a lo largo de toda su historia de vida y su exitoso establecimiento en el planeta. Con ayuda de la teoría de la evolución podemos señalar que cada especie tiene adaptaciones para el lugar donde vive. Y vive donde vive, como resultado del trabajo de la selección natural que ha actuado a través del tiempo donde el ambiente ha estado implicado (Smith & Smith, 2014)

Los organismos interactúan todo el tiempo con su entorno, las condiciones físicas que rodean a un ser vivo impactan en sus procesos fisiológicos. Todas las señales que se dan durante el ciclo de vida de un individuo, que están implicadas en momentos tan importantes como el desarrollo, crecimiento y reproducción, son afectadas por el ambiente.

Denominamos como ambiente a la suma de todos los factores externos, bióticos y abióticos, que actúan sobre un individuo, población o comunidad. Los factores bióticos incluyen a los seres vivos, mientras que los factores abióticos incluyen todas las condiciones físicas de un determinado lugar. Los factores abióticos son principalmente mediados por condiciones geológicas y climáticas del planeta. Entre los factores abióticos tenemos la temperatura, pH, luz, gases atmosféricos, cantidad de agua disponible, además de otras condiciones que son necesarias para que sobrevivan las especies. Los factores abióticos son importantes porque proporcionan el medio físico para que los organismos crezcan y se reproduzcan, dando condiciones requeridas para que la células y el organismo en general realicen las funciones necesarias para la vida, como la luz durante la fotosíntesis o el agua para las reacciones metabólicas. Las condiciones físicas, al ser parte del ambiente, también pueden actuar como una fuerza que guía a la selección natural por lo que han sido importantes a lo largo de la historia evolutiva de los organismos siendo una pieza fundamental para establecer la estructura y dinámica de los ecosistemas. Estas condiciones son proporcionadas por la naturaleza del lugar y se mantienen (o no cambian de manera drástica) a menos que exista algún fenómeno que altere las condiciones naturales, como ha ocurrido en las extinciones masivas del pasado, donde algún suceso catastrófico altera drásticamente los

factores abióticos, cambiando las condiciones necesarias para los individuos, lo que conlleva, en los casos más extremos, a la muerte y pérdida masiva de la biodiversidad (Smith & Smith, 2014).

### **Cambios en el ambiente: estrés abiótico**

Para que un organismo se desarrolle, crezca y se reproduzca de forma adecuada es necesario un ambiente favorable, sin embargo, las condiciones no siempre son óptimas, y al existir en tiempo y espacio en un lugar físico, una especie está sujeta a posibles cambios, donde algunos pueden comprometer su ciclo de vida, ya que no se tienen las condiciones para realizar las funciones celulares necesarias. Cuando las condiciones en el ambiente cambian y dicho cambio resulta en algo perjudicial para el organismo, se considera una condición de estrés. El estrés abiótico se define como condiciones ambientales que disminuyen el rendimiento y crecimiento por debajo de los niveles óptimos causadas por agentes no biológicos, como puede ser cualquier cambio en la disponibilidad de agua, temperatura, pH, características del suelo, etcétera (Cramer et al., 2011).

Es importante considerar el impacto que tiene el estrés abiótico en los organismos principalmente porque, con el cambio climático y el calentamiento global, se han generado drásticas perturbaciones en el ambiente, lo que afecta directamente a una infinidad de especies. Estudios enfocados en la agricultura han demostrado que en la actualidad solo el 3.5 % de la superficie terrestre mundial no se ve afectada por alteraciones en el ambiente (FAO, 2016) lo cual resulta alarmante para la conservación de la biodiversidad y la producción óptima de los cultivos. Además, se han predicho futuros escenarios de las condiciones ambientales físicas del planeta. Desgraciadamente se espera que en los próximos años exista una pérdida masiva de especies, contaminación atmosférica, una insuficiente producción de cultivos para abastecer a la población, y escasez de agua.

### **Limitación del agua y estrés osmótico**

Por sus propiedades vitales inigualables, debemos considerar al agua como el recurso más valioso que tenemos, ya que sin él, la vida como ocurre en la Tierra no sería posible. Por desgracia, la baja disponibilidad de agua es una de las mayores problemáticas ambientales en la actualidad. Tal problema es resultado de actividades antropogénicas que contaminan el agua y alteran su ciclo (principalmente como consecuencia del cambio climático), y como tal la accesibilidad (Kummu et al., 2016), ocasionando un impacto en muchas especies que comienzan a carecer del recurso, y en áreas como la agricultura, la escasez de este líquido ya está comenzando a ser un reto para la óptima producción de los cultivos.

El agua por sus propiedades fisicoquímicas ha permitido la vida en nuestro planeta. Es el disolvente universal por excelencia y es el medio en el cual se llevan a cabo las reacciones bioquímicas dentro de la célula. Además, la estructura de las proteínas, ácidos nucleicos, y membranas biológicas, es una consecuencia directa de la interacción con las moléculas de agua (Rawn, 2010). Todo esto es gracias a las características intrínsecas del agua como la polaridad, su capacidad de formar puentes de hidrógeno, entre otras cualidades que hacen posibles ciertos fenómenos biológicos, como la capilaridad



que se relaciona con el transporte vascular a través de los tallos de las plantas o simplemente el transporte de agua por el suelo.

Existen varios tipos de estrés relacionados a la baja disponibilidad de agua, como la sequía, el congelamiento, la salinidad y el estrés osmótico. En general, durante el estrés hídrico no existe suficiente cantidad de agua para realizar de manera óptima todas las funciones celulares, lo que compromete la supervivencia de los organismos si el estrés se extiende de una manera crónica.

El estrés hiperosmótico es ocasionado por un aumento en la cantidad de solutos en el medio externo de la célula, lo que provoca una salida del agua como resultado del rompimiento de la homeostasis osmótica que ocurre dentro de la célula y en toda la planta, por lo que cambios drásticos originan daño molecular afectando el crecimiento y eventualmente la muerte (Zamora-Briseño et al., 2021).

### **Plantas y su respuesta a estrés abiótico**

Las plantas son organismos sésiles cuya evolución está estrechamente relacionada con los factores abióticos, desde su colonización y especiación en el medio terrestre. El carecer de movilidad las compromete a estar constantemente expuestas a los cambios en el ambiente, incluso todos aquellos que pueden resultar en algún tipo de estrés. Afortunadamente, a lo largo de su trayectoria evolutiva, las plantas han adquirido mecanismos que les permiten tolerar estos períodos de tensión, e incluso les han ayudado en su establecimiento en la tierra. Dentro de sus estrategias se incluyen adaptaciones morfológicas, fisiológicas, bioquímicas y moleculares necesarias para la supervivencia de la planta (García-Morales et al., 2013).

Durante el estrés, deben ocurrir los procesos de detección (percepción) y transducción de señales, que finalmente conduzcan a una respuesta para conferir tolerancia a la planta. Todos estos procesos involucran cientos de moléculas como hormonas, especies reactivas de oxígeno, iones y proteínas, que forman complejas redes de interacción (He et al., 2018) . Las redes de interacción, suelen ser específicas para cada tipo de estrés, debido a que cada factor abiótico transmite diferente información a la célula, lo que causa respuestas extremadamente complejas y muy particulares. Sin embargo, también existen algunas moléculas que participan en la respuesta de varios tipos de estrés, lo que aumenta la dinámica provocando respuestas múltiples (Figura 1) (Joshi et al., 2016).

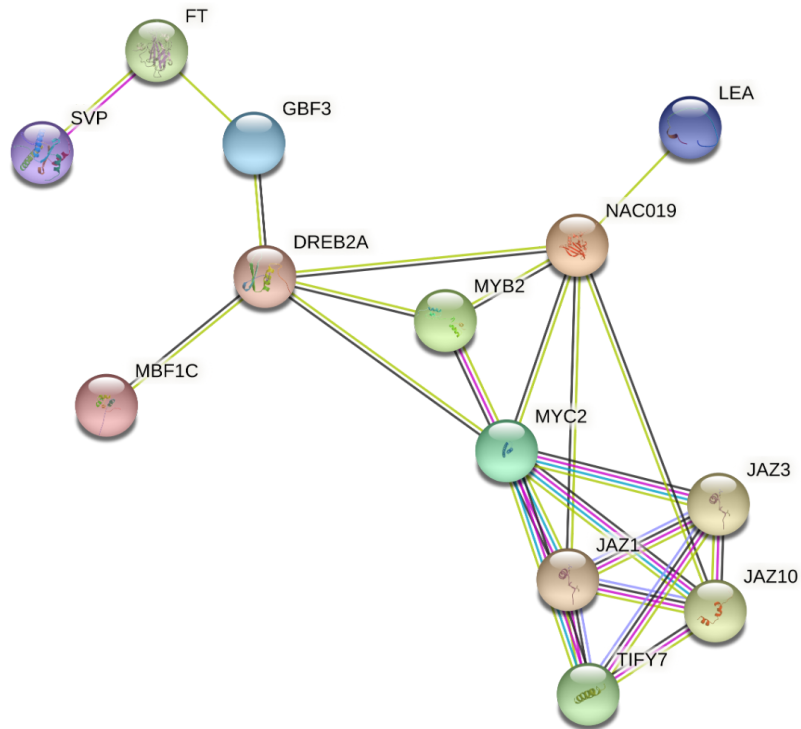


Figura 1. Posibles interacciones entre redes sinérgicas de factores de transcripción en plantas. DREB2A: se expresa durante estrés por sequía, MYB2: aumenta su expresión durante altas concentración de sal, MBF1C: ha demostrado tener un papel en múltiples tipos de estrés hídrico y por patógenos. Imagen generada en <https://string-db.org/>.

Para tratar de entender los mecanismos que se desencadenan durante el estrés, se han realizado diferentes análisis moleculares y bioinformáticos con el objetivo de dilucidar el proceso de respuesta. Gracias a esto ha sido posible trazar rutas y describir el papel de moléculas que participan en las redes y vías de señalización que ocurren dentro de la célula y que le permiten aclimatarse al estrés. Estos estudios también han demostrado que debido a que muchas veces los mecanismos no son lineales e involucran redes sinérgicas para responder dinámicamente, la generación de mapas moleculares que describan exactamente el proceso de respuesta ante un estrés particular o de manera general ha sido un reto para la ciencia de plantas, pero comprender los mecanismos podría ser clave para generar futuras aplicaciones que ayuden a crear cultivos resistentes (Wang et al., 2016).

### Factores de transcripción

Uno de los mecanismos de respuesta al estrés abiótico involucra a un tipo de proteínas llamadas factores de transcripción (FT), los cuales tienen la capacidad de reconocer y unirse a una secuencia nucleotídica específica en el promotor de un gen blanco. Dicha unión puede inhibir o promover la actividad de la RNA pol II, regulando de manera coordinada y eficiente la transcripción. Los FT poseen típicamente dos dominios principales, un dominio de unión al DNA que reconoce una secuencia

específica de nucleótidos (elementos en *Cis*), y un dominio de transactivación que le ayuda en la interacción con otras proteínas (Barberis & Petrascheck, 2003). Los FT pueden unirse de diversas maneras al DNA, como interacciones directas donde los aminoácidos forman puentes de hidrógeno con nucleótidos, o interacciones indirectas según la topología del DNA. Además, típicamente hacen múltiples interacciones con otras proteínas como coactivadores u otros FT, y poseen la habilidad de reconocer más de un motivo de DNA, lo que ayuda a regular la expresión génica mediante interacciones complejas. Estas proteínas son consideradas un punto de control de la expresión genética debido a que su unión puede controlar la actividad de la RNA pol II, y participar directamente regulando la transcripción (Siggers & Gordân, 2014).

Los factores de transcripción participan en casi todos los procesos biológicos, desde el ciclo celular hasta la floración y crecimiento foliar. En todos los organismos estos se ven representados como una parte importante del proteoma y han demostrado su relevancia funcional en procariontes y eucariontes, donde se incluyen las plantas. El genoma de las plantas tiene cerca de 2,000 genes que codifican para FT, lo que representa alrededor del 5.6 % de su totalidad (Mitsuda & Ohme-Takagi, 2009). Los genes que codifican para FT han demostrado una mayor expansión en plantas que en animales, es decir, que las familias de FT que tienen en común los dos grupos tienen muchos más miembros en plantas debido a fenómenos como la duplicación génica que causan un aumento en el número de genes, y además se ha demostrado que existen muchas otras familias que son específicas de plantas, lo que nos habla de la gran diversidad de FT en este grupo, lo que biológicamente podría tener un significado funcional y relacionarse con un rol fundamental en plantas (Shiu et al., 2005).

Los FT se han agrupado en familias de acuerdo con los elementos en *Cis* que son capaces de reconocer. En el caso de la planta modelo *Arabidopsis thaliana*, es posible agrupar a familias como WYRK, NAC, MYB, bHLH, bZIP entre muchas otras, cuyos miembros están involucrados en varios procesos donde se incluye la respuesta a estrés abiótico y biótico (Yao et al., 2021). Durante el estrés abiótico, los FT se unen a los promotores de genes inducibles para regular positivamente la transcripción. El reconocimiento de estos genes es gracias a elementos en *Cis* que funcionan como sitios de unión para FT (Hrmova & Hussain, 2021). Los genes que se activan durante el estrés codifican para diferentes tipos de proteínas de respuesta a estrés, por ejemplo, transportadores de membrana, enzimas antioxidantes, inhibidores de proteasas, proteínas LEA, cinasas, fosfatasas y proteínas reguladoras (Hirayama & Shinozaki, 2010).

Estudios en mutantes de sobre-expresión y de pérdida de función han demostrado que los FT son claves durante la tolerancia al estrés abiótico en las plantas. Se ha demostrado que los FT participan en la respuesta y la permisividad de resistir al estrés mediante la regulación de múltiples vías fisiológicas y moleculares que le confieren a la planta la capacidad de sobrevivir cuando es sometida a tensiones demostrando que el papel de estas proteínas es necesario en la respuesta y tolerancia (Khan et al., 2018).

Un enfoque importante es tratar de entender cómo es qué trabajan exactamente los FT para actuar como reguladores maestros durante la respuesta al estrés. Se ha observado que su papel depende de señales ambientales e interacciones con otras moléculas dentro del entorno celular. En algunos casos pueden unirse a distintos elementos *Cis* en diferentes condiciones celulares, y pueden interaccionar en diferentes complejos transcripcionales según el blanco genético deseado (Purohit et al., 2019; Wang et al., 2021). Ciertamente los FT son proteínas muy complejas que aún no entendemos del todo a causa de las enormes y complejas redes moleculares que pueden desarrollarse dentro de la célula, y a su posible plasticidad en respuesta al ambiente.

Una característica que se ha observado en los FT, es que muchos son proteínas que contienen altos niveles de “desorden estructural”, el cual les podría conferir ciertas ventajas mediante otras formas de regulación. Se ha demostrado que las regiones desordenadas tienen un papel clave en la respuesta al estrés abiótico (Sun et al., 2013), por lo que estas regiones podrían ayudar a explicar por qué es que los factores de transcripción poseen la capacidad de responder dinámicamente a cambios ambientales.

### **Regiones intrínsecamente desordenadas**

Los aminoácidos que componen a las proteínas pueden formar cadenas entre sí que se pliegan y que incluso, con otros péptidos, se asocian para formar estructuras complejas. En bioquímica se creía que la estructura tridimensional estable era necesaria para que una proteína pudiera realizar su función. Sin embargo, algunas proteínas pueden poseer regiones que carecen de estructura tridimensional estable y son biológicamente funcionales. Este tipo de proteínas son conocidas como proteínas intrínsecamente desordenadas (Hsiao, 2022). Este grupo puede carecer total o parcialmente de una estructura secundaria bien definida, por lo que hay proteínas que pueden contener múltiples regiones desordenadas.

Las regiones intrínsecamente desordenadas (IDR, del inglés *Intrinsic Disordered Region*) son secuencias de aminoácidos dentro de una proteína que se caracterizan por carecer de una estructura secundaria bien definida (Cuevas-Velázquez & Covarrubias-Robles, 2011). Están enriquecidas en aminoácidos hidrofílicos y pequeños (Gln, Ser, Pro, Glu, Lys, Arg, Gly, Ala), es decir, son regiones de baja complejidad de secuencia. El desorden estructural ha demostrado tener una enorme relevancia biológica y es necesario como cualquier otro arreglo en las proteínas. Las IDR están particularmente presentes en proteínas implicadas en procesos de señalización y regulación celular (Pietrosemoli et al., 2013). En el caso de los factores transcripcionales, su relevancia se ha demostrado con análisis bioinformáticos, biofísicos y bioquímicos, donde se sugiere que tal enriquecimiento en estas proteínas se debe a que las IDR son capaces de realizar funciones como mediar interacciones proteína-proteína o bien, reconocer secuencias de DNA, lo que representa la principal función de factor de transcripción (Salladini et al., 2020).

Las regiones desordenadas participan en muchos procesos como el ensamblaje del ribosoma, la remodelación de la cromatina, en la dinámica de los microfilamentos y microtúbulos, el transporte a

través del poro nuclear, e incluso en el funcionamiento de chaperonas. Todas las funciones que realizan estas proteínas sin estructura son gracias a su composición sesgada de aminoácidos, que le confieren diversas propiedades, como una alta especificidad pero baja afinidad durante la interacción con otras proteínas. Además, las IDR suelen ser blancos de modificaciones postraduccionales, también importante para el dinamismo y promiscuidad proteica (Wright & Jane Dyson, 2015).

Un papel importante en el que han tomado relevancia las IDR es en la capacidad de responder ante un estrés ambiental, ya que existe evidencia clara de su participación en la respuesta al estrés (Sun et al., 2013). Diversas familias de FT de plantas poseen miembros con IDR en su estructura, las cuales parecen ser claves ya que eliminarlas compromete la función del factor, como es el caso de algunos miembros de las familias NAC, bZIP, AP2/ERF, bHLH, MYB, TCP y ARF, que ya se sabe que le confieren tolerancia a la planta frente a diversas tensiones abióticas (estrés osmótico, salino y genotóxico) (Salladini et al., 2020).

El desorden también es importante en otro tipo de proteínas que le confieren tolerancia a las plantas. Las proteínas abundantes en la embriogénesis tardía (LEA, *Late Embryogenesis Abundant*) son proteínas desordenadas que se han relacionado con la tolerancia a condiciones de falta de agua, ya que se acumulan en etapas secas de la semilla y en tejidos vegetativos cuando la planta se somete a estrés hídrico (Amara et al., 2014). Un ejemplo sobresaliente de la sensibilidad ambiental de este tipo de proteínas es la IDR de la proteína AtLEA4-5, la cual responde ante condiciones de estrés osmótico plegándose para conferir su actividad protectora. Dicho plegamiento depende del medio fisicoquímico que al cambiar, provoca un cambio conformacional en la región desordenada, llevándola a formar estructuras de alfa-hélice y activando su función (Cuevas-Velazquez et al., 2016).

Gracias a su flexibilidad estructural, las IDR han demostrado ser muy sensibles a las propiedades fisicoquímicas del entorno, resultando en regiones desordenadas que en cierto momento cambian a estructuras ordenadas en respuesta a cambios de temperatura, pH o salinidad, lo que conduce a regiones más estructuradas, ya que algunos cambios en las propiedades físicas del ambiente pueden causar cambios en las cargas intramoleculares y conducir a un plegamiento impulsado por la hidrofobicidad a una forma más ordenada (Uversky, 2019).

### ***Arabidopsis thaliana* como organismo modelo en biología de plantas**

*Arabidopsis thaliana* es una herbácea, perteneciente a la familia Brassicaceae. Es una planta pequeña cuyas hojas crecen en roseta, presenta una inflorescencia, un fruto tipo silicua y una raíz primaria en la que se producen pelos radicales (Passardi et al., 2007). Es el organismo modelo más utilizado para estudios de biología molecular, genética, y ciencia de plantas en general, incluyendo estudios ecológicos y evolutivos debido a la variedad de accesiones (pequeñas poblaciones de *Arabidopsis thaliana* aisladas geográficamente con variaciones genotípicas y fenotípicas como resultado del ambiente) (Krämer, 2015), todo esto se debe a ciertas características que presenta esta planta, como su rápido ciclo de vida, la viabilidad de sus semillas, su capacidad de autopolinizarse, incluyendo el relativamente pequeño tamaño

de su genoma. Fue la primera planta en tener su genoma secuenciado, y aunque en la actualidad también se conoce el de otras especies, *Arabidopsis* se ha convertido en el modelo principal para estudios bioinformáticos. Esto ha llevado a la generación de plataformas y herramientas computacionales para generar datos de calidad que permitan interpretar información de valor biológico, y gracias a esto hoy en día existen datos de transcriptómica, proteómica, epigenética, entre otros, que facilitan las interpretaciones; y aunque grandes conjuntos de datos pueden dificultar entender una cuestión biológica determinada, nos pueden ayudar a generar hipótesis *in silico* que permitan planificar estudios experimentales más objetivos (De Lucas et al., 2014).

## ANTECEDENTES

Las regiones intrínsecamente desordenadas (IDR, *Intrinsically Disordered Regions*) han mostrado una elevada presencia en todos los organismos eucariontes. En el caso de las plantas, se han estudiado una enorme diversidad de proteomas de varias especies de monocotiledóneas y dicotiledóneas, donde se ha demostrado que efectivamente el desorden está presente en muchas proteínas, especialmente en aquellas que se relacionan con una unión o interacción con ácidos nucleicos y una respuesta a estímulos bióticos y abióticos, así como procesos regulatorios (Zamora-Briseño et al., 2021). Además, se analizó un pequeño grupo de FT de *Arabidopsis thaliana* y se observó que alrededor del 92 % poseen regiones desordenadas o de baja complejidad (Liu et al., 2006).

Varias familias de FT se relacionan con una respuesta al estrés abiótico, donde posiblemente existe alguna función directa mediada por las IDR, ya que se ha demostrado que pueden ser importantes al momento de participar en la regulación de la expresión genética con roles como ayudar directamente a la unión del FT con el DNA, cambiar su estructura e interactuar con su blanco a manera de llave-candado, pueden ser modificadas postraduccionalmente, formar condensados líquidos que modulan la transcripción y favorecen las reacciones bioquímicas, e incluso interactúan directamente con la RNA Pol II y pueden actuar como reguladores epigenéticos acetilando histonas para modificar la arquitectura de la cromatina. Por lo tanto, claramente las IDR ejercen un papel clave implicado en la regulación de la expresión génica (Figura 2) (Davey, 2019; Wright & Jane Dyson, 2015; Yang et al., 2020).

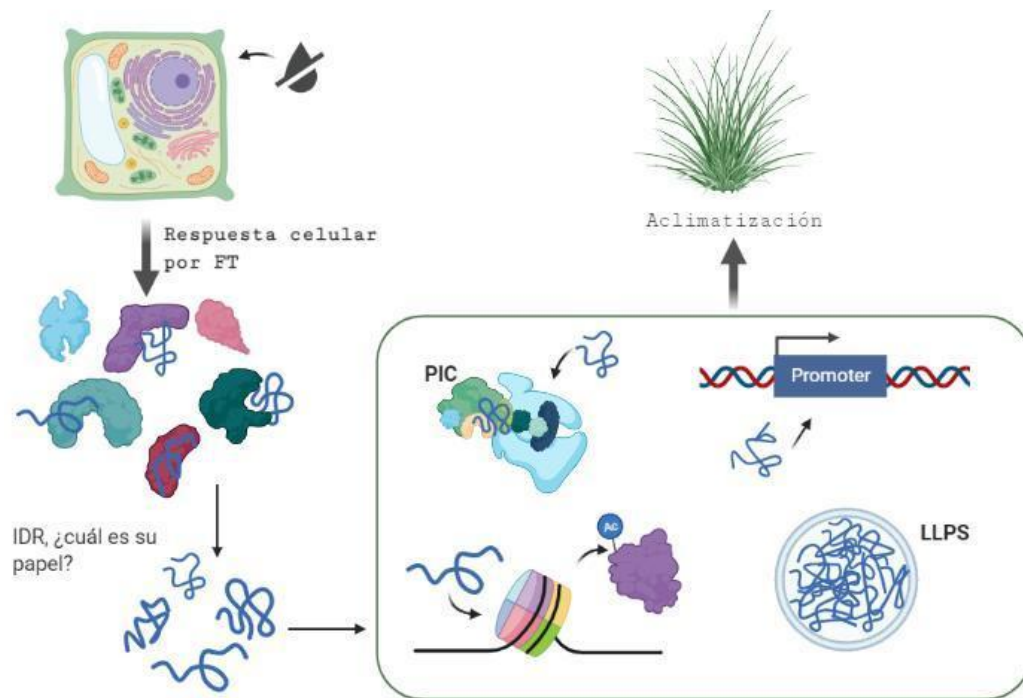


Figura 2. Algunas funciones conocidas de las regiones intrínsecamente desordenadas. Creada con Biorender.

Por otra parte, se ha demostrado que la delección o el cambio en las regiones desordenadas compromete la función del FT, y por lo tanto, su rol en la respuesta al estrés abiótico, como se observó en SOG1, donde la eliminación de su extremo N-terminal desestructurado compromete su estabilidad durante estrés salino (Mahapatra & Roy, 2019).

Existen más evidencias que demuestran que las IDR son sensibles a cambios en el ambiente, específicamente a los factores abióticos, como es el caso del FT ANAC019, el cual posee una región desordenada que sufre una desprotonación al aumentar el pH, lo que hace posible que ANAC019 pueda unirse con otro factor, lo que es importante ya que su unión al DNA solo es posible en un estado de dimerización (Kang et al., 2018). Otro ejemplos son algunos de FT de las familias bZIP, los cuales se ha demostrado que pueden formar interacciones dependientes de la temperatura (Wiese et al., 2021). Así como este ejemplo, hay algunos otros trabajos que han demostrado que las IDR experimentan cambios conformacionales en respuesta a distintos tipos de estrés.

Es posible dilucidar un posible cambio conformacional dentro de las IDR utilizando predictores bioinformáticos que nos ayuden a interpretar las propiedades codificadas en la secuencia de aminoácidos para determinar si la región podría ser capaz de formar algún tipo de estructura, como se observó en la proteína AtLEA4-5. Esta proteína posee una región desordenada que es capaz de plegarse durante condiciones de alta osmolaridad, y dicho plegamiento fue predecido utilizando herramientas bioinformáticas, que mostraron que a pesar de ser una región desordenada, esta tiene el potencial de formar estructuras que provocan un estímulo en la proteína, que en el caso de la AtLEA4-5, le confiere su actividad en respuesta a estrés osmótico (Cuevas-Velazquez et al., 2016), demostrando nuevamente que las regiones desordenadas son claramente sensibles a las propiedades fisicoquímicas del entorno.

Por lo tanto, la evidencia demuestra que la presencia de IDR en los FT podría relacionarse con su respuesta al estrés abiótico, en casos donde los cambios en el entorno como la sequía, salinidad o estrés osmótico conduzcan a un plegamiento en las IDR que causen la activación de su participación en la respuesta, accionando vías de transducción de señales al interactuar con otras proteínas o unirse a diferentes genes blanco en respuesta a dicho cambio conformacional.



## **HIPÓTESIS**

Los factores de transcripción de la planta modelo *Arabidopsis thaliana* contienen regiones intrínsecamente desordenadas con el potencial de experimentar cambios conformacionales modulados por el ambiente.

## **OBJETIVO GENERAL**

Buscar y caracterizar las regiones intrínsecamente desordenadas de los factores de transcripción de la planta modelo *Arabidopsis thaliana*

## **OBJETIVOS ESPECÍFICOS**

- Determinar los niveles de desorden en los factores de transcripción de *Arabidopsis thaliana*.
- Determinar la localización, tamaño y proporción de regiones intrínsecamente desordenadas de los factores de transcripción de *Arabidopsis thaliana*.
- Identificar la región intrínsecamente desordenada de mayor tamaño de cada factor de transcripción.
- Analizar el potencial de plegamiento a estructura secundaria tipo alfa-hélice de las regiones desordenadas más prominentes de cada factor de transcripción.
- Comparar los niveles de expresión en respuesta al estrés osmótico de los distintos factores de transcripción.
- Identificar factores de transcripción que se expresan en respuesta a estrés osmótico y que posean regiones desordenadas con potencial a formar estructuras secundarias.
- Determinar el porcentaje de desorden en las familias de factores de transcripción de *Arabidopsis thaliana*.

## MATERIALES Y MÉTODOS

### **Construcción de la base de datos de los factores de transcripción de *Arabidopsis thaliana***

Para construir la base de datos se utilizó una biblioteca que contiene 1,982 FT que representan el 96 % de todos los factores que se conocen en *Arabidopsis thaliana*, esta biblioteca es resultado del trabajo de Pruneda-Paz *et al* donde utilizaron cuatro bases de datos (PlnTFDB: <http://plntfdb.bio.uni-potsdam.de> DATF, <http://datf.cbi.pku.edu.cn>; DBD, <http://www.transcriptionfactor.org> y REGIA Consortium, <http://www.jicgenomelab.co.uk>) para compilar una lista completa de FT, los cuales fueron clonados en un vector (pENTR/D) y transferidos a un plásmido compatible con Y1H (pDEST22) que contiene el dominio de activación GAL4 (GAL4 \_AD) (Pruneda-Paz *et al.*, 2014). Los autores desarrollaron esta biblioteca con el fin de crear un recurso versátil para descubrir funciones biológicas de los FT. Se tomó la lista compilada de FT para crear la base de datos a utilizar en esta tesis. La lista contaba previamente con los identificadores génicos de *Arabidopsis* (AGI, *Arabidopsis Gene Identifier*) de los genes que codifican para cada factor de transcripción, así como la familia a la que pertenecen y descripciones generales de su función. Para poder iniciar con los análisis, primero se extrajo la secuencia peptídica de cada uno de los FT con el AGI hasta completar la lista utilizando ePLANT de BAR.utoronto (<http://bar.utoronto.ca/eplant/>) y NCBI (<https://www.ncbi.nlm.nih.gov/>). ePLANT es un visualizador de información que integra datos de varios servidores bioinformáticos para proporcionar descripciones, representaciones visuales de estructuras, información genética sobre secuencias de DNA y proteínas sin la necesidad de consultar múltiples bases de datos (Waese *et al.*, 2017). Las secuencias peptídicas presentadas por ePLANT son obtenidas de Araport a partir de la base de datos UniprotKB (<https://www.uniprot.org/uniprot/>) donde se recupera información funcional de las proteínas además de datos de entrada obligatorios como la secuencia de aminoácidos. En UniProt el 95 % de las secuencias proteicas provienen de traducciones computacionales de secuencias codificantes (CDS), además de integrar datos experimentales para una mayor calidad de la base de datos (Krishnakumar *et al.*, 2015)

El Centro Nacional de Información Biotecnológica (NCBI, *National Center for Biotechnology Information*) es un recurso sobre biología molecular que tiene como fin recolectar datos para ayudar a la comprensión de procesos moleculares y genéticos de los organismos vivos. La base de datos de proteínas es una colección de varias fuentes como GenBank, RefSeq y TPA que integra datos de traducción de secuencias codificantes (NCBI, 2021).

Los datos de la lista (AGI, familia) junto con la secuencia de aminoácidos forman la primera parte de la base de datos

### **Predicción de desorden**

A partir de la secuencia de aminoácidos se realizaron análisis individuales para predecir los niveles de desorden utilizando el servidor PONDR (<http://www.pondr.com/>). Cada secuencia peptídica se colocó en el buscador usando el predictor VL-XT. Se tomaron algunos datos arrojados como el porcentaje total de

desorden, número de regiones desordenadas, el tamaño y la localización de la región desordenada más larga. El algoritmo de PONDR (*Predictor Of Natural Disordered Regions*) utiliza algunas propiedades como la complejidad de secuencia, composición particular de aminoácidos y la hidropatía para predecir posibles estructuras desordenadas en secuencias de aminoácidos. PONDR contiene varios predictores entre ellos el VL-XT, el cual integra tres redes neuronales, el VL1 con datos obtenidos por resonancia magnética nuclear y cristalografía de rayos X de regiones desordenadas largas, el XC y XN, que son predictores de los extremos C-terminal y N-terminal, que usan datos obtenidos de cristalografía de rayos X. El predictor VL-XT utiliza salidas de números reales entre 1 y 0, y aplica un umbral con desorden a valores  $>0.5$ . De todos los predictores que tiene PONDR, VL-XT es el más preciso, los entrenamientos mostraron un menor porcentaje de falsos negativos y positivos con comparación del resto de los predictores (Dunker et al., 2001). Los valores de PONDR indican datos generales de desorden de las proteínas analizadas y proporciona valores específicos de cada aminoácido.

Además de los datos ya mencionados, en todos los casos se seleccionaba individualmente la secuencia desordenada más larga de cada FT indicada por el servidor, esta secuencia fue importante para análisis de predicción de alfa-hélice posteriores.

#### **Análisis de predicción de alfa-hélice**

Una vez obtenidos los datos de desorden, se procedió a analizar el porcentaje de predicción de estructuras alfa-hélice utilizando el algoritmo de AGADIR (<http://agadir.crg.es/>), el cual puede predecir la tendencia a formar estructuras alfa-hélice de péptidos monoméricos considerando valores experimentales de más de 1,200 secuencias peptídicas cortas. El análisis se hizo calculando el porcentaje para todas las secuencias de aminoácidos utilizando los parámetros fisicoquímicos (temperatura= 278, pH= 7 y fuerza iónica= 0.1) estándar de la plataforma.

#### **Determinación de los niveles de expresión génica**

Se utilizó eFP Browser de BAR.utoronto (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>) para obtener datos de los niveles de expresión de mRNA durante condiciones de estrés osmótico de cada FT a 0 y 24 h de tratamiento. eFP Browser (electronic Fluorescent Pictograph) es un visualizador que integra un conjunto de herramientas para datos de niveles de expresión génica de *Arabidopsis thaliana*, además permite conocer los sitios (órganos o tejidos) donde se expresa un gen durante varios procesos biológicos, así como durante diferentes condiciones ambientales. Para cuantificar la expresión genética se extrajo el RNA de plantas de 18 días que fueron sometidas a condiciones de estrés osmótico usando manitol (300 mM), los datos fueron normalizados, y curados para formar parte del AtGenE project (Kilian et al., 2007), dichos datos, son los arrojados por eFP Browser.

## RESULTADOS

### Generación de una base de datos de los factores de transcripción de *Arabidopsis thaliana*

Con el objetivo de caracterizar la secuencia primaria de los factores de transcripción (FT) de *Arabidopsis thaliana*, se utilizó una lista de FT que forma parte de una biblioteca que contiene el 96 % de los FT conocidos en esta especie. Se tomaron los 1,982 AGI de los FT contenidos en la biblioteca para extraer la secuencia primaria de cada uno.

Se construyó una nueva biblioteca con la secuencia de cada FT, la que se utilizó para todos los análisis posteriores. En la Tabla 1 se muestran algunos ejemplos de cómo está organizada la biblioteca.

Tabla 1. Ejemplo de la organización de la de la base de datos de FT de *Arabidopsis thaliana*

<b><i>Arabidopsis</i> Gene Identifier (AGI)</b>	<b>TF family</b>	<b>Secuencia primaria</b>
AT3G17100	bHLH	MESISPVSNQLLQPTTSSNSDRSRRKRKKKSSPSSV
AT4G36570	MYB	MASNSMSSASWTRKENKLFERALATYDQDTPDRWHNVARAVGGKS AEEVRRHYEPPH
AT4G23800	HMG	MATNADPAPTKKPRNSRKALKQKNELVETPPSPVSVKGSAKSFEQD LMEMQTMLEKM
AT5G44080	bZIP	MTSFQVMRSSNSRNSDLSRRISSASTSSSSIRPQQFRRDLTSVGYG GRNDGLYSSNSMTVEGILHDTFAS
AT5G18037		MSMVKYYDKEEHVGFWFHPSAQELIIRYIGPKVTKRSNKEFELEDKFDI YAKEPWRLSHTETNFLEPNEWF

### Análisis de los niveles de desorden de los factores de transcripción de *Arabidopsis thaliana*

Una vez obtenidas todas las secuencias peptídicas, se realizaron los análisis de desorden con PONDR-VLXT (Predictor of Natural Disordered Regions) (<http://www.pondr.com/>). Los análisis se hicieron de manera individual para cada uno de los FT. Al inicio se tomó el número de residuos para obtener el tamaño de cada proteína, porque aunque no tiene relación con el desorden directamente, nos ayuda a

entender mejor la distribución de los porcentajes orden-desorden. Estos datos nos dejaron observar que los factores tienen una amplia variedad de tamaños que van desde 77 hasta 1,724 residuos, sin embargo, la gran mayoría se encuentra entre 200 y 400 aminoácidos (Figura 3).

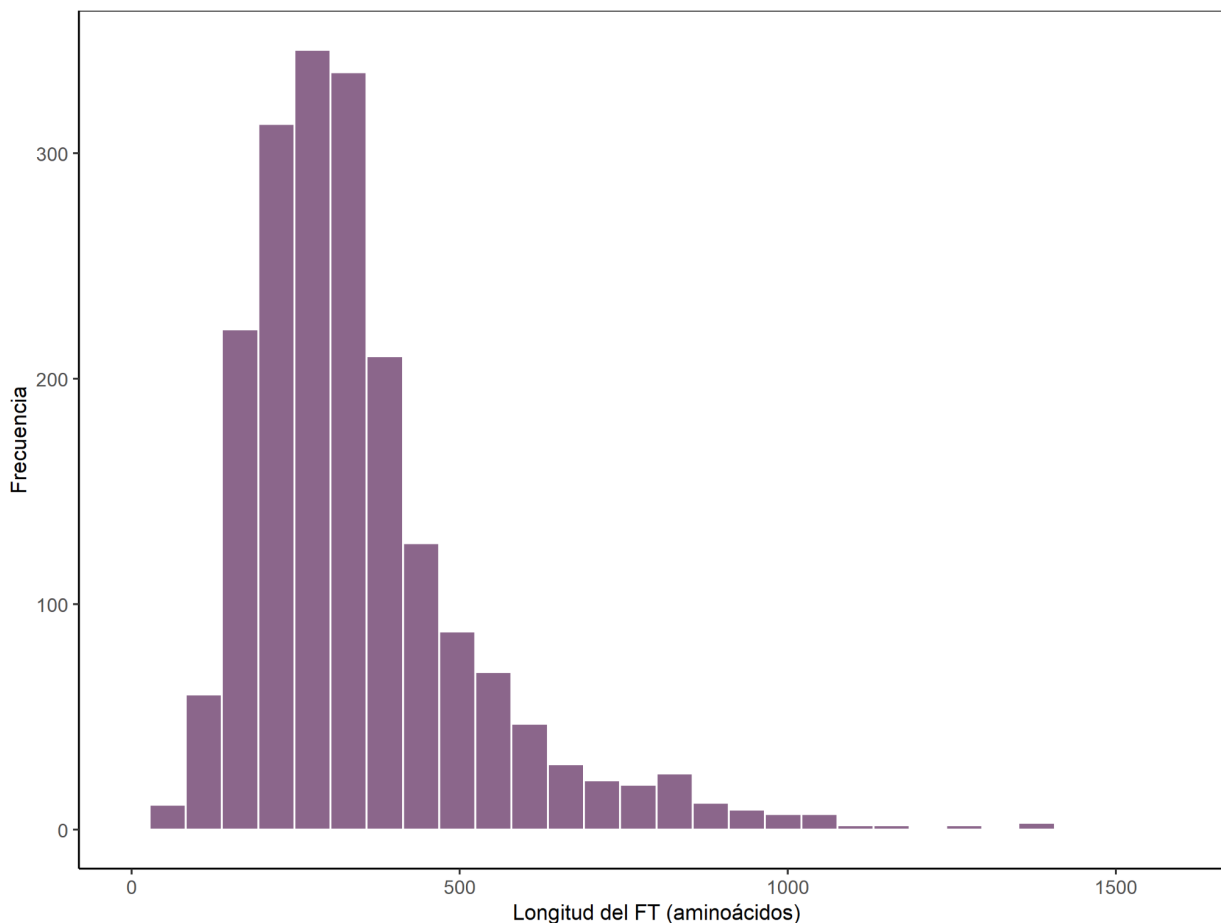


Figura 3. Distribución de la longitud (aminoácidos) de los FT de *Arabidopsis thaliana*

Debido a que las proteínas desordenadas pueden ser total o parcialmente desordenadas, primero se utilizó PONDR para calcular el número de regiones desordenadas de cada FT. Con el análisis de datos se logró determinar que los FT tienen entre 1 y 37 regiones desordenadas cada uno, con una media de 7 regiones (Figura 4), indicando que los FT de *Arabidopsis* contienen altos niveles de desorden, lo cual podría ser importante en su función.

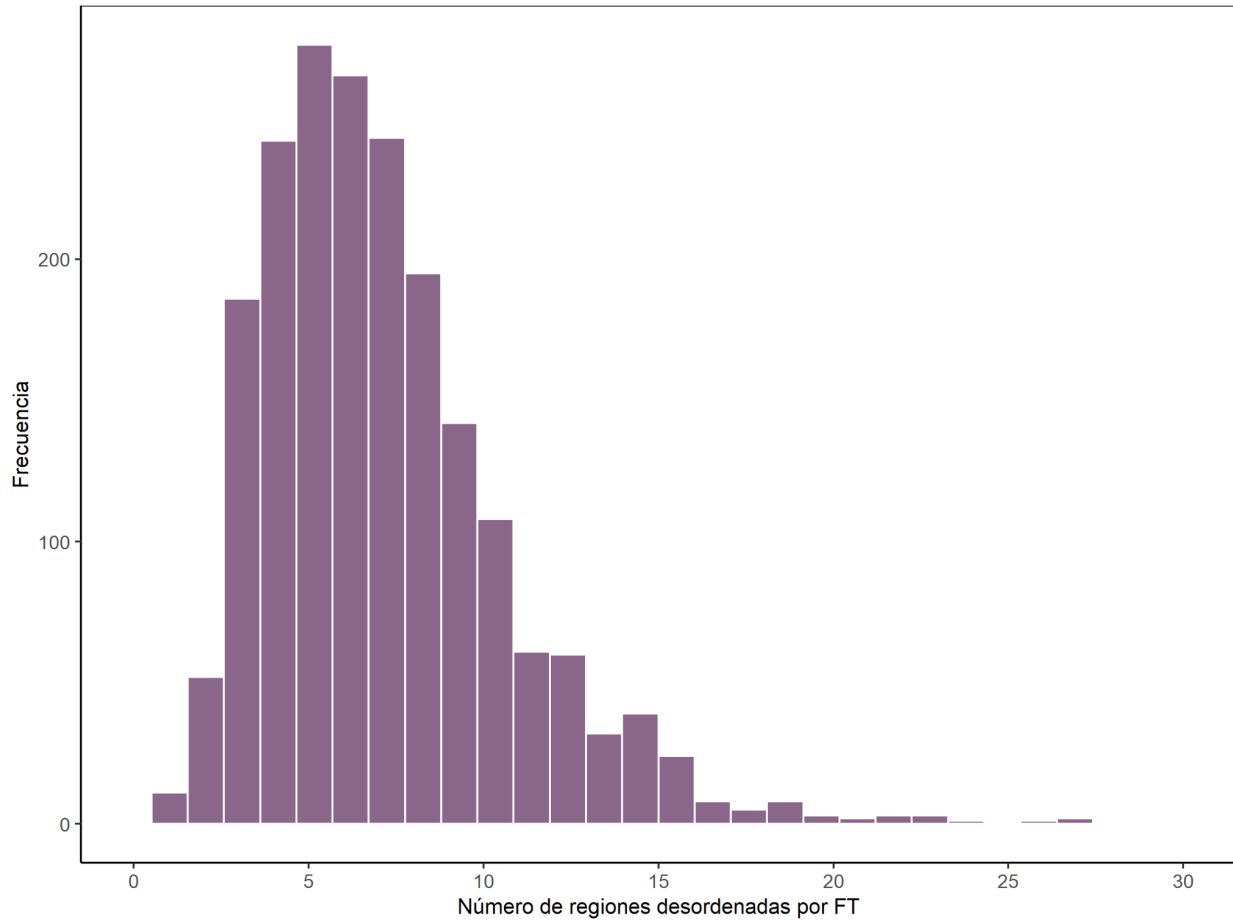


Figura 4. Distribución del número de regiones desordenadas presentes en los factores de transcripción de *Arabidopsis thaliana*

A causa del gran número de regiones desordenadas, se decidió analizar el tamaño de la región desordenada más larga de cada FT. Gracias a esto se pudo determinar que el 92 % de los FT analizados poseen regiones intrínsecamente desordenadas de por lo menos 30 residuos. Interesantemente, encontramos FT que contienen regiones de hasta 300 residuos. La longitud promedio de la región desordenada más grande fue de 29 aminoácidos (Figura 5). Para los siguientes análisis, se tomó la secuencia de la región desordenada de mayor tamaño (LDR, Longest Disordered Region).

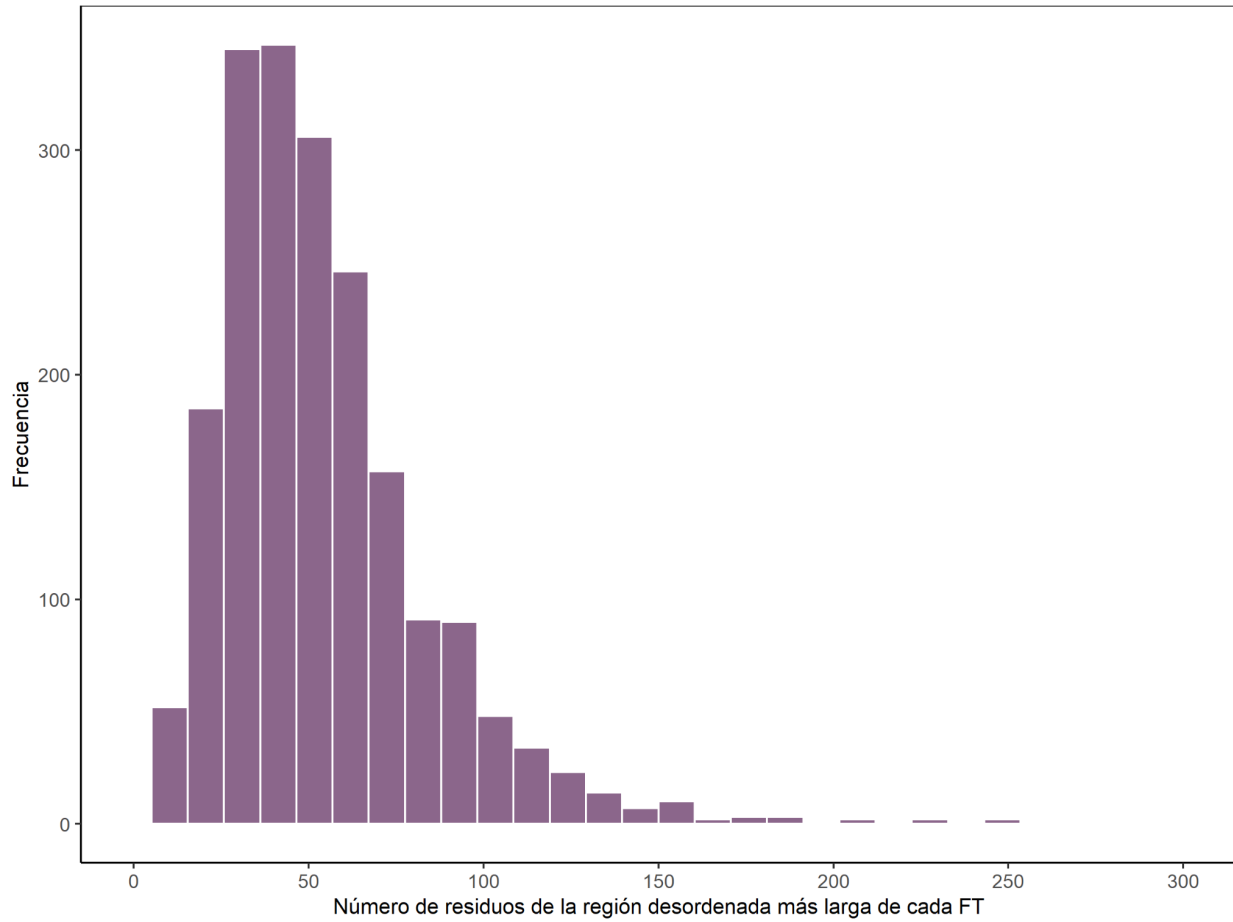


Figura 5. Distribución del tamaño de la región desordenada más larga de cada factor de transcripción de *Arabidopsis thaliana*

La LDR, en su mayoría corresponde a una fracción del 0.11 al 0.14 de todo el FT, lo que sería equivalente a un 11 a 14 % de todo el FT. Sin embargo, existen FT que tienen regiones tan largas que representan hasta el 50 % de la proteína completa. En algunos casos, la LDR representa casi todo el FT (Figura 6).

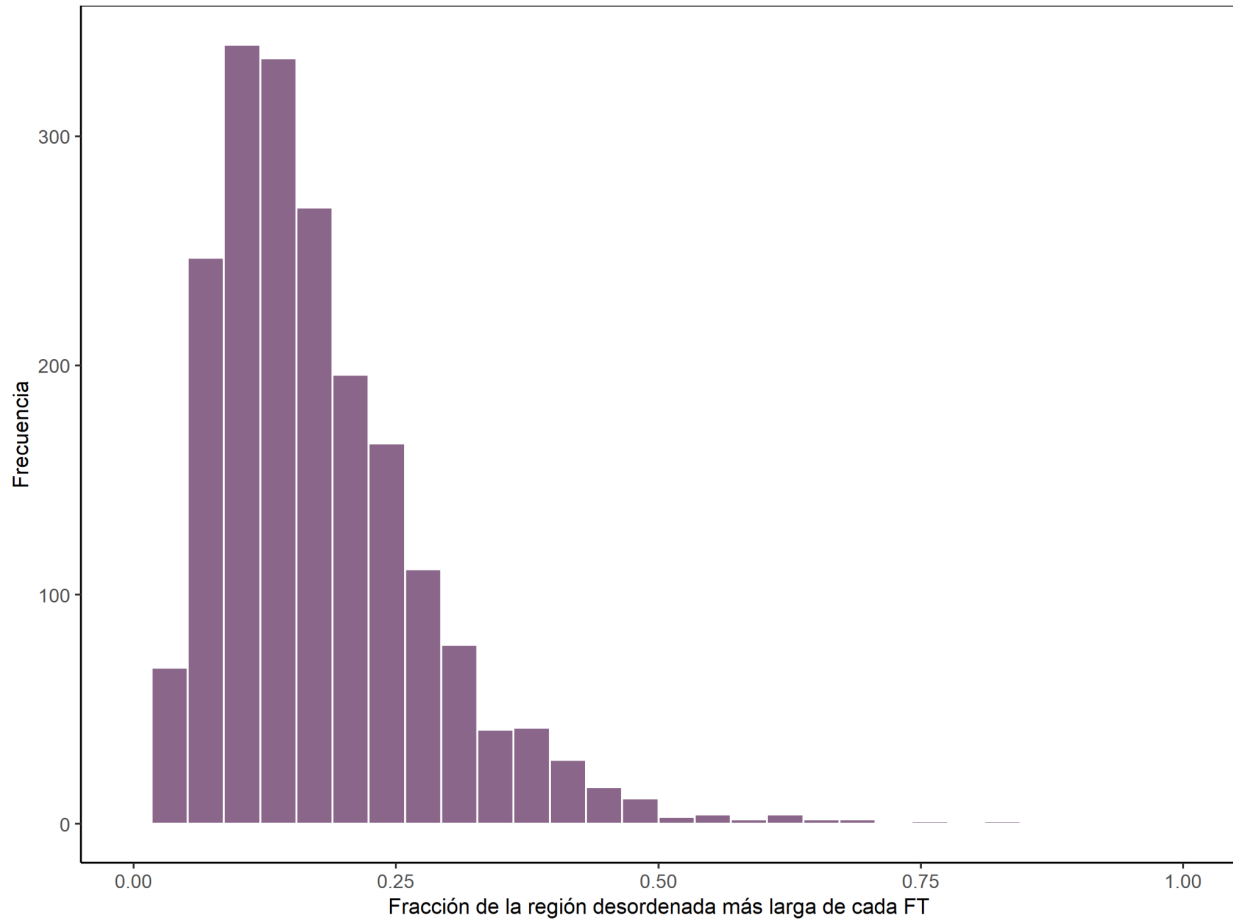


Figura 6. Distribución de la fracción correspondiente a la región desordenada más grande de cada factor de transcripción de *Arabidopsis thaliana*

Finalmente, se calculó el porcentaje total de desorden con PONDR. De forma interesante, encontramos una enorme distribución, pues existen algunos FT que se predicen completamente estructurados o que tienen niveles muy bajos de desorden, y algunos que presentan niveles de más del 60 % de desorden (Figura 7). Sin embargo, la gran mayoría de los FT poseen cierto nivel de desorden, lo cual las convierte en un grupo rico en esta característica estructural.



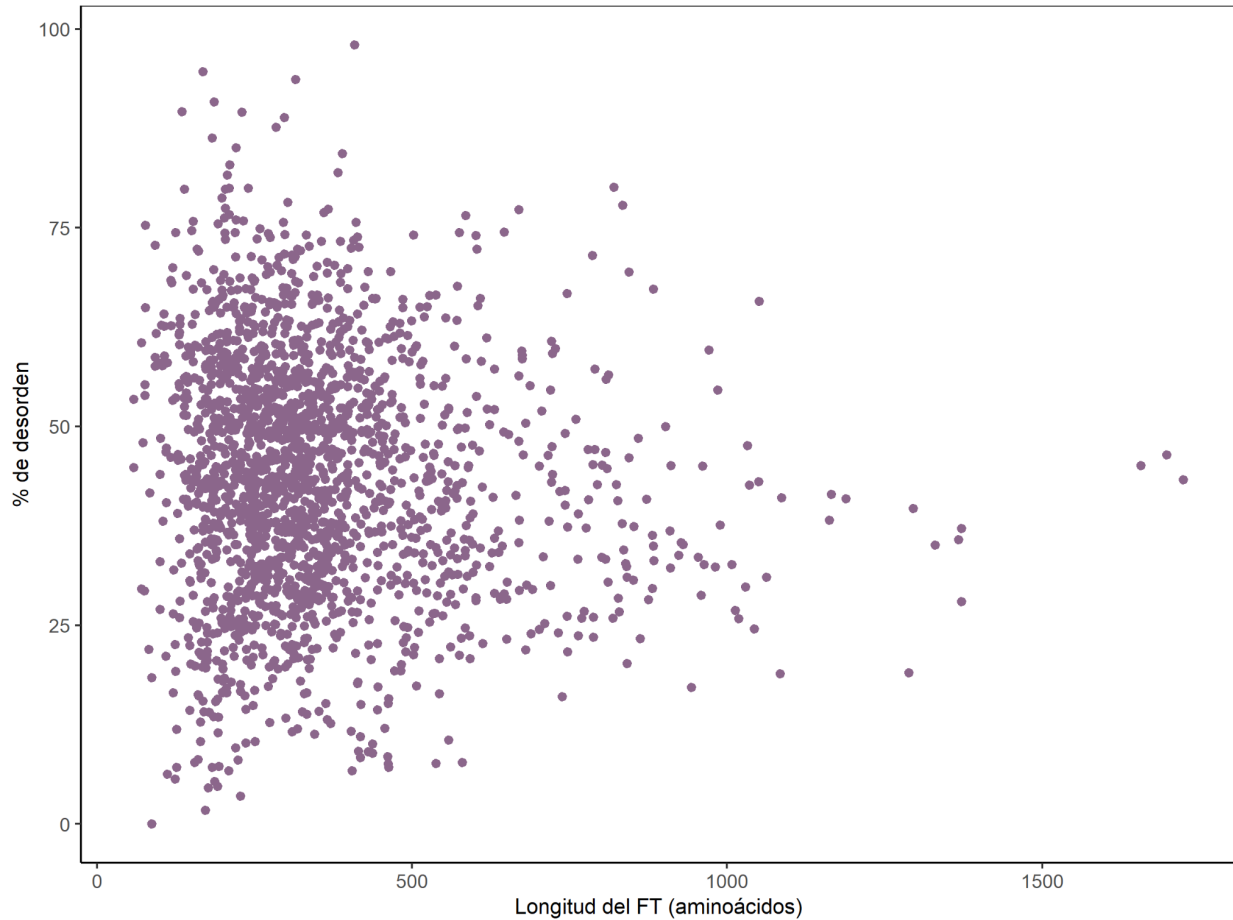


Figura 7. Porcentaje total de desorden de los FT, cada punto representa un FT

Dentro del grupo de los FT que presentan los niveles más altos de desorden tenemos a algunos muy particulares que podrían ayudarnos a entender más el papel de las regiones desordenadas en la regulación de la expresión génica. Un claro ejemplo es bZIP16 (AT2G35530) cuyo porcentaje de desorden es del 98.04 % (Figura 8). Además, se encuentran otros como HY5 (AT5G11260) con un porcentaje de desorden de 94.64% (Figura 9), GBF1 (AT4G36730) con 93.64 % (Figura 10) y NF-YA7 (AT1G30500) con 90.86 % (Figura 11). Los altos niveles de desorden encontrados sugieren alguna función regulatoria en los FT, por lo que comprender el rol de las regiones desordenadas podría ser un punto clave para entender más a profundidad cómo participa el desorden estructural en la expresión génica.

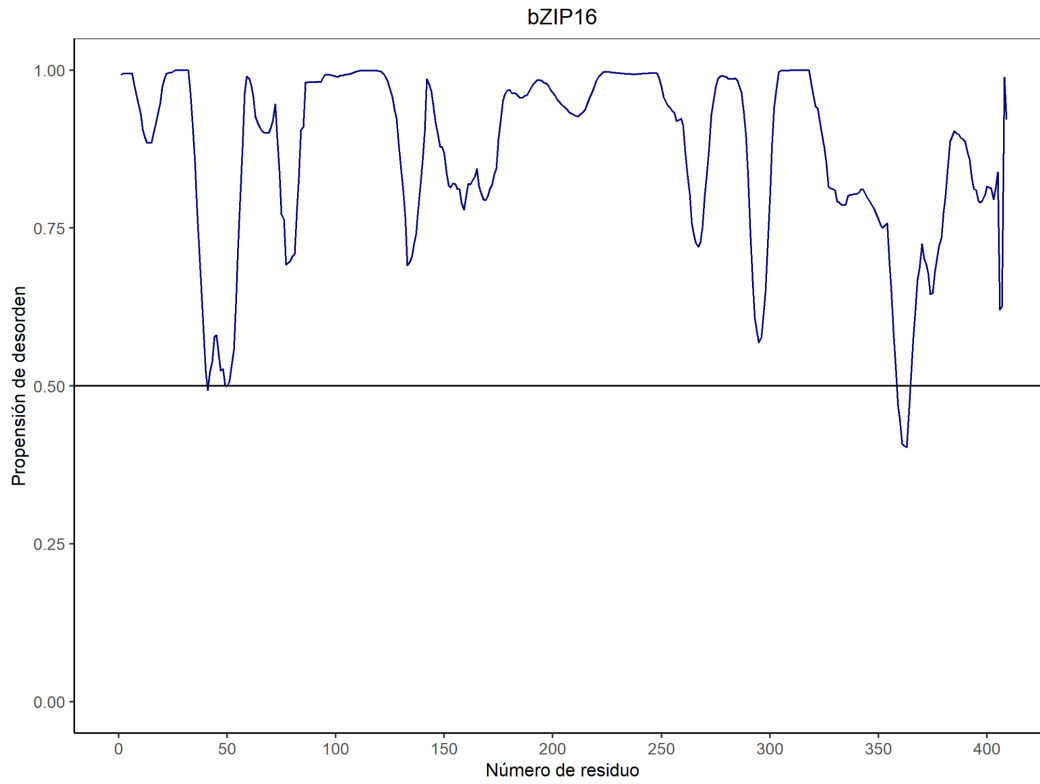


Figura 8. Predicción de desorden por residuo del FT bZIP16. Una propensión mayor a 0.50 indica alta confiabilidad en la predicción de desorden

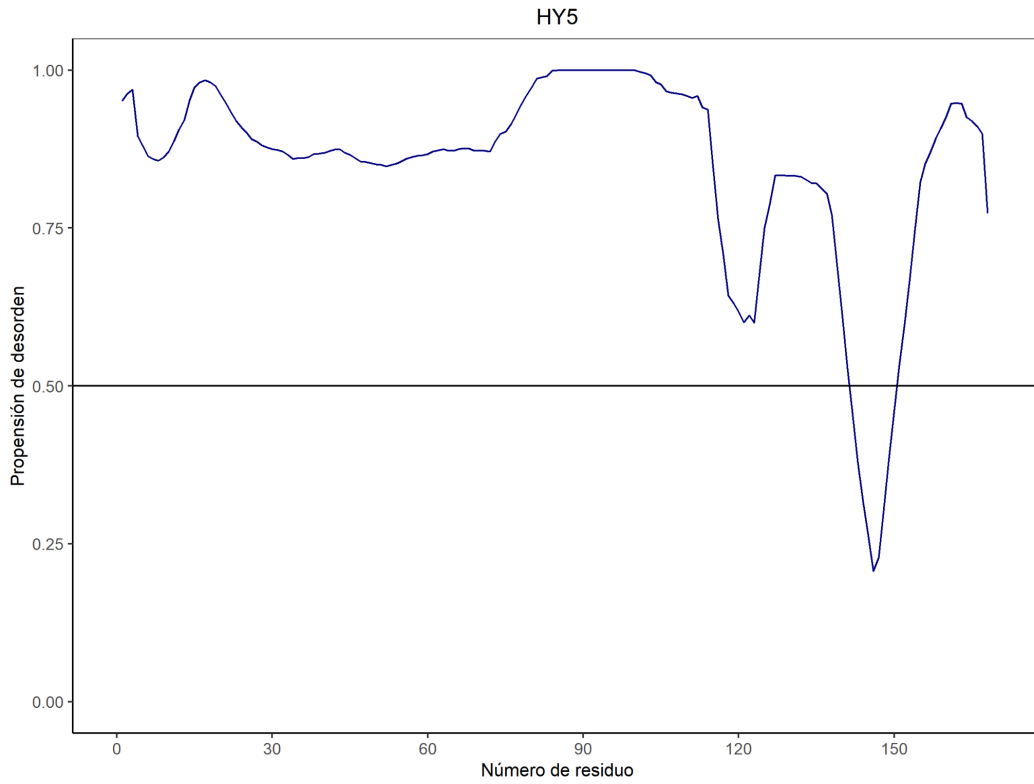


Figura 9. Predicción de desorden por residuo del FT HY5. Una propensión mayor a 0.50 indica alta confiabilidad en la predicción de desorden

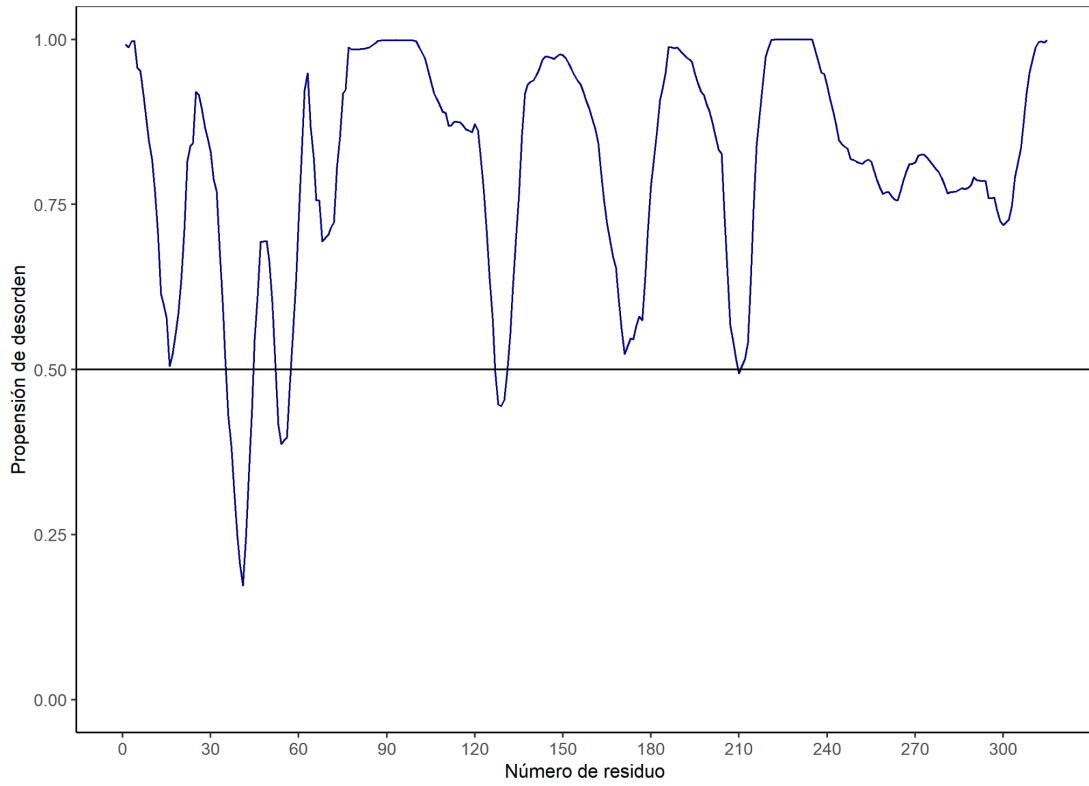


Figura 10. Predicción de desorden por residuo del FT GBF1. Una propensión mayor a 0.50 indica alta confiabilidad en la predicción de desorden.

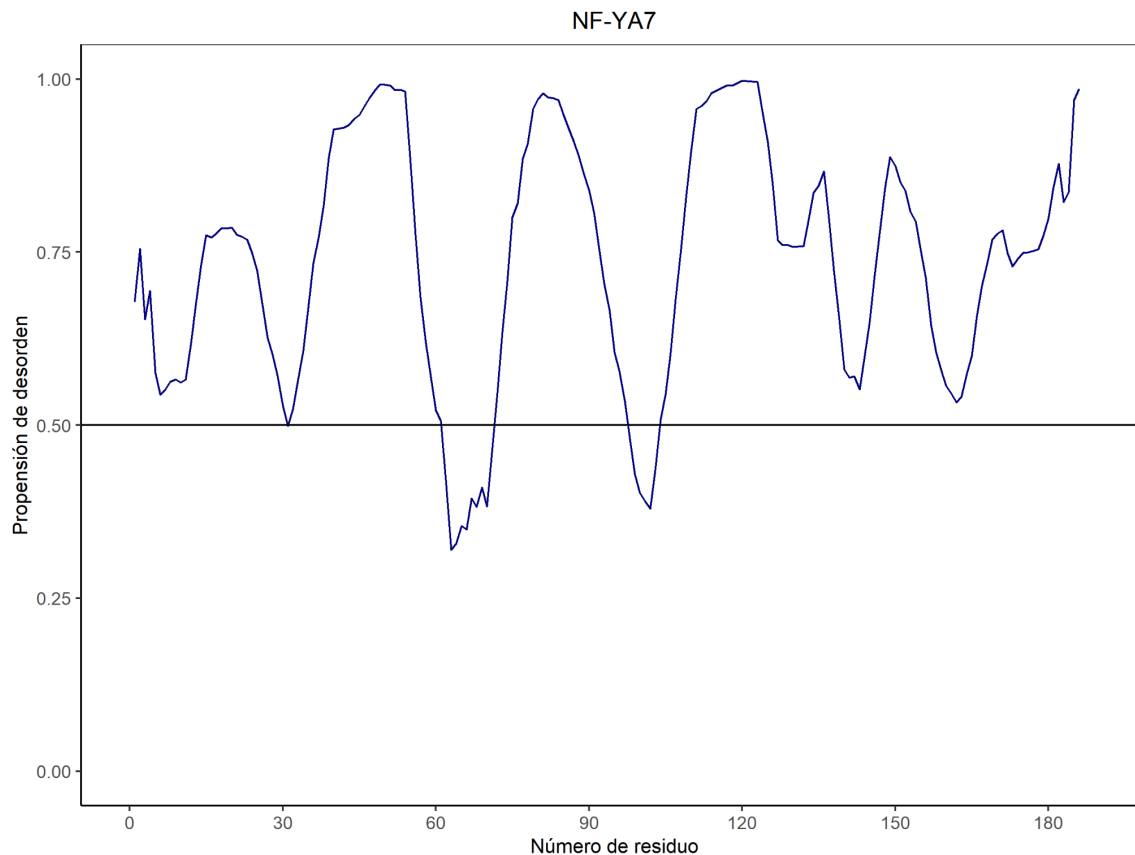


Figura 11. Predicción de desorden por residuo del FT NF-YA7. Una propensión mayor a 0.50 indica alta confiabilidad en la predicción de desorden.

### **Potencial a formar estructuras tipo alfa-hélice de las regiones desordenadas de los factores de transcripción de *Arabidopsis thaliana*.**

En estudios previos en nuestro laboratorio, se ha observado que las regiones desordenadas pueden experimentar cambios conformacionales, los cuales se pueden predecir utilizando servidores bioinformáticos. Para estudiar esta posibilidad en los FT, se utilizó AGADIR, un predictor de estructura alfa-hélice. Se realizó un análisis masivo de los FT, calculando la predicción de alfa-hélice únicamente de la región desordenada de mayor tamaño (LDR) de cada FT, sin importar el número de aminoácidos o de regiones. El resultado mostró que existen regiones desordenadas que tienen un alto potencial para formar estructuras secundarias a pesar de que también se predicen como desordenadas. Encontramos regiones desordenadas cuyo potencial a formar alfa-hélice es incluso mayor al 60 %, lo cual es sumamente relevante debido a la fidelidad del predictor (Figura 12). Estos análisis contribuyen a seleccionar FT con regiones desordenadas que sean capaces de plegarse bajo ciertas condiciones, lo cual es importante por la evidencia que demuestra que efectivamente algunos FT se pliegan al

interaccionar con su blanco, o bien, donde el plegamiento activa una función alternativa en distintos ambientes abióticos.

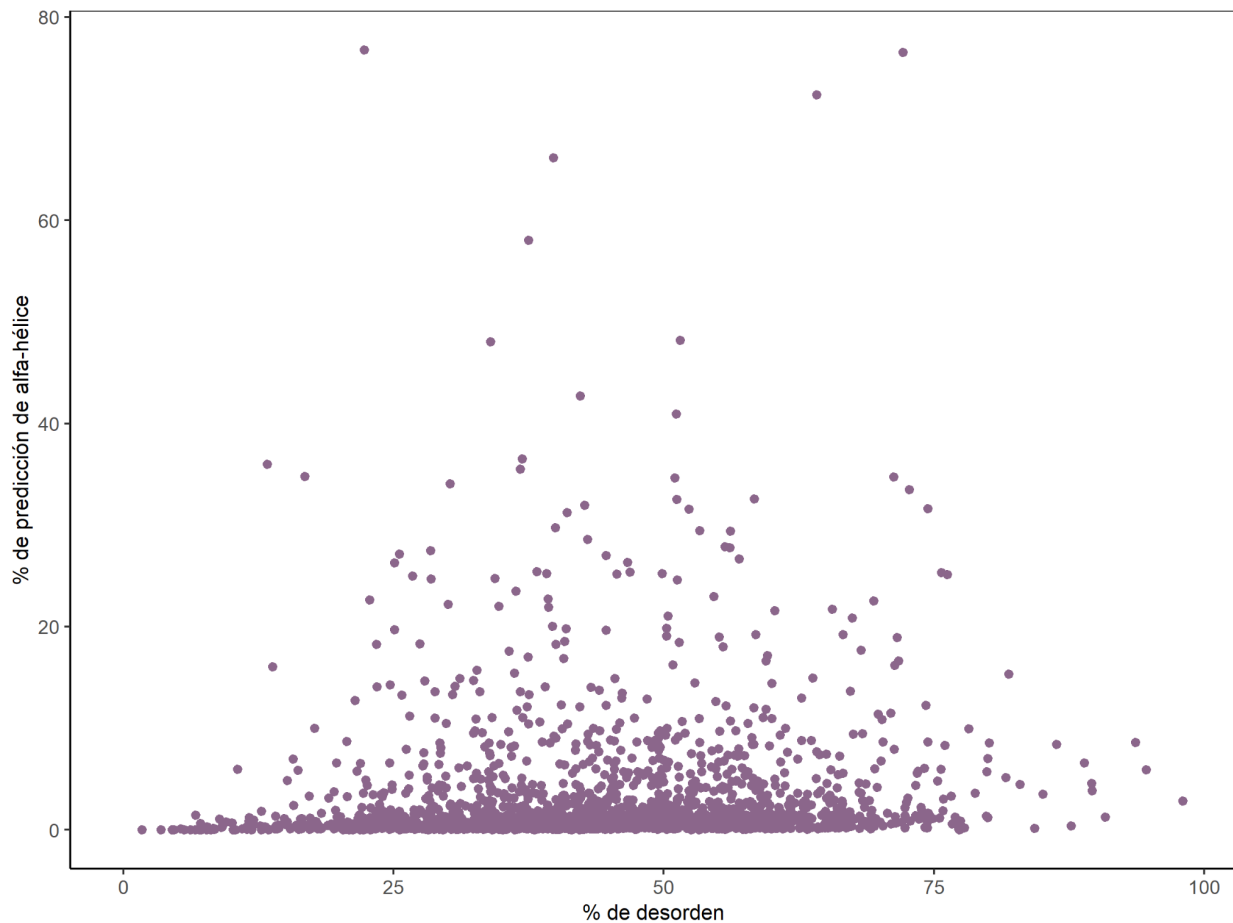


Figura 11. Predicción de estructuras alfa-hélice de la región desordenada de mayor tamaño de cada FT de *Arabidopsis thaliana*

Determinar el porcentaje de predicción de alfa-hélice nos permitió agrupar a algunos FT que además de tener un alto potencial a formar estructuras secundarias, también tuvieron un alto porcentaje total de desorden (Tabla 2). Este grupo de FT, es importante porque posee dos características de relevancia en las proteínas desordenadas, y facilitará la selección de regiones desordenadas para estudios estructurales y funcionales en distintas condiciones abióticas.

Tabla 2. Factores de transcripción con un niveles de desorden mayores a 60% y una predicción alfa-hélice mayor al 15 %

<b>Factor de transcripción</b>	<b>% de desorden</b>	<b>% de predicción alfa-h</b>
C2H2 and C2HC zinc fingers superfamily protein (AG1)	76,24	25,13
sequence-specific DNA binding transcription factor	75.68	25.34
AGAMOUS-LIKE 19	74.43	31.62
ATBZIP52	72.70	33.49
Homeodomain-like superfamily protein	72.14	76.56
Homeodomain-like superfamily protein	71.28	34.72
ARABIDOPSIS THALIANA CELL DIVISION CYCLE 5	69.43	22.53
MYB DOMAIN PROTEIN 77	67.44	20.88
HAT1	65.60	21.70
ubiquitin-associated (UBA)/TS-N domain-containing protein	64.16	72.35

### **Algunos factores de transcripción con regiones desordenadas se expresan en respuesta a estrés osmótico**

Para relacionar el desorden con la respuesta a estrés abiótico, decidimos obtener los niveles de expresión de mRNA ante estrés osmótico de aquellos factores de transcripción con un alto porcentaje de desorden (> 60 %) y una predicción de estructura alfa-hélice mayor al 20 % (Tabla 2) utilizando eFP Browser. Se tomaron los valores a 0 y 24 h de datos obtenidos de plantas de 18 días de *Arabidopsis thaliana* sometidas a tratamiento de manitol (300 mM). Con este análisis, encontramos dos FT que cambian su expresión significativamente durante condiciones de estrés: MYB77 y HAT1 (Figura 12).

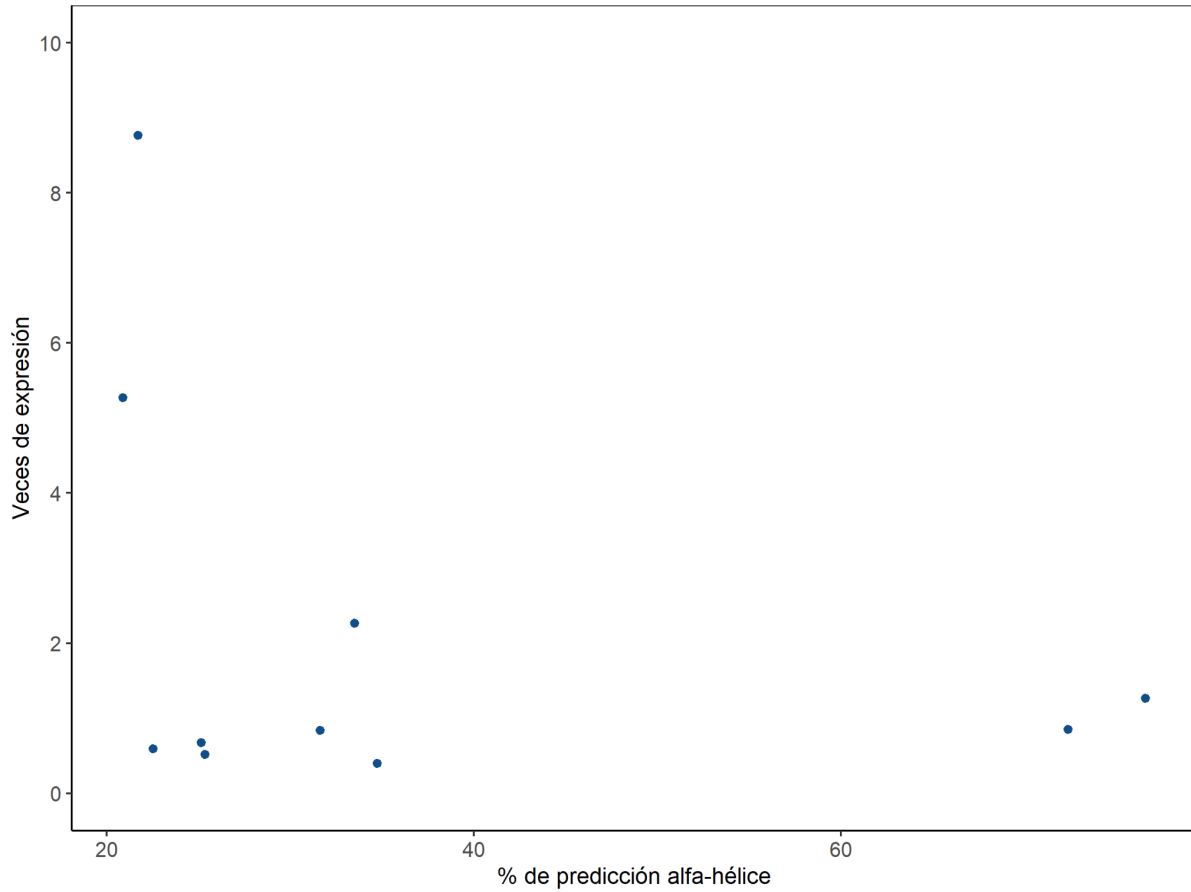


Figura 12. Cambio en los niveles de expresión de transcritos de FT con un porcentaje de desorden mayor al 60 % y una predicción alfa-hélice de más del 20 %

Por otro lado, se analizó el cambio en la expresión de transcritos de los FT cuya región desordenada tuviera una predicción de formar alfa-hélice mayor del 15 % sin importar su porcentaje de desorden total (Figura 13). De esta manera pudimos encontrar a cuatro FT que aumentan significativamente su expresión en respuesta a estrés osmótico (Tabla 3).



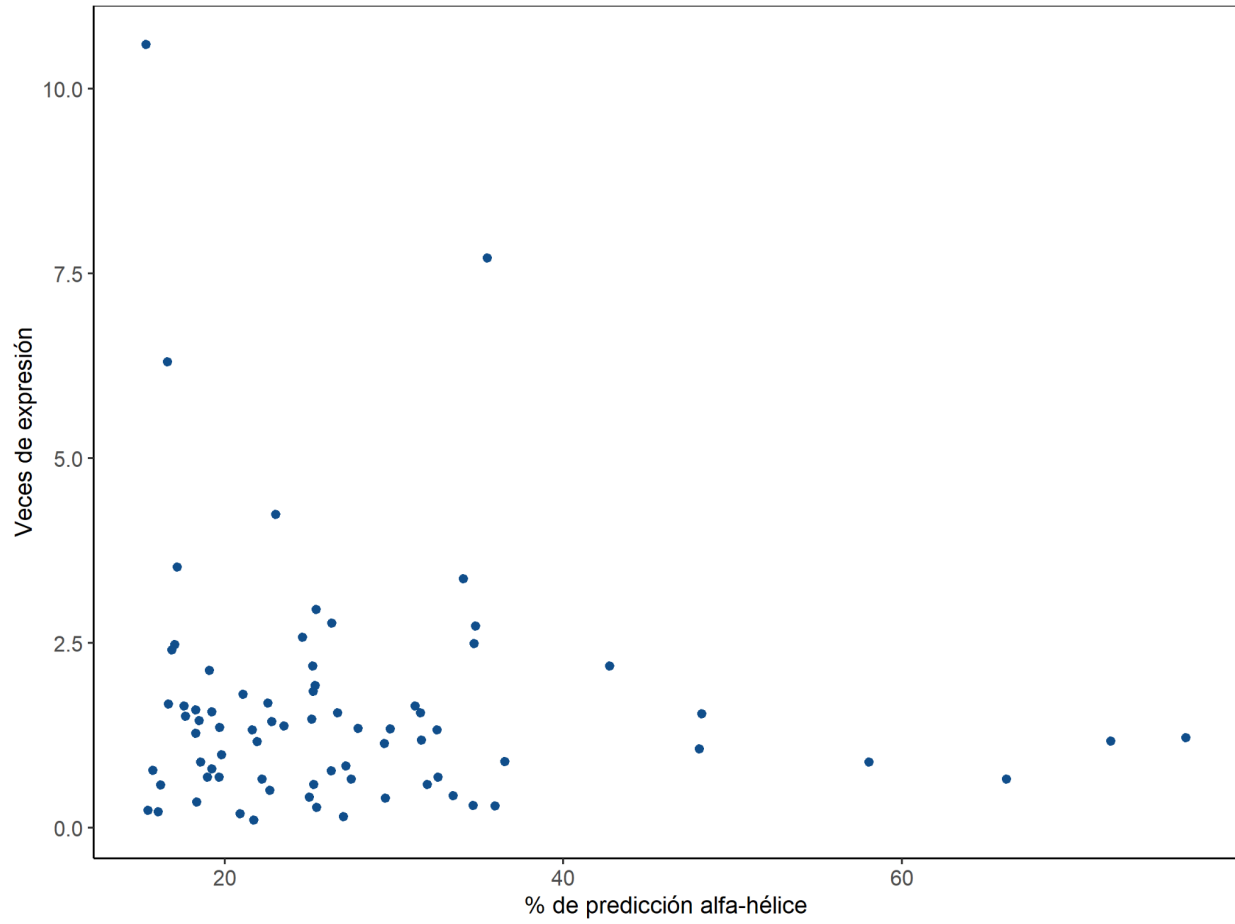


Figura 13. Cambio en los niveles de expresión de transcritos de FT con predicción alfa-hélice mayor al 15 %

De esta manera pudimos encontrar a cuatro FT que aumentan significativamente su expresión en respuesta a estrés osmótico (Tabla 3).

Tabla 3. Factores de transcripción que cambian significativamente su expresión ante estrés osmótico

<b>Nombre</b>	<b>% total de desorden</b>	<b>% predicción alfa-hélice</b>	<b>Veces de expresión</b>	<b>Descripción</b>
PAN	36.73	35.5	7.7	Implicado en desarrollo floral
HSFA6A	39.72	19.01	798.6	Proteína de choque térmico
MBF1C	59.46	16.61	6.3	Participa en respuesta a estrés (osmótico, salino, calor)
GBF3	81.94	15.32	10.6	Se expresa en respuesta a ABA (estrés hídrico por sequía y congelamiento)

#### **Familias con alto porcentaje de desorden**

Los 1,986 FT analizados pertenecen a 107 familias diferentes. Con el objetivo de determinar los niveles de desorden en las familias de FT de *Arabidopsis thaliana*, procedimos a agrupar y calcular los niveles de desorden. Encontramos que las familias con los mayores niveles de desorden son SART-1, IWS1, YL1, GIF y CCAAT, sin embargo, existen familias que poseen FT que tienen alto porcentaje de desorden pero a la vez, también tienen algunos miembros con porcentajes de desorden muy bajos (Figura 14).

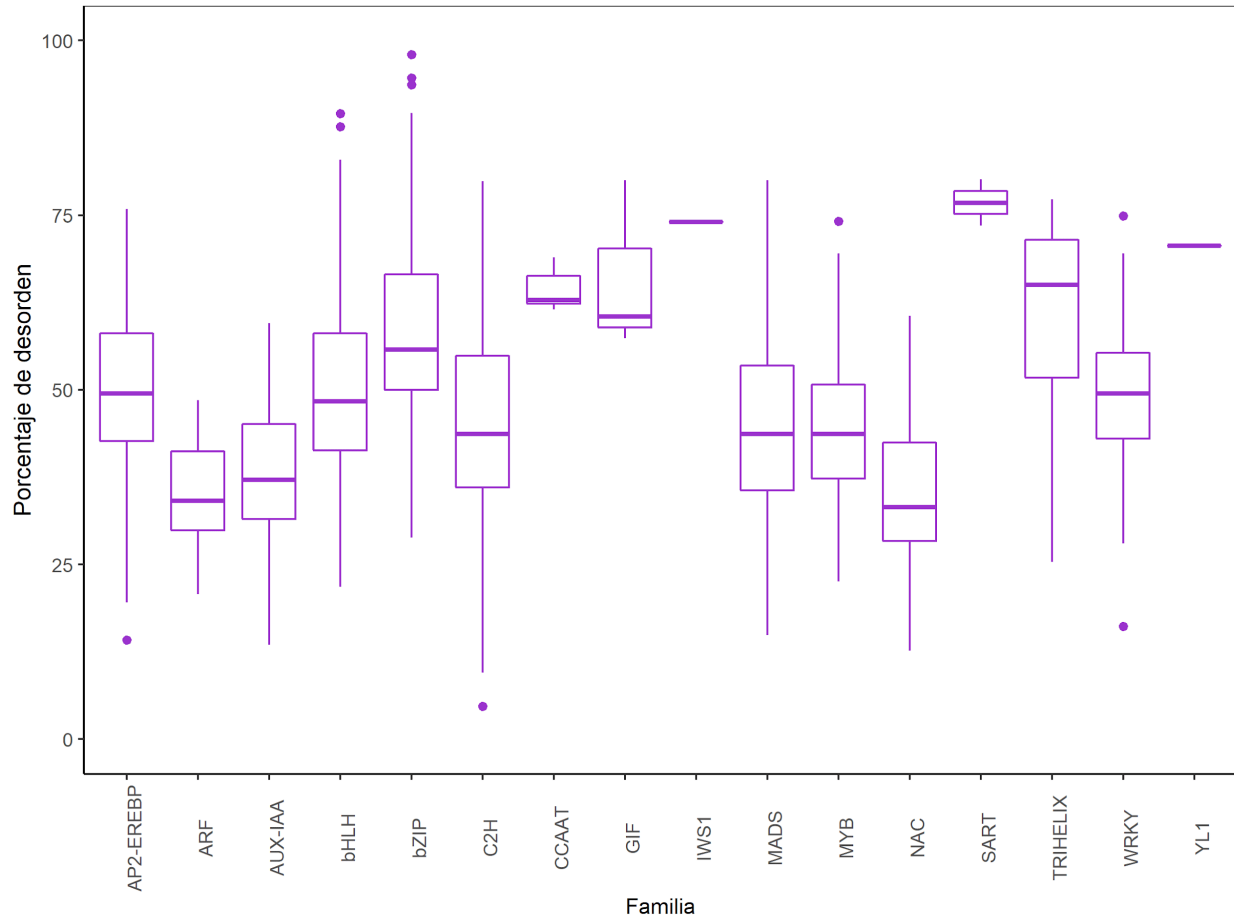


Figura 14. Porcentaje de desorden de las familias de FT de *Arabidopsis thaliana*

## DISCUSIÓN

Recientemente se ha estudiado el porcentaje de desorden dentro de los proteomas de diversos organismos. Interesantemente, *Arabidopsis thaliana* posee un nivel de desorden global relativamente bajo con respecto a otros organismos eucariotas, sin embargo, posee un mayor enriquecimiento de desorden en algunos procesos relacionados con la detección y respuesta a estímulos ambientales, lo que podría relacionarse con su plasticidad y tolerancia al estrés debido su carácter sésil (Zamora-Briseño et al., 2021). La célula necesita de una excelente regulación transcripcional para mantener el equilibrio homeostático, y parte de esta regulación está dada por los FT. Independientemente de la vía o proceso, las regiones desordenadas dentro de los FT podrían modular la red de interacciones proteína-proteína o proteína-DNA debido a sus características fisicoquímicas que les confieren una alta especificidad pero baja afinidad al momento de unirse a su blanco. Por ende, estas regiones se asocian frecuentemente con múltiples interacciones transitorias y como consecuencia, un incremento en el número de posibles uniones entre proteínas y/o DNA. Esta propiedad aumenta la complejidad de una red genética sin necesidad de aumentar el número de proteínas implicadas, y con ello promover una mayor plasticidad en la respuesta general (Pietrosemoli et al., 2013). Las regiones desordenadas cada vez demuestran tener una mayor importancia debido a los múltiples mecanismos que pueden mediar dentro de los FT. La promiscuidad molecular que se asocia a estas regiones, además de estar ligadas a las interacciones transitorias y débiles, también se puede explicar por los múltiples plegamientos que ocurren en las regiones desordenadas. Existen algunos casos donde la región desordenada puede plegarse de diversas maneras, donde cada conformación permite una unión con una proteína específica (Kannan et al., 2016). Además, las regiones desordenadas son importantes para guiar la especificidad de unión y ayudar a reconocer diferentes sitios en el DNA, y así unirse a diversos promotores (Brodsky et al., 2021).

Las regiones desordenadas también se asocian con modificaciones postraduccionales, lo cual puede ser importante en los FT, porque uno de los mecanismos que desencadenan su activación son las fosforilaciones (Ng et al., 2018). Aparentemente, las regiones desordenadas pueden utilizar estos mecanismos, ya que las modificaciones postraduccionales suelen provocar cambios conformacionales, lo cual como se ha mencionado, es clave en la promiscuidad de los FT (Darling & Uversky, 2018). Esto se debe a que las regiones desordenadas, al estar mayoritariamente desplegadas, experimentan una mayor frecuencia de modificaciones postraduccionales que las regiones ordenadas (Gao et al., 2021). Por lo tanto, estar desordenado puede representar una ventaja para transducir señales ambientales que estén reguladas por fosforilaciones, como es el caso de algunos FT.

Otra característica de las regiones desordenadas es su gran flexibilidad, esto puede ser importante para la unión de diferentes ligandos que se acoplan según la topología, ya que ser flexible

permite que se den los contactos entre los sitios de reconocimiento molecular (Shammas, 2017). Como se puede ver el desorden apunta a ser fundamental en los FT, tanto que podría mediar mecanismos específicos de cada factor o de cada red genética, e incluso, han demostrado ser tan dinámicos que posiblemente cada región de un factor con múltiples regiones, podría fácilmente mediar ciertas funciones específicas.

Debido a que la presencia de regiones desordenadas dentro de los FT podría relacionarse con múltiples funciones, aquellos factores que se predicen con un alto porcentaje de desorden podrían sin duda tener alguna función mediada por dicha región. Este podría ser el caso de algunos FT que se identificaron con altos porcentajes de desorden, como bZIP16, HY5 y GBF1. De manera interesante, estos tres FT actúan juntos en varios momentos del desarrollo para regular el crecimiento y otras respuestas fototróficas. En el caso de HY5 se ha descrito que su región N-terminal desordenada es fosforilada como parte de su regulación (Hardtke et al., 2000). Mientras que bZIP16 y GBF1, podrían ser regulados de forma similar, ya que suelen relacionarse con fotorreceptores y cinasas (Mallappa et al., 2008). Además, existe evidencia que demuestra que estos tres FT tiene un papel en la respuesta al estrés abiótico. Un estudio con mutantes de sobre-expresión y pérdida de función, demostró que GBF1, es necesario para la resistencia a la sequía (Sun et al., 2015). Mientras que HY5, está relacionado con una respuesta a ácido abscísico (ABA) y con algunos genes como DREB2A, y mutaciones en HY5 causan alteraciones en los niveles de proteínas LEA, por lo que este factor demuestra una clara participación durante el estrés hídrico (Chen et al., 2008; Sato et al., 2015). En el caso de bZIP16, también puede actuar en respuesta a ABA y giberelinas (Hsieh et al., 2012).

Finalmente, NF-YA7 forma parte de un heterotrímero llamado NF-Y, este complejo participa en procesos de crecimiento, floración y desarrollo de la semilla (Siefers et al., 2009), aunque al parecer sus miembros, especialmente la familia NF-YA, están relacionados con la tolerancia al estrés por sequía, frío, calor, incluso bajos niveles de potasio y nitrógeno. Al inicio se creía que los miembros del complejo sólo podían actuar juntos, pero al parecer un punto clave para que estos participen durante diferentes estímulos depende de con qué otros FT actúen para dirigir una respuesta específica (Leyva-González et al., 2012). Particularmente, la sobreexpresión de NF-YA7 con otros miembros del complejo aumenta la resistencia a la sequía (Zhao et al., 2017).

Todo esto indica que dichos FT pueden actuar durante una respuesta al estrés abiótico, sin embargo, no es posible relacionar de manera directa su papel, ni mucho menos definir específicamente cual es la función de las regiones desordenadas durante la percepción de los cambios ambientales. Entender la relación entre el desorden estructural y la sensibilidad al ambiente podría ayudar a explicar el papel de las regiones desordenadas en la regulación de la expresión genética en respuesta al estrés.

Dado que las proteínas desordenadas pueden plegarse en distintas conformaciones para responder dinámicamente al ambiente, utilizar un predictor de estructuras alfa-hélice nos permite hallar

FT que podrían ser capaces de funcionar de esta manera, es decir, a partir de un plegamiento inducido por cambios en el entorno. Desde esta perspectiva, podemos enfatizar el papel de los FT que están ligados a cambios ambientales para tratar de interpretar las posibles funciones de las regiones desordenadas. Dentro de las regiones desordenadas de los FT que poseen una alta predicción de formar estructuras alfa-hélice (Tabla 2) encontramos a algunas cuyo rol en las redes genéticas de los procesos biológicos en los que participan aún se desconoce. Desafortunadamente, no existen estudios directos sobre la función particular de cada uno de dichos FT, sin embargo, podemos relacionarlos a ciertos procesos gracias a análisis transcriptómicos. Los FT AT3G05760, AGL19 y bZIP52 se relacionan con el desarrollo floral, AGL19 promueve la floración después de la vernalización mediante tratamientos con frío (Suter et al., 2014); mientras que AT3G54390 y AT1G09770 (CCD5), se relacionan con desarrollo vegetativo (Meinke, 2020); AT2G38250 con la germinación de la semilla (Dekkers et al., 2013); MYB77 participa en la regulación del crecimiento de raíces laterales mediante la vía de auxinas (Shin et al., 2007); y HAT1 está implicado en la formación del gineceo, el desarrollo del fruto, y el crecimiento vegetativo (Zúñiga-Mayo et al., 2012). A pesar de que las funciones descritas no se relacionan directamente con el estrés abiótico, aun así no podemos descartar la posibilidad de que estos FT puedan plegarse en respuesta al ambiente, ya que al tratarse de plantas, muchos de sus procesos como la floración, el crecimiento vegetativo, el desarrollo de órganos, y la germinación son mediados también por factores abióticos como la luz, gravedad y temperatura (O'Connor et al., 2020; Yavari et al., 2021). Y además, las regiones desordenadas ya han demostrado su capacidad para responder por medio de plegamientos a factores como la temperatura.

Dentro del grupo de FT con una predicción dual de desorden/alfa-hélice (Tabla 2) también se encuentran algunos que posiblemente están relacionados con el estrés, como AT1G06850 (bZIP52) y AT2G38250. Un análisis transcriptómico demostró que estos FT se expresan durante condiciones de sequía (Ding et al., 2013), mientras que AT5G01380 (gt-3A) durante estrés salino (Ma et al., 2006), e interesantemente gt-3a y bZIP52 durante tratamientos de frío y calor, respectivamente (Lee et al., 2005). Durante el estrés por calor bZIP52 se acumula en el núcleo gracias a una desfosforilación que desencadena su transporte desde el citoplasma, esto permite que regule genes relacionados con el metabolismo energético, estrés, y traducción (Wiese et al., 2021), este mecanismo sugiere que posiblemente pueda ocurrir un cambio conformacional causado por la desfosforilación, lo cual podría permitir el transporte de este FT al núcleo; GT-3a, es un FT no muy conocido y forma parte de un grupo exclusivo de FT de plantas (Ayadi et al., 2004). Por último en este grupo tenemos a HAT1, este factor es importante para el crecimiento, defensa por patógenos, y en la respuesta a estrés por sequía mediada por ABA, donde funciona como un regulador negativo de la síntesis y señalización, por lo cual debe ser inhibido durante el estrés, y para esto debe ser fosforilado (Tan et al., 2018). Al igual que HAT1, CCD5 y F13M7.16 (AT1G04850) también están relacionados con la respuesta a patógenos (Ascencio-Ibáñez et al., 2008).

El grupo de FT que aumentan su expresión durante estrés osmótico y que poseen regiones desordenadas con una predicción de alfa-hélice mayor al 15 % (Tabla 3) se convierte en un blanco importante, ya que la sobreexpresión de MBF1c y GBF3 confieren tolerancia al estrés hídrico (por sequía, salinidad y congelamiento), pero en ambos casos, no existe una caracterización de su estructura. Aún así, estos FT resultan bastante llamativos debido al alto porcentaje de desorden que poseen y la posible capacidad de adquirir estructuras plegadas.

Un detalle interesante acerca de MBF1c, es que existen tres genes homólogos (MBF1a, MBF1b y MBF1c) de este tipo en *Arabidopsis thaliana*, pero interesantemente solo MBF1c se relaciona con la respuesta al estrés abiótico por múltiples factores como salinidad, estrés osmótico, calor, e infección bacteriana (Suzuki et al., 2005), sin embargo, el mecanismo es desconocido. Una posibilidad radica en que una proteína desordenada podría regular las señales ambientales por medio de su desorden. La función biológica de GBF3 por la cual participa regulando el estrés hídrico se cree que puede ser por la vía de señalización de ABA debido a que regula la expresión de genes relacionados con esa vía durante el estrés (Ramegowda et al., 2017). Existe también dentro de este grupo un factor de choque térmico, HSFA6A, el cual confiere tolerancia a la sequía (Hwang et al., 2014), posiblemente por medio de la regulación en la vía de señalización de ROS (Wenjing et al., 2020). El hecho de que algunos FT estén bien relacionados con la tolerancia al déficit hídrico nos permite proponer una posible relación entre el estrés abiótico y el desorden, así como el potencial de formar estructuras secundarias tipo alfa-hélice.

Es importante destacar que el modelo de cambio conformacional relacionado con el entorno es un mecanismo nuevo que podría fungir como un factor relevante que le confiere una mayor plasticidad a la respuesta de las plantas, por todo el dinamismo molecular que representa, ya que un cambio estructural estaría asociado con percibir y responder ante cambios en el ambiente. Este tipo de respuesta sería muy útil debido a que las plantas son sésiles, y diversificar la forma de percibir el entorno seguramente estaría ligado a su éxito evolutivo. El trabajo realizado en esta tesis permitió encontrar y caracterizar algunos FT que indican tener alguna función mediada por regiones desordenadas. Estudios posteriores serán necesarios para demostrar realmente este hecho.

## CONCLUSIONES

La caracterización global de los niveles de desorden en los FT de *Arabidopsis thaliana* demuestra que son proteínas que se predicen con un enriquecimiento de desorden. El desorden estructural puede relacionarse con una función directa mediada por las regiones desordenadas.

Este estudio permitió identificar a MBF1c, GBF3 y H8SFA6A, proteínas que tienen un papel en la tolerancia al estrés hídrico, como FT ricos en desorden que puede formar estructuras alfa-hélice. Además de otro grupo de FT con las mismas características que se relacionan con una respuesta al estrés abiótico mediante análisis transcriptómicos. Estos FT podrían ayudar a entender el papel de las regiones desordenadas y su relación con el entorno, por lo que se convierten en un blanco de estudios experimentales.



## REFERENCIAS

- Amara, I., Zaidi, I., Masmoudi, K., Ludevid, M. D., Pagès, M., Goday, A., Brini, F., Amara, I., Zaidi, I., Masmoudi, K., Ludevid, M. D., Pagès, M., Goday, A., & Brini, F. (2014). Insights into Late Embryogenesis Abundant (LEA) Proteins in Plants: From Structure to the Functions. *American Journal of Plant Sciences*, 5(22), 3440–3455. <https://doi.org/10.4236/AJPS.2014.522360>
- Ascencio-Ibáñez, J. T., Sozzani, R., Lee, T. J., Chu, T. M., Wolfinger, R. D., Cella, R., & Hanley-Bowdoin, L. (2008). Global Analysis of Arabidopsis Gene Expression Uncovers a Complex Array of Changes Impacting Pathogen Response and Cell Cycle during Geminivirus Infection. *Plant Physiology*, 148(1), 436–454. <https://doi.org/10.1104/PP.108.121038>
- Ayadi, M., Delaporte, V., Li, Y. F., & Zhou, D. X. (2004). Analysis of GT-3a identifies a distinct subgroup of trihelix DNA-binding transcription factors in Arabidopsis. *FEBS Letters*, 562(1–3), 147–154. [https://doi.org/10.1016/S0014-5793\(04\)00222-4](https://doi.org/10.1016/S0014-5793(04)00222-4)
- Barberis, A., & Petrascheck, M. (2003). Transcription Activation in Eukaryotic Cells. *ELS*. <https://doi.org/10.1038/NPG.ELS.0003303>
- Brodsky, S., Jana, T., & Barkai, N. (2021). Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. *Current Opinion in Structural Biology*, 71, 110–115. <https://doi.org/10.1016/J.SBI.2021.06.011>
- Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D. K., Carmi, M., & Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Molecular Cell*, 79(3), 459–471.e4. <https://doi.org/10.1016/J.MOLCEL.2020.05.032>
- Chen, H., Zhang, J., Neff, M. M., Hong, S. W., Zhang, H., Deng, X. W., & Xiong, L. (2008). Integration of light and abscisic acid signaling during seed germination and early seedling development. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11), 4495–4500. [https://doi.org/10.1073/PNAS.0710778105/SUPPL\\_FILE/INDEX.HTML](https://doi.org/10.1073/PNAS.0710778105/SUPPL_FILE/INDEX.HTML)
- Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M., & Shinozaki, K. (2011). Effects of abiotic stress on plants: A systems biology perspective. *BMC Plant Biology*, 11(1), 1–14. <https://doi.org/10.1186/1471-2229-11-163/FIGURES/2>
- Cuevas-Velázquez, C. L., & Covarrubias-Robles, A. A. (2011). Las proteínas desordenadas y su función: una nueva forma de ver la estructura de las proteínas y la respuesta de las plantas al estrés. *Revista Especializada En Ciencias Químico-Biológicas*, 14(2), 97–105. [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1405-888X2011000200004&lng=es&nr m=iso&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-888X2011000200004&lng=es&nr m=iso&tlng=es)
- Cuevas-Velazquez, C. L., Saab-Rincón, G., Reyes, J. L., & 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. <https://doi.org/10.1074/JBC.M116.720318>
- Darling, A. L., & Uversky, V. N. (2018). Intrinsic disorder and posttranslational modifications: The darker side of the biological dark matter. *Frontiers in Genetics*, 9(MAY), 158. <https://doi.org/10.3389/FGENE.2018.00158/BIBTEX>
- Davey, N. E. (2019). The functional importance of structure in unstructured protein regions. *Current Opinion in Structural Biology*, 56, 155–163. <https://doi.org/10.1016/J.SBI.2019.03.009>
- De Lucas, M., Provart, N. J., & Brady, S. M. (2014). Bioinformatic tools in arabidopsis research. *Methods in Molecular Biology*, 1062, 97–136. [https://doi.org/10.1007/978-1-62703-580-4\\_5/FIGURES/000520](https://doi.org/10.1007/978-1-62703-580-4_5/FIGURES/000520)
- Dekkers, B. J. W., Pearce, S., van Bolderen-Veldkamp, R. P., Marshall, A., Widera, P., Gilbert, J., Drost, H. H., Bassel, G. W., Müller, K., King, J. R., Wood, A. T. A., Grosse, I., Quint, M., Krasnogor, N., Leubner-Metzger, G., Holdsworth, M. J., & Bentsink, L. (2013). Transcriptional Dynamics of Two Seed Compartments with Opposing Roles in Arabidopsis Seed Germination. *Plant Physiology*, 163(1), 205–215. <https://doi.org/10.1104/PP.113.223511>
- Ding, Y., Liu, N., Virlouvet, L., Riethoven, J. J., Fromm, M., & Avramova, Z. (2013). Four distinct types of dehydration stress memory genes in Arabidopsis thaliana. *BMC Plant Biology*, 13(1), 1–11. <https://doi.org/10.1186/1471-2229-13-229/FIGURES/4>
- Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C. H., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., & Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, 19(1), 26–59.

- [https://doi.org/10.1016/S1093-3263\(00\)00138-8](https://doi.org/10.1016/S1093-3263(00)00138-8)
- Gao, C., Ma, C., Wang, H., Zhong, H., Zang, J., Zhong, R., He, F., & Yang, D. (2021). Intrinsic disorder in protein domains contributes to both organism complexity and clade-specific functions. *Scientific Reports* 2021 11:1, 11(1), 1–18. <https://doi.org/10.1038/s41598-021-82656-9>
- García-Morales, S., Gómez-Merino, F. C., Trejo-Téllez, L. I., & Herrera-Cabrera, É. B. (2013). Factores de transcripción involucrados en respuestas moleculares de las plantas al estrés osmótico. *Revista Fitotecnia Mexicana*, 36(2), 105–115. [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0187-73802013000200003&lng=es&nrm=iso&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-73802013000200003&lng=es&nrm=iso&tlng=es)
- Hardtke, C. S., Gohda, K., Osterlund, M. T., Oyama, T., Okada, K., & Deng, X. W. (2000). HY5 stability and activity in arabidopsis is regulated by phosphorylation in its COP1 binding domain. *The EMBO Journal*, 19(18), 4997–5006. <https://doi.org/10.1093/EMBOJ/19.18.4997>
- He, M., He, C. Q., & Ding, N. Z. (2018). Abiotic stresses: General defenses of land plants and chances for engineering multistress tolerance. *Frontiers in Plant Science*, 871, 1771. <https://doi.org/10.3389/FPLS.2018.01771/BIBTEX>
- Hirayama, T., & Shinozaki, K. (2010). Research on plant abiotic stress responses in the post-genome era: past, present and future. *The Plant Journal*, 61(6), 1041–1052. <https://doi.org/10.1111/J.1365-313X.2010.04124.X>
- Hrmova, M., & Hussain, S. S. (2021). Plant Transcription Factors Involved in Drought and Associated Stresses. *International Journal of Molecular Sciences* 2021, Vol. 22, Page 5662, 22(11), 5662. <https://doi.org/10.3390/IJMS22115662>
- Hsiao, A. S. (2022). Plant Protein Disorder: Spatial Regulation, Broad Specificity, Switch of Signaling and Physiological Status. *Frontiers in Plant Science*, 13, 1799. <https://doi.org/10.3389/FPLS.2022.904446/BIBTEX>
- Hsieh, W. P., Hsieh, H. L., & Wu, S. H. (2012). Arabidopsis bZIP16 Transcription Factor Integrates Light and Hormone Signaling Pathways to Regulate Early Seedling Development. *The Plant Cell*, 24(10), 3997–4011. <https://doi.org/10.1105/TPC.112.105478>
- Hwang, S. M., Kim, D. W., Woo, M. S., Jeong, H. S., Son, Y. S., Akhter, S., Choi, G. J., & Bahk, J. D. (2014). Functional characterization of ArabidopsisHsfA6a as a heat-shock transcription factor under high salinity and dehydration conditions. *Plant, Cell and Environment*, 37(5), 1202–1222. <https://doi.org/10.1111/PCE.12228/SUPPINFO>
- Joshi, R., Wani, S. H., Singh, B., Bohra, A., Dar, Z. A., Lone, A. A., Pareek, A., & Singla-Pareek, S. L. (2016). Transcription factors and plants response to drought stress: Current understanding and future directions. *Frontiers in Plant Science*, 7(2016JULY), 1029. <https://doi.org/10.3389/FPLS.2016.01029/BIBTEX>
- Kang, M., Kim, S., Kim, H. J., Shrestha, P., Yun, J. hye, Phee, B. K., Lee, W., Nam, H. G., & Chang, I. (2018). The C-Domain of the NAC Transcription Factor ANAC019 Is Necessary for pH-Tuned DNA Binding through a Histidine Switch in the N-Domain. *Cell Reports*, 22(5), 1141–1150. <https://doi.org/10.1016/J.CELREP.2018.01.002>
- Kannan, S., Lane, D. P., & Verma, C. S. (2016). Long range recognition and selection in IDPs: the interactions of the C-terminus of p53. *Scientific Reports* 2016 6:1, 6(1), 1–13. <https://doi.org/10.1038/srep23750>
- Khan, S. A., Li, M. Z., Wang, S. M., & Yin, H. J. (2018). Revisiting the Role of Plant Transcription Factors in the Battle against Abiotic Stress. *International Journal of Molecular Sciences* 2018, Vol. 19, Page 1634, 19(6), 1634. <https://doi.org/10.3390/IJMS19061634>
- Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., Angelo, C. D. ', Bornberg-Bauer, E., Rg Kudla, J., & Harter, K. (n.d.). *The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses*. <https://doi.org/10.1111/j.1365-313X.2007.03052.x>
- Krämer, U. (2015). Planting molecular functions in an ecological context with arabidopsis thaliana. *ELife*, 2015(4). <https://doi.org/10.7554/ELIFE.06100.001>
- Krishnakumar, V., Hanlon, M. R., Contrino, S., Ferlanti, E. S., Karamycheva, S., Kim, M., Rosen, B. D., Cheng, C. Y., Moreira, W., Mock, S. A., Stubbs, J., Sullivan, J. M., Krampis, K., Miller, J. R., Micklem, G., Vaughn, M., & Town, C. D. (2015). Araport: the Arabidopsis Information Portal. *Nucleic Acids Research*, 43(D1), D1003–D1009. <https://doi.org/10.1093/NAR/GKU1200>
- Kummu, M., Guillaume, J. H. A., De Moel, H., Eisner, S., Flörke, M., Porkka, M., Siebert, S., Veldkamp, T.

- I. E., & Ward, P. J. (2016). The world's road to water scarcity: shortage and stress in the 20th century and pathways towards sustainability. *Scientific Reports* 2016 6:1, 6(1), 1–16. <https://doi.org/10.1038/srep38495>
- Lacroix, E., Viguera AR & Serrano, L. (1998). Elucidating the folding problem of a-helices: Local motifs, long-range electrostatics, ionic strength dependence and prediction of NMR parameters. *J. Mol. Biol.* 284, 173-191.
- Lee, B. H., Henderson, D. A., & Zhu, J. K. (2005). The Arabidopsis Cold-Responsive Transcriptome and Its Regulation by ICE1. *The Plant Cell*, 17(11), 3155–3175. <https://doi.org/10.1105/TPC.105.035568>
- Leyva-González, M. A., Ibarra-Laclette, E., Cruz-Ramírez, A., & Herrera-Estrella, L. (2012). Functional and Transcriptome Analysis Reveals an Acclimatization Strategy for Abiotic Stress Tolerance Mediated by Arabidopsis NF-YA Family Members. *PLOS ONE*, 7(10), e48138. <https://doi.org/10.1371/JOURNAL.PONE.0048138>
- Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., & Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry*, 45(22), 6873–6888. <https://doi.org/10.1021/BI0602718/ASSET/IMAGES/LARGE/BI0602718F00009.JPEG>
- Li, X., Romero, P., Rani, M., Dunker, A. K., & Obradovic, Z. (1999). Predicting protein disorder for N-, C-and internal regions. *Genome informatics*, 10, 30-40.
- Ma, S., Gong, Q., & Bohnert, H. J. (2006). Dissecting salt stress pathways. *Journal of Experimental Botany*, 57(5), 1097–1107. <https://doi.org/10.1093/JXB/ERJ098>
- Mahapatra, K., & Roy, S. (2019). An insight into the folding and stability of Arabidopsis thaliana SOG1 transcription factor under salinity stress in vitro. *Biochemical and Biophysical Research Communications*, 515(4), 531–537. <https://doi.org/10.1016/J.BBRC.2019.05.183>
- Mallappa, C., Singh, A., Ram, H., & Chattopadhyay, S. (2008). GBF1, a transcription factor of blue light signaling in Arabidopsis, is degraded in the dark by a proteasome-mediated pathway independent of COP1 and SPA1. *The Journal of Biological Chemistry*, 283(51), 35772–35782. <https://doi.org/10.1074/JBC.M803437200>
- Meinke, D. W. (2020). Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. *New Phytologist*, 226(2), 306–325. <https://doi.org/10.1111/NPH.16071>
- Mitsuda, N., & Ohme-Takagi, M. (2009). Functional Analysis of Transcription Factors in Arabidopsis. *Plant and Cell Physiology*, 50(7), 1232–1248. <https://doi.org/10.1093/PCP/PCP075>
- Muñoz, V. & Serrano, L. (1994a). Elucidating the folding problem of helical peptides using empirical parameters. *Nature: Struct. Biol.* 1, 399-409.
- Muñoz, V. & Serrano, L. (1994b). Elucidating the folding problem of a-helical peptides using empirical parameters, II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol* 245, 275-296.
- Muñoz, V. & Serrano, L. (1994b). Elucidating the folding problem of a-helical peptides using empirical parameters, II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol* 245, 275-296.
- Muñoz, V. & Serrano, L. (1994c). Elucidating the folding problem of a-helical peptides using empirical parameters III: Temperature and pH dependence. *J. Mol. Biol* 245, 297-308.
- Muñoz, V. & Serrano, L. (1997). Development of the Multiple Sequence Approximation within the Agadir Model of a-Helix Formation. Comparison with Zimm-Bragg and Lifson-Roig Formalisms. *Biopolymers* 41, 495-509.
- Ng, D. W. K., Abeyasinghe, J. K., & Kamali, M. (2018). Regulating the Regulators: The Control of Transcription Factors in Plant Defense Signaling. *International Journal of Molecular Sciences* 2018, Vol. 19, Page 3737, 19(12), 3737. <https://doi.org/10.3390/IJMS19123737>
- O'Connor, K., González-Suárez, P., & Dixon, L. E. (2020). Temperature Control of Plant Development. *Annual Plant Reviews Online*, 3(4), 563–606. <https://doi.org/10.1002/9781119312994.APR0745>
- Passardi, F., Dobias, J., Valério, L., Guimil, S., Penel, C., & Dunand, C. (2007). Morphological and physiological traits of three major Arabidopsis thaliana accessions. *Journal of Plant Physiology*, 164(8), 980–992. <https://doi.org/10.1016/J.JPLPH.2006.06.008>

- Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., & Dunker, A. K. (2005). Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins: Structure, Function, and Bioinformatics*, 61(S7), 176-182.
- Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K., & Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC bioinformatics*, 7, 1-17.
- Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., Dunker, A. K., & Obradovic, Z. (2005). Optimizing long intrinsic disorder predictors with protein evolutionary information. *Journal of bioinformatics and computational biology*, 3(01), 35-60.
- Pietrosemoli, N., García-Martín, J. A., Solano, R., & Pazos, F. (2013). Genome-Wide Analysis of Protein Disorder in *Arabidopsis thaliana*: Implications for Plant Environmental Adaptation. *PLOS ONE*, 8(2), e55524. <https://doi.org/10.1371/JOURNAL.PONE.0055524>
- Pruneda-Paz, J. L., Breton, G., Nagel, D. H., Kang, S. E., Bonaldi, K., Doherty, C. J., Ravelo, S., Galli, M., Ecker, J. R., & Kay, S. A. (2014). A Genome-Scale Resource for the Functional Characterization of *Arabidopsis* Transcription Factors. *Cell Reports*, 8(2), 622–632. <https://doi.org/10.1016/j.celrep.2014.06.033>
- Purohit, A., Ganguly, S., Chaudhuri, R. K., & Chakraborti, D. (2019). Understanding the Interaction of Molecular Factors During the Crosstalk Between Drought and Biotic Stresses in Plants. *Molecular Plant Abiotic Stress: Biology and Biotechnology*, 427–446. <https://doi.org/10.1002/9781119463665.CH21>
- Ramegowda, V., Gill, U. S., Sivalingam, P. N., Gupta, A., Gupta, C., Govind, G., Nataraja, K. N., Pereira, A., Udayakumar, M., Mysore, K. S., & Senthil-Kumar, M. (2017). GBF3 transcription factor imparts drought tolerance in *Arabidopsis thaliana*. *Scientific Reports 2017 7:1*, 7(1), 1–13. <https://doi.org/10.1038/s41598-017-09542-1>
- Rawn, D. (2010). *Biochemistry* (1989 N. Patterson Publishers (ed.)).
- Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Structure, Function, and Bioinformatics*, 42(1), 38-48.
- Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J. E., & Dunker, A. K. (1997, June). Identifying disordered regions in proteins from amino acid sequence. In *Proceedings of International Conference on Neural Networks (ICNN'97)* (Vol. 1, pp. 90-95). IEEE
- Salladini, E., Jørgensen, M. L. M., Theisen, F. F., & Skriver, K. (2020). Intrinsic Disorder in Plant Transcription Factor Systems: Functional Implications. *International Journal of Molecular Sciences 2020, Vol. 21, Page 9755*, 21(24), 9755. <https://doi.org/10.3390/IJMS21249755>
- Sato, H., Mizoi, J., Tanaka, H., Maruyama, K., Qin, F., Osakabe, Y., Morimoto, K., Ohori, T., Kusakabe, K., Nagata, M., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2015). *Arabidopsis* DPB3-1, a DREB2A Interactor, Specifically Enhances Heat Stress-Induced Gene Expression by Forming a Heat Stress-Specific Transcriptional Complex with NF-Y Subunits. *The Plant Cell*, 26(12), 4954–4973. <https://doi.org/10.1105/TPC.114.132928>
- Shammas, S. L. (2017). Mechanistic roles of protein disorder within transcription. *Current Opinion in Structural Biology*, 42, 155–161. <https://doi.org/10.1016/J.SBI.2017.02.003>
- Shin, R., Burch, A. Y., Huppert, K. A., Tiwari, S. B., Murphy, A. S., Guilfoyle, T. J., & Schachtman, D. P. (2007). The *Arabidopsis* Transcription Factor MYB77 Modulates Auxin Signal Transduction. *The Plant Cell*, 19(8), 2440–2453. <https://doi.org/10.1105/TPC.107.050963>
- Shiu, S. H., Shih, M. C., & Li, W. H. (2005). Transcription Factor Families Have Much Higher Expansion Rates in Plants than in Animals. *Plant Physiology*, 139(1), 18–26. <https://doi.org/10.1104/PP.105.065110>
- Siefers, N., Dang, K. K., Kumimoto, R. W., Bynum IV, W. E., Tayrose, G., & Holt, B. F. (2009). Tissue-Specific Expression Patterns of *Arabidopsis* NF-Y Transcription Factors Suggest Potential for Extensive Combinatorial Complexity. *Plant Physiology*, 149(2), 625–641. <https://doi.org/10.1104/PP.108.130591>
- Siggers, T., & Gordán, R. (2014). Protein–DNA binding: complexities and multi-protein codes. *Nucleic Acids Research*, 42(4), 2099–2111. <https://doi.org/10.1093/NAR/GKT1112>
- Smith, T., & Smith, R. L. (2014). *Elements of Ecology* (Pearson Education (ed.); 9th ed.).
- Sun, X., Rikkerink, E. H. A., Jones, W. T., & Uversky, V. N. (2013). Multifarious Roles of Intrinsic Disorder in Proteins Illustrate Its Broad Impact on Plant Biology. *The Plant Cell*, 25(1), 38–55.

- <https://doi.org/10.1105/TPC.112.106062>
- Sun, Y., Xu, W., Jia, Y., Wang, M., & Xia, G. (2015). The wheat TaGBF1 gene is involved in the blue-light response and salt tolerance. *The Plant Journal*, *84*(6), 1219–1230. <https://doi.org/10.1111/TPJ.13082>
- Suter, L., Rüegg, M., Zemp, N., Hennig, L., & Widmer, A. (2014). Gene Regulatory Variation Mediates Flowering Responses to Vernalization along an Altitudinal Gradient in Arabidopsis. *Plant Physiology*, *166*(4), 1928–1942. <https://doi.org/10.1104/PP.114.247346>
- Suzuki, N., Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., & Mittler, R. (2005). Enhanced Tolerance to Environmental Stress in Transgenic Plants Expressing the Transcriptional Coactivator Multiprotein Bridging Factor 1c. *Plant Physiology*, *139*(3), 1313–1322. <https://doi.org/10.1104/PP.105.070110>
- Tan, W., Zhang, D., Zhou, H., Zheng, T., Yin, Y., & Lin, H. (2018). Transcription factor HAT1 is a substrate of SnRK2.3 kinase and negatively regulates ABA synthesis and signaling in Arabidopsis responding to drought. *PLOS Genetics*, *14*(4), e1007336. <https://doi.org/10.1371/JOURNAL.PGEN.1007336>
- Uversky, V. N. (2019). Intrinsically disordered proteins and their “Mysterious” (meta)physics. *Frontiers in Physics*, *7*(FEB), 10. <https://doi.org/10.3389/FPHY.2019.00010/BIBTEX>
- Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., Cumming, M., Kelley, L. A., Sternberg, M. J., Krishnakumar, V., Ferlanti, E., Miller, J., Town, C., Stuerzlinger, W., & Provart, N. J. (2017). ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. *The Plant Cell*, *29*(8), 1806–1821. <https://doi.org/10.1105/TPC.17.00073>
- Wang, H., Wang, H., Shao, H., & Tang, X. (2016). Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology. *Frontiers in Plant Science*, *7*(FEB2016), 67. <https://doi.org/10.3389/FPLS.2016.00067/BIBTEX>
- Wang, N., Lefaudeux, D., Mazumder, A., Li, J. J., & Hoffmann, A. (2021). Identifying the combinatorial control of signal-dependent transcription factors. *PLOS Computational Biology*, *17*(6), e1009095. <https://doi.org/10.1371/JOURNAL.PCBI.1009095>
- Wenjing, W., Chen, Q., Singh, P. K., Huang, Y., & Pei, D. (2020). CRISPR/Cas9 edited HSFA6a and HSFA6b of Arabidopsis thaliana offers ABA and osmotic stress insensitivity by modulation of ROS homeostasis. *Plant Signaling and Behavior*, *15*(12). [https://doi.org/10.1080/15592324.2020.1816321/SUPPL\\_FILE/KPSB\\_A\\_1816321\\_SM1739.DOCX](https://doi.org/10.1080/15592324.2020.1816321/SUPPL_FILE/KPSB_A_1816321_SM1739.DOCX)
- Wiese, A. J., Steinbachová, L., Timofejeva, L., Čermák, V., Klodová, B., Ganji, R. S., Limones-Mendez, M., Bokvaj, P., Hafidh, S., Potěšil, D., & Honys, D. (2021). Arabidopsis bZIP18 and bZIP52 Accumulate in Nuclei Following Heat Stress where They Regulate the Expression of a Similar Set of Genes. *International Journal of Molecular Sciences 2021, Vol. 22, Page 530*, *22*(2), 530. <https://doi.org/10.3390/IJMS22020530>
- Wright, P. E., & Jane Dyson, H. (2015). *Intrinsically disordered proteins in cellular signalling and regulation*. <https://doi.org/10.1038/nrm3920>
- Wu, S., & Gallagher, K. L. (2012). Transcription factors on the move. *Current Opinion in Plant Biology*, *15*(6), 645–651. <https://doi.org/10.1016/J.PBI.2012.09.010>
- Yang, X., Yan, J., Zhang, Z., Lin, T., Xin, T., Wang, B., Wang, S., Zhao, J., Zhang, Z., Lucas, W. J., Li, G., & Huang, S. (2020). Regulation of plant architecture by a new histone acetyltransferase targeting gene bodies. *Nature Plants* *2020* *6*:7, *6*(7), 809–822. <https://doi.org/10.1038/s41477-020-0715-2>
- Yao, T., Zhang, J., Xie, M., Yuan, G., Tschaplinski, T. J., Muchero, W., & Chen, J. G. (2021). Transcriptional Regulation of Drought Response in Arabidopsis and Woody Plants. *Frontiers in Plant Science*, *11*, 2044. <https://doi.org/10.3389/FPLS.2020.572137/BIBTEX>
- Yavari, N., Tripathi, R., Wu, B. Sen, MacPherson, S., Singh, J., & Lefsrud, M. (2021). The effect of light quality on plant physiology, photosynthetic, and stress response in Arabidopsis thaliana leaves. *PLOS ONE*, *16*(3), e0247380. <https://doi.org/10.1371/JOURNAL.PONE.0247380>
- Zamora-Briseño, J. A., Pereira-Santana, A., Reyes-Hernández, S. J., Cerqueda-García, D., Castaño, E., & Rodríguez-Zapata, L. C. (2021). Towards an understanding of the role of intrinsic protein disorder on plant adaptation to environmental challenges. *Cell Stress and Chaperones*, *26*(1), 141–150. <https://doi.org/10.1007/s12192-020-01162-5>
- Zhao, H., Wu, D., Kong, F., Lin, K., Zhang, H., & Li, G. (2017). The Arabidopsis thaliana nuclear factor Y transcription factors. *Frontiers in Plant Science*, *7*, 2045. <https://doi.org/10.3389/FPLS.2016.02045/BIBTEX>
- Zúñiga-Mayo, V. M., Marsch-Martínez, N., & De Folter, S. (2012). JAIBA, a class-II HD-ZIP transcription factor involved in the regulation of meristematic activity, and important for correct gynoecium and fruit development in Arabidopsis. *The Plant Journal*, *71*(2), 314–326.

<https://doi.org/10.1111/J.1365-313X.2012.04990.X>