



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**FACULTAD DE QUÍMICA**

---

---

Estudio de la sensibilidad estructural al ambiente de  
regiones intrínsecamente desordenadas en factores de  
transcripción de *Arabidopsis*

**T E S I S**

QUE PARA OBTENER EL TÍTULO DE  
**QUÍMICO FARMACÉUTICO BIÓLOGO**

**P R E S E N T A**

**CÉSAR ANTONIO PONCE DIEGO**



**Ciudad Universitaria, CD. MX. 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.

**JURADO ASIGNADO:**

**PRESIDENTE:** Profesor: CRUZ GARCIA FELIPE  
**VOCAL:** Profesor: GONZALEZ SEGURA LILIAN  
**SECRETARIO:** Profesor: CUEVAS VELAZQUEZ CESAR LUIS  
**1er. SUPLENTE:** Profesor: DIMITROVA DINKOVA TZVETANKA  
**2° SUPLENTE:** Profesor: GUTIERREZ AGUILAR MANUEL

**SITIO DONDE SE DESARROLLÓ EL TEMA:**

FACULTAD DE QUÍMICA, UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**ASESOR DEL TEMA:**

**DR. CESAR LUIS CUEVAS VELAZQUEZ**



**SUSTENTANTE (S):**

**CESAR ANTONIO PONCE DIEGO**



## **AGRADECIMIENTOS INSTITUCIONALES**

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por el donativo de Ciencia de Frontera CONAHCYT CF-2029/252952 al Dr. César Luis Cuevas Velázquez. Agradezco al CONAHCYT por la beca recibida.

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM IA203422.

Este material está basado en trabajo realizado gracias al financiamiento de la University of California Institute for Mexico and the United States (UC MEXUS) y el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT). 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.



*Dos de mis pasatiempos favoritos son explorar diferentes lugares y jugar videojuegos. Esta imagen es una representación de mí dentro de un videojuego futurista en el Ángel de la independencia de la Ciudad de México. La imagen fue creada por una inteligencia artificial llamada Midjourney.*

*Para despertar tu potencial oculto, es necesario estar cómodo con la incomodidad.*

***-Adam Grant, Hidden Potential***

## AGRADECIMIENTOS PERSONALES

A mi padre Antonio Ponce y a mi madre, Martina Diego, que con su apoyo incondicional y sacrificios que hicieron posible la conclusión de mis estudio de licenciatura y mi desarrollo como profesional.

A mis hermanos, Edgar y Fernanda por su constante apoyo y compañía, junto a nuestra querida familia.

A mi tío Alejandro Ponce, agradezco su constante apoyo, esperanza y motivación los cuales fueron fundamentales para mi progreso.

A mi abuela, Lucha, por su inquebrantable atención y apoyo a lo largo de mi carrera académica.

A mis primas, Yarezi y Andrea, cuyo apoyo incondicional ha sido un pilar fundamental para seguir adelante.

A mis tíos Eduardo y Rosaura, por su respaldo constante y la confianza que siempre depositaron en mí.

A mis primos, Isaac y Karen, por su apoyo y confianza en mí, y por años que compartimos en nuestra infancia.

A mi tía Gaby, por su apoyo incondicional, motivación, confianza y esperanza en mi progreso profesional, y por nunca abandonarme en momentos difíciles. **Que en paz descanse.**

A mi abuelo Fructuoso, por sus sabios consejos y el ánimo constante para seguir luchando por nuestros sueños, sin importar las dificultades. **Que en paz descanse.**

A mis amigos de la carrera, agradezco sus valiosos consejos y la compañía a lo largo de este camino, así como la confianza que depositaron en mí.

A mis compañeros y compañeras de laboratorio, por los momentos compartidos, sus consejos y las enseñanzas que me brindaron en este recorrido académico.

Al Dr. Cesar Luis Cuevas Velázquez, mi tutor y guía en este proyecto, quien me transmitió valiosas cualidades científicas que han impulsado mi crecimiento profesional.

<b>RESUMEN</b> .....	<b>7</b>
<b>ABSTRACT</b> .....	<b>8</b>
<b>ABREVIATURAS</b> .....	<b>9</b>
<b>1. INTRODUCCIÓN</b> .....	<b>10</b>
1.1. Las regiones intrínsecamente desordenadas y el paradigma estructura función..	10
1.2. Niveles de desorden en los factores de transcripción de Arabidopsis .....	12
1.3. Propiedades fisicoquímicas de las proteínas y regiones intrínsecamente desordenadas.....	14
1.4. Técnicas utilizadas en el estudio de regiones intrínsecamente desordenadas .....	19
<b>2. PLANTEAMIENTO DEL PROBLEMA</b> .....	<b>30</b>
<b>3. HIPÓTESIS</b> .....	<b>31</b>
<b>4. OBJETIVO</b> .....	<b>31</b>
<b>4.1. OBJETIVOS PARTICULARES.</b> .....	<b>31</b>
<b>5. MATERIALES Y MÉTODOS</b> .....	<b>32</b>
5.1. Obtención de plásmidos, transformación y genotipificación.....	<b>32</b>
5.2. Construcción de bases de datos y caracterización bioinformática de las regiones intrínsecamente desordenadas .....	33
5.3. Generación de las construcciones de la biblioteca de regiones desordenadas. ....	34
5.4. Preparación de inóculos de levadura.....	38
5.5. Determinación de la sensibilidad estructural mediante ensayos FRET .....	38
5.6. Validación y cuantificación de los ensayos FRET.....	38
<b>6. RESULTADOS</b> .....	<b>40</b>
6.1. Los factores de transcripción poseen mayores niveles de desorden que el resto de las proteínas del proteoma de Arabidopsis.....	40
6.2. Caracterización fisicoquímica de las regiones intrínsecamente desordenadas más largas de los factores de transcripción .....	42
6.3. Determinación de los niveles de dualidad de la IDR más largas de los factores de transcripción de Arabidopsis.....	46
6.4. Determinación de los cambios conformacionales de las IDRs de los factores de transcripción de Arabidopsis mediante ensayos FRET .....	53
<b>7. DISCUSIÓN</b> .....	<b>64</b>
<b>8. CONCLUSIONES</b> .....	<b>68</b>
<b>8.1. PERSPECTIVAS</b> .....	<b>69</b>
<b>9. BIBLIOGRAFÍA</b> .....	<b>70</b>

## RESUMEN

Las propiedades fisicoquímicas del ambiente intracelular se ven afectadas por cambios en las composición de macromoléculas. Entre las macromoléculas, las proteínas poseen una función basada en su estructura; sin embargo, algunos dominios proteicos pueden carecer parcialmente de una estructura secundaria o tridimensional definida, siendo nombrados regiones intrínsecamente desordenadas (IDR, *Intrinsically Disordered Regions*). Las IDRs destacan por su composición de aminoácidos que les permiten percibir cambios en el ambiente, implicando un posible cambio estructural. Los factores de transcripción (FTs) de *Arabidopsis thaliana* (Arabidopsis) son ejemplos de proteínas con altos niveles de desorden estructural. Sin embargo, aún se desconocen los niveles de sensibilidad estructural de las IDRs de los FT de Arabidopsis ante cambios en el ambiente fisicoquímico. En este proyecto, determinamos los niveles de sensibilidad estructural *in vivo* de 21 IDRs de FT de Arabidopsis. Los FTs fueron seleccionados del proteoma de Arabidopsis, el cual contenía 27,411 proteínas, de las cuales seleccionamos 1,958 FTs. Con base en una caracterización *in silico* basada en la predicción de formación de hélice- $\alpha$ , el nivel de desorden y un parámetro que mide el posible cambio de hélice- $\alpha$ /desorden, denominado dualidad, seleccionamos 69 FTs través del uso de funciones de probabilidad con una significancia estadística de 0.99. De este análisis, seleccionamos 21 FTs que estaban incluidos en una biblioteca de biosensores desarrollada previamente en el grupo de investigación, la cual consiste en IDRs transformadas en levaduras. Las levaduras transformadas fueron sometidas a choques hiperosmóticos usando cloruro de sodio (NaCl), con el objetivo de evaluar el posible cambio estructural de las IDRs. Utilizamos la técnica de transferencia de energía por resonancia de Forster (FRET, *Forster Resonance Energy Transfer*) para analizar los cambios conformacionales. Los resultados del análisis conformacional revelaron que la distribución homogénea de cargas positivas y negativas a lo largo de la secuencia de las IDRs de los FT podría ser un predictor de la alta respuesta ante cambios en el ambiente. En contraste, los factores de baja respuesta mostraron distribuciones de cargas menos homogéneas, caracterizadas por contener residuos con grupos con cargas positivas. Los resultados obtenidos proporcionan un punto de partida para el análisis de otros parámetros dentro de las secuencias de las IDRs de los FTs para caracterizar las propiedades que determinan la sensibilidad estructural *in vivo* de las IDR de los FTs ante cambios ambientales.



## ABSTRACT

The physicochemical properties of the intracellular environment are affected by changes in the composition of macromolecules. Within macromolecules, proteins have a function based on their structure, however, some protein domains partially lack a well-defined secondary or tridimensional structure, being named Intrinsically Disordered Regions (IDRs). IDRs have a specific composition of amino acids that allows them to perceive changes in the environment, implying a possible structural change. The *Arabidopsis thaliana* (Arabidopsis) transcription factors (TF) are examples of proteins with high levels of structural disorder. However, little is known about the structural sensitivity of the IDRs in TFs of Arabidopsis in response to changes in the physicochemical environment. In this work, we determined the *in vivo* structural sensitivity of 21 IDRs of TFs from Arabidopsis. The transcription factors were selected from the proteome of Arabidopsis, which had 27,411 proteins, and from which we selected 1,958 TFs. Based on an *in silico* characterization of the percentage of formation of  $\alpha$ -helix, disorder level, and a parameter that measures the possible change of  $\alpha$ -helix/disorder, called duality, we obtained 69 TFs using probabilistic functions with a statistical significance of 0.99. From this analysis, 21 TFs we selected from an IDR-based library of biosensors previously created in the research group, which consist of IDRs transformed in yeasts. Transformed yeasts were subjected to osmotic shocks using sodium chloride (NaCl), with the purpose of determining the possible structural change of the IDRs. We use the Forster Resonance Energy Transfer technique, called FRET, to analyze conformational changes. The conformational analysis results showed that the homogeneous distribution of positive and negative charges along the IDR sequence of the transcription factors could be a predictor of a high response against changes in the environment. In contrast, the low response transcription factors showed less homogeneity in their charge distribution, characterized by groups of positive and negative charges in a single side of the sequence. The results obtained provide a starting point for the analysis of other parameters inside the IDR sequences of TFs to characterize the properties that confer the structural sensitivity *in vivo* of the IDRs of TFs against environmental changes.

## ABREVIATURAS

FRET	Transferencia de energía por resonancia de fluorescencia ( <i>Fluorescence Resonance Energy Transfer</i> )
IDRs	Regiones Intrínsecamente Desordenadas ( <i>Intrinsically Disordered Regions</i> )
FTs	Factores de transcripción ( <i>Transcription Factors</i> )
CIDER	Classification of Intrinsically Disordered Ensemble Regions
PONDR	Predictor of Natural Disordered Regions
NMR	Resonancia Magnética Nuclear ( <i>Nuclear Magnetic Resonance</i> )
SAXS	Dispersión de Rayos X por Ángulo Pequeño ( <i>Small-angle X-ray scattering</i> )
FCR	Fracción de residuos cargados ( <i>Fraction of Charged Residues</i> )
NCPR	Carga neta por residuo ( <i>Net Charge Per Residue</i> )

## 1. INTRODUCCIÓN

### 1.1. Las regiones intrínsecamente desordenadas y el paradigma estructura función

A lo largo de las décadas, se han descubierto proteínas que carecen parcial o totalmente de una estructura secundaria o tridimensional estable y bien definida. Su relevancia comenzó a destacarse poco después de completar la secuenciación del genoma, cuando se observó que dichas proteínas son bastante comunes en organismos eucariotas (Wright, 2015). Estas proteínas se denominan proteínas intrínsecamente desordenadas (IDPs, *Intrinsically Disordered Proteins*) cuando se trata de una proteína completa. Sin embargo, las proteínas estructuradas pueden contener dominios desordenados, conocidos como regiones intrínsecamente desordenadas (IDRs, *Intrinsically Disordered Regions*) (Wright, 2015). La proporción de regiones desordenadas aumenta con la complejidad de los organismos, siendo similar en arqueas y bacterias, pero significativamente mayor en eucariotas (Uversky, 2019).

Las IDPs presentan diversas características que les otorgan un control significativo sobre los procesos celulares. Entre estas características, su alto grado de flexibilidad les permite interactuar con múltiples blancos moleculares (Wright, 2015). Los dominios desordenados de las proteínas tienen sensibilidad estructural a diversos osmolitos, lo que proporciona información valiosa acerca de la regulación de las IDRs en sistemas biológicos. Esto incluye su papel en la señalización celular, la regulación transcripcional y en el mantenimiento de la homeostasis (Moses *et al.*, 2020).

Dado el carácter altamente dinámico de las IDPs/IDRs, es más apropiado considerarlas como un ensamble de diferentes configuraciones estructurales. El ambiente fisicoquímico puede aumentar la probabilidad de que ciertas IDPs/IDRs adopten estructuras secundarias, lo que se ha asociado con funciones específicas de estas proteínas (Arai, 2015). Debido a su constante variabilidad estructural, las IDPs/IDRs responden a las modificaciones en su entorno. En general, estas regiones exponen residuos de aminoácidos a la solución que la rodea, lo que conlleva a que diferentes IDRs de distintas proteínas presenten diversos grados de flexibilidad

estructural, provocando las IDPs/IDRs respondan de manera única a la presencia de solutos (Moses *et al.*, 2020).

Cuando se habla de sensibilidad estructural de las IDRs, se refiere a la idea de que cambios en el entorno que rodea a estas regiones pueden influir en su actividad. Esto es especialmente relevante, ya que, en muchas proteínas, la función está intrínsecamente ligada a su cambio estructural (Moses *et al.*, 2020).

El tradicional paradigma estructura-función es aplicable para gran variedad de proteínas, como enzimas, proteínas transmembranales y dominios de señalización. Sin embargo, existe una diversidad considerable de proteínas no estructuradas que desempeñan funciones cruciales en las células y donde este paradigma ya no es aplicable. Las IDRs y IDPs logran sus funciones a través de interacciones cooperativas, tanto intra como intermolecularmente, con proteínas estructuradas y, en ocasiones, con otras IDRs/IDPs (Bondos, 2021).

Uno de los aspectos fascinantes de las IDPs/IDRs, es su estructura desordenada en condiciones fisiológicas. Este comportamiento está relacionado con la composición de aminoácidos de las IDPs/IDRs, la cual difiere de la que se encuentra en las proteínas ordenadas. En las proteínas desordenadas, se observa una mayor cantidad de aminoácidos con cargas positivas y negativas, lo que provoca un desequilibrio de cargas. Además, tienen una carga neta mayoritariamente negativa a pH de 7.0 y una mayor prevalencia de aminoácidos hidrofóbicos (Bondos, 2021; Uversky, 2000).

Los análisis realizados en secuencias desordenadas permiten predecir el grado de desorden existente en los proteomas de diversas especies. Se observa que el nivel de desorden en los proteomas de eucariotas es más alto en comparación con el de procariontas. En los proteomas de arqueas y eubacterias, se estima que alrededor del 15% al 30% de sus residuos de aminoácidos se encuentran en IDPs e IDRs, mientras que en los proteomas de eucariotas se predice que un rango de entre el 30% y el 50% de sus residuos pertenece a IDPs e IDRs. Curiosamente, los virus exhiben una mayor variación en cuanto al grado de desorden en sus genomas (Xue, 2012; Bondos, 2021).

Gracias a las predicciones y los esfuerzos dedicados al análisis de secuencias desordenadas en diferentes organismos, se ha logrado caracterizar la funcionalidad de las IDPs/IDRs en diversos procesos biológicos. Entre ellos, se destacan procesos como la diferenciación, la transcripción, la regulación de la transcripción, la espermatogénesis, la condensación del DNA, el ciclo celular, el procesamiento de RNA, el proceso de splicing del RNA, la mitosis, la apoptosis, entre otros. La gran mayoría de los procesos en los que participan las IDPs/IDRs se relacionan con sistemas de señalización y regulación (Bondos, 2021; Xie, 2007).

## **1.2. Niveles de desorden en los factores de transcripción de Arabidopsis**

Dentro de los sistemas de regulación, se encuentran los factores de transcripción (FT), que son una de las clases más extensas de proteínas funcionales codificadas en el genoma de los organismos eucariotas (Pruneda, 2014). Estos FTs desempeñan un papel crucial en la regulación de la expresión génica al interactuar con otros factores de transcripción que se unen a regiones reguladoras del DNA (Pruneda, 2014). Las células llevan a cabo sus procesos biológicos a través de interacciones moleculares dinámicas con diversas proteínas involucradas en las vías de señalización y regulación de procesos biológicos. Los FTs son especialmente relevantes en estos procesos, ya que contribuyen a la inducción o represión de genes al unirse a regiones promotoras (Salladini, 2020).

Los FTs de los organismos eucariotas poseen dominios que se consideran como IDRs, lo que les confiere cierto grado de flexibilidad necesario para interactuar con otras proteínas a través de estas regiones. Esto mejora la eficiencia de los procesos de regulación celular. Muchas de estas interacciones entre IDRs y proteínas son mediadas por SliM (*short linear sequence motifs*, por sus siglas en inglés), que sirven como motivos regulatorios mediadores de la interacción de una gran variedad de proteínas. Además, los denominados MoRFs (*molecular recognition fragments*, por sus siglas en inglés) también desempeñan un papel importante al facilitar la unión y el plegamiento de estas regiones, ya que múltiples SliMs y los MoRFs han sido clasificados como IDRs (Basu et al., 2023). En consecuencia, las IDRs presentes en los factores de transcripción poseen un alto nivel de flexibilidad y dinamismo, lo que les permite responder a condiciones de estrés y eventos de señalización de manera efectiva (Salladini, 2020).

Los FTs de *A. thaliana* han sido objeto de extensos estudios, lo que ha permitido dilucidar su participación en una amplia variedad de procesos celulares. Parte de su funcionalidad radica en su capacidad de unión al DNA, la cual está mediada por dominios específicos, como los dominios bZIP o HD-ZIP (Hrmova, 2021). En particular el dominio bZIP presenta una región básica que facilita la unión al DNA, así como repeticiones de leucina que promueven la dimerización de este dominio (Hrmova, 2021; Agarwal, 2019).

Dado que los FTs desempeñan un papel importante en el desarrollo de las plantas, se ha investigado su implicación en la regulación de respuestas frente a diversas condiciones de estrés abiótico (Hrmova, 2021). Entre los FTS caracterizados que contribuyen a la mediación de respuestas de estrés se encuentran bZIP (basic-leucine zipper), HD-ZIP (Homeodomain-Leucine Zipper), MYB (Myeloblastoma), AP2 (Apetala2)/ERF (Ethylene Responsive Factor), DREB (Drought-Response Elements Binding)/CBF (C-Repeat Binding Factor), SHN (Shine) y WXLP (Wax Production-Like) (Hrmova, 2021). Algunos FTs que poseen dominios bZIP contribuyen a la generación de tolerancia de las plantas frente a diversos tipos de estrés, incluido el salino (Agarwal, 2019).

Para desempeñar su papel en la regulación de la expresión génica, algunos de estos FTs adquieren una conformación estructural secundaria estable, como ocurre con los factores de la familia bZIP (Hrmova, 2021; Sornaraj, 2016; Das, 2012). Estos factores bZIP presentan un motivo básico conservado y una estructura de cremallera de leucina (Leucine Zipper) menos conservada (Hrmova, 2021). Por otro lado, los FTs de la familia HD-ZIP desempeñan un papel crucial en la respuesta a condiciones de estrés, como temperaturas extremas, ambientes hiperosmóticos, desecación, y biosíntesis de ácido abscísico (Li *et al.*, 2022; Hrmova *et al.*, 2021). Estos factores de transcripción HD-ZIP, al igual que los bZIP, cuentan con un dominio altamente conservado de unión a DNA y un motivo de leucina que facilita la homo- y hetero dimerización (Tang *et al.*, 2019).

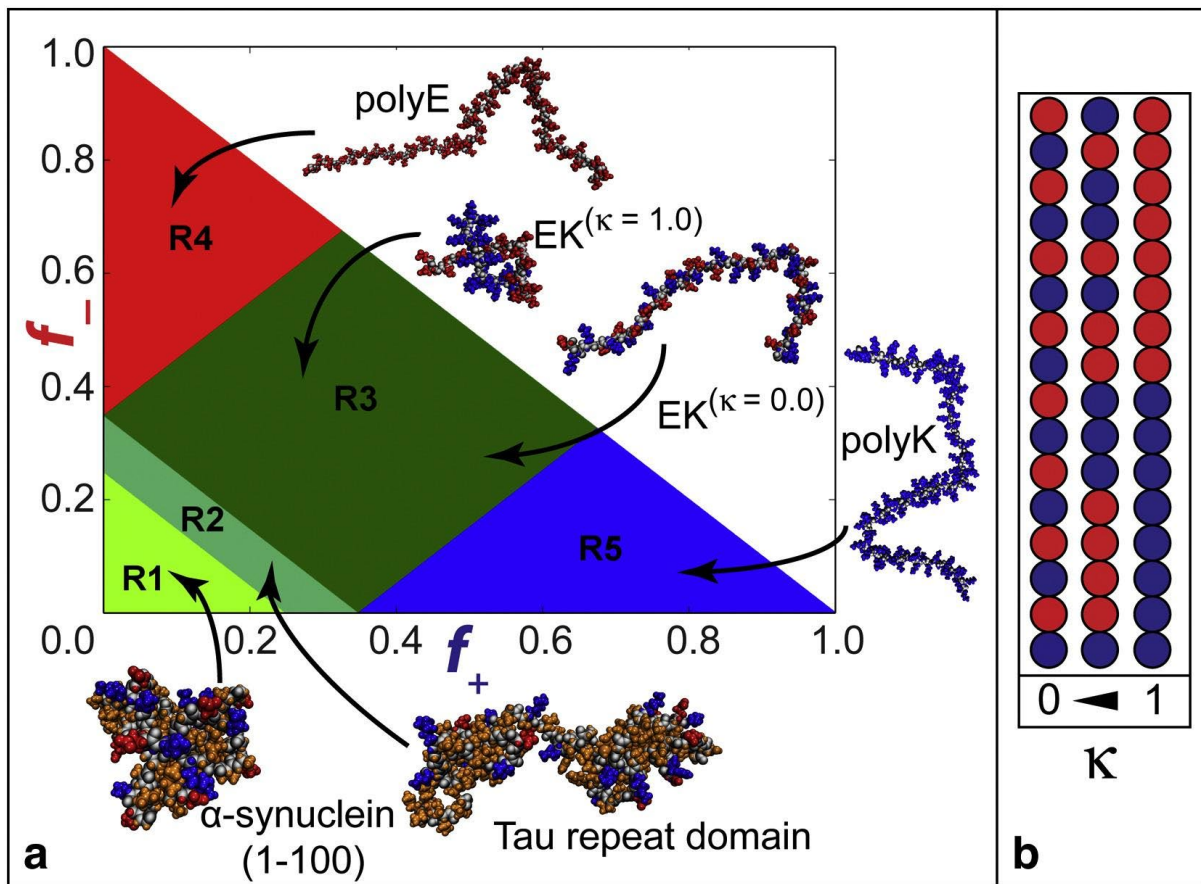
### 1.3. Propiedades fisicoquímicas de las proteínas y regiones intrínsecamente desordenadas

Las IDRs, debido a su falta de una estructura secundaria estable, plantean desafíos significativos para su estudio con las técnicas tradicionales de biología estructural (Metskas & Rhoades, 2020). Durante décadas, el estudio de la relación estructura y función de las proteínas se benefició enormemente de la resolución atómica de las estructuras proteicas. Inicialmente, la cristalografía de rayos X desempeñó un papel fundamental al proporcionar información detallada sobre la arquitectura tridimensional de proteínas, ácidos nucleicos e incluso virus, contribuyendo a la identificación de blancos terapéuticos (Zheng *et al.*, 2015).

Las proteínas desordenadas al carecer de una estructura secundaria o terciaria estable que facilite las interacciones existen en una colección altamente dinámica de conformaciones que responden a las condiciones del entorno. La capacidad de cambiar de conformación, como la transición de estados plegados y desplegados, es una característica destacada de estas proteínas. Además, la conformación de estas proteínas desordenadas está estrechamente vinculada a propiedades fisicoquímicas específicas en sus secuencias, como la distribución de cargas. Las proteínas desordenadas con una alta carga neta tienden a exhibir una expansión estructural más pronunciada debido a la repulsión de cargas (Muller *et al.*, 2010).

Me parece que esta información ya fue descrita en párrafos anteriores

La influencia de las cargas en la conformación de las IDPs/IDRs ha llevado al desarrollo de un diagrama que permite la clasificación general de las proteínas según su fracción de cargas positivas y negativas (**Figura 1**). Esto ofrece un marco para comprender cómo la carga neta por residuo en una proteína desordenada modula su conformación (Mao *et al.*, 2010; Holehouse *et al.*, 2017). Asimismo, la presencia de clústeres, al igual que la presencia de clústeres o grupos de carga, ha sido observada al diseñar IDRs sintéticas (Tedeschi *et al.*, 2018).



**Figura 1.** a) Diagrama de estados mostrando la conformación de algunas IDPs representativas de acuerdo con su posición en el diagrama. Cada secuencia posee residuos coloreados de azul (carga positiva) y rojo (carga negativa), los cuales representan la distribución de sus cargas y cómo se asocia este parámetro a su conformación. El diagrama representa la fracción de cargas positivas ( $f_+$ ) y negativas ( $f_-$ ) de las secuencias desordenadas, lo cual permite clasificarlas en cinco regiones de acuerdo con su conformación b) Esquema representativo de la distribución de cargas a lo largo de una secuencia desordenada cuantificada a través del parámetro de kappa ( $\kappa$ ); valores cercanos a 0 indican una distribución de cargas homogénea a lo largo de la secuencia, mientras que valores cercanos a 1 representan una distribución sesgada hacia los extremos N-terminal, C-terminal o ambos. Tomado de Holehouse *et al.*, 2017.

El estudio de la influencia de las cargas en la compactación de las IDPs/IDRs ha demostrado que existen otros parámetros fisicoquímicos que afectan su conformación. La baja hidropatía, en combinación con una alta carga neta, que usualmente es negativa, permite a las IDPs/IDRs adoptar una conformación extendida (Uversky, 2019). En algunas IDPs/IDRs, la carga neta promedio y la hidropatía



promedio son suficientes para distinguir entre una IDR extendida y una compactada (Uversky, 2019).

Las propiedades específicas de las IDPs e IDR incluyen su simplicidad en la secuencia de aminoácidos y la presencia de residuos promotores del desorden, como Arg, Pro, Gln, Gly, Glu, Ser, Ala y Lys. Algunos de estos residuos se encuentran en repeticiones en tándem, lo que influye en la estructura de la proteína (Jorda *et al.*, 2010). Además, secuencias de polialanina, que suelen ser desordenadas, se han relacionado con enfermedades hereditarias (Darling & Uversky, 2017).

Dada las propiedades únicas de las IDPs e IDRs, se han desarrollado parámetros de predicción para definir sus características conformacionales. Softwares como CIDER y localCIDER proporcionan información biofísica de las IDPs/IDRs, y se basan en análisis biofísicos realizados a lo largo de la décadas (Holehouse *et al.*, 2017). Estos estudios han identificado parámetros fisicoquímicos como hidropatía, NCPR (Net Charge Per Residue), FCR (Fraction of Charged Residues) y kappa, que ayudan a caracterizar las IDPs/IDRs y a comprender la relación entre secuencia, estructura y función (Holehouse *et al.*, 2017).

La hidropatía mide la hidrofobicidad de la secuencia y, en general, las IDPs se caracterizan por tener una baja hidropatía y carecer de residuos hidrofóbicos voluminosos como Ile, Leu y Val (Jorda *et al.*, 2010). El NCPR proporciona información sobre la compacidad de la cadena polipeptídica y contribuye a determinar las dimensiones globales de las IDPs/IDR (Uversky, 2019). Un NCPR por debajo de 0.25 se asocia con una conformación globular compacta, mientras que valores superiores a 0.25 indican una conformación extendida (Holehouse *et al.*, 2017).

Estos parámetros no son los únicos factores determinantes de la conformación de las IDPs/IDRs. La fracción de residuos cargados (FCR) y el patrón de cargas (kappa) también influye en el tamaño, la forma y la dinámica de estas proteínas clasificadas como polianfolitos (Holehouse *et al.*, 2017). La segregación de cargas opuestas (**Figura 2**) a lo largo de las secuencias desordenadas contribuye a las fluctuaciones conformacionales (Das & Pappu, 2013).

Etiqueta	Secuencia	$\kappa$
sv1	EKE	0.0009
sv2	EEEEKKKKEEEKKKEEEKKKKEEEKKKKEEEKKKKEEEKKKKEEKKKE	0.0025
sv3	KEKKKEKKEEKKEKEKEKEKEKEKEKEKEKEKEKKEKKEKEKEEKKKEE	0.0139
sv4	KEKEKKEEKEKKEEKEKEKEKEKKEKEKKEKKEKKEKKEKKEEKEEKEE	0.0140
sv5	KEKEEKEKKEEEEEEKKEKKEEKEKEKEKEEKEKKEEKKKKEKKEKEKE	0.0245
sv6	EEEEKKEKKEEKEKKEKKEKEEKEKKEEKEKKEEKEKKEEKEEKKKEE	0.0273
sv7	EEEEKKKKKEEEKKKKKEEEKKKKKEEEKKKKKEEEKKKKKEEEKKKKKE	0.0450
sv8	KKKKEEEEEKKKKEEEKKKKEEEKKKKEEEKKKKEEEKKKKEEEKKKKEE	0.0450
sv9	EKKKEEKEKEKEEKEEKEEKEKKEEKEKKEEKEEKEKKEEKEEKEEKEE	0.0624
sv10	EKKKKKKEEKKKEEEEEEKKKEEKKKEEKEEKEEKEEKEEKEEKEEKEE	0.0834
sv11	EKEKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE	0.0841
sv12	EKKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.0864
sv13	KEKKKEKEKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE	0.0951
sv14	EKKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.1311
sv15	KKEKKEKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE	0.1354
sv16	EKEEKEEKKKEKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.1458
sv17	EKEKKEEKKKEKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.1643
sv18	KEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.1677
sv19	EEEEKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE	0.1941
sv20	EEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE	0.2721
sv21	EEEEEEEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE	0.2737
sv22	KEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.3218
sv23	EEEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.3545
sv24	EEEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.4456
sv25	EEEEEEEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE	0.5283
sv26	KEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.6101
sv27	KKKEKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE	0.6729
sv28	EKKKKKKKKKKKKKKKKKKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.7666
sv29	KEEEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKE	0.8764
sv30	EEEEEEEEEEEEEEEEEEEEEEEEKKKKKKKKKKKKKKKKKKKKKKKKKKKK	1.0000

**Figura 2.** Esquema representativo del parámetro de distribución de cargas ( $\kappa$ ) de 30 secuencias diferentes. Los valores de  $\kappa$  se muestran a la derecha. Un valor de  $\kappa$  de 1.0 indica una segregación de cargas a cada lado opuesto de la secuencia, mientras que valores bajos de  $\kappa$  indican una segregación a lo largo de la secuencia. Modificado de Das & Pappu *et al.*, 2013.

Las IDPs/IDRs se han clasificado en cinco grupos distintos según su fracción de cargas negativas y positivas. Esta clasificación se basa en el parámetro FCR que se calcula sumando las fracciones de cargas negativas y positivas presentes en una IDP/IDR (Holehouse *et al.*, 2017; Das & Pappu, 2013). La influencia de los residuos cargados como D (ácido aspártico) E (ácido glutámico), K (lisina) y R (arginina), en el parámetro FCR es específica para un valor de pH dado (Holehouse *et al.*, 2017; Lermyte, 2020).

La relación entre las cargas en las propiedades de las IDPs/IDRs se ha estudiado a través de la correlación con el parámetro  $\kappa$ . El valor de  $\kappa$  se calcula comparando la carga local de seis residuos de aminoácidos con la distribución

de cargas de toda la secuencia (Holehouse *et al.*, 2017). Esto indica que, cuando las cargas positivas y negativas están homogéneamente mezcladas en la secuencia, kappa tiende a ser cercano a 0.0, lo que sugiere una mayor probabilidad de expansión de la cadena de aminoácidos. Por otro lado, cuando las cargas se segregan de forma opuesta, los valores de kappa tienden a 1.0, lo que indica una mayor probabilidad de que la secuencia se compacte debido a un mayor rango de interacción entre los residuos (Sherry *et al.*, 2017).

Es esencial destacar que el valor de kappa debe correlacionarse con el valor de FCR, ya que kappa describe la distribución de las cargas en la secuencia, mientras que FCR indica la densidad de residuos cargados en la misma secuencia. Esto proporciona una idea más completa del efecto de kappa en la conformación de una IDP/IDR (Das *et al.*, 2015).

El análisis bioinformático que considera las propiedades biofísicas de los aminoácidos ha revelado que las secuencias clasificadas como polianfolitos (R3) poseen mayor cantidad de bloques de carga. Además, se ha demostrado que la distribución de cargas, dada por el valor de kappa, es un componente importante en la determinación del ensamble de una IDP/IDR, aunque no es el único factor que influye en su conformación (Holehouse *et al.*, 2020).

La distribución de FCR y kappa varía de un organismo a otro, y estos dos parámetros están relacionados entre sí. Por lo general, la mayoría de los organismos tienen un rango de FCR que oscila entre 0.2 y 0.4, lo que coincide con las regiones R2 y algunas de las regiones R3 del diagrama de estados. Estas regiones son donde el valor de kappa tiene una mayor influencia en las propiedades conformacionales de las proteínas desordenadas. Además, la mayoría de las IDPs/IDRs se encuentran en la región R2, que generalmente es donde se localizan la mayoría de las proteínas desordenadas. Esto sugiere que el patrón de cargas podría desempeñar un papel crítico en el ensamble conformacional de las IDPs/IDRs (Holehouse *et al.*, 2020).

A pesar de la importancia del valor de kappa en la conformación de las IDPs/IDRs, no es el único factor que influye. Se han establecido rangos de FCR en los que la distribución de cargas influye significativamente. En un rango de  $FCR \geq 0.2$  y  $FCR \leq 0.4$ , las secuencias de IDPs/IDRs tienden a tener secuencias con valores

bajos de kappa lo que indica una distribución de cargas mezcladas homogéneamente. Cuando el FCR disminuye, en el rango de  $FCR \geq 0$  y  $FCR \leq 0.2$ , la influencia del patrón de cargas sobre la conformación de las IDPs/IDRs disminuye, y las secuencias experimentan una menor presión selectiva. Otros, como la presencia de motivos conservados, la composición de aminoácidos o la estructura secundaria residual también pueden influir en la conformación de las IDPs/IDRs (Holehouse *et al.*, 2020).

Además, se ha observado que existe una falta de polielectrolitos en los proteomas de organismos en las IDPs/IDRs. Al comparar parámetros como NCPR y FCR, nota que, a medida que FCR aumenta, los valores de NCPR también aumentan de manera negativa, lo que sugiere que la carga neta por residuo tiende a ser ácida en estas secuencias (Holehouse *et al.*, 2020). Sin embargo, en secuencias de diferentes organismos, se ha observado una ausencia de secuencias con una carga neta positiva. Los análisis biofísicos y bioinformáticos utilizan estos parámetros para predecir cambios conformacionales en las IDPs/IDRs. Aunque estos parámetros son importantes, puede haber otros determinantes de la conformación, como la presencia de IDPs ricas en prolina con valores específicos de NCPR, ya que pueden ser los determinantes en los niveles de compactación de las IDPs (Holehouse *et al.*, 2020; Kuravi *et al.*, 2010), o la influencia del efecto hidrofóbico en la compactación de la proteína (Holehouse *et al.*, 2020).

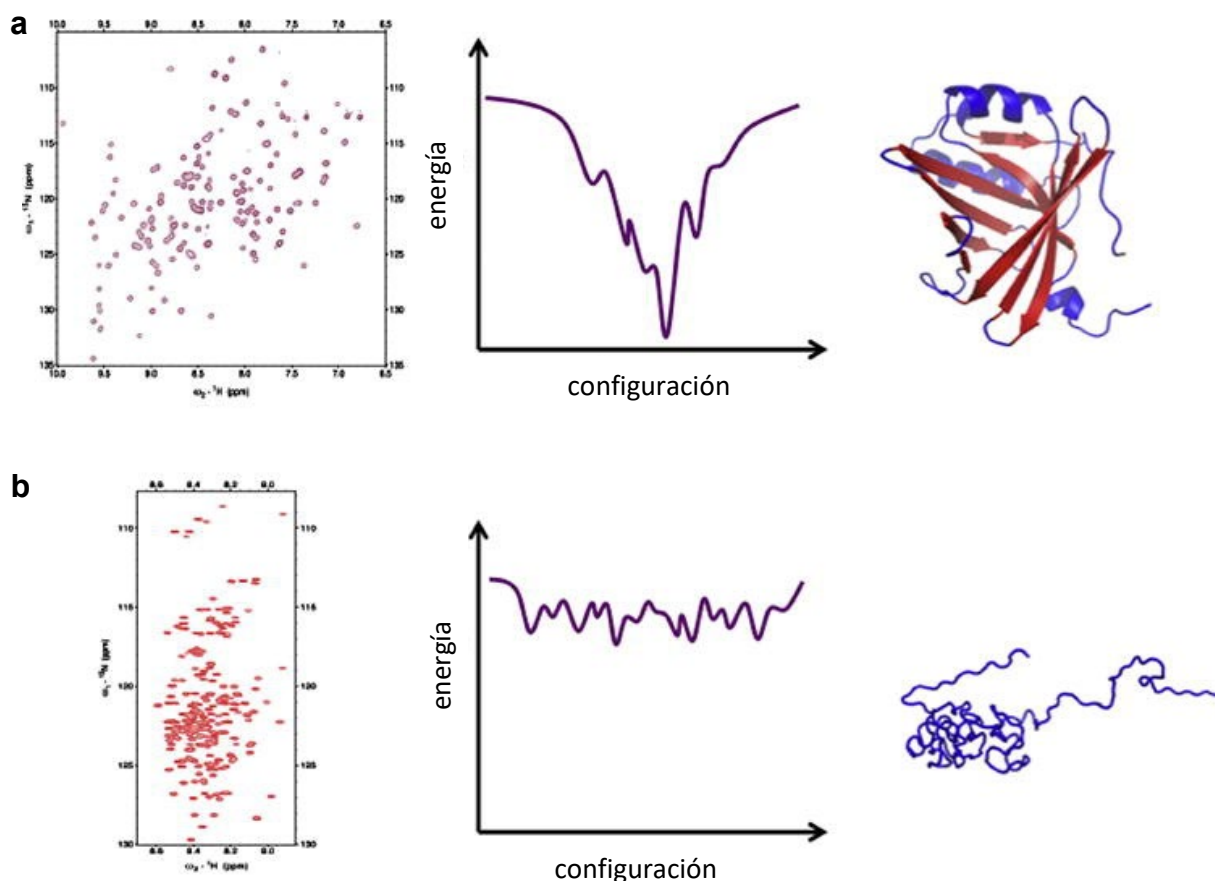
#### **1.4. Técnicas utilizadas en el estudio de regiones intrínsecamente desordenadas**

Se sabe que las proteínas intrínsecamente desordenadas juegan un papel importante durante el desempeño de las funciones celulares. La falta de una estructura secundaria estable y su capacidad de cambiar de conformación ante cambios en el ambiente químico que las rodea limita su estudio con las técnicas de la biología estructural (Metskas & Rhoades, 2020). Durante las últimas décadas, el estudio de la relación entre estructura y función de las proteínas ha sido gracias a la posibilidad de visualizar a resolución atómica, las estructuras de las proteínas, comenzando con el uso de la cristalografía de rayos X, que ha contribuido a la resolución de gran variedad de estructuras tridimensionales de proteínas, ácidos nucleicos e incluso virus, lo cual ha contribuido a la búsqueda de blancos terapéuticos (Zheng *et al.*, 2015). Sin embargo, las investigaciones han encontrado que cerca del

30% - 50% del proteoma de eucariotas contiene IDR (van der Lee, *et al.*, 2014; Moses *et al.*, 2020; Bondos *et al.*, 2021), y muchas de las regiones desordenadas contribuyen a la función de los procesos celulares. Por ejemplo, péptidos de 4 - 12 residuos denominados *short linear motifs*, clusters de motivos desordenados dentro de una secuencia o las propiedades electrostáticas e hidrofóbicas de las IDPs/IDRs para permitir una localización celular específica (Bauer *et al.*, 2015), así como algunas proteínas que sufren separación de fases líquido-líquido y que contienen repeticiones de R/G para facilitar los procesos de formación de condensados biomoleculares (Chong *et al.*, 2018), o incluso la presencia de cargas alternadas en una región desordenada de una proteína que regula el ciclo celular (Das *et al.*, 2016). Todas estas características de las IDPs/IDRs involucradas en procesos celulares son solo algunas de las propiedades que han sido descubiertas en las IDPs/IDRs; sin embargo, aún existen algunas por ser descubiertas y caracterizadas. Ante ello, se han desarrollado técnicas dirigidas para estudiar la dinámica de las IDPs/IDRs, la cual es incompatible con técnicas como la cristalografía de rayos X.

Las IDPs/IDRs pueden ser caracterizadas mediante métodos directos o indirectos. Los métodos directos proporcionan información estructural de la IDP/IDR, caso contrario a los métodos indirectos, que dependen de propiedades específicas de las IDPs, así como de su composición de aminoácidos (Ferreon *et al.*, 2022). Una de las técnicas que se ha utilizado desde hace algunas décadas para caracterizar a las IDPs ha sido la resonancia magnética nuclear (RMN), siendo una técnica versátil que permite estudiar moléculas dinámicas como lo son las IDPs/IDRs, ya que la técnica de NMR permite detectar el ensamble conformacional de las IDPs/IDRs (Tompa, 2002; Adamski *et al.*, 2019). Lo que la técnica de NMR permite es alinear el spin nuclear de los átomos para lograr caracterizar a la IDP/IDR. Los núcleos atómicos alcanzan un nivel de alta energía posterior a la radiación, y obtener al final otro estado de baja energía que permite determinar mediante una transformada de Fourier el espectro de NMR de los diferentes núcleos atómicos, así como el arreglo espacial de los átomos. Para el caso de las IDPs/IDRs o una proteína en general, cada residuo experimenta un campo magnético distinto, por lo que esta propiedad puede emplearse para determinar la estructura secundaria de una proteína; el espectro final obtenido mediante NMR para una IDP/IDR es la suma de los espectros obtenidos del ensamble conformacional de la IDP/IDR de cada residuo de la proteína, dando

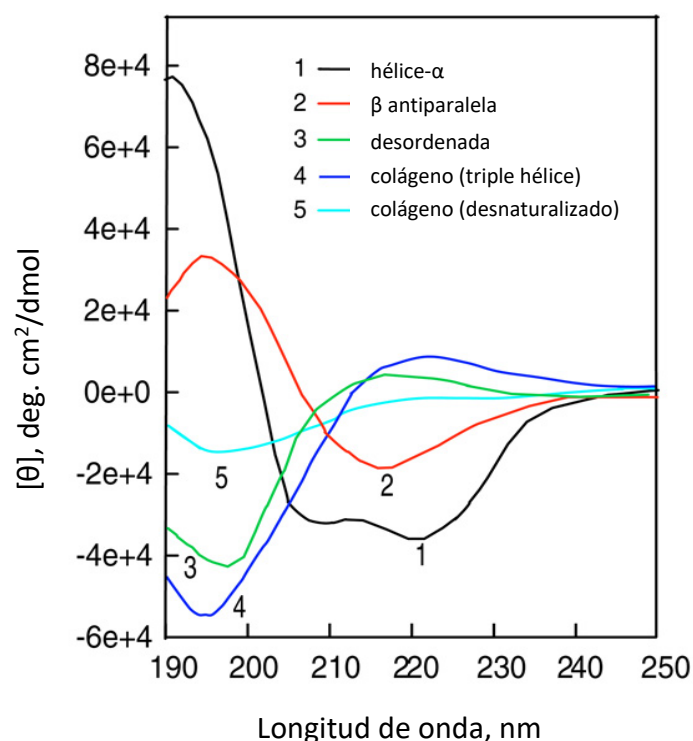
información a nivel de residuo (**Figura 3**). Por ello, NMR ha sido una de las técnicas más utilizadas para estudiar IDPs/IDRs (Ferreon *et al.*, 2022).



**Figura 3.** Espectro de NMR de una proteína estructurada y una IDP. a) Espectro de NMR de una proteína estructurada. b) Los espectros de las IDPs/IDRs presentan una señal más estrecha respecto a aquellas proteínas estructuradas. Las señales del espectro de las IDPs/IDRs se muestran mayormente sobrelapados (puntos rojos). Modificado de Konrat, 2014.

Otra de las técnicas de elección para estudiar a las IDPs/IDRs es el dicroísmo circular, basado en la propiedad diferencial de absorción de la luz polarizada, y es utilizada generalmente para determinar las propiedades estructurales de las proteínas en solución. Se emplea ampliamente para determinar de forma cuantitativa el contenido de hélices- $\alpha$ , estructuras  $\beta$  y *loops* en distintas condiciones e incluso en proteínas formando complejos (Ferreon *et al.*, 2022; Miles *et al.*, 2023), sin embargo, el uso de esta técnica puede encontrarse limitada si la proteína por analizar incluye residuos de aminoácidos que no están presentes en estructuras secundarias canónicas; estos residuos se han clasificado como “otros”, “irregulares”, “desordenados” o “random coil” (Miles *et al.*, 2023). Cuando se analizan a las

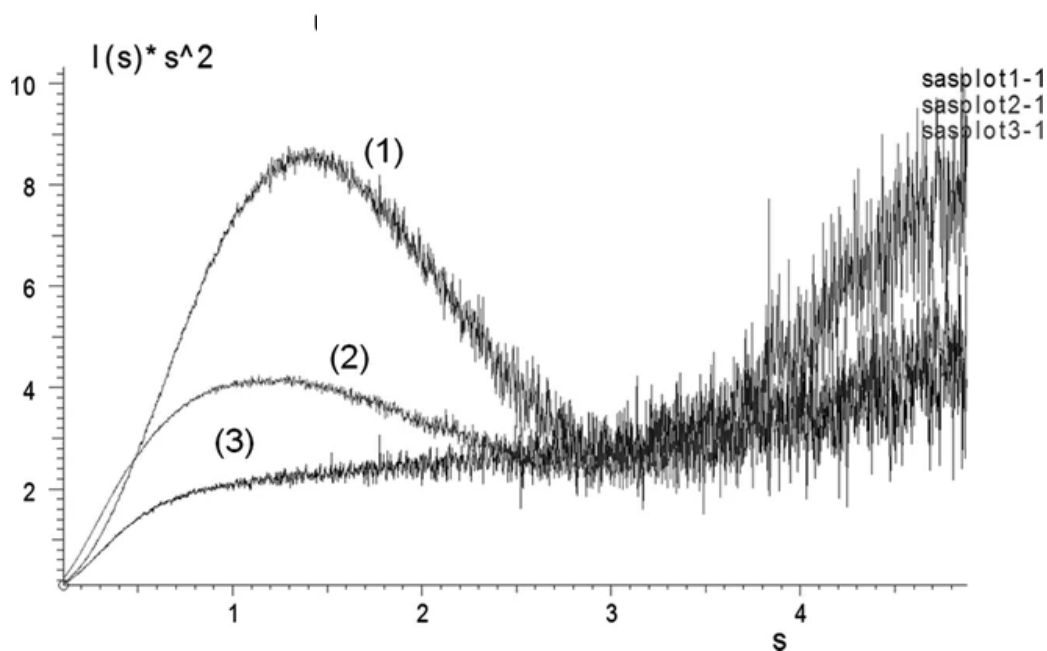
IDPs/IDRs mediante dicroísmo circular, es posible obtener un espectro específico de absorción en el UV-lejano que se diferencia de los espectros de dicroísmo de una estructura de hélice- $\alpha$  o una estructura  $\beta$  plegada (**Figura 4**), sin embargo, este método es semicuantitativo, y no proporciona información a nivel de residuo, esto implica que no sería posible identificar de forma clara proteínas que contengan tanto regiones ordenadas como desordenadas (Dunker *et al.*, 2001).



**Figura 4.** Espectros de dicroísmo circular de estructuras proteicas representativas. El espectro de dicroísmo de las IDPs/IDRs puede diferenciarse de aquellos espectros de hélices- $\alpha$  o estructuras  $\beta$  plegada. Modificado de Greenfield, 2006.

Dentro de los métodos de caracterización estructural directos, se encuentra la dispersión de rayos X de ángulo reducido (SAXS, por sus siglas en inglés), la cual es una técnica de baja resolución que no solo proporciona información sobre la forma, el estado oligomérico e información de estructuras cuaternarias, así como de complejos proteicos, sino que permite estudiar cuantitativamente sistemas flexibles o altamente dinámicos como las IDPs/IDRs (Bernadó & Svergun, 2012). En la técnica de SAXS, las macromoléculas por analizar se encuentran en solución, para que sean expuestas a un láser de rayos X; la intensidad se medirá a través de la dispersión de luz del láser en un detector como función del ángulo disperso. Con esta información, muchos

parámetros pueden ser obtenidos mediante las curvas que proporciona SAXS, permitiendo además analizar proteínas de una gran variedad de tamaños, y es posible validar sus estructuras en condiciones nativas o durante cambios en el ambiente que las rodea (pH, temperatura, fuerza iónica, adición de ligandos, etc), sin embargo, es importante destacar que para realizar experimentos de SAXS, la solución de la proteína debe ser bien preparada, conteniendo a la proteína altamente purificada, y además, se debe emplear un equipo altamente especializado para ello (Bernardó & Svergun, 2012). Los datos obtenidos por SAXS deben ser analizados de forma correcta para obtener las curvas espectrales correspondientes a cada estructura analizada. Para ello, se emplea el gráfico de Kratky que utiliza la intensidad de la señal dispersada y la longitud de onda en función de la transferencia del momento. Este análisis permite detectar la flexibilidad de las proteínas e identificar cuantitativamente a las regiones desordenadas de aquellas que son proteínas globulares (**Figura 5**) (Bernadó & Severgun, 2012; kikhney & Svergun, 2015).



**Figura 5.** Representación de las curvas de SAXS. Se muestran las mediciones realizadas para una proteína (1) globular, (2) parcialmente desordenada y (3) completamente desordenada. Tomado de (Bernardó & Svergun, 2012)

Las curvas proporcionadas por SAXS mediante el gráfico de Kratky, muestra a las curvas de las proteínas globulares como una campana con un máximo, mientras que a las IDPs/IDRs las exhibe con una meseta inicial seguido de un aumento gradual de la señal (**Figura 5**). La caracterización conformacional de las IDPs/IDRs puede ser



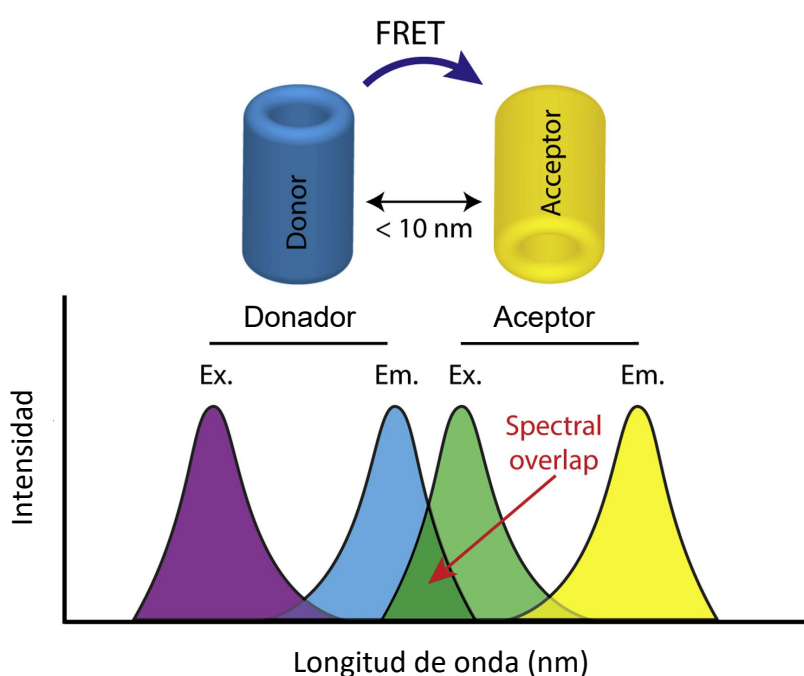
validada mediante el cálculo del radio de giro a través de los datos de SAXS, y comparar los valores teóricos con los experimentales para validar las estructuras. Por todo lo anterior, SAXS proporciona una caracterización rápida de las proteínas en solución, con una metodología de preparación de la muestra fácil de realizar. Es importante destacar que para obtener una adecuada caracterización de las conformaciones de las IDPs/IDRs, no basta con emplear una sola técnica, sino que, es posible validar los experimentos con técnicas complementarias como difracción de rayos X, mediciones espectroscópicas o incluso con métodos indirectos como una cromatografía (Bernadó & Svergun, 2012). De igual manera, pueden emplearse herramientas de predicción de estructuras como I-TASSER, CORAL y AlphaFold 2.0 (Bernadó & Svergun, 2012; Zhang, 2008; Alderson *et al.*, 2023).

Existe otra técnica llamado microscopía crio-electrónica (Cryo-EM, por sus siglas en inglés), que permite una caracterización estructural de las macromoléculas con una resolución atómica menor a 4 Å, sin embargo, esta técnica se ha adaptado bien a las proteínas estructuradas/globulares, ya que la IDPs/IDRs es una de las limitantes principales de esta metodología de caracterización estructural, inclusive, aquellas regiones de proteínas que no logran ser resueltas por Cryo-EM coinciden con regiones predichas como IDRs, por lo que se muestra un comportamiento similar a la cristalografía de rayos X (Bernadó & Svergun, 2012; Nwanochie & Uversky, 2019). La limitación en el uso de Cryo-EM para estudiar IDPs/IDRs no significa que no pueda ser utilizada para comprender otros fenómenos de las proteínas desordenadas, ya que, esta técnica se ha empleado de forma complementaria con NMR de alta resolución para lograr un mejor entendimiento de las propiedades estructurales y funcionales de las IDPs/IDRs, ya que NMR ofrece la capacidad de investigar la dinámica conformacional o el mecanismo de unión de una proteína, pero se limita cuando se tienen proteínas de gran peso molecular; por otro lado, Cryo-EM sí puede ser utilizada para analizar proteínas con un peso molecular grande, incluso complejos, por lo que el uso de NMR y Cryo-EM podría adaptarse para estudiar el mecanismo de proteínas con una función específica, tales como aquellas involucradas en la regulación de la cromatina (Musselman & Kutateladze, 2021).

La combinación de técnicas experimentales para caracterizar la conformación estructural de las IDPs/IDRs permite asegurar cuál es el ensamble conformacional que una IDP/IDR posee en una condición específica, además de prevenir el

sobreajuste por el uso de una única técnica experimental (Thomassen & Lindorff, 2022). Uno de los métodos que permiten determinar el ensamble conformacional de las IDPs/IDRs de forma integrativa es la transferencia de energía por resonancia de fluorescencia de una sola molécula (smFRET, por sus siglas en inglés), el cual tiene las capacidades únicas y complementarias de poder mapear la distribución poblacional de la conformación de las IDPs/IDRs, puede combinarse con otros métodos basados en fluorescencia y tolerar un amplio rango de perturbaciones experimentales, incluyendo el tamaño de la proteína. La técnica de smFRET ha sido probada ampliamente de forma experimental y actualmente es utilizada para caracterizar una gran variedad de sistemas biofísicos y biológicos (Metskas & Rhoades, 2020).

Las técnicas basadas en fluorescencia han aumentado su uso dramáticamente en los últimos años (Algar *et al.*, 2019), una de las más ampliamente utilizadas para el análisis conformacional de las IDPs/IDRs es FRET. FRET es la transferencia de energía no radiante que ocurre cuando existe un acoplamiento de los dipolos de dos proteínas fluorescentes llamadas donador (D) y aceptor (A). Estas proteínas fluorescentes tienen la peculiaridad de contar con un solapamiento en su espectro de fluorescencia, el cual es uno de los principales requisitos que debe existir para que se genere la transferencia de energía del donador hacia el aceptor, por lo que, FRET es una técnica que depende fuertemente de la distancia o proximidad entre los fluoróforos (Algar *et al.*, 2019) (**Figura 6**).

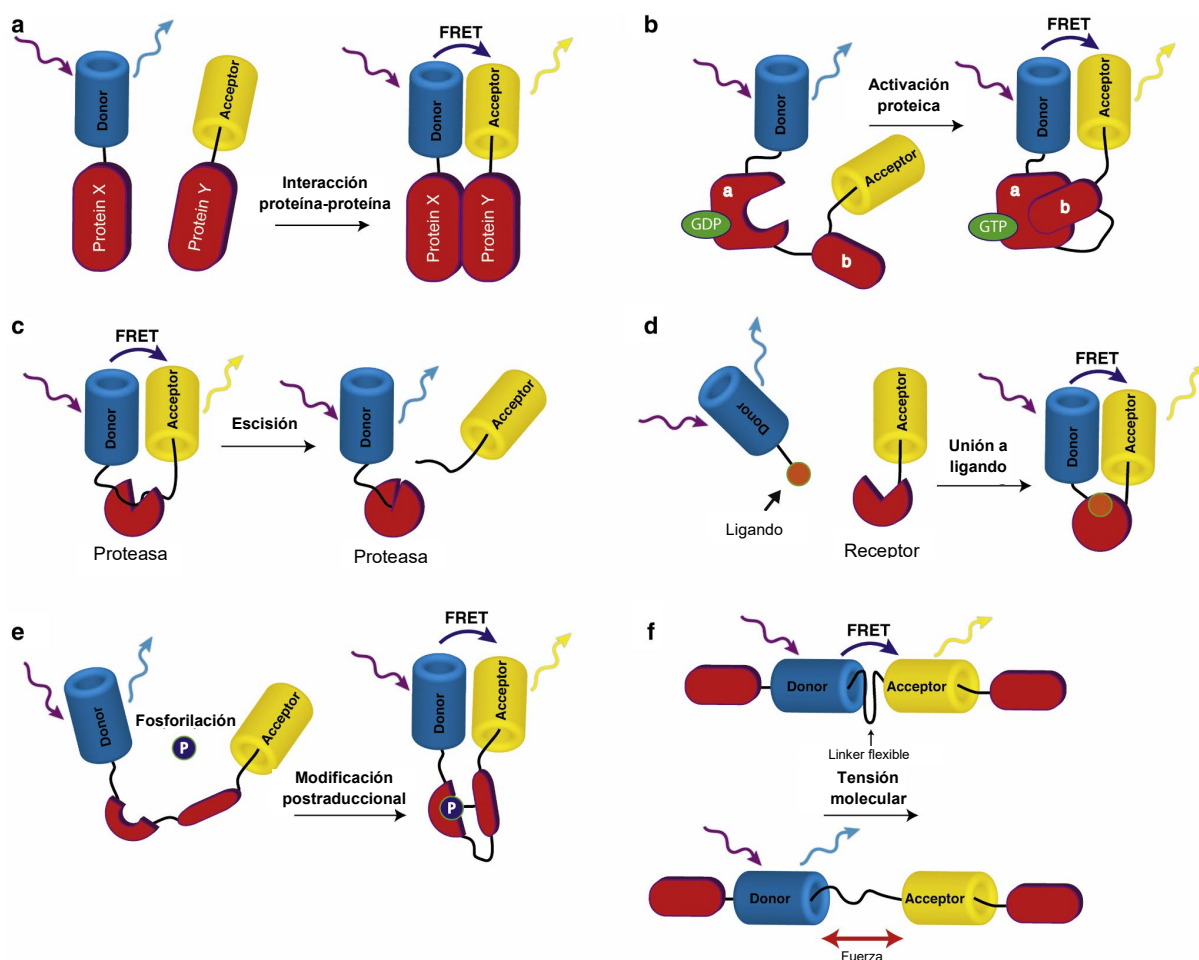


**Figura 6.** Diagrama de los requerimientos para FRET. El espectro de emisión de fluorescencia de la proteína donadora (azul) y el espectro de excitación de la proteína aceptora (verde) deben tener un solapamiento. La eficiencia de FRET es alta cuando se cumple una distancia menor o igual a 10 nm, y los dipolos de ambos fluoróforos están orientados de forma paralela. Modificado de Broussard & Green, 2017.

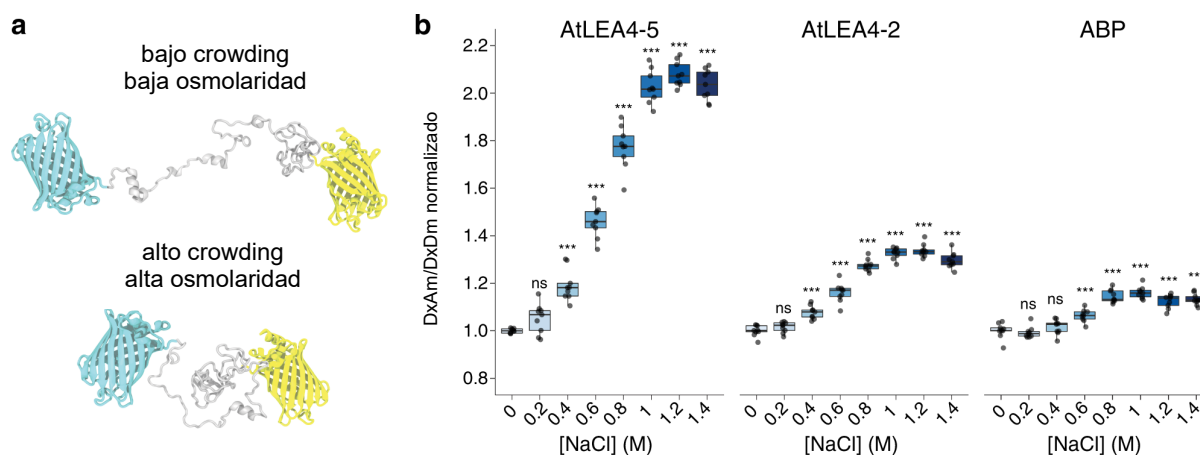
Una de las aplicaciones ampliamente utilizadas para FRET es en los biosensores codificados genéticamente, en donde se emplea una proteína aceptora y otra donadora. Generalmente se utilizan las proteínas *cyan* como donadora y *yellow* como aceptora debido al solapamiento en sus espectros, sin embargo, es posible utilizar otro par de fluoróforos, tales como los espectros de las proteínas fluorescentes verdes y rojas (Broussard & Green, 2017). FRET posee restricciones mínimas para su uso, incluso, puede medirse el fenómeno de FRET dentro de células vivas. El rango de aplicaciones de FRET es amplio; una de sus aplicaciones ha sido en la elucidación de estructuras bio-macromoleculares, pero puede utilizarse para otro tipo de aplicaciones, como el monitoreo de la dinámica de interacción proteína-proteína dentro de una célula, o cuando la bioquímica tradicional no permite obtener información espacial del comportamiento de un sistema proteico (**Figura 7**) (Broussard & Green, 2017; Algar *et al.*, 2019).

Debido a su amplia versatilidad como herramienta para estudiar diferentes propiedades de las proteínas, FRET se emplea para estudiar los cambios conformacionales de las IDPs/IDRs ante cambios en el ambiente químico que las rodea. El diseño de biosensores basados en IDRs se debe a que las IDPs/IDRs poseen propiedades específicas que les permiten responder ante cambios en el ambiente químico, presentando una alta sensibilidad a los cambios (Moses *et al.*, 2020, Cuevas *et al.*, 2021). Al utilizar FRET de forma complementaria con un linker que corresponda con una IDP/IDR es posible estudiar el posible cambio conformacional de las proteínas, pero, además, lograr monitorear los efectos de alguna perturbación en el ambiente dentro de una célula. Incluso, cuando se emplea el biosensor FRET con una proteína globular como referencia, no se muestra la misma intensidad de cambio de FRET, lo cual sugiere la alta capacidad de cambio conformacional de las IDPs/IDRs en este sistema (Cuevas *et al.*, 2021; **Figura 8**).

De forma complementaria, la técnica de single-molecule FRET (smFRET, por sus siglas en inglés) se ha empleado para estudiar el ensamble conformacional de las IDPs/IDRs (Metskas & Rhoades, 2020; Gomes *et al.*, 2020), una técnica que puede utilizarse para estudiar de forma específica la agregación IDPs/IDRs que, cuando son mezcladas en una solución, la molécula donadora transfiere su energía a la molécula aceptora debido a que las proteínas se encuentran formando agregados. Otra aplicación de smFRET es estudiar las interacciones proteicas, por ejemplo, es posible que una IDP/IDR se una a un blanco debido a la fosforilación de la IDP/IDR, lo cual indicaría una alta afinidad de la proteína con sus receptor; esta técnica ha permitido estudiar redes de interacciones de proteínas. smFRET puede emplearse al igual que los biosensores genéticamente modificados basados en FRET para estudiar a las IDPs/IDRs dentro de ambientes complejos, donde las IDPs/IDRs al ser altamente sensibles pueden cambiar su conformación desplegada a una estructura secundaria estable, o incluso cuantificar el fenómeno de separación de fases líquido-líquido (Metskas & Rhoades, 2020).

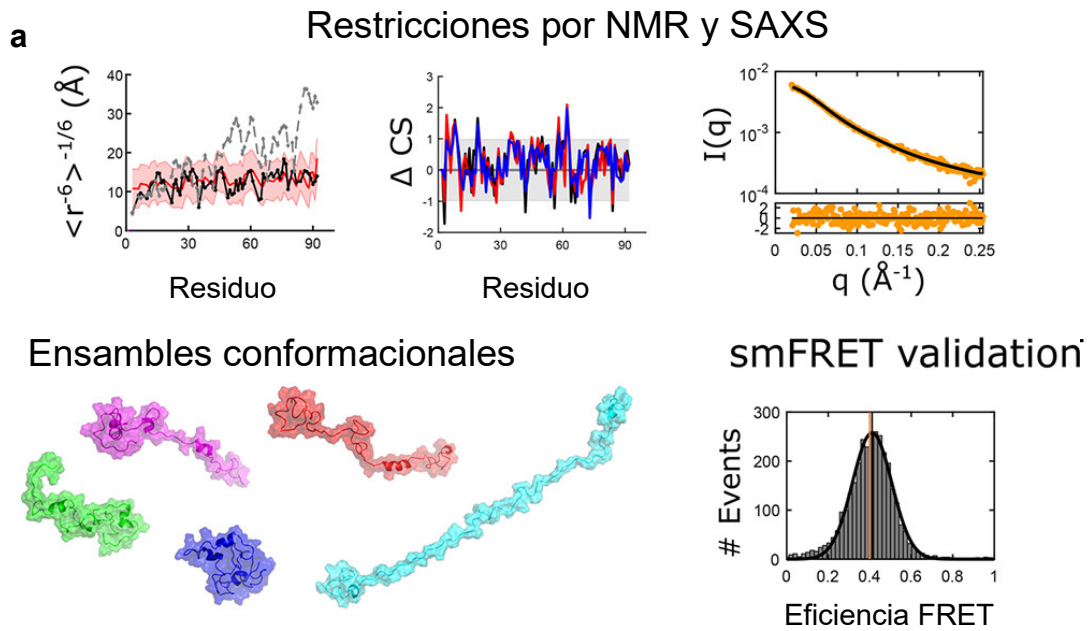


**Figura 7.** Aplicaciones comunes de los sensores FRET. La técnica de FRET puede utilizarse para estudiar diferentes mecanismos, *in vitro* o *in vivo*, entre ellos, se tiene a la a) interacción proteína-proteína, b) activación proteica cuando se une un sustrato a una enzima, cambia la conformación de la enzima y los fluoróforos se acercan; c) el mecanismo de acción de una proteasa; d) la unión de un ligando a un receptor para estudiar su posible afinidad; e) modificaciones postraduccionales como las fosforilaciones o f) la tensión molecular de un linker, determinando de forma indirecta la fuerza aplicada al linker que une a dos sistemas proteicos. Modificado de Broussard & Green, 2017.



**Figura 8.** Representación esquemática de un biosensor FRET. a) Diagrama del biosensor FRET utilizando la proteína CFP como molécula donadora y YFP como molécula aceptora. b) Valores de fluorescencia normalizada del biosensor FRET expresado en células de levadura bajo diferentes concentraciones de NaCl. Aquí, se determinó el cambio estructural de la proteína AtLEA4-5, una proteína desordenada; AtLEA4-2 una proteína parcialmente desordenada y ABP (arabinose-binding protein), una proteína globular. Modificado de Cuevas *et al.* 2021.

La gran ventaja de las técnicas basadas en fluorescencia como lo es FRET o smFRET o variaciones, es la flexibilidad que posee la metodología para caracterizar las propiedades conformacionales y biofísicas de las IDPs/IDRs, logrando resolver preguntas de investigación complejas relacionadas a la dinámica conformacional de las IDPs/IDRs (Metskas & Rhoades, 2020). Una de las aproximaciones con mayor ventaja es la combinación de técnicas experimentales para caracterizar las propiedades conformacionales de las IDPs/IDRs, logrando una mejor descripción y clasificación de las proteínas desordenadas (Gomes *et al.*, 2020; **Figura 9**)



**Figura 9.** Representación de la combinación de técnicas experimentales para estudiar las IDPs/IDRs. Es posible estudiar las propiedades del ensamble conformacional de las IDPs/IDRs a través del uso integrativo de técnicas como NMR, SAXS y smFRET, logrando proporcionar un alto grado de confiabilidad en los resultados. Modificado de Gomes *et al.*, 2020.

## 2. PLANTEAMIENTO DEL PROBLEMA

Los factores de transcripción (FT) poseen dominios de unión a DNA y regiones intrínsecamente desordenadas vinculadas a motivos funcionales, elementos esenciales para llevar a cabo un control transcripcional (Már *et al.*, 2023). En el contexto de plantas como *A. thaliana*, se han identificado regiones desordenadas en diversas familias de FT (Salladini, 2020). Estas regiones desordenadas han sido propuestas como una vía de reconocimiento molecular, dada su plasticidad de unión, lo cual facilita la regulación transcripcional de los dominios estructurales de los FT (Hsiao, 2022).

En Arabidopsis, los FT exhiben un enriquecimiento en regiones desordenadas en comparación con su proteoma, sugiriendo que los niveles de desorden podrían desempeñar un papel crucial en la función de los FT (Strader *et al.*, 2022). Las regiones desordenadas de FT, junto con los mecanismos de transducción de señales, pueden formar redes de interacción para responder ante diversos estímulos, incluido el estrés abiótico (Hsiao, 2022). Sin embargo, aún no se ha realizado una caracterización experimental de la sensibilidad de las IDRs de los FT de Arabidopsis ante condiciones de estrés hiperosmótico. El objetivo fundamental de este proyecto es explorar la sensibilidad estructural de las IDRs más largas de algunos FT de Arabidopsis in vivo bajo condiciones crecientes de NaCl mediante el empleo del sistema FRET. Este enfoque proporcionará información crucial sobre la sensibilidad de las IDRs en situaciones de estrés hiperosmótico y su posible asociación con la funcionalidad durante las respuestas adaptativas de las plantas.

### 3. HIPÓTESIS

La estructura de la región intrínsecamente desordenada más larga de algunos factores de transcripción de *A. thaliana* es sensible a los efectos producidos por choques hiperosmóticos con NaCl en células de levadura vivas.

### 4. OBJETIVO

Determinar los niveles de sensibilidad estructural de la región desordenada más larga de algunos de los factores de transcripción de *Arabidopsis thaliana* mediante ensayos FRET en células de levadura vivas ante choques hiperosmóticos con cloruro de sodio.

#### 4.1. OBJETIVOS PARTICULARES.

- Predecir los niveles de desorden de todos los factores de transcripción del proteoma de *Arabidopsis thaliana*.
- Determinar las propiedades fisicoquímicas de la región desordenada más larga de algunos factores de transcripción de *Arabidopsis thaliana*.
- Seleccionar regiones desordenadas de los factores de transcripción de *Arabidopsis thaliana* con una alta predicción dual de desorden/hélice- $\alpha$ .
- Determinar los cambios conformacionales de las IDR's en los factores de transcripción mediante FRET en células de levaduras vivas.

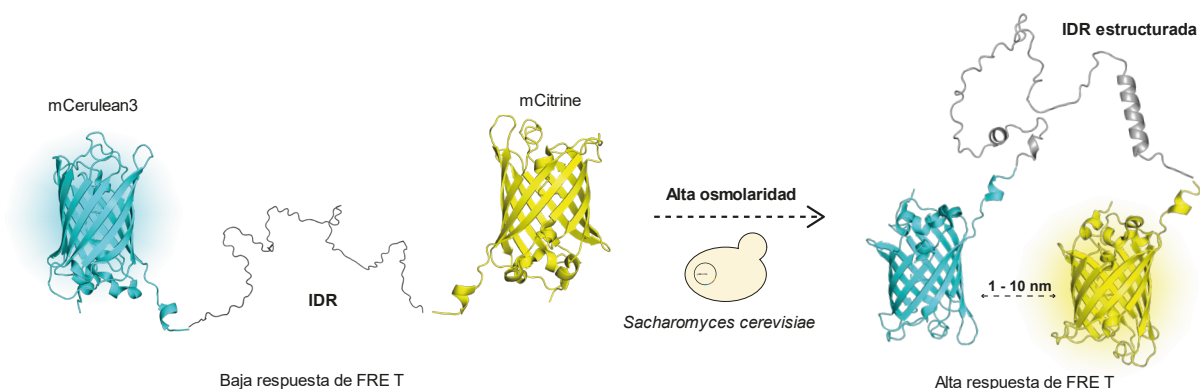


## 5. MATERIALES Y MÉTODOS

### 5.1. Obtención de plásmidos, transformación y genotipificación

Los plásmidos conteniendo las construcciones de las regiones desordenadas más largas de los FT de Arabidopsis fueron adquiridos a través de la empresa GenScript. En total, se generaron 200 plásmidos que contienen regiones desordenadas de distintas proteínas de una gran variedad de organismos. Los plásmidos pertenecen a una biblioteca parte de un proyecto de posgrado, en la cual se encuentran las 21 construcciones empleadas.

Las células de levadura se cultivaron de manera regular a una temperatura de 30°C, en medio de cultivo denominado SD-Ura, el cual carece de uracilo. Posteriormente, para la selección de levaduras portadoras de fluorescencia, se tomaron cinco colonias aisladas que se cultivaron en secciones individuales dentro del medio SD-Ura. Se realizó la genotipificación de estas colonias y se seleccionaron aquellas que demostraron tener los plásmidos deseados.



**Esquema.** Representación del fenómeno de FRET al someter a las células de levadura *Sacharomyces cerevisiae* a un choque hiperosmótico con NaCl. Las IDRs al ser sometidas a una alta osmolaridad, cambian su conformación extendida a estructurada, permitiendo acercar a las proteínas fluorescentes, ocurriendo una transferencia de energía de la proteína donadora (mCerulean3) hacia la proteína aceptora (mCitrine).

## 5.2. Construcción de bases de datos y caracterización bioinformática de las regiones intrínsecamente desordenadas

Se calculó el porcentaje de hélice- $\alpha$  para un conjunto de 27,411 proteínas correspondientes al proteoma de Arabidopsis. Estos datos se obtuvieron a partir de una base de datos previamente creada en colaboración con el estudiante Manuel Andrés Velasco Saavedra. Adicionalmente, se emplearon los valores de hélice- $\alpha$  de 1,958 FT de Arabidopsis contenidos en una base de datos desarrollada en su totalidad por la estudiante María del Rocío Alcántara.

Se llevó a cabo un proceso de filtrado de los 1,958 FT en la base de datos de 27,411 proteínas. Este filtro se basó en los valores de hélice- $\alpha$  y se utilizó una función de probabilidad implementada en el lenguaje de programación R. Se generaron histogramas considerando niveles de significancia de 0.90, 0.95 y 0.99. Cada nivel de significancia proporcionó tres valores de hélice- $\alpha$ : 11.76, 14.19 y 18.73, correspondientes a la distribución de hélice- $\alpha$  de los 1,958 FT. Para el análisis subsiguiente, se empleó exclusivamente los valores con una significancia de 0.99, lo que resultó en un conjunto de datos más manejable.

La base de datos de 1,958 FT contiene tanto las secuencias completas de las proteínas como las secuencias de las regiones desordenadas más largas. Estas últimas se utilizaron para calcular el porcentaje de hélice- $\alpha$  por residuo de aminoácido, haciendo uso del software AGADIR (<http://agadir.crg.es/agadir.jsp>). Además, se determinó el porcentaje de desorden por cada residuo de aminoácido en las secuencias de las regiones desordenadas más largas mediante el predictor de desorden PONDR (<http://www.pondr.com/>).

Los datos obtenidos del porcentaje de hélice- $\alpha$  y desorden se emplearon para calcular un valor denominado “dualidad”, que se utilizó en el laboratorio. Antes de realizar el cociente de dualidad, se aplicaron dos filtros: cada residuo de aminoácido debía tener un porcentaje de desorden igual o superior al 50% y un porcentaje de hélice- $\alpha$  de al menos 10%. Los residuos que cumplieron con ambos criterios se utilizaron para determinar su valor de dualidad. Los valores de dualidad para cada residuo se utilizaron para llevar a cabo un análisis gráfico de las IDRs de interés.

Las 69 IDR's seleccionadas se sometieron a un análisis bioinformático, en el cual se calcularon los valores de hidropatía, carga neta media, fracción de desorden promovido, fracción de residuos cargados (FCR) y carga neta por residuo (NCPR), utilizando el software de CIDER (<http://pappulab.github.io/localCIDER/>).

### 5.3. Generación de las construcciones de la biblioteca de regiones desordenadas.

Se hizo uso de una base de datos que contenía 200 IDR's desarrollada en el laboratorio del Dr. Cesar Luis Cuevas Velázquez. Las IDR's previamente caracterizadas de los FT fueron sometidas a un proceso de filtrado dentro de esta biblioteca de 200 IDR's. Como resultado de este filtrado, se obtuvo un conjunto de 21 IDR's correspondientes a los FT de Arabidopsis. Estos 21 IDR's fueron los seleccionados para su posterior utilización en los experimentos.

**Tabla 1.** Secuencias de las IDR's empleadas en los ensayos.

Identificador	AGI	Uniprot	Secuencia de la IDR
IDRBS-045	At1g68640	Q9SX27	SSDQRTLRLAQNREAARKSR LRKKAYVQQLENSRIRLAQLEE ELKRARQQGSLVER
IDRBS-054	At2g41130	O80674	TTMQSLCGSSSSTSSYYPLAIS GIGETMAQDRALAALRNHKEA ERRRRERINSHLNKLR
IDRBS-064	At3g13040	Q949U2	GAIQLTEALRMQMEVQKQLHE QLEVQRVLQLRIEEHAKYLEKM LEEQRKTGRWISSSSQTVLSP SDDSIQDSQNMSKTKASSPQP PLPAENKASETEDDKCESPQK RRRLENIAESEDPKR
IDRBS-077	At1g21200	Q9LMN5	LEEQKLQIQVELLELEKQRFRW QRFSKKRDQELERMRENER MKLENDRMGLELQRELQVE

IDRBS-081	At5g45710	Q9FK72	LQAQNPLTESERRSMEDQIER LKNEKEGLLAELQNQEQRKE FELQVTTLKD
IDRBS-082	At1g62990	Q9FPQ8	FGPLLPTESERSLMERVQEL KLELKQGFKSRIEDVREEIMRK RRAGKLPGDTT
IDRBS-136	At3g05760	F4J8I3	RALGMSMRVERSSLEQVQER FEVLKKRKAPGTFTEQDLDERI RKQQEEEEELKRQRREKKKEK KKGKVVEEPEMDPEVAEMM GFGGFGSSKKS
IDRBS-137	At3g54390	A0A1I9LT W1	NKPEKDAEMDTSSTPVVGTK VRGKKVKRRYKEEKKEIAGSIR WLAEVVMRSEARMETMKEIE RMRAEAEAKRGELDLKRTEIM ANT
IDRBS-138	At4g22950	O82743	RKLLGEGIDACSIEELQQLNQ LDRSLSRIRAKKYQLLREEIEKL KAEERNLVKENK
IDRBS-139	At1g06850	Q8VWG0	LSELWNIDPKRAKRILANRQSA ARSKERKARYIQELERKVQSL QTEATTLQAQLTLYQRDTNGLA NENTELKLRLQAMEQQAQLRN ALNEALRKEVERMKMETGEIS GNSDSFDM
IDRBS-140	At5g01380	Q9SDW0	NTLKDILEEFMRQTVKMEKEW RDAWEMKEIEREKREKEWRR RMAELEEEERAATERWMERE EERRLREEARAQKRDSLIDALL
IDRBS-141	At2g38250	O80450	DPKILNPKKNIKKRKGGSNSS

			NSNNGVREVLEEFMRHQVRM ESEWREGWEAREKERAKEE EWRRKMEELEKERLAMERMW RDREEQRRSREEMRAEKRDS LINAL
IDRBS-142	At1g09770	P92948	PFRDELHINEDMDMHESAKLE RQRREEARRSLRSGLTGLPQP KNEYQIVAQPPPEESEPEEKI EEDMSDRIAREKAEERQQA LLKKRSKVLQRDLPRPPAASLA VIRNSLLSADGDKSSVVPPTPI EVADKMVREELLQ
IDRBS-143	At3g50060	Q9SN12	MADRVKGPWSQEEDEQLRRM VEKYGPRNWSAISKIPGRSG KSCR
IDRBS-144	At4g17460	P46600	VEKLTEENRRLEKEAAELRALK LSPRLYGQMSPPTLLMCPSC ERVAGPSSSNHNQRSVSL
IDRBS-145	At1g04850	A0A178W 2G7	TPEEVKLKAQELRERARKKKE EEEKRMEREREKERIRIGKELL EAKRMEEVNERKRLMFLRKA KEEEKRAREKIRQKLEEDKAER RRKLGLPPEDPATAAAKPSV
IDRBS-146	At1g68240	Q5XVH0	PLQEIFGVTKNNDHEKHDEEP DEESYRMAKRQRSMEYRMMM EKKRRKEIKDKVDIL
IDRBS-147	At5g04150	Q9FYE6	KRKL SIPMTVARVVKYIPEQKQ ELQRLSRRKEELLKRISRKTHQ EQLRNKAMMDSID
IDRBS-148	At1g22070	Q39234	YTSVYNNSLEAEPSSNNDQDE

			DRINDKMKRRLAQNREAARKS RLRKKAHVQQLEESRLKLSQL EQELV
IDRBS-149	At3g02860	Q9M8S7	LQVVDSRMEEEEVDAAETIEEE EQREQRSYKEKVEILKRKKME LKAARLAKRSKTSEGSVKKPK KTEEEPSDEEDEDEDSAVDWR
IDRBS-165	At3g24490	Q9LV59	MEKSKKEQMKELEKMRADFQ RDLELQKKQIVDRAQSEIARLR EEEENHHGGGDDDESEDEEM ENDSDVNLSDE

En varios casos, se empleó una estructura gráfica consistente para visualizar las 21 IDR experimentales. Estas IDR se distribuyeron en un formato de 3 filas y 7 columnas, creando un conjunto de 21 paneles (**Figura 10**).

1 IDRBS-045	2 IDRBS-054	3 IDRBS-064	4 IDRBS-077	5 IDRBS-081	6 IDRBS-082	7 IDRBS-136
8 IDRBS-137	9 IDRBS-138	10 IDRBS-139	11 IDRBS-140	12 IDRBS-141	13 IDRBS-142	14 IDRBS-143
15 IDRBS-144	16 IDRBS-145	17 IDRBS-146	18 IDRBS-147	19 IDRBS-148	20 IDRBS-149	21 IDRBS-165

**Figura 10.** Representación gráfica de las 21 IDR experimentales. El identificador IDRBS-número, corresponde a la posición de la IDR dentro de la biblioteca de 200 IDR.

#### **5.4. Preparación de inóculos de levadura**

Las células de levadura se cultivaron en 3 mL de medio SD-Ura 1X a 30°C durante un periodo de 12 a 15 horas. La densidad óptica a 600 nm ( $OD_{600}$ ) de los inóculos se midió tomando 900 mL de medio SD-Ura 1X estéril y añadiendo 100  $\mu$ L del inóculo de levadura en una celda de cuarzo espectrofotométrica. Los inóculos con una  $OD_{600}$  en el rango de 1 a 2 de absorbancia fueron procesados para los ensayos. Aquellos inóculos con una  $OD_{600}$  menor a 1, se incubaron a 30°C durante unos minutos. Los inóculos con una  $OD_{600}$  superior a 2 se descartaron para el ensayo, y se preparó un nuevo inóculo.

#### **5.5. Determinación de la sensibilidad estructural mediante ensayos FRET**

Los inóculos con una  $OD_{600}$  entre 1 y 2 se procesaron colocando 1 mL del inóculo en tubos eppendorf de 1.5 mL y se realizaron lavados con buffer MES a pH 6. Las células de levadura utilizadas en el ensayo se mantuvieron resuspendidas en buffer MES pH 6.

Para inducir choques osmóticos, las células de levadura se sometieron a concentraciones de NaCl de 0 M, 0.5 M, 1 M y 1.5 M. Se utilizaron placas de 96 pocillos, donde en cada fila (designadas como A, B, C hasta la H), se colocaron 50  $\mu$ L del cultivo de células de levadura previamente lavadas con buffer MES pH 6 en volúmenes variables de MES y NaCl para obtener las concentraciones deseadas. El volumen total de cada pocillo en la placa fue de 200  $\mu$ L.

Posteriormente, las placas de 96 pocillos se leyeron utilizando un lector de placas, el cual efectuó una medición directa de la fluorescencia en los pocillos. El rango de lectura se ajustó desde 460 nm hasta 550 nm.

#### **5.6. Validación y cuantificación de los ensayos FRET**

Los datos de fluorescencia recabados del lector de placas se normalizaron para generar espectros de fluorescencia que se utilizaron para validar los experimentos. El valor del punto isobéptico a 515 nm del par FRET de la construcción se empleó como referencia para normalizar todos los datos de fluorescencia

adquiridos en un rango de 460 nm hasta 550 nm. Se desarrolló un código en el lenguaje de programación R para automatizar el proceso de validación de los ensayos de las 21 IDRs de los FT. El script está disponible en Github ([https://github.com/Kaz-bits/Project\\_IDP\\_LAB103](https://github.com/Kaz-bits/Project_IDP_LAB103)).

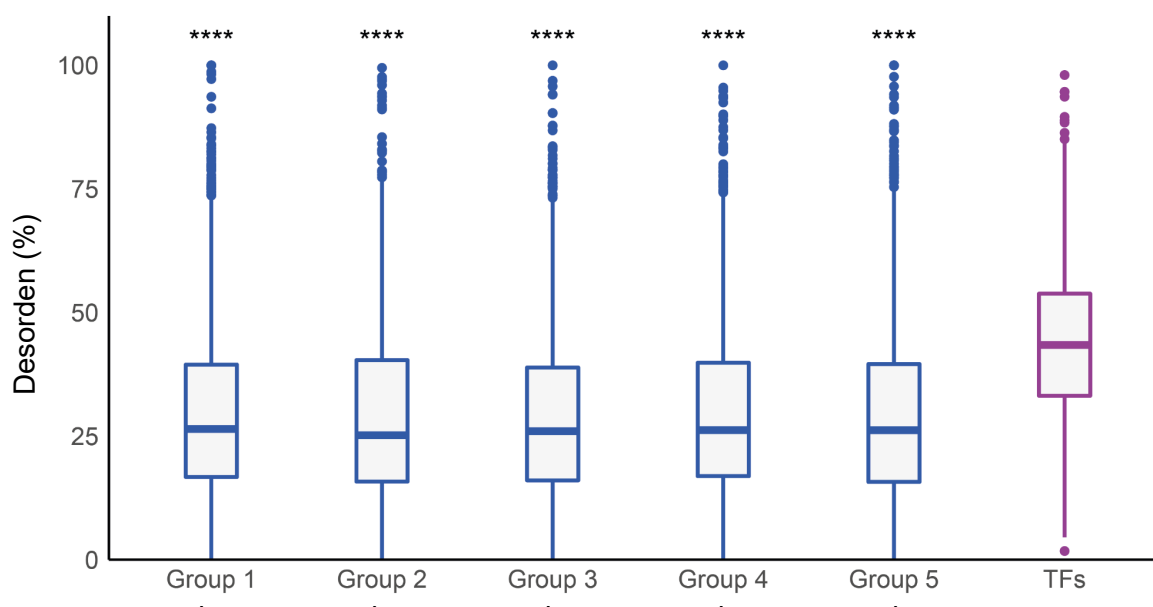
La cuantificación de los ensayos FRET de las 21 IDRs de los FT se llevó a cabo mediante la creación de gráficos de caja. Se utilizaron los valores de fluorescencia correspondiente a la máxima fluorescencia de la proteína donadora (mCerulean) y aceptora (citrina), que son 475 nm y 525 nm, respectivamente. A partir de estos datos de fluorescencia, se generó un cociente  $Dx_{Am}/Dx_{Dm}$ , denominado cociente FRET. Este cociente se obtiene al dividir el valor de fluorescencia a 525 nm entre el valor a 475 nm. El promedio de las réplicas técnicas a las concentraciones de 0 M se utilizó para normalizar los cocientes correspondientes a las concentraciones de 0.5 M, 1 M y 1.5 M. Para cada construcción se realizaron tres réplicas biológicas.



## 6. RESULTADOS

### 6.1. Los factores de transcripción poseen mayores niveles de desorden que el resto de las proteínas del proteoma de Arabidopsis

Utilizando una base de datos que abarcó los modelos más representativos del proteoma de Arabidopsis y un conjunto de 1,958 FT, se observó que los FT presentaron niveles significativamente más altos de desorden en comparación con el proteoma. El proteoma exhibió un nivel de desorden promedio de 27.09 %, mientras que los FT mostraron un nivel superior, alcanzando el 43.44 %. Esta misma base de datos fue utilizada para llevar a cabo un análisis que utilizó la distribución de diferentes parámetros, tales como longitud de las secuencias de aminoácidos, la fracción de desorden más larga, el número de IDRs, y la cantidad de residuos de aminoácidos presentes en las IDRs más largas. La **Tabla 1** desglosa los valores correspondientes a cada uno de estos parámetros.



**Figura 11.** Distribución del porcentaje de desorden entre las proteínas del proteoma de Arabidopsis en comparación con los FT. Se eligieron cinco grupos al azar del proteoma de Arabidopsis y se calcularon sus niveles de desorden promedio. Los grupos de proteínas elegidos se compararon con los FT de Arabidopsis. Se eliminaron los TFs del proteoma de Arabidopsis analizados para evitar sesgos en la comparación estadística.

Los resultados revelaron que el nivel de desorden en los grupos del proteoma de Arabidopsis exhibe una distribución similar. La media de porcentaje de desorden

para cada grupo se mantiene en un rango estrecho, con valores 25.4%, 25.6%, 25.9%, 26.9% y 25.1%, respectivamente (**Figura 11**).

En un esfuerzo por complementar los análisis de los niveles de desorden de los FT, se evaluó la fracción de residuos desordenados. Es importante destacar que las regiones desordenadas presentan una prevalencia de ciertos residuos en comparación con las regiones ordenadas. Los residuos considerados promotores del desorden incluyen Arg, Pro, Gln, Gly, Ser, Ala y Lys; por otro lado, se identifican residuos que promueven la formación de estructuras ordenadas, como Trp, Cys, Ile, Phe, Val, Asn y Leu (Uversky, 2019).

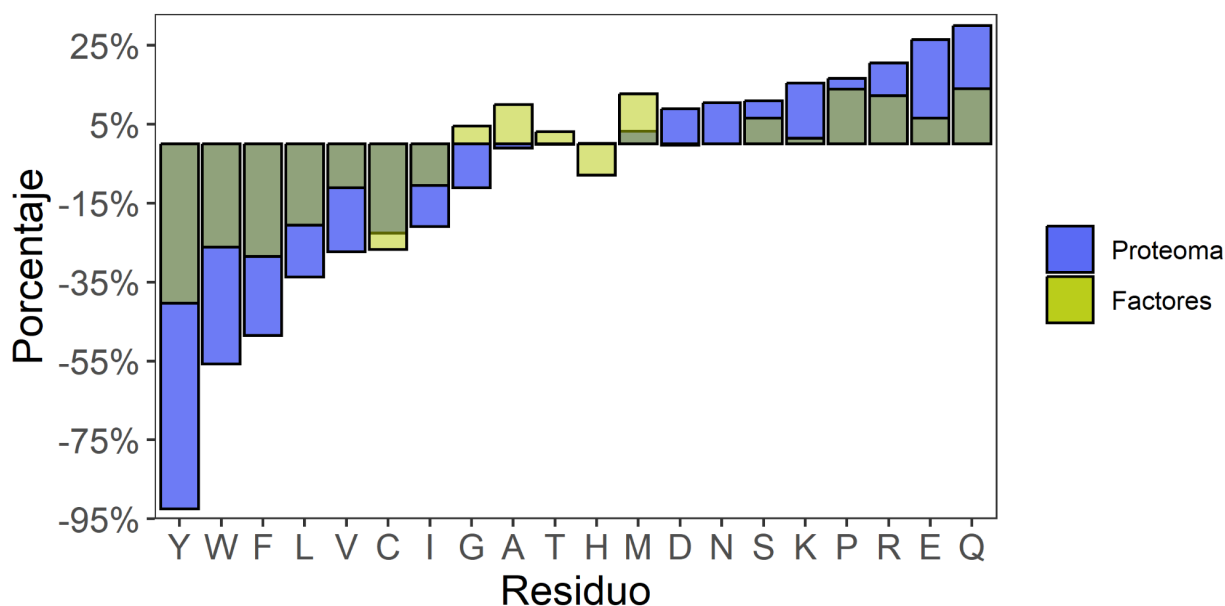
**Tabla 1.** Comparación promedio del proteoma y FT de Arabidopsis

	<b>Desorden (%)</b>	<b>Longitud (aa)</b>	<b>Fracción de desorden</b>	<b>Número de IDRs</b>	<b>Número de aa en IDRs más largas</b>
<b>Proteoma</b>	27.09	348	0.09	6	32
<b>Factores</b>	43.44	313	0.15	6	49

Al analizar la composición de residuos de aminoácidos en las secuencias de los FT en comparación con el proteoma, se observaron diferencias significativas en la frecuencia de ciertos residuos. En los FT, se encontró una mayor presencia de glicina (G), alanina (A), treonina (T), metionina (M), serina (S), lisina (K), prolina (P), arginina (A) ácido glutámico (E) y glutamina (Q). Todos estos aminoácidos son considerados promotores del desorden, con la excepción de treonina (T), metionina (M) y ácido glutámico (E) (Uversky, 2019). En contraste, se observó una menor presencia de tirosina (Y), triptófano (W), fenilalanina (F), leucina (L), valina (V), cisteína (C), isoleucina (I) e histidina (H) en los FT (**Figura 12**).

En lo que respecta al proteoma, se destacó la mayor presencia de metionina (M), ácido aspártico (D), asparagina (N), serina (S), lisina (K), prolina (P), arginina (R), ácido glutámico (E) y glutamina (Q) (**Figura 12**). Sin embargo, se encontró una menor

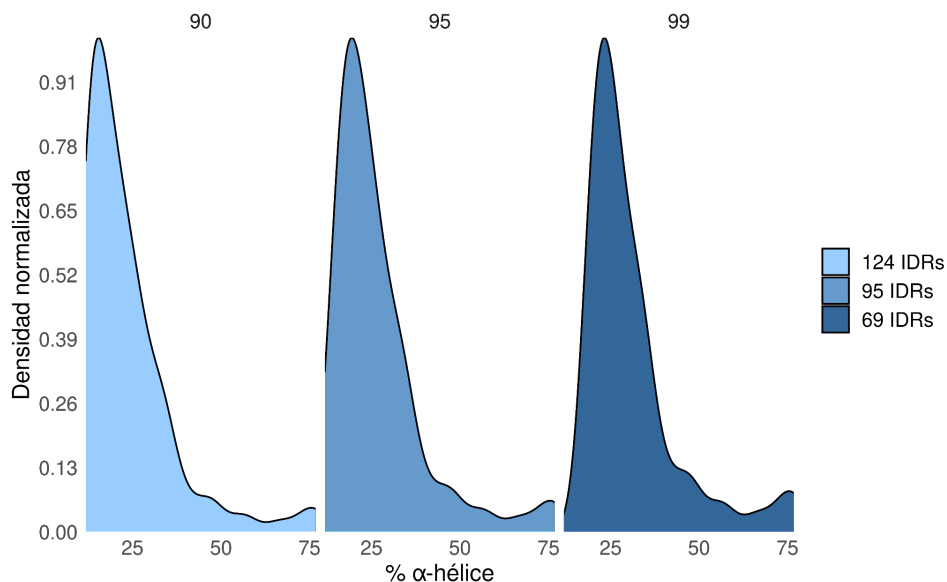
cantidad de tirosina (Y), triptófano (W), fenilalanina (F), leucina (L), valina (V), cisteína (C), glicina (G) y alanina (A) en el proteoma (**Figura 12**).



**Figura 12.** Distribución de los residuos de aminoácidos en el proteoma y FT de Arabidopsis. Las barras color amarillo corresponden a los residuos dentro de las secuencias de los FT, mientras que las barras azules pertenecen a los residuos encontrados en las proteínas del proteoma de Arabidopsis. Para el análisis, se removieron los FT del proteoma para evitar sesgos en la distribución de residuos.

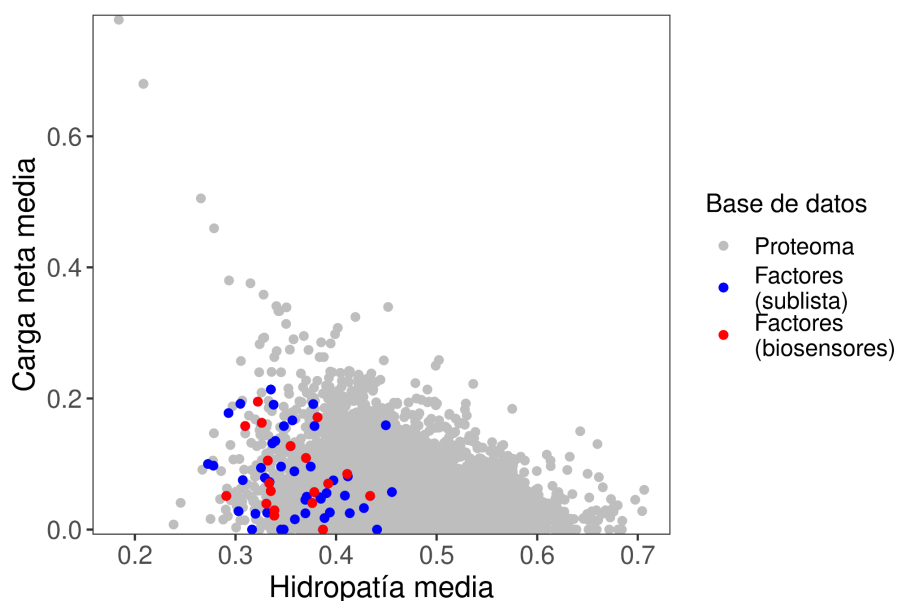
## 6.2. Caracterización fisicoquímica de las regiones intrínsecamente desordenadas más largas de los factores de transcripción

Utilizando la base de datos que comprendía 1,958 FT, se calculó el porcentaje de propensión a formar hélice- $\alpha$  para cada una de las secuencias de las IDR's más largas de los FT de Arabidopsis. Se aplicó un análisis estadístico para evaluar la distribución de los datos de porcentaje de hélice- $\alpha$  en todas las IDR's, empleando niveles de significancia de 0.90, 0.95 y 0.99 (**Figura 13**).



**Figura 13.** Función de distribución de probabilidades para el parámetro de hélice- $\alpha$  de los FT de Arabidopsis. Este análisis permitió elegir un grupo reducido de IDRs de los FT en donde el parámetro predicho de formación de hélice- $\alpha$  fuera mayor para las IDRs.

Se calcularon los parámetros de hidropatía, carga neta media y fracción de desorden promovido para evaluar las características de las 21 IDRs utilizadas en los ensayos FRET. Estos mismos parámetros se calcularon tanto para las 27,411 proteínas del proteoma de Arabidopsis como para la sublista de 69 FT seleccionados (**Figura 13** y **Figura 14**; **Métodos 4.2**).



**Figura 14.** Caracterización fisicoquímica de las IDRs más largas de los FT de Arabidopsis. El fondo gris representa las 27,411 proteínas del proteoma; los puntos azules los 69 FT

seleccionados con las funciones de probabilidad, y los puntos rojos las IDRs seleccionadas para los ensayos experimentales.

La información obtenida del proteoma completo de *Arabidopsis* se utilizó para recabar los datos relativos de los FT, así como de las 21 IDRs utilizadas en este proyecto. Los valores promedio para cada grupo de proteínas se detallan en la **Tabla 2**.

Con el fin de lograr una caracterización más completa de las regiones desordenadas en los FT, así como del proteoma y las secuencias completas de los FT de *Arabidopsis*, se empleó el parámetro de carga neta media. Las IDRs más largas presentaron valores de 0.08 y 0.09, considerados significativamente altos según la literatura. En dichos estudios se ha establecido que las proteínas no estructuradas generalmente tienen valores de carga neta media alta, aproximadamente alrededor de 0.12 (Uversky, 2000; Uversky, 2019; Bondos, 2021).

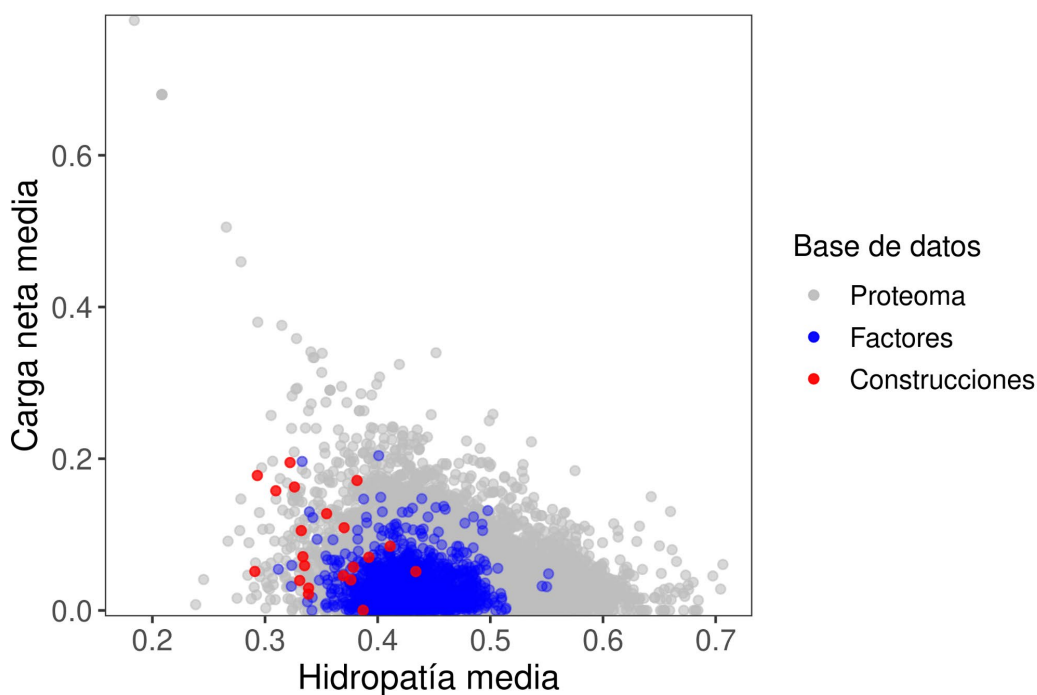
Se calcularon los valores de hidropatía para el proteoma, los factores y las IDRs empleadas en los experimentos. Los resultados revelaron que tanto el proteoma como los FT presentaban niveles similares de carga neta media, con valores de 0.03 y 0.02, respectivamente. En cuanto a la hidropatía, los valores fueron de 0.47 y 0.43, respectivamente (**Tabla 2**).

Respecto al subgrupo de IDRs más largas y las 21 IDRs utilizadas experimentalmente, los valores de carga neta media e hidropatía fueron de 0.08 y 0.09, y 0.36 y 0.35. Estos resultados permitieron identificar a las IDRs experimentales como regiones desordenadas, ya que se ha reportado que las IDRs carentes de estructuras secundarias suelen tener valores elevados de carga neta media y bajos valores de hidropatía (Uversky, 2019). En resumen, análisis bioinformáticos y estadísticos demuestran que las proteínas nativamente desplegadas suelen exhibir una carga neta media promedio de  $0.12 \pm 0.09$  y un valor de hidropatía de  $0.39 \pm 0.05$  (Uversky, 2000).

**Tabla 2.** Características fisicoquímicas de los grupos de proteínas usados en el proyecto.

Grupo	Carga neta media	Hidropatía	Fracción de desorden promovido	FCR
Proteoma	0.03	0.47	0.61	0.24
FT	0.02	0.43	0.65	0.25
Sublista de 69 IDRs	0.08	0.36	0.73	0.43
IDRs de 21 FT	0.09	0.35	0.72	0.43

Las secuencias completas de los FT y del proteoma mostraron similitudes en lo que respecta a los parámetros evaluados (**Figura 15.**, **Tabla 2**). Sin embargo, las 21 IDRs utilizadas en el experimento se comportaron de manera diferente. Esto se debe al hecho de que solo se consideró a la IDR más larga de los 21 FT. Esta diferencia sugiere que las IDRs empleadas podrían experimentar un cambio conformacional.



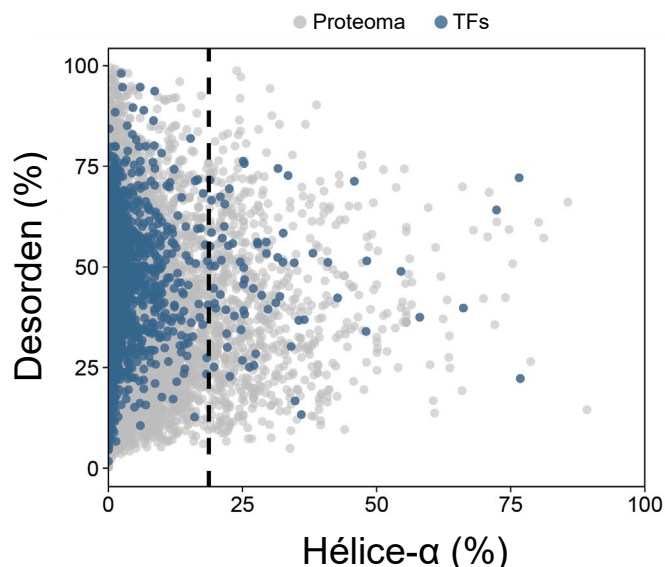
**Figura 15.** Gráfico de hidropatía media vs carga neta media. El fondo gris indica las 27,411 proteínas del proteoma; los puntos azules los 1,958 FT estudiados y, los puntos rojos las IDRs experimentales de Arabidopsis.

Otro de los parámetros considerados fue la fracción de residuos cargados (FCR, *Fraction of Charged Residues*). Aquí también se observaron similitudes entre el proteoma y los factores de transcripción, con las IDRs experimentales mostrando un valor más alto en comparación con los otros dos grupos (**Tabla 2**). El FCR indica la suma de cargas positivas y negativas de una secuencia específica, y en este contexto, puede utilizarse para evaluar el grado de compactación de una secuencia en función de la distribución de sus cargas (Uversky, 2019). Sin embargo, se ha observado que los valores de la distribución de cargas y el FCR por sí solos no son suficientes para determinar la hidrodinámica de las proteínas desordenadas (Firman, 2018; Uversky, 2019). Lo más apropiado sería determinar la posición de las cargas positivas y negativas dentro de la secuencia de aminoácidos, ya que esto puede influir en la conformación de la secuencia (Uversky, 2019).

Es importante señalar que la fracción de residuos desordenados para los FT es ligeramente mayor que la del proteoma, lo que indica la presencia de residuos promotores del desorden (Uversky, 2019). En el caso de las IDRs sometidas a experimentos de sensibilidad estructural, estas presentan un valor de 0.72, que es mayor que el del proteoma y los FT. Este valor ha permitido identificar otro parámetro que podría influir en la posibilidad de un cambio estructural de las IDRs experimentales (**Tabla 2**).

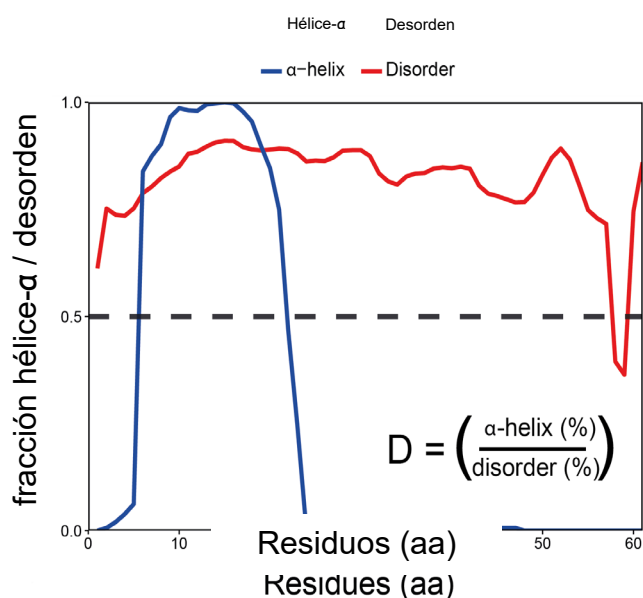
### **6.3. Determinación de los niveles de dualidad de la IDR más largas de los factores de transcripción de Arabidopsis**

Con base en los valores de porcentaje de hélice- $\alpha$  y desorden, se identificaron a las IDRs de los FT que podrían experimentar un posible cambio estructural. Esto se realizó utilizando los parámetros de CIDER mencionados anteriormente. Para ello, se observó la distribución del porcentaje de hélice- $\alpha$  y desorden, con un enfoque en el valor de 18.73% de hélice- $\alpha$  (**Figura 13**), y se seleccionaron las 69 IDRs que cumplieran con este criterio (**Figura 17, Métodos Sección 4.2**).



**Figura. 17.** Distribución del porcentaje de hélice- $\alpha$  y desorden de las secuencias del proteoma y de los FT de Arabidopsis. Los puntos grises representan a las 27,411 proteínas del proteoma, mientras que los puntos azules a los FT.

Después de filtrar las 69 IDRs, se calculó un parámetro denominado dualidad. La dualidad se define como el cociente del cambio de porcentaje de hélice- $\alpha$  entre el porcentaje de desorden por residuo de aminoácido. Este parámetro permitió una evaluación más detallada de las 69 IDRs seleccionadas. Los resultados de la evaluación de la dualidad se representan visualmente mediante gráficos de líneas (**Figura 18; Figura 19**). Estos resultados mostraron que la gran mayoría de las IDRs previamente seleccionadas cumplen con valores de dualidad significativamente altos (**Tabla 3**).





**Figura 18.** Parámetro de dualidad del proteoma y FT de *Arabidopsis thaliana*. El gráfico corresponde a una de las 21 IDRs analizadas. La línea color rojo representa el nivel de desorden, y la línea azul el nivel de propensidad de formación de hélice- $\alpha$ , ambos cuantificados por cada residuo de la secuencia.

**Tabla 3.** Valores de dualidad para las 69 IDRs filtradas y las 21 IDRs experimentales.

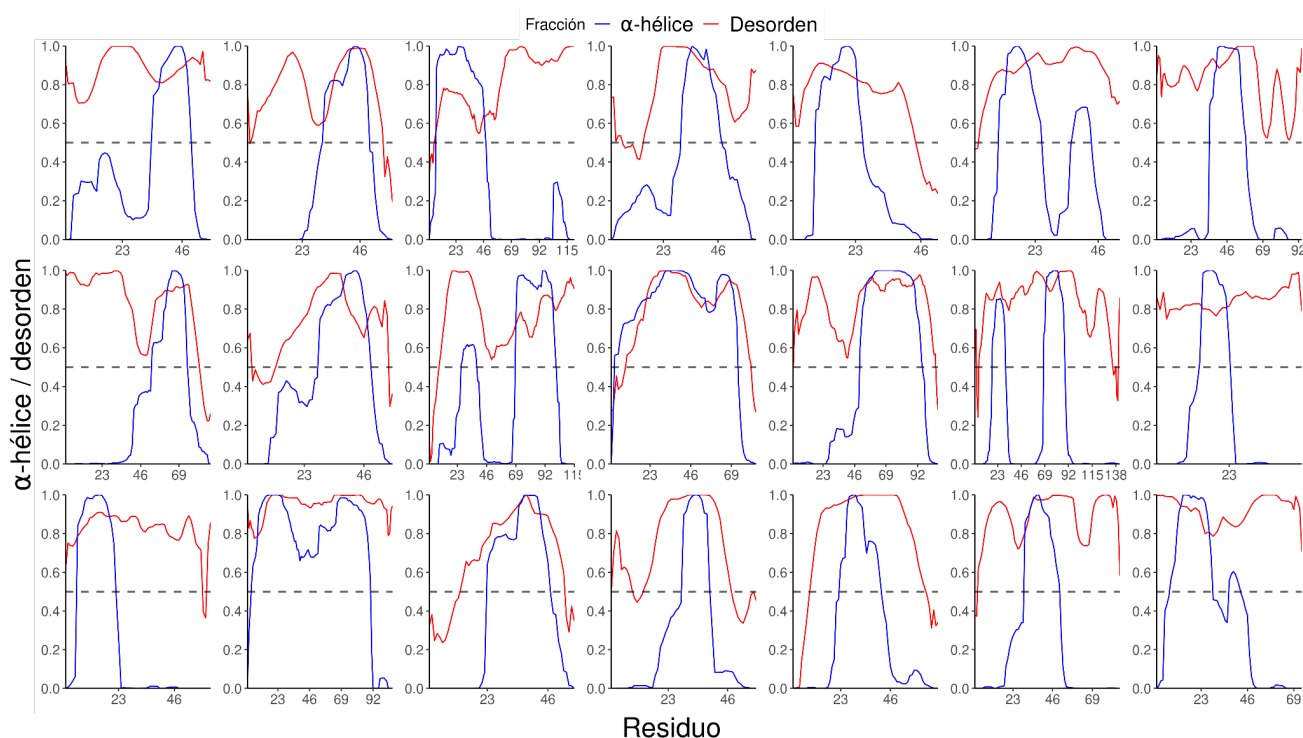
<b>IDR</b>	<b>1*</b>	<b>2</b>	<b>3*</b>	<b>4*</b>	<b>5</b>	<b>6*</b>	<b>7*</b>	<b>8</b>	<b>9</b>
Dualidad	84.91	54.29	50.43	70.42	65.08	66.13	63.24	74.29	68.42
<b>IDR</b>	<b>10</b>	<b>11*</b>	<b>12</b>	<b>13</b>	<b>14*</b>	<b>15*</b>	<b>16</b>	<b>17</b>	<b>18</b>
Dualidad	NA	66.67	62.50	NA	51.79	77.19	58.97	60.98	68.29
<b>IDR</b>	<b>19</b>	<b>20</b>	<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25*</b>	<b>26*</b>	<b>27*</b>
Dualidad	NA	75.00	52.56	51.06	64.00	59.46	67.29	52.54	62.35
<b>IDR</b>	<b>28*</b>	<b>29</b>	<b>30</b>	<b>31</b>	<b>32</b>	<b>33*</b>	<b>34</b>	<b>35</b>	<b>36*</b>
Dualidad	70.21	50.98	64.10	74.29	63.16	53.72	51.32	67.50	58.90
<b>IDR</b>	<b>37</b>	<b>38</b>	<b>39*</b>	<b>40</b>	<b>41</b>	<b>42*</b>	<b>43</b>	<b>44</b>	<b>45</b>
Dualidad	68.18	NA	68.89	71.32	53.13	55.68	NA	53.33	85.51
<b>IDR</b>	<b>46*</b>	<b>47*</b>	<b>48</b>	<b>49</b>	<b>50</b>	<b>51*</b>	<b>52</b>	<b>53</b>	<b>54</b>
Dualidad	70.49	70.18	69.91	57.69	72.82	89.16	62.79	NA	NA
<b>IDR</b>	<b>55*</b>	<b>56</b>	<b>57</b>	<b>58</b>	<b>59</b>	<b>60</b>	<b>61</b>	<b>62</b>	<b>63</b>
Dualidad	63.16	52.73	70.00	54.39	81.58	56.76	83.33	NA	61.82
<b>IDR</b>	<b>64*</b>	<b>65</b>	<b>66</b>	<b>67</b>	<b>68</b>	<b>69</b>			
Dualidad	50.00	65.52	75.00	62.86	57.89	86.36			

\* IDRs utilizadas en los ensayos de sensibilidad estructural

**Tabla 4.** Porcentaje de dualidad de las IDR's experimentales.

IDRBS	045	054	064	077	081	082	136
Dualidad	77.19	52.54	53.72	66.13	50.00	66.67	70.21
IDRBS	137	138	139	140	141	142	143
Dualidad	55.68	70.18	50.43	89.16	67.29	70.42	68.89
IDRBS	144	145	146	147	148	149	165
Dualidad	70.49	84.91	51.79	63.16	63.24	62.35	58.90

Es importante notar que los porcentajes de dualidad para cada una de las IDR's se calcularon bajo dos criterios. Primero, se identificaron los residuos de aminoácidos que cumplieran con un porcentaje de desorden mayor o igual al 50%, según el predictor de desorden PONDR. Luego, se aplicó un segundo filtro basado en el porcentaje de hélice- $\alpha$ , donde solo se consideraron a los residuos con un porcentaje superior al 10%, de acuerdo con el software AGADIR. Los valores de dualidad se calcularon solo para los residuos que cumplieran ambos criterios, lo que indica que algunos residuos carecen de valores de dualidad (**Figura18; Figura 19**).



**Figura. 19.** Gráficos de fracción hélice- $\alpha$  y desorden. Las líneas rojas representan el nivel de desorden, mientras que las líneas azules los niveles de propensidad a formar hélices- $\alpha$ . Los gráficos muestran a las 21 IDR seleccionadas para los experimentos FRET. El orden de los gráficos corresponde al diagrama mostrado en **Métodos 5.3, Figura 10**.

El valor de 0.5 para el nivel desorden se seleccionó porque el predictor utilizado, PONDR, clasifica como residuos desordenados a aquellos que superen este umbral. En cuanto al valor de 0.1 de hélice- $\alpha$ , se eligió debido a la variabilidad en la propensidad de los residuos a formar estructuras secundarias de hélice- $\alpha$ . El valor más bajo observado es de 0 para alanina, seguido de 0.06 para arginina, y el más alto de 1.0 para glicina (Pace, 1998). Por lo tanto, 0.1 se consideró como un umbral adecuado para abarcar una amplia gama de propiedades de residuos de aminoácidos.

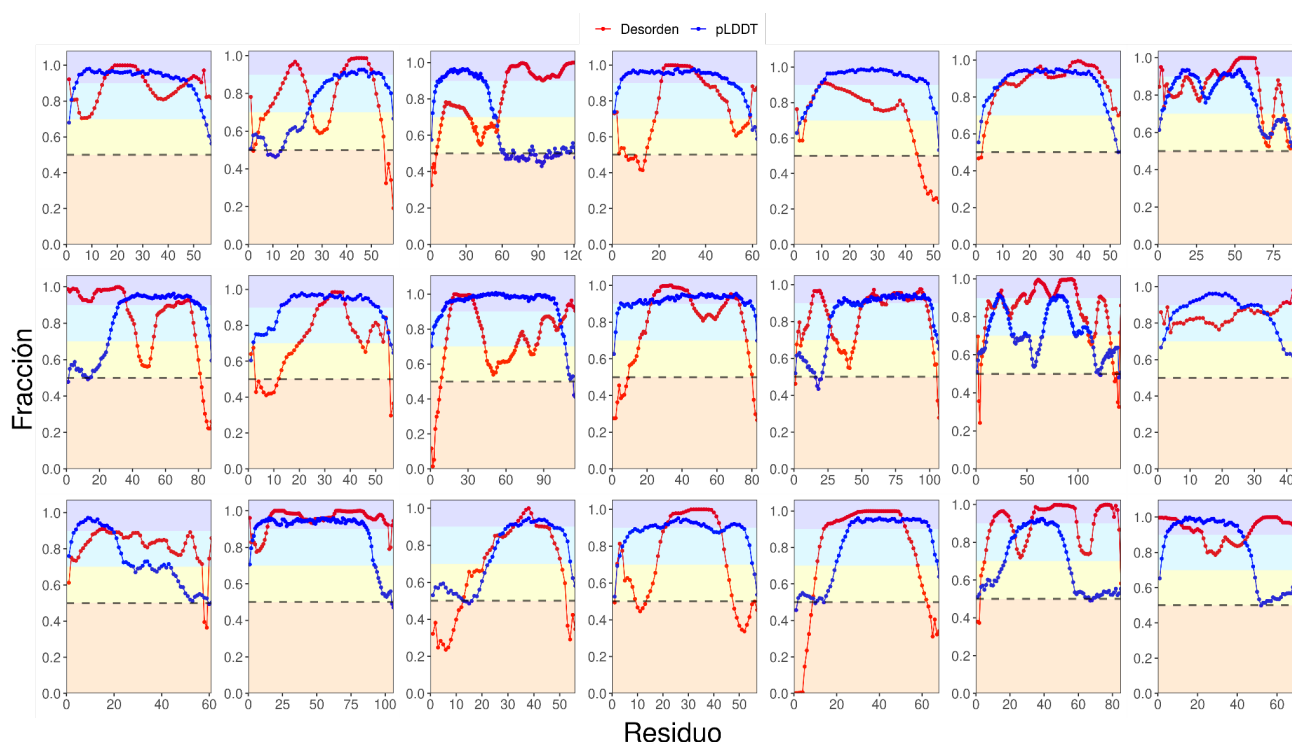
Los resultados de las visualizaciones de la dualidad de cada IDR indican que la mayoría de sus residuos son desordenados (tal como se esperaba), aunque se observaron regiones variables con propensidad a formar hélice- $\alpha$  (**Figura 19**). Esto sugiere que, a pesar de tener regiones altamente desordenadas, existe la posibilidad de un cambio estructural, respaldado por los datos de AGADIR y los gráficos de dualidad (hélice- $\alpha$  / desorden) (**Figura 19**).

Dado que los sistemas bioinformáticos para predecir la estructura de proteínas han ganado popularidad, se optó por usar la base de datos de AlphaFold 2.0 a través de códigos de Python y un meta-predictor de desorden llamado Metapredict V2 (Emenecker, 2021; Emenecker, 2022). AlphaFold 2.0 ha demostrado una alta precisión al predecir las estructuras de proteínas del proteoma humano, alcanzando un nivel del 98.5% (Tunyasuvunakool, 2021; Jumper, 2021). Por lo tanto, se emplearon los valores de pLDDT (predicted Local Distance Difference Test) de AlphaFold 2 para observar el posible cambio estructural de las IDRs.

Se calcularon los valores de pLDDT para cada residuo de aminoácido de las 21 IDRs. Estos valores se agruparon en cuatro categorías según la confiabilidad de la predicción de AlphaFold 2.0. Valores superiores al 90% se consideran altamente confiables; valores inferiores a 90% pero superiores al 70% se consideran confiables; valores inferiores al 70% pero superiores al 50% se consideran poco confiables, y

valores inferiores al 50% se consideran muy poco confiables (Tunyasuvunakool, 2021).

Al realizar las mismas visualizaciones gráficas como en la **Figura 19**, se observaron que ciertas zonas de las IDR, los residuos de aminoácidos presentaban valores superiores al 70%. Esto podría indicar una alta confiabilidad en la posibilidad de un cambio estructural en los residuos con estos valores de pLDDT. Es importante destacar que todas las IDRs analizadas contienen al menos una región con valores de pLDDT de alta confiabilidad (**Figura 20**).

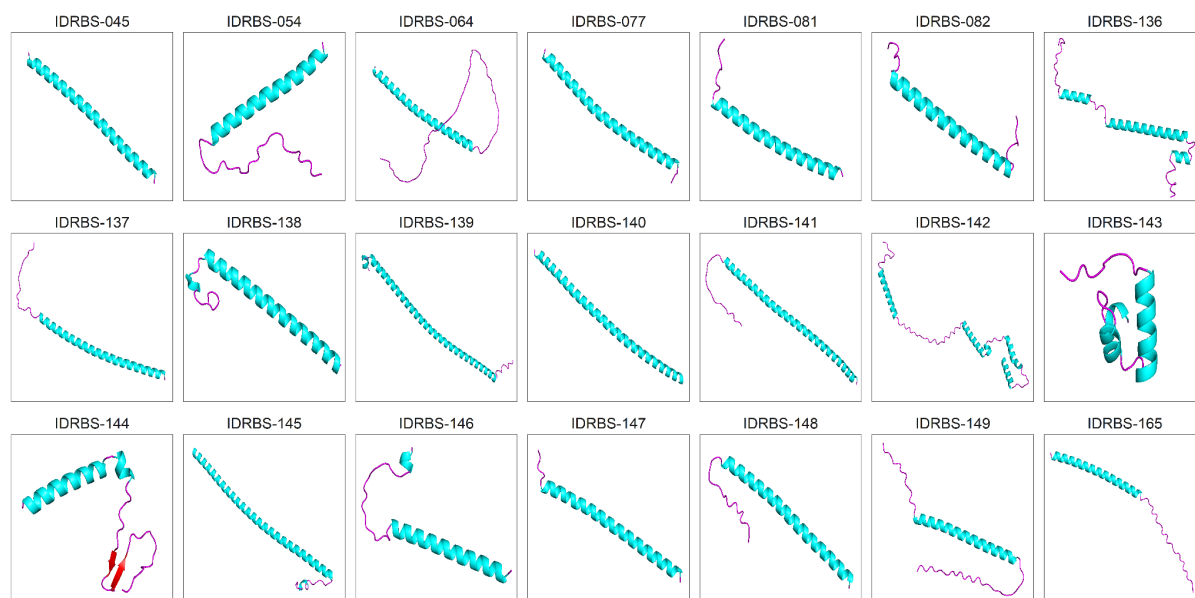


**Figura 20.** Gráficos de pLDDT y desorden de las 21 IDRs experimentales. Los colores indican la confiabilidad de la predicción por residuo de acuerdo con el valor de pLDDT. Las regiones morada y azul son las de mayor confiabilidad y en las que se encuentran la mayoría de los residuos de las IDRs analizadas. Las regiones amarilla y naranja son las de menor confiabilidad en la predicción de estructura. El orden de los gráficos corresponde al diagrama mostrado en **Métodos 5.3, Figura 10**.

Es importante mencionar que AlphaFold 2 realiza predicciones de las regiones desordenadas de forma condicional. Esto significa que las IDRs pueden mostrar un plegamiento estructural en función de ciertas condiciones, como interacciones o la unión a otras proteínas. En algunos casos, las IDRs pueden presentar estructuras que

son el resultado de varias condiciones combinadas (Alderson, 2022). Debido a esta observación, se decidió utilizar AlphaFold 2.0 para explorar posibles funciones implícitas que podrían influir en los ensayos de sensibilidad estructural posteriores.

Con la ayuda del algoritmo de AlphaFold 2 y el repositorio de ColabFold, que se basa en alineamientos múltiples de secuencias, se llevaron a cabo predicciones de las estructuras de las 21 IDRs experimentales (Mirdita, 2022) (**Figura 21**).



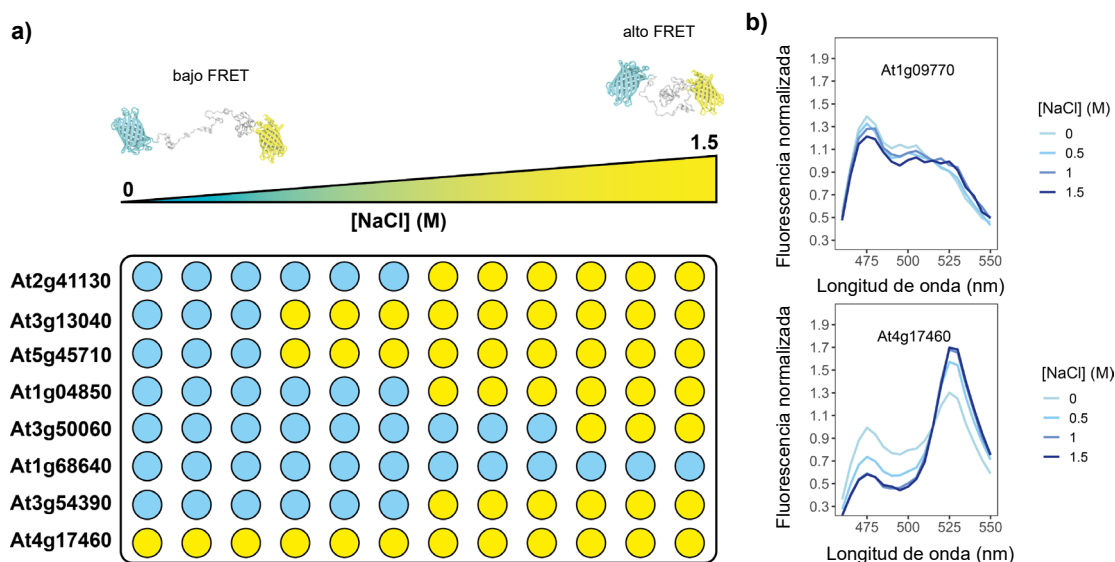
**Figura 21.** Estructuras de las 21 IDRs experimentales predichas por AlphaFold 2. El algoritmo de AlphaFold 2 predijo en cada una de las 21 IDR una estructura de hélice- $\alpha$  con valores de alta confiabilidad, mostrados en el gráfico de la figura 20. El orden de los gráficos corresponde al diagrama mostrado en **Métodos 5.3, Figura 10**.

Las estructuras predichas por el algoritmo de AlphaFold 2.0 revelaron que todas las IDRs presentan estructuras secundarias de hélice- $\alpha$ . En el caso de una de ellas, se observaron tanto estructuras de hélice- $\alpha$  como hojas  $\beta$ . Además, todas las IDRs exhibieron regiones desordenadas, aunque algunas de ellas tenían regiones más extensas en comparación con otras (**Figura 21**).

## 6.4. Determinación de los cambios conformacionales de las IDR de los factores de transcripción de Arabidopsis mediante ensayos FRET

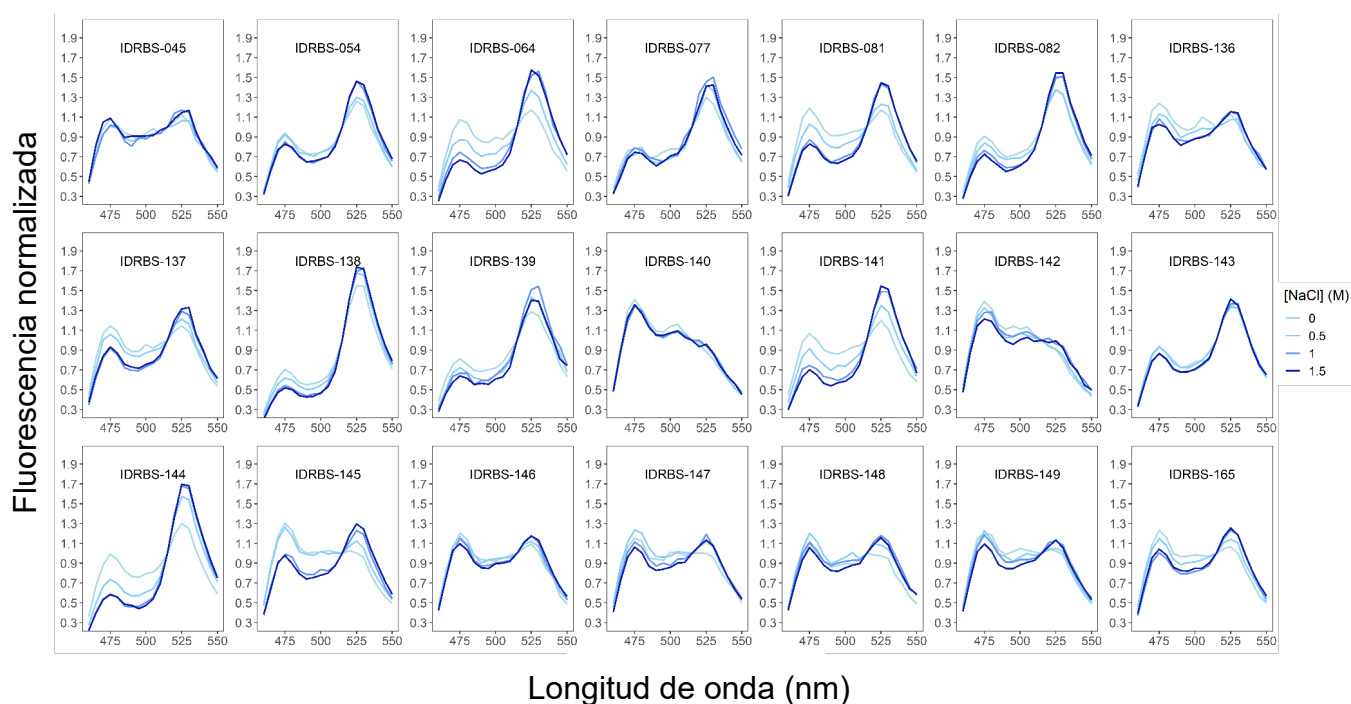
Los ensayos de sensibilidad estructural se llevaron a cabo utilizando la técnica de FRET (Fluorescence Resonance Energy Transfer, por sus siglas en inglés), donde las levaduras transformadas con un plásmido que contenía dos proteínas fluorescentes unidas por una de las 21 IDR se utilizaron como sistema de prueba. (Figura 22). Estos ensayos se realizaron en placas de 96 pocillos, con cada fila que contiene una IDR distinta. A lo largo de la placa, se dispusieron tres columnas para cada concentración de NaCl utilizada. Las concentraciones de NaCl abarcaron un rango desde 0 hasta 1.5 M, con valores intermedios de 0.5 M y 1.0 M (Figura 22).

La validación de los ensayos FRET de sensibilidad estructural se realizó mediante la visualización de espectros de fluorescencia (Figura 22). Estos espectros se normalizaron en función del punto isobéptico del par FRET utilizado, que consiste en las proteínas mCerulean3 como donadora y mCitrine como aceptora, con el punto isobéptico en 515 nm.



**Figura 22.** Diagrama del ensayo FRET empleado en la caracterización de las IDR de los FT. a) Los ensayos fueron realizados en placas de 96 pocillos, y b) los resultados se validaron mediante espectros de fluorescencia desde 450 nm a 550 nm de longitud de onda. Los picos a aumentados a 525 nm son indicativos de la transferencia de energía.

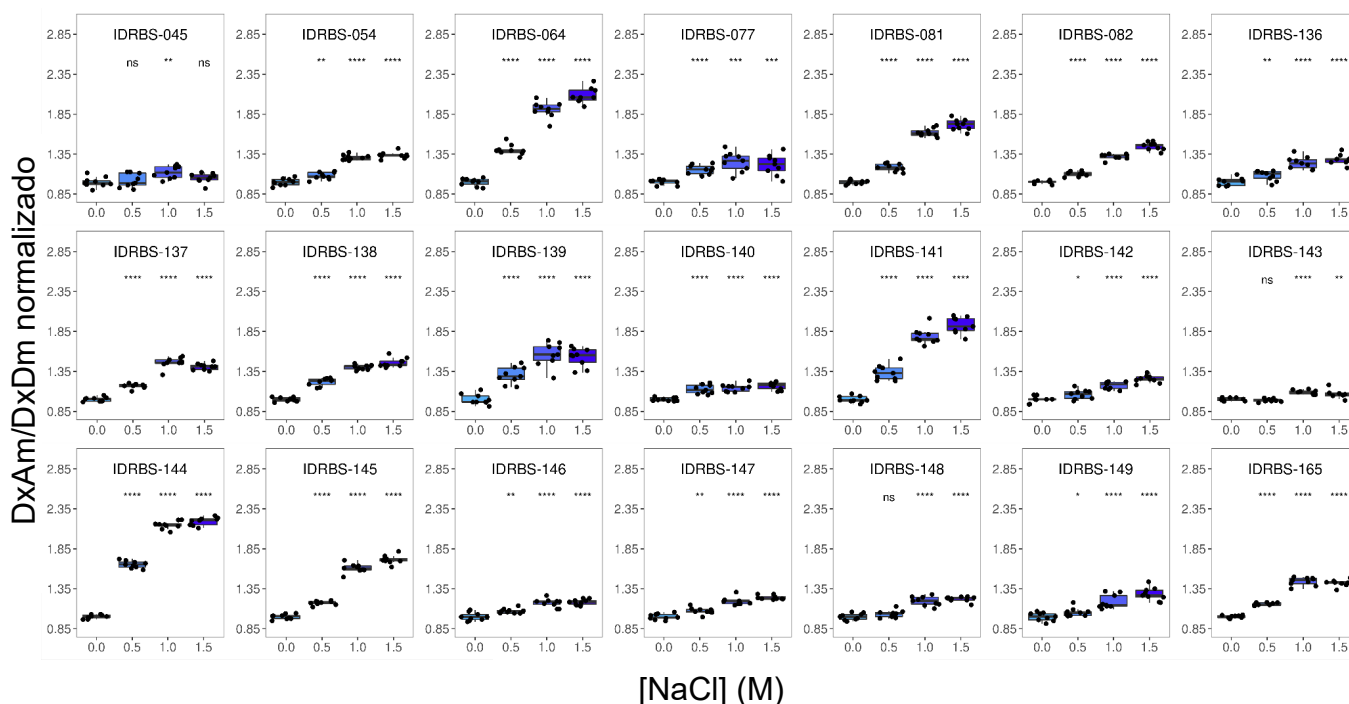
Se observaron una amplia variedad de resultados en los espectros de fluorescencia. Estos espectros revelaron dos picos significativos en cada una de las construcciones. El primer pico, a 475 nm, indica la fluorescencia de emisión de la proteína donadora, mientras que el pico a 525 nm corresponde a la fluorescencia de emisión de la proteína aceptora (**Figura 23**). Algunas construcciones mantuvieron niveles de fluorescencia similares en ambas longitudes de onda, mientras que otras mostraron un aumento significativo a 525 nm. En algunos casos, los niveles de fluorescencia se mantuvieron más altos (**Figura 23**). Estos resultados demostraron la variabilidad en la respuesta de las construcciones a los cambios en las condiciones de osmolaridad.



**Figura. 23.** Espectros de fluorescencia de las 21 IDRs experimentales de los FT. Los picos observados a 475 nm son indicativos de la fluorescencia del donador, mientras que, a 525 nm, indican la fluorescencia del aceptor. El orden de los gráficos corresponde al diagrama mostrado en **Métodos 5.3, Figura 10**.

Con el fin de obtener una comprensión más completa de la sensibilidad estructural, se realizaron visualizaciones gráficas utilizando gráficos de caja (**Figura 24**). Para ello, se recopilaron los valores de fluorescencia a 475 nm y 525 nm de cada una de las 21 IDRs y se calculó el cociente  $525 / 475$ . Estos datos obtenidos se normalizaron en relación con el promedio de los valores de fluorescencia a una concentración de 0 M. Este enfoque proporcionó una representación visual de la

variación en la sensibilidad estructural de las IDRs en respuesta a las condiciones de osmolaridad.



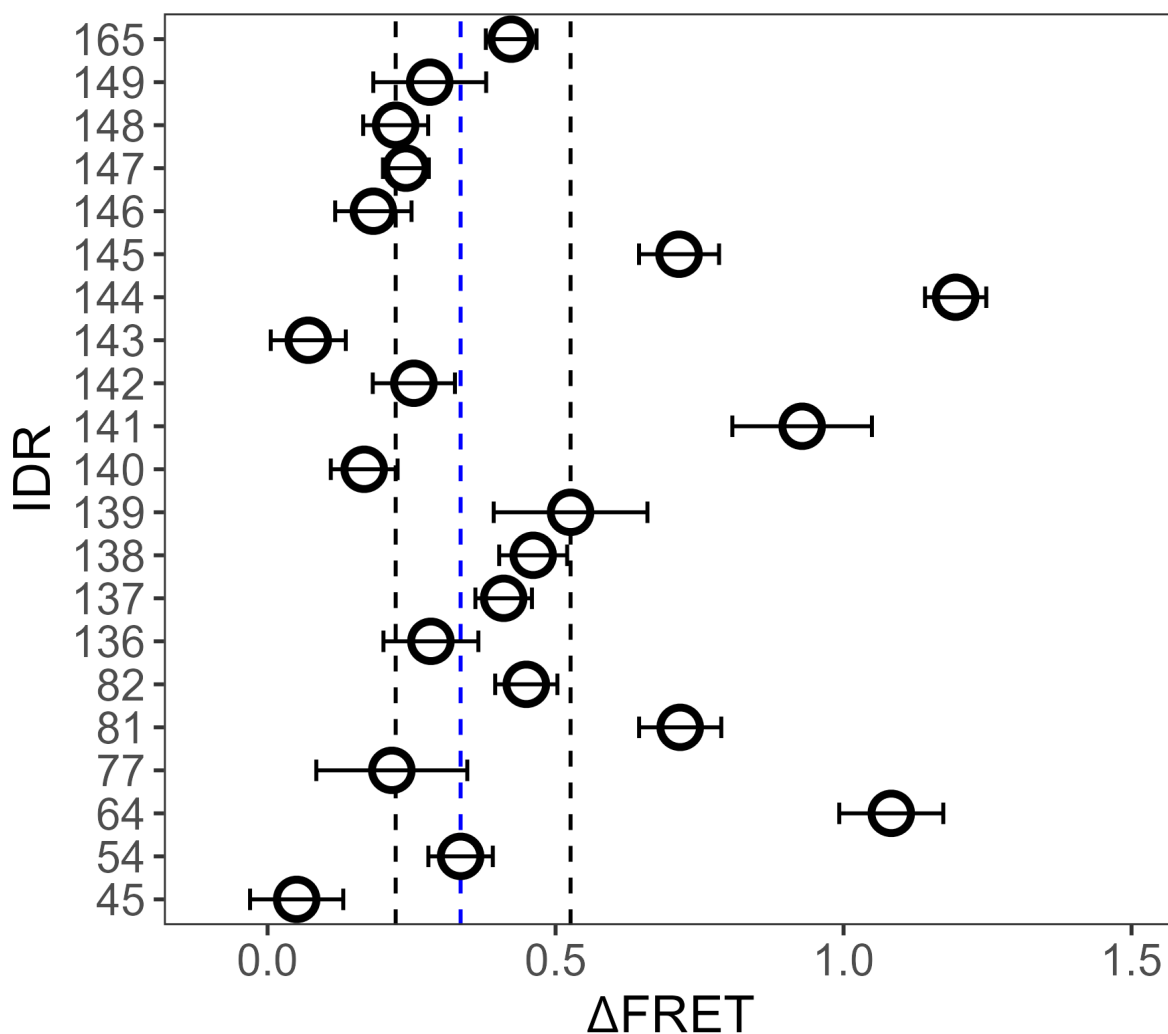
**Figura 24.** Cuantificación del cambio estructural de las IDRs. El nivel de respuesta de cada IDR de los FT se muestra para cada concentración de NaCl utilizada. Cada IDR mostró un nivel de respuesta variable. El orden de los gráficos corresponde al diagrama mostrado en **Métodos 5.3, Figura 10.**

Los valores del cociente FRET ( $DxAm/DxDm$ ) se utilizaron para calcular el valor de Delta FRET. Delta FRET se refiere a la diferencia entre los valores de  $DxAm/DxDm$  a 1.5 M de NaCl y los valores de  $DxAm/DxDm$  a 0 M de NaCl. Esta métrica nos permitió identificar las construcciones que mostraron la mayor sensibilidad estructural (**Figura 25**).

Para identificar de manera efectiva las construcciones con la mayor y menor respuesta de sensibilidad estructural, se calcularon los cuartiles al 75% y 25%. Esto condujo a la identificación de cinco construcciones con la mayor respuesta, designadas IDRBS-144, IDRBS-141, IDRBS-139, IDRBS-138 e IDRBS-064 (**Figura 25**). Asimismo, se identificaron cinco construcciones con la menor respuesta, identificadas como IDRBS-045, IDRBS-140, IDRBS-142, IDRBS-143 e IDRBS-146 (**Figura 25**). Adicionalmente, se detectaron 11 construcciones con una respuesta intermedia, es decir, que se ubicaron en los cuartiles correspondientes. Cabe destacar



que cada construcción contiene una IDR correspondiente a un factor de transcripción (**Tabla 6**). Esta IDR representada en cada construcción es la más larga encontrada en cada uno de los factores.



**Figura 25.** Diagrama de clasificación estadística de las 21 IDRs. En el eje “x” se encuentra el valor de Delta FRET calculado para cada una de las IDRs. El eje “y” representa el identificador de las IDRs. El valor de  $\Delta FRET$  fue calculado restando los valores de fluorescencia de cada IDR a 1.5 M de NaCl menos el valor de fluorescencia a 0 M de NaCl.

**Tabla 6.** Valores de delta FRET de las 21 construcciones

Construcción	Proteína	Familia	$\Delta$ FRET	Desorden (%)	Hélice- $\alpha$ (%)
IDRBS-045	Transcription factor PERIANTHIA	bZIP	0.051	36.73	35.50
IDRBS-054	Transcription factor bHLH106	bHLH	0.335	56.13	27.77
IDRBS-064	Myb family transcription factor PHL6	G2-like	1.083	52.34	31.56
IDRBS-077	F16F4.11 protein	Trihelix	0.216	51.24	24.59
IDRBS-081	Heat stress transcription factor A-4c	HSF	0.716	55.65	27.85
IDRBS-082	Homeobox protein knotted-1-like 7	TALE	0.449	51.20	32.53
IDRBS-136	C2H2 and C2HC zinc fingers superfamily protein	Other	0.284	76.24	25.13
IDRBS-137	Sequence-specific DNA binding transcription	Trihelix	0.410	75.68	25.34

	factor				
IDRBS-138	AGAMOUS-like 19	MIKC_M ADS	0.461	74.43	31.62
IDRBS-139	AtbZIP transcription factor	bZIP	0.526	72.70	33.49
IDRBS-140	Trihelix transcription factor GT-3a	Trihelix	0.168	72.14	76.56
IDRBS-141	Trihelix transcription factor GT-3b	Trihelix	0.928	71.28	45.84
IDRBS-142	Cell division cycle 5-like protein	MYB	0.254	69.43	22.53
IDRBS-143	Transcription factor MYB77	MYB	0.071	67.44	20.88
IDRBS-144	Homeobox-leucine zipper protein HAT1	HD-ZIP	1.195	65.60	21.70
IDRBS-145	Ubiquitin-associated (UBA)/TS-N domain-containing	Other	0.714	64.16	72.35
IDRBS-146	Transcription factor bHLH109	bHLH	0.184	58.38	32.59
IDRBS-147	Transcription	bHLH	0.240	51.67	21.07

	factor bHLH101				
IDRBS-148	Transcription factor TGA3	bZIP	0.222	50.26	19.08
IDRBS-149	Coiled-coil domain- containing protein 16	Other	0.281	60.06	21.60
IDRBS-165	Alcohol dehydrogenas e transcription factor Myb/SANT- like	Trihelix	0.423	56.16	29.43

Para una comprensión más detallada de la respuesta FRET y la sensibilidad, se analizó la distribución de la respuesta en función de las familias de FT por familia. Existen 58 familias de FT en Arabidopsis, con un total de 2296 FTs reportados en diversas bases de datos (Tian, 2020; Jin, 2017; Jin, 2015; Jin, 2014). Cada familia de FT de Arabidopsis incluye varias proteínas con funciones específicas (**Tabla 7**).

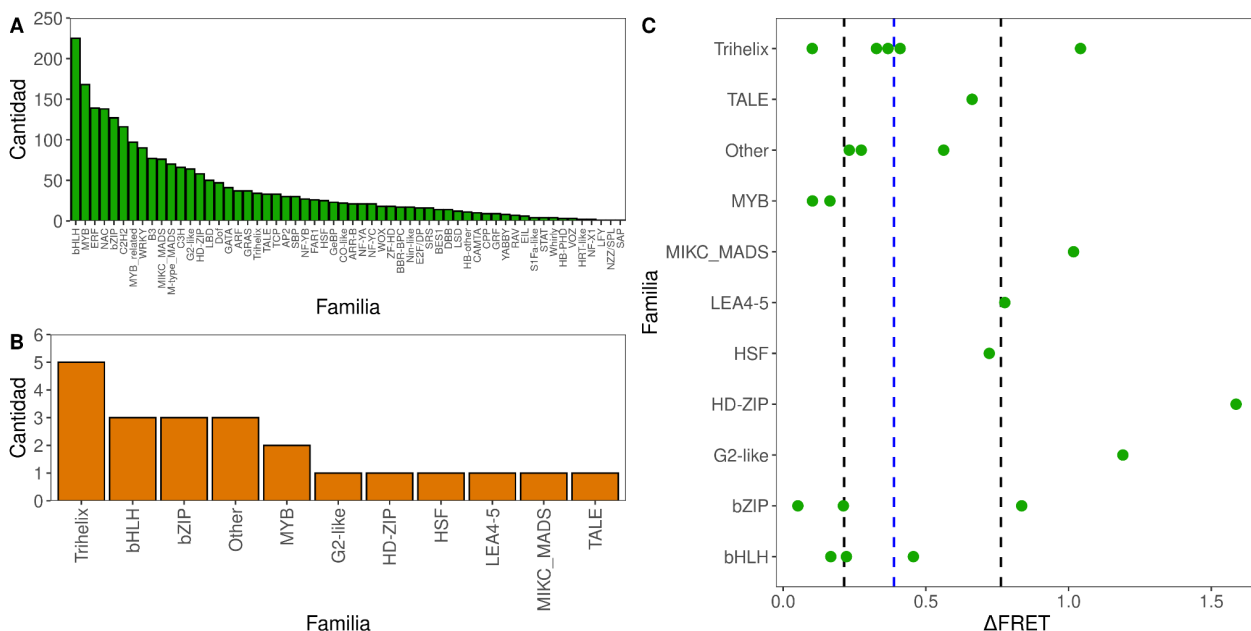
**Tabla 7.** Agrupación de las 58 familias de los FT de Arabidopsis\*

AP2 (30)	ARF (37)	ARR-B (21)	B3 (77)	CO-like (22)
BES1 (14)	C2H2 (116)	C3H (66)	CAMTA (10)	EIL (6)
CPP (9)	DBB (14)	Dof (47)	E2F/DP (16)	GRAS (37)
ERF (139)	FAR1 (26)	G2-like (64)	GATA (41)	HD-ZIP (58)
GRF (9)	GeBP (23)	HB-PHD (3)	HB-other (11)	LSD (12)
HRT-like (2)	HSF (25)	LBD (50)	LFY (1)	NAC (138)
M-type-	MIKC_MADS	MYB (168)	MYB_related	NZZ/SPL (1)

MADS(70)	(76)		(97)	
NF-X1 (2)	NF-YA (21)	NF-YB (27)	NF-YC (21)	SBP (30)
Nin-like (17)	RAV (7)	S1Fa-like (4)	SAP (1)	Trihelix (34)
SRS (16)	STAT (4)	TALE (33)	TCP (33)	YABBY (8)
VOZ (3)	WOX (18)	WRKY (90)	Whirly (4)	
ZF-HD (18)	bHLH (225)	bZIP (127)	BBR-BPC (17)	

\*Los números entre paréntesis indican la cantidad de FTs por cada familia

De todos los FT reportados, la familia más grande de ellos es bHLH, nombrada como basic/helix-loop-helix, con 225 proteínas. La segunda familia más representativa es MYB, con 168 proteínas (**Figura 26A y Figura 26B**). Cabe destacar que, dentro de las 21 IDR s analizadas, estas fueron tomadas de 9 familias diferentes: Trihelix (5); bHLH (3); bZIP (3); MYB (2); G2-like (1), HD-ZIP (1); HSF (1); MICK\_MADS (1); TALE (1) y Otros (3) (**Figura 26C**).

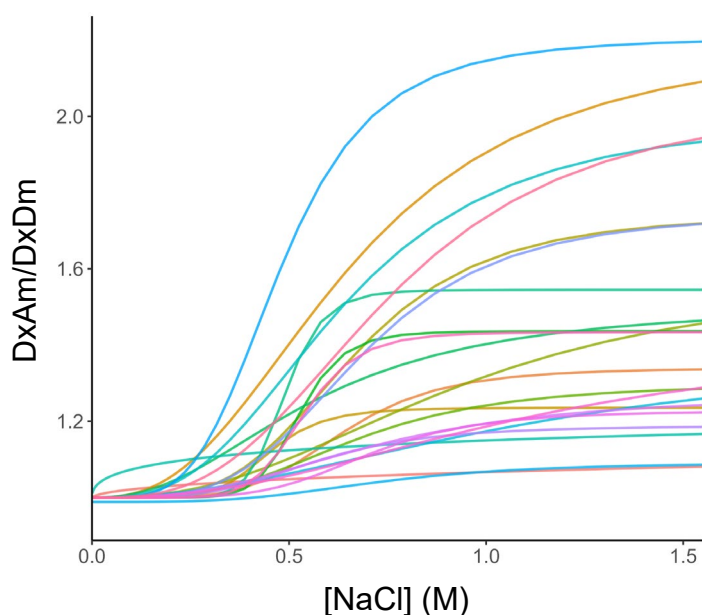


**Figura 26.** Distribución de las familias de los FT de Arabidopsis. **A)** Agrupación de las 58 familias de FT de Arabidopsis, **B)** Familias pertenecientes a las 21 IDRs experimentales. **C)** Respuesta de Delta FRET clasificado por familia de las 21 IDRs experimentales.

Los análisis revelaron una amplia gama de respuestas entre las diferentes familias de FT. Sin embargo, es importante destacar que el número de IDRs analizadas por familia no es representativo debido a que, en algunos casos, solo se evaluó un IDR por familia (**Figura 26C**).

Al considerar solo las familias con al menos dos IDRs evaluadas, se obtuvieron resultados interesantes. Por ejemplo, en la familia Trihelix, se encontró una IDR con baja respuesta y tres con respuestas intermedias. En la familia bHLH, se identificó una IDR con baja respuesta y dos con respuestas intermedias. La familia bZIP incluyó dos IDRs con baja respuesta y una con alta respuesta. En contraste, la familia MYB demostró tener ambas IDRs con baja respuesta. Esta clasificación se basó en los valores de los cuartiles al 25% y 75%, los cuales fueron 0.21 y 0.76, respectivamente. Los valores de Delta FRET asociados a cada IDR se pueden consultar en la **Tabla 6**.

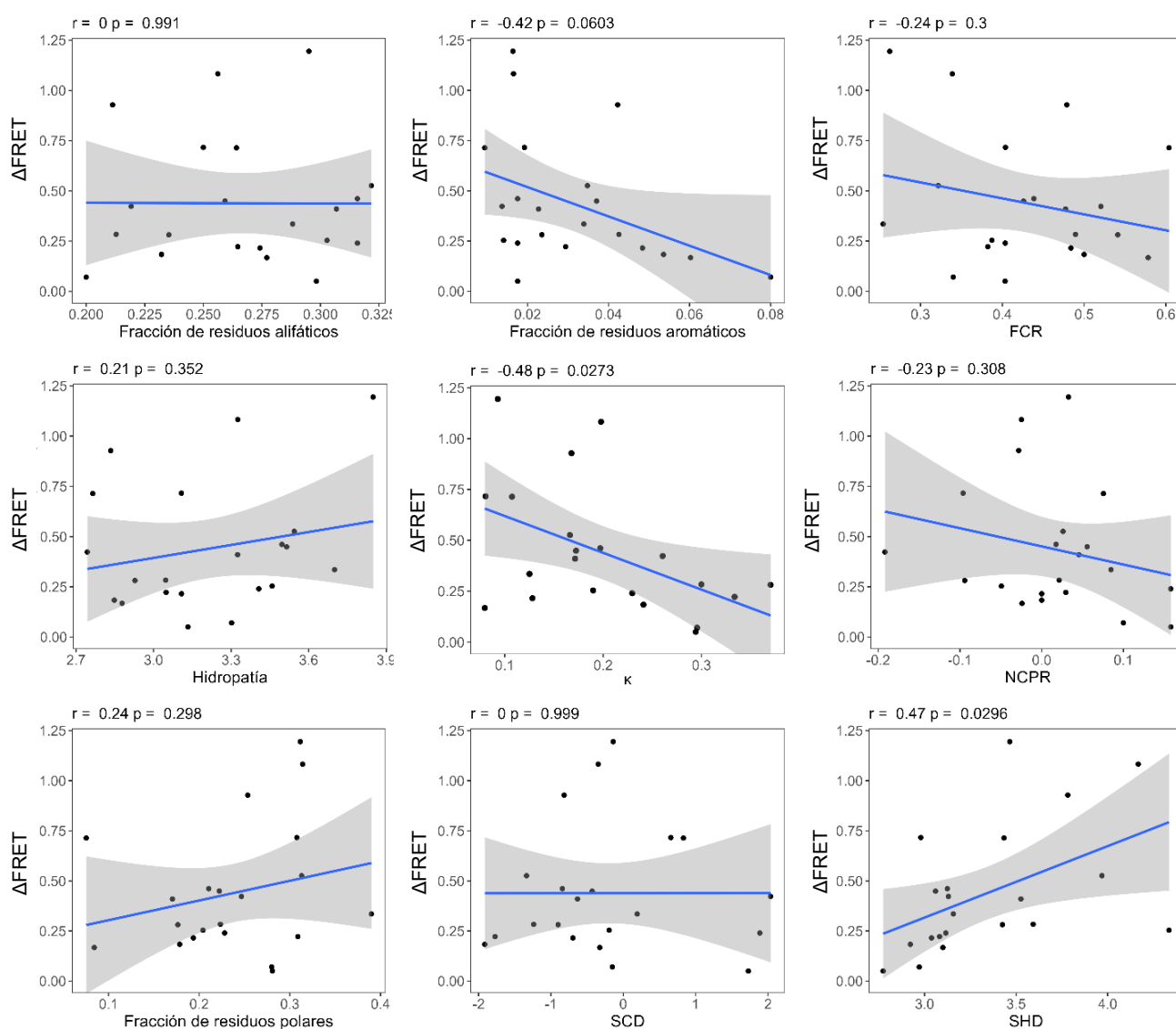
Por otro lado, al examinar la respuesta de las IDRs a los choques hiperosmóticos en células de levadura, se observó una relación de concentración-respuesta variable. Algunas IDRs demostraron ser sensibles incluso a bajas concentraciones de NaCl, mientras que otras, requirieron concentraciones más altas para mostrar su respuesta, o incluso no mostraron una respuesta apreciable (**Figura 27**).



**Figura 27.** Curva-dosis respuesta de las 21 IDRs de FT de Arabidopsis. Cada color representa una IDR. El ajuste de las curvas se realizó mediante un modelo de Hill log-logístico de cuatro parámetros. El color de las curvas representa a cada una de las 21 IDRs estudiadas.

Siendo la IDR de mayor respuesta de FRET la IDRBS-144, representada con el color azul claro (parte superior).

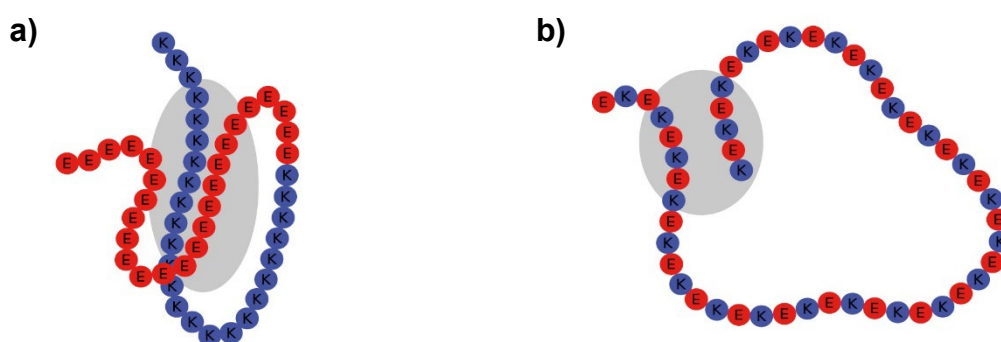
Se utilizaron parámetros fisicoquímicos para definir propiedades de las IDRs con su posible nivel de compactación ( $\Delta$ FRET). Posteriormente, se realizaron correlaciones con diferentes parámetros fisicoquímicos previamente definidos para las IDRs. Estos parámetros incluyen kappa, fracción de residuos cargados (FCR, *Fraction of Charged Residues*), hidropatía, carga neta por residuo (NCPR, *Net Charge Per Residue*), decoración de cargas (SCD, *Sequence Charge Decoration*), decoración de hidropatía (SHD, *Sequence Hidropathy Decoration*) y la fracción de residuos aromáticos, alifáticos y polares (**Figura 28**).



**Figura 28.** Correlaciones entre parámetros fisicoquímicos asociados a las IDRs vs  $\Delta$ FRET. Los parámetros de FCR, NCPR, hidropatía y kappa han sido ampliamente estudiados en las

IDRs, y recientemente se han propuesto parámetros como la decoración de cargas (SCD) y la decoración de hidropatía (SHD) como indicativos del cambio conformacional de las IDRs. Se graficaron los valores de las 21 IDRs de los FT de Arabidopsis; cada punto representa una IDR.

En contraste, en el caso del  $\Delta$ FRET se observaron correlaciones con kappa ( $r = -0.48$ ;  $p = 0.027$ ) y SHD ( $r = 0.47$ ;  $p = 0.03$ ) (**Figura 28**). Estos hallazgos proporcionan información valiosa sobre las posibles relaciones entre las propiedades fisicoquímicas de las IDRs y su respuesta en los ensayos FRET. La correlación con el parámetro de kappa indica que aquellas IDRs con una distribución de cargas menos homogénea a lo largo de la secuencia, provoca un aumento en la respuesta de FRET. Esta información, podría atribuirse a las fuerzas de Coulomb entre los residuos cargados de la secuencia de una proteína desordenada y su distribución a lo largo de la misma (**Figura 29**). En la **Figura 29a**, la secuencia mostrada tiene un kappa más alto, mientras que la secuencia en la **Figura 29b**, posee un kappa más bajo. Las IDRs de los FT estudiados en este trabajo muestran que un kappa más bajo provoca un mayor cambio conformacional (**Figura 28**; **Figura 27**).



**Figura 29.** Efectos electrostáticos relacionados en la compactación de las proteínas desordenadas. **a)** Secuencia con un valor de kappa alto, indicando una distribución de cargas positivas o negativas hacia los extremos N-terminal y C-terminal. La cadena de aminoácidos es más compacta debido a la fuerte interacción intracadena (región gris) entre los residuos positivos y negativos. **b)** Secuencia con un valor de kappa bajo, indicando una distribución homogénea de cargas positivas y negativas a lo largo de la secuencia. La fuerza de atracción electrostática es menor (región gris) debido a la cantidad de bloques de cargas en la parte media de la secuencia, ocasionando una mayor repulsión entre los residuos de aminoácido. Modificado de Lin *et al.*, 2017.



## 7. DISCUSIÓN

Nuestro estudio contribuye al conocimiento de las características de las IDRs en los FT de *Arabidopsis*. Previa investigación sugieren que estas poseen niveles significativos de desorden en sus secuencias (Salladini *et al.*, 2020). Este desorden alto en los FT juega un papel esencial en la biología de las plantas al permitirles interactuar con diversas proteínas y crear redes de regulación que son cruciales para la transducción de señales y, por lo tanto, para el desarrollo de las plantas (Millard, Kragelund y Burow, 2019) favoreciendo el desarrollo de estas.

Nuestros resultados también revelan que las IDRs más largas en los FT de *Arabidopsis*, exhiben una variabilidad en su sensibilidad estructural. A pesar de esta variabilidad, es importante destacar, que las respuestas al cambio conformacional no son homogéneas para las IDRs de los FT, ya que las proteínas desordenadas pueden adoptar múltiples estados conformacionales, dependiendo del entorno químico circundante (Tompa, 2005; Tompa, 2012; Dunker, 2001; Moses, 2020). En este estudio, hemos demostrado que, las IDRs de los FT son sensibles a los cambios en la osmolaridad, especialmente ante choques hiperosmóticos, con algunas IDRs respondiendo incluso a concentraciones de NaCl menores a 0.5 M.

La metodología FRET utilizada en este trabajo permitió la identificación de IDRs con diferentes niveles de sensibilidad estructural: alta, media y baja. Es importante señalar que, aunque comprobamos la presencia del fenómeno de FRET a través de los espectros de fluorescencia, es esencial realizar una validación adicional utilizando las proteínas fluorescentes individuales para eliminar posibles señales de ruido. Esta validación es crucial para determinar la eficiencia FRET, lo que a su vez es fundamental para evaluar la magnitud de los ensayos FRET (Algar *et al.*, 2019).

Los FT desempeñan un papel vital en la regulación de procesos biológicos en plantas, especialmente en respuesta a condiciones adversas. Dado que se sabe que estos FT poseen niveles significativos de desorden, las IDRs de los FT pueden jugar un papel clave en la respuesta y desarrollo de las plantas (Salladini, 2020). En consecuencia, nuestras observaciones respaldan la hipótesis de que las IDRs en los FT de *Arabidopsis* podrían desempeñar un papel crucial en la regulación de procesos biológicos ante condiciones de estrés.

Además, los resultados obtenidos en este estudio proporcionan una base para investigar las propiedades que influyen en la sensibilidad de las IDR's en diferentes entornos. Dado que las secuencias de las IDR's son poco conservadas y presentan niveles altos de divergencia (Zarin *et al.*, 2019), la determinación de patrones específicos puede ser un enfoque prometedor. Las IDR's con alta sensibilidad estructural podrían exhibir patrones característicos en sus propiedades fisicoquímicas, lo mismo aplicado a las IDR's con sensibilidad media y baja. Esta investigación puede llevar a la identificación de características que influyen en la sensibilidad estructural de las IDR's en respuesta a diferentes entornos *in vivo*.

La distribución de cargas en las IDR's de los FT de Arabidopsis, también puede desempeñar un papel importante en la sensibilidad estructural. La formación de condensados biomoleculares, influenciados por patrones de carga específicos dentro de las IDR's, se ha observado en FT humanos (Lyons *et al.*, 2023). Por lo tanto, es fundamental explorar patrones de carga similares en las IDR's con alta sensibilidad y examinar cómo se relacionan estas cargas con el entorno circundante. En este sentido, la distribución de cargas puede ser un factor determinante en la sensibilidad estructural de las IDR's y debe investigarse en relación con otros parámetros fisicoquímicos para comprender su función y comportamiento.

La determinación de patrones fisicoquímicos que les permiten a las IDR's responder ante cambios en el ambiente permitiría diseñar IDR's sintéticas de novo con propiedades únicas, capaces de percibir cualquier ambiente químico. Para ello, es importante determinar de forma precisa el ensamble conformacional de las IDR's ante condiciones ambientales específicas. Actualmente, es posible predecir de forma precisa la conformación que adquiere una IDR basada en propiedades fisicoquímicas como la distribución de cargas negativas y positivas (Pesce *et al.*, 2023), inclusive, diseñar cientos o miles de secuencias desordenadas completamente sintéticas (Pesce *et al.*, 2023; Lotthammer *et al.*, 2023) con una conformación única, ya que se ha observado que, aquellas IDR's con una distribución de cargas homogénea a lo largo de la secuencia, poseen una conformación más compacta (Pesce *et al.*, 2023), por lo que, diseñar una IDR que actúe como un biosensor y que se compacte ante un cambio en el ambiente químico, sería posible conociendo sus propiedades fisicoquímicas.

Los ensayos FRET han emergido como herramientas fundamentales para el estudio y caracterización de diversas moléculas, incluidas las proteínas. El FRET es un fenómeno electromagnético que implica la transferencia de energía no radiante entre proteínas fluorescentes donadoras yceptoras en función de la proximidad de los fluoróforos (Algar, 2019; Ma *et al.*, 2014). Aunque nuestras observaciones experimentales podrían deberse a un FRET intramolecular, la presencia de IDRs podría dar lugar a fenómenos de FRET intermolecular. Dado que los FT a menudo interactúan con proteínas correguladoras y tienen alta afinidad por secuencias similares (Frieze & Farnham, 2011; Lambert *et al.*, 2018) es probable que ocurra un fenómeno de FRET intermolecular, lo que podría influir en nuestras observaciones sobre la compactación de las IDRs en el sistema del biosensor.

En los análisis de un biosensor FRET, es crucial abordar considerar no solo las interacciones generadas por fenómenos como FRET intermolecular, sino también corregir la fluorescencia de fondo, la autofluorescencia y el fotoblanqueo, factores que pueden afectar las mediciones. Estas correcciones son esenciales para optimizar a los biosensores y garantizar una caracterización molecular precisa (Gordon *et al.*, 1998; Algar, 2019).

Más allá de los desafíos asociados con el biosensor, las respuestas adaptativas de las células de *Sacharomyces cerevisiae* pueden influir en las observaciones de FRET. El pH intracelular, crucial para las funciones celulares normales, varía en respuesta a cambios osmóticos. La activación de la vía HOG, que ocurre en minutos, incluye en la respuesta observada y estopa vinculada a cambios en el pH intracelular (Saito & Posas, 2012).

El aumento en la osmolaridad externa provoca la salida de agua de la célula, lo que resulta en un aumento en la concentración de iones, especialmente sodio ( $\text{Na}^+$ ) (Saito & Posas, 2012), y un mayor hacinamiento macromolecular (Miermont *et al.*, 2013). Este fenómeno puede afectar la señal de fluorescencia observada, ya que los biosensores genéticamente codificados podrían empaquetarse y generar FRET intermolecular. La respuesta de FRET puede depender del tiempo de exposición al estrés osmótico, ya que un choque severo puede causar una mayor compresión de la célula y retardar la respuesta de la vía HOG (Miermont *et al.*, 2012).

Además de los factores internos inducidos por el estrés osmótico, la cantidad de moléculas de biosensor y su expresión controlada a través de promotores regulables son aspectos que considerar. Aunque no podemos determinar la cantidad durante los ensayos FRET, el control de la expresión es posible mediante promotores GAL (Peng *et al.*, 2018).

Nuestros resultados proponen una metodología para determinar los posibles cambios conformacionales de las IDRs y su uso como biosensores para monitorear el ambiente intracelular ante choques hiperosmóticos. También podría esclarecer el papel de las IDRs en los FT de Arabidopsis ante estas condiciones de estrés.

Sugerimos que, al realizar los ensayos FRET, es necesario considerar las posibles interacciones intracelulares, el sistema de fluoróforos y las respuestas adaptativas del modelo celular. Quedan cuestiones por resolver, y otra aproximación podría ser estudiar a las proteínas completas de los FT para comprender mejor el papel de las IDRs en las respuestas transcripcionales de Arabidopsis.

## 8. CONCLUSIONES

- Los FT de la planta modelo Arabidopsis son proteínas que presentan mayores niveles de desorden comparado con su proteoma.
- En su mayoría, los FT de Arabidopsis poseen un promedio de 6 regiones desordenadas por proteína.
- La cantidad promedio de residuos de aminoácidos que poseen las IDRs de los FT respecto a su proteoma es mayor, con una prevalencia mayor de aminoácidos como glicina, alanina, metionina, serina, lisina, prolina, arginina, ácido glutámico y glutamina, considerados promotores del desorden.
- Las IDRs más largas de los FT de Arabidopsis, poseen una diversidad de grados de sensibilidad estructural en condiciones de choque hiperosmótico.
- La respuesta de sensibilidad estructural de las IDRs está influenciado por el patrón específico de cargas en combinación con otras propiedades fisicoquímicas relevantes.
- Los altos niveles de desorden de los FT de Arabidopsis y su posible mecanismo de sensibilidad estructural permite proponer a las IDRs de los FT, como un caso de estudio para buscar alguna función asociada con la sensibilidad estructural de las IDRs. Nuestros hallazgos proporcionan información valiosa sobre las IDRs en los FT de Arabidopsis y sus sensibilidad estructural ante condiciones de estrés hiperosmótico.
- El estudio de la sensibilidad estructural en correlación con la distribución de cargas de las IDRs de los FT, permiten determinar los parámetros fisicoquímicos idóneos para identificar y diseñar IDRs sintéticas con un alto grado de sensibilidad estructural ante condiciones ambientales adversas.

## 8.1. PERSPECTIVAS

Los resultados obtenidos en esta tesis sugieren que la sensibilidad estructural de las IDRs presentes en los FT de *Arabidopsis* podría estar influenciada por la combinación de parámetros fisicoquímicos que les permitan responder ante cambios bruscos en el ambiente osmótico, y para ello, será necesario generar IDRs sintéticas provenientes de FT en las que se modifiquen las propiedades fisicoquímicas de las secuencias.

La técnica de FRET fue exitosamente utilizada para estudiar la sensibilidad estructural de las IDRs de los FT de *Arabidopsis*, por lo que se plantea aumentar la validez de las mediciones mediante la obtención de la eficiencia FRET, lo que podría ayudar a descartar la presencia de fenómenos de inter-FRET entre dos proteínas fluorescentes de dos biosensores distintos.

En esta tesis se caracterizaron 21 IDRs de 21 FT de la planta modelo *Arabidopsis thaliana in vivo*, pero aún permanece pendiente el asociar esta sensibilidad estructural con la función específica de cada factor de transcripción *in vivo*. Para ello, se puede utilizar el sistema de expresión transitoria en hojas de *Nicotiana benthamiana* empleando la proteína completa del factor de transcripción y realizar modificaciones en la secuencia de la IDR de la proteína para detectar la expresión de un gen reportero en función de la sensibilidad estructural asociada a la IDR del factor de transcripción.

## 9. REFERENCIAS

1. Adamski, W., Salvi, N., Maurin, D., Magnat, J., Milles, S., Jensen, M. R., Abyzov, A., Moreau, C. J., & Blackledge, M. (2019). A Unified Description of Intrinsically Disordered Protein Dynamics under Physiological Conditions Using NMR Spectroscopy. *Journal of the American Chemical Society*, *141*(44), 17817–17829. <https://doi.org/10.1021/JACS.9B09002>
2. Agarwal, P., Baranwal, V. K., & Khurana, P. (2019). Genome-wide Analysis of bZIP Transcription Factors in wheat and Functional Characterization of a TabZIP under Abiotic Stress. *Scientific Reports 2019 9:1*, *9*(1), 1–18. <https://doi.org/10.1038/s41598-019-40659-7>
3. Alderson, T. R., Pritišanac, I., Moses, A. M., & Forman-Kay, J. D. (2022). Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *BioRxiv*, 2022.02.18.481080. <https://doi.org/10.1101/2022.02.18.481080>
4. Algar, W. R., Hildebrandt, N., Vogel, S. S., & Medintz, I. L. (2019). FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls. *Nature Methods 2019 16:9*, *16*(9), 815–829. <https://doi.org/10.1038/s41592-019-0530-8>
5. Arai, M., Sugase, K., Dyson, H. J., & Wright, P. E. (2015). Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(31), 9614–9619. <https://doi.org/10.1073/pnas.1512799112>
6. Bauer, N. C., Doetsch, P. W., & Corbett, A. H. (2015). Mechanisms Regulating Protein Localization. *Traffic (Copenhagen, Denmark)*, *16*(10), 1039–1061. <https://doi.org/10.1111/TRA.12310>
7. Basu, S., Kihara, D., & Kurgan, L. (2023). Computational prediction of disordered binding regions. *Computational and Structural Biotechnology Journal*, *21*, 1487–1497. <https://doi.org/10.1016/J.CSBJ.2023.02.018>
8. Bernadó, P., & Svergun, D. I. (2012). Analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Methods in Molecular Biology (Clifton, N.J.)*, *896*, 107–122. [https://doi.org/10.1007/978-1-4614-3704-8\\_7](https://doi.org/10.1007/978-1-4614-3704-8_7)
9. Bondos, S. E., Dunker, A. K., & Uversky, V. N. (2021). On the roles of intrinsically disordered proteins and regions in cell communication and

- signaling. In *Cell Communication and Signaling* (Vol. 19, Issue 1, pp. 1–9). BioMed Central Ltd. <https://doi.org/10.1186/s12964-021-00774-3>
10. Broussard, J. A., & Green, K. J. (2017). Research Techniques Made Simple: Methodology and Applications of Förster Resonance Energy Transfer (FRET) Microscopy. *Journal of Investigative Dermatology*, *137*(11), e185–e191. <https://doi.org/10.1016/J.JID.2017.09.006>
  11. Chong, P. A., Vernon, R. M., & Forman-Kay, J. D. (2018). RGG/RG Motif Regions in RNA Binding and Phase Separation. *Journal of Molecular Biology*, *430*(23), 4650–4665. <https://doi.org/10.1016/J.JMB.2018.06.014>
  12. 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>
  13. Cuevas-Velazquez, C. L., Velloso, T., Guadalupe, K., Schmidt, H. B., Yu, F., Moses, D., Brophy, J. A. N., Cosio-Acosta, D., Das, A., Wang, L., Jones, A. M., Covarrubias, A. A., Sukenik, S., & Dinneny, J. R. (2021). Intrinsically disordered protein biosensor tracks the physical-chemical effects of osmotic stress on cells. *Nature Communications* 2021 12:1, *12*(1), 1–12. <https://doi.org/10.1038/s41467-021-25736-8>
  14. Darling, A. L., & Uversky, V. N. (2017). Intrinsic Disorder in Proteins with Pathogenic Repeat Expansions. *Molecules* 2017, Vol. 22, Page 2027, *22*(12), 2027. <https://doi.org/10.3390/MOLECULES22122027>
  15. Das, R. K., & Pappu, R. v. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(33), 13392–13397. [https://doi.org/10.1073/PNAS.1304749110/SUPPL\\_FILE/SAPP.PDF](https://doi.org/10.1073/PNAS.1304749110/SUPPL_FILE/SAPP.PDF)
  16. Das, R. K., Crick, S. L., & Pappu, R. v. (2012). N-Terminal Segments Modulate the  $\alpha$ -Helical Propensities of the Intrinsically Disordered Basic Regions of bZIP Proteins. *Journal of Molecular Biology*, *416*(2), 287–299. <https://doi.org/10.1016/J.JMB.2011.12.043>
  17. Das, R. K., Huang, Y., Phillips, A. H., Kriwacki, R. W., & Pappu, R. v. (2016). Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. *Proceedings of the National Academy of Sciences of the*



- United States of America*, 113(20), 5616–5621.  
<https://doi.org/10.1073/PNAS.1516277113>
18. Das, R. K., Ruff, K. M., & Pappu, R. v. (2015). Relating sequence encoded information to form and function of intrinsically disordered proteins. *Current Opinion in Structural Biology*, 32, 102–112.  
<https://doi.org/10.1016/J.SBI.2015.03.008>
19. 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)
20. Emenecker, R. J., Griffith, D., & Holehouse, A. S. (2021). Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophysical Journal*, 120(20), 4312–4319.  
<https://doi.org/10.1016/J.BPJ.2021.08.039>
21. Emenecker, R. J., Griffith, D., & Holehouse, A. S. (2022). Metapredict V2: An update to metapredict, a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *BioRxiv*, 2022.06.06.494887.  
<https://doi.org/10.1101/2022.06.06.494887>
22. Fang, Y., Xie, K., Hou, X., Hu, H., & Xiong, L. (2010). Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses. *Molecular Genetics and Genomics : MGG*, 283(2), 157–169.  
<https://doi.org/10.1007/S00438-009-0507-X>
23. Ferreon, C., Chris, A., Ferreon, M., Trivedi, R., & Nagarajaram, H. A. (2022). Intrinsically Disordered Proteins: An Overview. *International Journal of Molecular Sciences 2022, Vol. 23, Page 14050*, 23(22), 14050.  
<https://doi.org/10.3390/IJMS232214050>
24. Firman, T., & Ghosh, K. (2018). Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *The Journal of Chemical Physics*, 148(12). <https://doi.org/10.1063/1.5005821>
25. Fietze, S., & Farnham, P. J. (2011). Transcription Factor Effector Domains. *Sub-Cellular Biochemistry*, 52, 261. [https://doi.org/10.1007/978-90-481-9069-0\\_12](https://doi.org/10.1007/978-90-481-9069-0_12)

26. Gomes, G. N. W., Krzeminski, M., Namini, A., Martin, E. W., Mittag, T., Head-Gordon, T., Forman-Kay, J. D., & Gradinaru, C. C. (2020). Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET. *Journal of the American Chemical Society*, *142*(37), 15697–15710. <https://doi.org/10.1021/JACS.0C02088>
27. Gordon, G. W., Berry, G., Liang, X. H., Levine, B., & Herman, B. (1998). Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy. *Biophysical Journal*, *74*(5), 2702–2713. [https://doi.org/10.1016/S0006-3495\(98\)77976-7](https://doi.org/10.1016/S0006-3495(98)77976-7)
28. Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, *1*(6), 2876–2890. <https://doi.org/10.1038/NPROT.2006.202>
29. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G., & Pappu, R. v. (2017). CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophysical Journal*, *112*(1), 16–21. <https://doi.org/10.1016/j.bpj.2016.11.3200>
30. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G., & Pappu, R. v. (2017). CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophysical Journal*, *112*(1), 16–21. <https://doi.org/10.1016/j.bpj.2016.11.3200>
31. 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>
32. Hsiao, A. S. (2022). Plant Protein Disorder: Spatial Regulation, Broad Specificity, Switch of Signaling and Physiological Status. *Frontiers in Plant Science*, *13*, 904446. <https://doi.org/10.3389/FPLS.2022.904446>
33. Hummel, N. F. C., Zhou, A., Li, B., Markel, K., Ornelas, I. J., & Shih, P. M. (2022). The trans-regulatory landscape of gene networks in plants. *BioRxiv*, 2022.10.23.513368. <https://doi.org/10.1101/2022.10.23.513368>
34. Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J., & Gao, G. (2015). An Arabidopsis Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors. *Molecular Biology and Evolution*, *32*(7), 1767–1773. <https://doi.org/10.1093/MOLBEV/MSV058>

35. Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, *45*(D1), D1040–D1045. <https://doi.org/10.1093/NAR/GKW982>
36. Jin, J., Zhang, H., Kong, L., Gao, G., & Luo, J. (2014). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research*, *42*(D1), D1182–D1187. <https://doi.org/10.1093/NAR/GKT1016>
37. Jorda, J., Xue, B., Uversky, V. N., & Kajava, A. v. (2010). Protein tandem repeats – the more perfect, the less structured. *The FEBS Journal*, *277*(12), 2673–2682. <https://doi.org/10.1111/J.1742-4658.2010.07684.X>
38. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* *2021* *596*:7873, *596*(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
39. Kikhney, A. G., & Svergun, D. I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*, *589*(19), 2570–2577. <https://doi.org/10.1016/J.FEBSLET.2015.08.027>
40. Konrat, R. (2014). NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance*, *241*(1), 74–85. <https://doi.org/10.1016/J.JMR.2013.11.011>
41. Kuravi, S., Lan, T. H., Barik, A., & Lambert, N. A. (2010). Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical Journal*, *98*(10), 2383–2390. <https://doi.org/10.1016/J.BPJ.2010.02.006>
42. Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650–665. <https://doi.org/10.1016/J.CELL.2018.01.029/ATTACHMENT/EDE37821-FD6F-41B7-9A0E-9D5410855AE6/MMC2.XLSX>
43. Lermyte, F. (2020). Roles, Characteristics, and Analysis of Intrinsically Disordered Proteins: A Minireview. *Life*, *10*(12), 1–19. <https://doi.org/10.3390/LIFE10120320>

44. Li, Y., Yang, Z., Zhang, Y., Guo, J., Liu, L., Wang, C., Wang, B., & Han, G. (2022). The roles of HD-ZIP proteins in plant abiotic stress tolerance. *Frontiers in Plant Science*, *13*. <https://doi.org/10.3389/FPLS.2022.1027071>
45. Lin, Y. H., & Chan, H. S. (2017). Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated. *Biophysical Journal*, *112*(10), 2043–2046. <https://doi.org/10.1016/J.BPJ.2017.04.021>
46. Lotthammer, J. M., Ginell, G. M., Griffith, D., Emenecker, R. J., & Holehouse, A. S. (2023). Direct Prediction of Intrinsically Disordered Protein Conformational Properties From Sequence. *BioRxiv*, 2023.05.08.539824. <https://doi.org/10.1101/2023.05.08.539824>
47. Lyons, H., Veetil, R. T., Pradhan, P., Fornero, C., de La Cruz, N., Ito, K., Eppert, M., Roeder, R. G., & Sabari, B. R. (2023). Functional partitioning of transcriptional regulators by patterned charge blocks. *Cell*, *186*(2), 327–345.e28. <https://doi.org/10.1016/J.CELL.2022.12.013>
48. Ma, L., Yang, F., & Zheng, J. (2014). Application of fluorescence resonance energy transfer in protein studies. *Journal of Molecular Structure*, *1077*, 87. <https://doi.org/10.1016/J.MOLSTRUC.2013.12.071>
49. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., & Pappu, R. v. (2010). Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(18), 8183–8188. <https://doi.org/10.1073/PNAS.0911107107>
50. Már, M., Nitsenko, K., & Heidarsson, P. O. (2023). Multifunctional Intrinsically Disordered Regions in Transcription Factors. *Chemistry – A European Journal*, *29*(21), e202203369. <https://doi.org/10.1002/CHEM.202203369>
51. Metskas, L. A., & Rhoades, E. (2020). Single-Molecule FRET of Intrinsically Disordered Proteins. *Annual Review of Physical Chemistry*, *71*, 391–414. <https://doi.org/10.1146/ANNUREV-PHYSCHEM-012420-104917>
52. Miermont, A., Waharte, F., Hu, S., McClean, M. N., Bottani, S., Léon, S., & Hersen, P. (2013). Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(14), 5725–5730. [https://doi.org/10.1073/PNAS.1215367110/SUPPL\\_FILE/SM08.WMV](https://doi.org/10.1073/PNAS.1215367110/SUPPL_FILE/SM08.WMV)

53. Miles, A. J., Drew, E. D., & Wallace, B. A. (2023). DichroIDP: a method for analyses of intrinsically disordered proteins using circular dichroism spectroscopy. *Communications Biology* 2023 6:1, 6(1), 1–8. <https://doi.org/10.1038/s42003-023-05178-2>
54. Millard, P. S., Kragelund, B. B., & Burow, M. (2019). R2R3 MYB Transcription Factors – Functions outside the DNA-Binding Domain. *Trends in Plant Science*, 24(10), 934–946. <https://doi.org/10.1016/j.tplants.2019.07.003>
55. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods* 2022 19:6, 19(6), 679–682. <https://doi.org/10.1038/s41592-022-01488-1>
56. Moses, D., Yu, F., Ginell, G. M., Shamoan, N. M., Koenig, P. S., Holehouse, A. S., & Sukenik, S. (2020). Revealing the Hidden Sensitivity of Intrinsically Disordered Proteins to their Chemical Environment. *Journal of Physical Chemistry Letters*, 11(23), 10131–10136. <https://doi.org/10.1021/acs.jpcclett.0c02822>
57. Moses, D., Yu, F., Ginell, G. M., Shamoan, N. M., Koenig, P. S., Holehouse, A. S., & Sukenik, S. (2020). Revealing the Hidden Sensitivity of Intrinsically Disordered Proteins to their Chemical Environment. *Journal of Physical Chemistry Letters*, 11(23), 10131–10136. <https://doi.org/10.1021/acs.jpcclett.0c02822>
58. Müller-Späth, S., Soranno, A., Hirschfeld, V., Hofmann, H., Rügger, S., Reymond, L., Nettels, D., & Schuler, B. (2010). Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), 14609–14614. <https://doi.org/10.1073/PNAS.1001743107>
59. Musselman, C. A., & Kutateladze, T. G. (2021). Characterization of functional disordered regions within chromatin-associated proteins. *iScience*, 24(2). <https://doi.org/10.1016/J.ISCI.2021.102070>
60. Naudi-Fabra S, Tengo M, Jensen MR, Blackledge M, Milles S. Quantitative Description of Intrinsically Disordered Proteins Using Single-Molecule FRET, NMR, and SAXS. *J Am Chem Soc*. 2021 Dec 8;143(48):20109-20121. doi: 10.1021/jacs.1c06264. Epub 2021 Nov 24. PMID: 34817999; PMCID: PMC8662727.

61. Nwanochie, E., & Uversky, V. N. (2019). Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) is the Limit. *International Journal of Molecular Sciences* 2019, Vol. 20, Page 4186, 20(17), 4186. <https://doi.org/10.3390/IJMS20174186>
62. Pace, C. N., & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal*, 75(1), 422–427. [https://doi.org/10.1016/S0006-3495\(98\)77529-0](https://doi.org/10.1016/S0006-3495(98)77529-0)
63. Peng, B., Wood, R. J., Nielsen, L. K., & Vickers, C. E. (2018). An Expanded Heterologous GAL Promoter Collection for Diauxie-Inducible Expression in *Saccharomyces cerevisiae*. *ACS Synthetic Biology*, 7(2), 748–751. <https://doi.org/10.1021/ACSSYNBIO.7B00355>
64. Pesce, F., Bremer, A., Tesei, G., Hopkins, J. B., Grace, C. R., Mittag, T., & Lindorff-Larsen, K. (2023). Design of intrinsically disordered protein variants with diverse structural properties. *BioRxiv*, 2023.10.22.563461. <https://doi.org/10.1101/2023.10.22.563461>
65. 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>
66. Saito, H., & Posas, F. (2012). Response to Hyperosmotic Stress. *Genetics*, 192(2), 289–318. <https://doi.org/10.1534/GENETICS.112.140863>
67. 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>
68. Sherry, K. P., Das, R. K., Pappu, R. v., & Barrick, D. (2017). Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 114(44), E9243–E9252. <https://doi.org/10.1073/PNAS.1706083114/-/DCSUPPLEMENTAL>
69. Sornaraj, P., Luang, S., Lopato, S., & Hrmova, M. (2016). Basic leucine zipper (bZIP) transcription factors involved in abiotic stresses: A molecular model of a wheat bZIP factor and implications of its structure in function. *Biochimica et*

- Biophysica Acta (BBA) - General Subjects*, 1860(1), 46–56.  
<https://doi.org/10.1016/J.BBAGEN.2015.10.014>
70. Strader, L., Weijers, D., & Wagner, D. (2022). Plant transcription factors - being in the right place with the right company. *Current Opinion in Plant Biology*, 65. <https://doi.org/10.1016/J.PBI.2021.102136>
71. Tang, Y., Bao, X., Wang, S., Liu, Y., Tan, J., Yang, M., Zhang, M., Dai, R., & Yu, X. (2019). A Physic Nut Stress-Responsive HD-Zip Transcription Factor, JcHDZ07, Confers Enhanced Sensitivity to Salinity Stress in Transgenic *Arabidopsis*. *Frontiers in Plant Science*, 10.  
<https://doi.org/10.3389/FPLS.2019.00942>
72. Tedeschi, G., Salladini, E., Santambrogio, C., Grandori, R., Longhi, S., & Brocca, S. (2018). Conformational response to charge clustering in synthetic intrinsically disordered proteins. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1862(10), 2204–2214.  
<https://doi.org/10.1016/J.BBAGEN.2018.07.011>
73. Thomasen, F. E., & Lindorff-Larsen, K. (2022). Conformational ensembles of intrinsically disordered proteins and flexible multidomain proteins. *Biochemical Society Transactions*, 50(1), 541–554. <https://doi.org/10.1042/BST20210499>
74. Tian, F., Yang, D. C., Meng, Y. Q., Jin, J., & Gao, G. (2020). PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Research*, 48(D1), D1104–D1113. <https://doi.org/10.1093/NAR/GKZ1020>
75. Tompa, P. (2002). Intrinsically unstructured proteins. *Trends in Biochemical Sciences*, 27(10), 527–533. [https://doi.org/10.1016/S0968-0004\(02\)02169-2](https://doi.org/10.1016/S0968-0004(02)02169-2)
76. Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*, 579(15), 3346–3354.  
<https://doi.org/10.1016/J.FEBSLET.2005.03.072>
77. Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends in Biochemical Sciences*, 37(12), 509–516.  
<https://doi.org/10.1016/J.TIBS.2012.08.004>
78. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., ... Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* 2021 596:7873, 596(7873), 590–596. <https://doi.org/10.1038/s41586-021-03828-1>

79. Uversky, V. N. (2019). Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Frontiers in Physics*, 7. <https://doi.org/10.3389/fphy.2019.00010>
80. Valli, M., Sauer, M., Branduardi, P., Borth, N., Porro, D., & Mattanovich, D. (2005). Intracellular pH Distribution in *Saccharomyces cerevisiae* Cell Populations, Analyzed by Flow Cytometry. *Applied and Environmental Microbiology*, 71(3), 1515. <https://doi.org/10.1128/AEM.71.3.1515-1521.2005>
81. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM. Classification of intrinsically disordered regions and proteins. *Chem Rev*. 2014 Jul 9;114(13):6589-631. doi: 10.1021/cr400525m. Epub 2014 Apr 29. PMID: 24773235; PMCID: PMC4095912.
82. Waadt, R., Seller, C. A., Hsu, P. K., Takahashi, Y., Munemasa, S., & Schroeder, J. I. (2022). Plant hormone regulation of abiotic stress responses. *Nature Reviews Molecular Cell Biology* 2022 23:10, 23(10), 680–694. <https://doi.org/10.1038/s41580-022-00479-6>
83. Wright, P. E., & Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. In *Nature Reviews Molecular Cell Biology* (Vol. 16, Issue 1, pp. 18–29). <https://doi.org/10.1038/nrm3920>
84. Yamasaki, K., Kigawa, T., Watanabe, S., Inoue, M., Yamasaki, T., Seki, M., Shinozaki, K., & Yokoyama, S. (2012). Structural basis for sequence-specific DNA recognition by an Arabidopsis WRKY transcription factor. *The Journal of Biological Chemistry*, 287(10), 7683–7691. <https://doi.org/10.1074/JBC.M111.279844>
85. Zarin, T., Strome, B., Nguyen Ba, A. N., Alberti, S., Forman-Kay, J. D., & Moses, A. M. (2019). Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *ELife*, 8. <https://doi.org/10.7554/ELIFE.46883>
86. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9(1), 1–8. <https://doi.org/10.1186/1471-2105-9-40/FIGURES/4>
87. Zheng, H., Handing, K. B., Zimmerman, M. D., Shabalin, I. G., Almo, S. C., & Minor, W. (2015). X-ray crystallography over the past decade for novel drug discovery – where are we heading next? *Expert Opinion on Drug Discovery*, 10(9), 975. <https://doi.org/10.1517/17460441.2015.1061991>



88. Zheng, W., Dignon, G., Brown, M., Kim, Y. C., & Mittal, J. (2020). Hydropathy Patterning Complements Charge Patterning to Describe Conformational Preferences of Disordered Proteins. *The Journal of Physical Chemistry Letters*, 11(9), 3408–3415. <https://doi.org/10.1021/ACS.JPCLETT.0C00288>