



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE QUÍMICA

Determinación de la gramática molecular en el fenómeno de separación de fases líquido-líquido en respuesta a estrés abiótico.

T E S I S

QUE PARA OBTENER EL TÍTULO DE
QUÍMICO FARMACÉUTICO BIÓLOGO

PRESENTA

DIEGO RUBÉN CASCO GARCÍA

DIRECTOR DE TESIS

Dr. César Luis Cuevas Velázquez



CIUDAD UNIVERSITARIA, CDMX 2022



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 Prof: **CERVANTES PEREDO ALICIA BEATRIZ**
- Vocal: Prof: **DIMITROVA DINKOVA TZVETANKA**
- Secretario Prof: **CUEVAS VELAZQUEZ CESAR LUIS**
- Suplente 1 Prof: **GUTIERREZ VENEGAS GLORIA**
- Suplente 2 Prof: **PLASENCIA DE LA PARRA FRANCISCO JAVIER**

SITIO DONDE SE DESARROLLÓ EL TEMA:

- Departamento de Bioquímica, Facultad de Química, Universidad Nacional Autónoma de México, Conjunto E, laboratorio 103.

ASESOR DEL TEMA:



Dr. César Luis Cuevas Velázquez.

SUSTENTANTE:



Diego Rubén Casco García.

AGRADECIMIENTOS.

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

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

ÍNDICE.

RESUMEN.	
ABREVIATURAS.	7
I. INTRODUCCIÓN.....	9
I.I Estrés osmótico.	9
I.II Condensados biomoleculares.....	10
I.III Gránulos de estrés.....	12
I.IV Cuerpos P.	13
I.V Proteínas y regiones intrínsecamente desordenadas.....	14
I.VI Separación de fases líquido-líquido.....	17
II. PLANTEAMIENTO DEL PROBLEMA.....	19
III. HIPÓTESIS.	22
IV. OBJETIVO GENERAL.	22
V. OBJETIVOS ESPECÍFICOS.	22
VI. MATERIALES Y MÉTODOS	24
VI.I Obtención de los grupos LLPS, IDR-LLPS y Delta-IDR-LLPS.	24
VI.II Obtención del grupo AtHomólogos e IDR-AtHomólogos.	24
VI.III Caracterización de grupos de estudio.	25
VI.IV Integración de parámetros en una base de datos.	25
VI.V Análisis de la base de datos utilizando Rstudio.	25
VII. RESULTADOS	26
VII.I Búsqueda de proteínas que experimentan separación de fases líquido-líquido durante condiciones de estrés osmótico.	26
VII.II Caracterización de los niveles de desorden y composición de aminoácidos de las regiones intrínsecamente desordenadas de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.	31
VII.III Determinación de los aminoácidos abundantes y carentes en las regiones intrínsecamente desordenadas de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.	32
VII.IV Búsqueda de proteínas de <i>Arabidopsis thaliana</i> ortólogas a las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.	49

VII.V Comparación de la composición de aminoácidos de las regiones intrínsecamente desordenadas entre las proteínas de humanos y los ortólogos de <i>Arabidopsis thaliana</i>	54
VIII. DISCUSIÓN.....	66
IX. CONCLUSIONES.....	70
X. PERSPECTIVAS.....	71
XI. BIBLIOGRAFÍA.....	72

ANEXOS.

Anexo I. Aminoácidos abundantes y carentes del grupo IDR-LLPS respecto al grupo Delta-IDR-LLPS.

Anexo II. Conservación de la IDR del grupo LLPS en el grupo AtHomólogos.

Anexo III. Aminoácidos conservados como abundantes y carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogos.

RESUMEN.

El interior celular además de estar conformado por el citosol y orgánulos delimitados por membranas también contiene compartimentos que carecen de membranas lipídicas, a los cuales se les conoce como condensados biomoleculares. Los condensados biomoleculares se forman a través de un proceso fisicoquímico conocido como separación de fases líquido-líquido. Estudios recientes han revelado que un grupo de proteínas conocidas como proteínas intrínsecamente desordenadas o aquellas que contienen regiones intrínsecamente desordenadas contienen la información necesaria para conferir la capacidad de experimentar separación de fases líquido-líquido y formar condensados biomoleculares. Las proteínas intrínsecamente desordenadas y las regiones intrínsecamente desordenadas tienen secuencias de complejidad baja, debido a que carecen de aminoácidos promotores del orden (Ile, Leu, Tyr, Trp, Phe, Asn, Val, y Cys) pero están enriquecidas en residuos promotores del desorden (Pro, Ser, Arg, Gly, Glu, Gln, Lys y Ala), lo que les confiere una notable flexibilidad conformacional y plasticidad estructural. En nuestro laboratorio se ha observado que cuando se expresa una proteína intrínsecamente desordenada conocida como AtLEA4-5 de la planta modelo *Arabidopsis thaliana* fusionada con dos proteínas fluorescentes en sus extremos (para formar un par de *Förster Resonance Energy Transfer*) en células vivas de levadura, esta experimenta el fenómeno de separación de fases líquido-líquido sólo en condiciones de estrés osmótico, lo que podría conferirle a la levadura la capacidad de responder de manera rápida y eficiente al estrés.

El objetivo de este proyecto de tesis fue identificar proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico, así como determinar la composición de aminoácidos de sus regiones intrínsecamente desordenadas y compararla con las regiones intrínsecamente desordenadas ortólogas de la planta modelo *Arabidopsis thaliana*. Se realizó una búsqueda en la literatura de proteínas que

experimentan el fenómeno de separación de fases líquido-líquido en respuesta a condiciones de estrés osmótico, y de las cuales se sabe que su región intrínsecamente desordenada es la responsable de dirigir el ensamblaje de los condensados biomoleculares. Todas las proteínas encontradas fueron de humano (*Homo sapiens*). Las proteínas fueron YAP, DCP1A, RAD23B y G3BP1. Usando herramientas bioinformáticas, se caracterizó este grupo de proteínas con respecto a ciertos parámetros (porcentaje de desorden, porcentaje de estructura y composición de aminoácidos), los cuales se integraron en una base de datos. Esta información nos permitió determinar cuáles eran los aminoácidos característicos en las regiones intrínsecamente desordenadas con respecto al resto de su secuencia, con la hipótesis de que la composición de aminoácidos es la responsable de conferirle a las regiones intrínsecamente desordenadas la capacidad de dirigir la separación de fases líquido-líquido en respuesta a estrés osmótico. Se encontró que los aminoácidos abundantes en las regiones intrínsecamente desordenadas analizadas fueron Asn, Asp, Glu, Val, Met, Gly, His, Pro y Arg. Además, se encontró que los aminoácidos carentes en las regiones intrínsecamente desordenadas en todas las proteínas analizadas fueron Cys y Trp. Por otro lado, también se buscaron las proteínas ortólogas a YAP, DCP1A, RAD23B y G3BP1, en la planta modelo *Arabidopsis thaliana*. Se encontró que la composición de aminoácidos abundantes en las regiones intrínsecamente desordenadas se conserva en los ortólogos de *Arabidopsis thaliana*. Estos datos sugieren que posiblemente los ortólogos en esta planta también tienen la capacidad de experimentar separación de fases líquido-líquido durante condiciones de estrés osmótico. Los resultados obtenidos en esta tesis contribuyen a entender la gramática molecular de la separación de fases líquido-líquido en respuesta a estrés osmótico en distintos organismos.

ABREVIATURAS.

AFM: Microscopia de fuerza atómica (del inglés *Atomic force microscopy*).

BFP: Sonda de fuerza de biomembrana (del inglés *Biomembrane Force Probe*).

CETN2: Centrin-2.

FRAP: Recuperación de fluorescencia después del fotoblanqueo (del inglés *Fluorescence Recovery After Photobleaching*).

GG-NER: Reparación por escisión de nucleótidos del genoma global (del inglés *Global genome nucleotide excision repair*).

IDP: Proteína intrínsecamente desordenada (del inglés *Intrinsecally disordered protein*).

IDR: Región intrínsecamente desordenada (del inglés *Intrinsically disordered region*).

LEA: Proteínas abundantes en embriogénesis tardía (del inglés *Late embryogenesis abundant proteins*).

LLPS: Separación de fases líquido-líquido (del inglés *Liquid-liquid phase separation*).

lncRNA: RNA largo no codificante (del inglés *long noncoding RNA*).

miRNA: micro RNA

mRNA: RNA mensajero (del inglés *Messenger RNA*).

ncRNA: RNA no codificante (del inglés *non-coding RNA*).

OT: Pinzas ópticas (del inglés *Optical tweezers*).

p97 (VCP): Proteína que contiene valosina (del inglés *Valosin-containing protein*).

PB: Cuerpo de procesamiento (del inglés *Processing body*).

PLD: Dominio similar a prion (del inglés *Prion-Like domain*).

RBP: Proteína de unión a RNA (del inglés *RNA-binding protein*).

RNA: Ácido ribonucleico (del inglés *Ribonucleic acid*).

RNAi: RNA interferente (del inglés *RNA interference*).

SG: Granulo de estrés (del inglés *Stress granule*).

TAD: Dominio de activación de la transcripción (del inglés *Transcription activation domain*).

TrD: Dominio de trimerización (del inglés *Trimerization domain*).

UBA1: Dominio 1 asociado a ubiquitina (del inglés *Ubiquitin-associated domain 1*).

UBA2: Dominio 2 asociado a ubiquitina (del inglés *Ubiquitin-associated domain 2*).

XPC: Xeroderma pigmentosum, grupo de complementación C (del inglés *Xeroderma pigmentosum, complementation group C*).

I. INTRODUCCIÓN.

I.I Estrés osmótico.

A lo largo de la evolución, células y organismos han desarrollado una gran variedad de estrategias y adaptaciones fisiológicas para poder desarrollarse en entornos específicos, en los que podrían estar presentes diferentes desafíos ambientales como por ejemplo cambios en el pH, en la osmolaridad y en la temperatura, que podrían provocar la alteración de las funciones del organismo afectando su supervivencia (Miermont *et al.*, 2013).

Existen perturbaciones en el equilibrio osmótico intracelular y extracelular, conocidas como ambiente hipoosmótico e hiperosmótico, que pueden provocar que las células inevitablemente se hinchen amenazando la integridad de la membrana (Grant, WD. 2004 ; Hohmann, S. 2002 ; Lang F, *et al.*, 1998) o se encojan provocando la contracción celular (Petelenz *et al.*, 2011 ; Hohmann, S. 2002), respectivamente, esto debido al flujo de agua impulsado osmóticamente. Las consecuencias de esto son perturbaciones en el equilibrio fisiológico y bioquímico de las células, amenazando su función y supervivencia, por lo cual estas han desarrollado distintas vías de señalización, así como respuestas transcripcionales para mediar la homeostasis del agua y su volumen celular (Miermont *et al.*, 2013).

Por lo general, los organismos (unicelulares y multicelulares) pueden encontrarse bajo condiciones de estrés hiperosmótico, que también es relevante para algunos tejidos en mamíferos, como es el caso de las células que conforman a los riñones, donde existe un ambiente hiperosmótico en relación con el resto del cuerpo (Brocker *et al.*, 2012 ; Marsh y Azen, 1975). Este tipo de estrés es un factor ambiental ubicuo que tiene la capacidad de

alterar diversos procesos celulares como, por ejemplo, inducir la remodelación de la cromatina, alterar la estructura del citoesqueleto, así como desencadenar la detención del ciclo celular y la apoptosis (Silva *et al.*, 2005 ; Escoté *et al.*, 2004 ; Kültz y Chakravarty, 2001 ; Slaninová *et al.*, 2000).

El estudio de este tipo de estrés en organismos como *C. elegans* y *S. cerevisiae*, ha permitido conocer cuáles son los principales cambios fisiológicos y moleculares que se producen inmediatamente después del estrés hiperosmótico debido al encogimiento de la célula provocado por el flujo de agua hacia el exterior y que conduce hacia el hacinamiento molecular. Se ha observado una reducción en la velocidad de difusión de proteínas debida a un aumento en la viscosidad citoplasmática, así como una afectación en el plegamiento, asociación y estabilidad de proteínas en el interior de la célula lo que provoca su agregación (Oeser *et al.*, 2016 ; Miermont *et al.*, 2013 ; Burkewitz *et al.*, 2011 ; Han y Emr, 2011).

I.II Condensados biomoleculares.

El interior celular, además de estar constituido por el citosol, también está conformado por distintos orgánulos que son los encargados de llevar a cabo diversos procesos metabólicos en áreas aisladas demarcadas por membranas para lograr especificidad y eficiencia (Mao *et al.*, 2011). Estas membranas son impermeables a la mayoría de las moléculas biológicas, lo que permite que el interior y el exterior de los orgánulos estén separados físicamente, siendo maquinarias especializadas de transporte las que llevan a cabo la regulación de la composición en su interior. Sin embargo, ahora se sabe que existen muchos compartimentos celulares que no están separados por una membrana, pero que tienen la capacidad de compartimentar y concentrar moléculas específicas al igual que los orgánulos. A este tipo de compartimentos, debido a que sus propiedades se asemejan a líquidos, geles o sólidos, se les ha denominado como condensados biomoleculares (u

orgánulos sin membrana) (Banani *et al.*, 2017 ; Shin y Brangwynne, 2017). La ventaja que tienen los condensados biomoleculares sobre los orgánulos con membrana es que tienen la capacidad de formarse para responder dinámicamente contra diversas condiciones de estrés, condiciones metabólicas alteradas o simplemente ante procesos fisiológicos básicos, permitiendo concentrar ácido ribonucleico (RNA, del inglés *Ribonucleic acid*) y proteínas en sitios discretos (Mao *et al.*, 2011).

La razón por la que los condensados pueden formarse y mantenerse unidos es debido a interacciones débiles entre sus componentes, que contienen colecciones de miles de moléculas cuya organización relativa es típicamente muy dinámica. Estos están conformados por dos tipos de componentes proteicos, por un lado están los clasificados como “andamios”, que son los encargados de impulsar su formación reversible, y por otro lado están los “clientes”, que se distribuyen preferentemente hacia el interior de los condensados (Banani *et al.*, 2017 ; Banani *et al.*, 2016). Además, el que la composición macromolecular de los condensados sea específica, implica que estos al formarse logran dividir clientes particulares en su interior, mientras que a otros los mantienen fuera (Banani *et al.*, 2016 ; Feric *et al.*, 2016 ; Zhang *et al.*, 2015).

Un tipo particular de condensados biomoleculares son aquellos que están constituidos por RNA y proteínas, por ejemplo, los gránulos de estrés (SG, del inglés *Stress granule*), los cuerpos de Cajal y los cuerpos de procesamiento (PB, del inglés *Processing body*) (Shin y Brangwynne, 2017).

Como dato histórico, alrededor del año 1830, el primer condensado biomolecular observado fue el nucleolo, que se encontraba ubicado en el interior del núcleo de células neuronales (Pederson, 2011). Desde entonces, se han descubierto muchos de estos condensados en el núcleo y en el citoplasma de prácticamente todas las células eucariotas. En la Figura 1 se representan estos condensados biomoleculares en una célula.

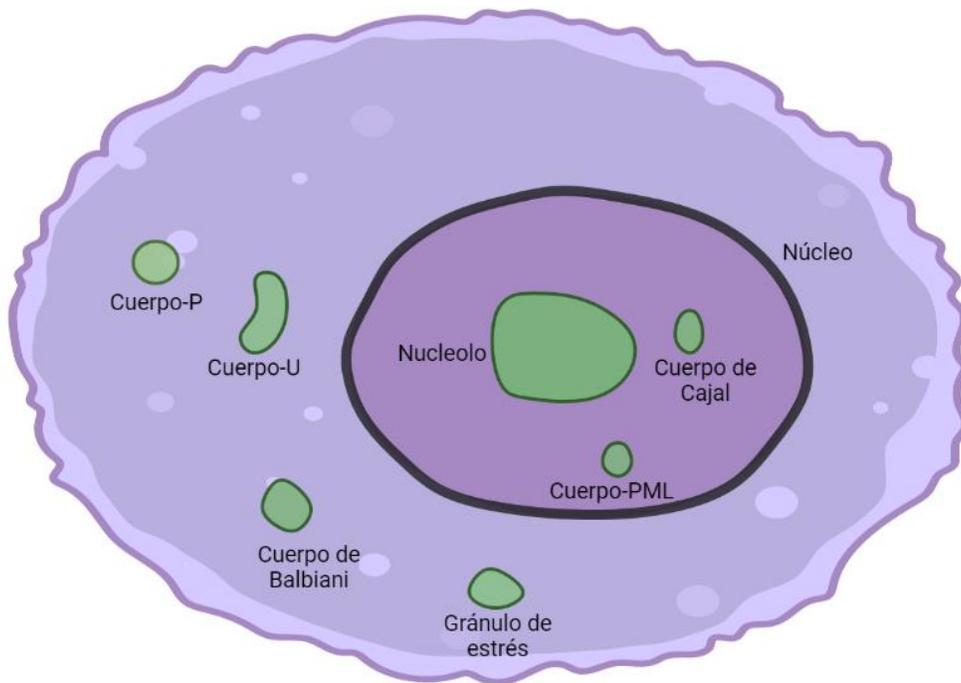


Figura 1. Condensados biomoleculares en células eucariotas. Cuerpo P: Cuerpo de procesamiento. Cuerpo U: Cuerpo que contiene ribonucleoproteínas nucleares ricas en uridina. Cuerpo PML: Cuerpo de leucemia promielocítica. Modificado de *Banani et al., 2017*.

I.III Gránulos de estrés.

Este es un tipo de condensado biomolecular cuya composición está dada por distintos tipos de moléculas de RNA (mRNA, del inglés messenger RNA, lncRNA, del inglés *long noncoding RNA*, y miRNA, del inglés *micro RNA*) y proteínas, el cual se forma bajo diversas condiciones de estrés (temperatura, oxidación, genotoxicidad, privación de nutrientes y osmótico) (Kedersha *et al.*, 2013 ; Thomas *et al.*, 2011 ; Pothof *et al.*, 2009), por lo que se le conoce como SG. Una característica importante es que este tipo de condensado puede contener elementos de la traducción de mRNA como subunidades ribosomales o factores de traducción. El gránulo de estrés tiene la capacidad de intercambiar proteínas y RNA con el citoplasma circundante u otros compartimentos (como los PB) gracias a sus propiedades dinámicas (Zhang *et al.*, 2011 ; Mollet *et al.*, 2008 ; Kedersha *et al.*, 2005).

La formación de los SG resulta muy importante para la fisiología y supervivencia celular bajo condiciones de estrés, ya que le permiten a la célula poder controlar y minimizar el gasto energético y la ribostasis (Panas *et al.*, 2016 ; Walters y Parker, 2015 ; Arimoto *et al.*, 2008). Sin embargo, la formación de estos gránulos resulta un arma de doble filo, ya que su ensamblaje incorrecto se ha relacionado con distintas enfermedades como las neurodegenerativas y autoinmunes, el glaucoma y las cataratas, por mencionar algunas (Shukla y Parker, 2016 ; Buchan, 2014 ; Lachke *et al.*, 2011).

El ensamblaje de los SG se debe a proteínas de unión a RNA que están constituidas por dominios similares a priones (PLD, del inglés *Prion-Like domain*) o por regiones intrínsecamente desordenadas (IDR, del inglés *Intrinsically disordered regions*) (Gilks *et al.*, 2004) que impulsan su formación a través del fenómeno de separación de fases líquido-líquido (LLPS, del inglés *Liquid-liquid phase separation*) mediante múltiples interacciones débiles gracias a su complejidad baja dada por su composición característica de aminoácidos, en particular polares, incluidos Ser, Tyr, Gln y Asn (Courchaine *et al.*, 2016 ; Lin *et al.*, 2015 ; Malinovska *et al.*, 2013).

Es importante mencionar que las modificaciones postraduccionales de las proteínas que constituyen al SG afectan su interacción y en consecuencia a su ensamblaje, coalescencia y desensamblado general (Xie y Denman, 2011).

I.IV Cuerpos P.

Otro ejemplo de condensados biomoleculares de ribonucleoproteínas citoplásmicos son los PB, de los cuales se sabe que llevan a cabo diversas funciones importantes como lo es la señalización celular, la homeostasis de proteínas, el almacenamiento y degradación de mRNA y la respuesta a estrés (Zhang y Herman, 2019 ; Brengues *et al.*, 2005).

Estos, al igual que los SG, están conformados por RNA y proteínas que contienen PLD e IDR en su secuencia, lo que les confiere la capacidad a dichos componentes para formar condensados a través del fenómeno de LLPS (Riggs *et al.*, 2020 ; Banani *et al.*, 2017). Cabe destacar que las proteínas que contienen los PB participan en ciertos procesos postranscripcionales como la represión traduccional y el silenciamiento génico mediado por interferencia de RNA (RNAi, del inglés *RNA interference*), por mencionar algunos (Eulalio *et al.*, 2020). Los PB no contienen subunidades ribosomales u otros elementos de la traducción de mRNA, aunque si pueden contener a la proteína de unión al 5' Cap.

Finalmente, respecto a la relación que existe entre el ensamblaje incorrecto de los PB y la enfermedad, se ha reportado que, en infecciones por virus de RNA, estos pueden reclutar componentes de estos condensados a los centros de replicación viral, afectando tanto su número como su estabilidad (Liu *et al.*, 2019 ; Chahar *et al.*, 2013 ; Emara y Brinton, 2007).

I.V Proteínas y regiones intrínsecamente desordenadas.

En la actualidad, la visión clásica de función-estructura de una proteína ha cambiado drásticamente, ya que ahora es bien sabido que las regiones carentes de estructura de las proteínas también cumplen una función importante. Estas regiones carentes de estructura se han denominado IDR, sin embargo, también existen proteínas completamente desordenadas que se conocen como proteínas intrínsecamente desordenadas (IDP, del inglés *Intrinsically disordered proteins*) (van der Lee *et al.*, 2014 ; Dunker *et al.*, 2013) (Figura 2).

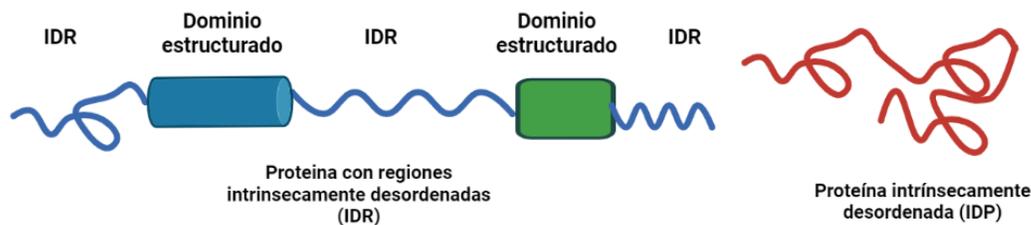


Figura 2. Proteínas y regiones intrínsecamente desordenadas. Modificado de *van der Lee et al., 2014*

Las IDP e IDR son frecuentes en eucariotas, por ejemplo, en *Homo sapiens* se sabe que alrededor de 44% de los genes codifican estas proteínas, las cuales están conformadas por regiones intrínsecamente desordenadas de más de 30 aminoácidos de longitud (Oates *et al.*, 2012), cuya composición carece de aminoácidos hidrófobos voluminosos (Tyr, Phe, Val, Trp, Cys, Asn, Ile y Leu), lo que les impide tener una estructura tridimensional bien definida en condiciones fisiológicas (Romero *et al.*, 2001 ; Uversky *et al.*, 2000).

En cambio, en diversos estudios sobre IDR e IDP, se ha determinado que están enriquecidas en aminoácidos promotores del desorden y que comúnmente están repetidos en tándem (Simon y Hancock, 2009) como por ejemplo Pro, Arg, Gly, Gln, Glu, Ser, Lys, Asp, Arg, His y Thr (Uversky y Dunker, 2010), confiriéndole a la secuencia una notable plasticidad y flexibilidad estructural (Uversky, 2019).

Anteriormente se mencionó que las IDP y las IDR no pueden formar estructuras estables, sin embargo, algunos autores señalan que, debido a la naturaleza de los aminoácidos y su distribución en la secuencia, bajo condiciones de apilamiento molecular, algunas fluctúan estocásticamente entre diferentes estados estructurales en forma transitoria, y que van desde estructuras desplegadas, hasta estructuras secundarias localizadas y estados globulares más compactos (Mao *et al.*, 2010 ; Marsh y Forman, 2010). Por ejemplo, en estudios recientes *in vitro* se ha caracterizado la estructura secundaria del grupo 4 de proteínas abundantes en embriogénesis tardía

(LEA, del inglés *Late Embryogenesis-Abundant*) de *Arabidopsis thaliana* (*AtLEA4-2* y *AtLEA4-5*), que tienen la composición característica de las IDP. Las LEA del grupo 4 están desordenadas en solución acuosa, sin embargo, se ha demostrado que los potenciales osmóticos bajos y el apiñamiento macromolecular pueden inducir niveles significativos de α -hélice, particularmente en la región N-terminal conservada de *AtLEA4-5*, mientras que la región C-terminal muestra un desorden estructural alto (Cuevas *et al.*, 2016).

Por otra parte, con respecto a la evolución y conservación del desorden en IDR e IDP, explican van der Lee *et al.*, 2014, que este suele ocurrir más rápido con respecto a los dominios globulares; señalando también que el desorden se conserva solo para las secuencias funcionales, pero que varía de acuerdo con sus funciones específicas.

Las funciones de las IDP e IDR se han distinguido en diferentes clases funcionales entre las que destacan las cadenas entrópicas (enlazadores flexibles y espaciadores de dominios), chaperonas (permiten interacciones macromoleculares rápidas) (Rogers *et al.*, 2013), efectores (interactúan con otras proteínas y modifican su actividad) (Wright y Dyson, 2009), ensambladores (promueven la formación de complejos de proteínas de orden superior como el ribosoma) (Cumberworth *et al.*, 2013 ; Wu, 2013), y carroñeros (almacenan y neutralizan ligandos pequeños) (Tompa, 2005 ; Tompa, 2002).

En relación a la participación de las IDP e IDR en la formación de los condensados biomoleculares, se ha observado que todos los condensados tienen numerosos IDP/IDR, caracterizándose por niveles generales muy altos de desorden estructural. En otras palabras, parece que la fluidez y la naturaleza altamente dinámica de los condensados están determinadas por la naturaleza intrínsecamente desordenada de las proteínas involucradas en su formación (Uversky, 2017).

I.VI Separación de fases líquido-líquido.

Cuevas y Dinneny (2018) mencionan que una de las estrategias que han desarrollado los organismos para contender contra las distintas formas de estrés abiótico, incluido el hiperosmótico, es la compartimentalización de biomoléculas a través de un fenómeno conocido como LLPS, que es un proceso en el que dos líquidos se mezclan en dos fases no miscibles (una fase rica y otra pobre en proteínas) según el umbral de concentración crítica de los componentes, que puede ajustarse mediante alteraciones fisicoquímicas del sistema (similar a la separación de gotas de aceite en un mezcla de aceite/agua) (Mitrea y Kriwacki, 2016). Estas gotas (condensados biomoleculares) pueden experimentar un proceso conocido como envejecimiento, en el cual la dinámica molecular del condensado, previamente similar a un líquido, se ve disminuida (Babinchak y Surewicz, 2020). Por otro lado, la investigación continua de este tipo de compartimentos que se forman a través de LLPS, ha dado lugar al descubrimiento de que sus propiedades materiales pueden estar moduladas por agentes de apiñamiento (Kaur *et al.*, 2019), cationes divalentes (Onuchic *et al.*, 2019) y presencia, tipo y concentración de RNA (Alshareedah *et al.*, 2019)

Para la investigación de LLPS se consideran tres características distintas de los condensados biomoleculares: fusión, reversibilidad y dinámica alta (Guo *et al.*, 2021). Los métodos desarrollados para el estudio de este fenómeno son: la recuperación de fluorescencia después del fotoblanqueo (FRAP, del inglés *Fluorescence Recovery After Photobleaching*) (Banani *et al.*, 2017 ; Hyman *et al.*, 2014), que permite determinar la dinámica molecular del condensado, y la microscopia de fuerza atómica (AFM, del inglés *Atomic force microscopy*), que ayuda a detectar la forma de los condensados a través de imágenes de alta resolución (Zemla *et al.*, 2018). También se ha desarrollado una sonda de fuerza de biomembrana (BFP, del inglés *Biomembrane Force Probe*) y las pinzas ópticas (OT, del inglés *Optical tweezers*), las cuales

permiten detectar la tensión superficial y la elasticidad de los condensados (Jawerth *et al.*, 2018).

Respecto a las características de secuencia e interacción de proteínas que están involucradas en el fenómeno de LLPS, Saar *et al.*, 2021, explican que diversos autores han sugerido la influencia y participación de distintos tipos de interacciones involucradas en la formación de los condensados biomoleculares, como lo son las interacciones tipo pi-pi, catión-pi, interacciones electrostáticas e hidrofóbicas. Además, también la valencia y la distribución de las IDR a lo largo de la secuencia son importantes para que ocurra LLPS (Gomes y Shorter *et al.*, 2019 ; Martin y Mittag, 2018).

Evidencias experimentales sustentan la capacidad de las IDR y los desplazados internos para dirigir y modular la LLPS de proteínas; sin embargo, cada vez más estudios revelan que la concentración de RNA puede tener un efecto regulador sobre la aparición de LLPS, esto a través de interacciones electrostáticas entre proteínas (carga positiva) y moléculas de RNA (carga negativa) presentes en el núcleo celular (Henninger *et al.*, 2021). Relacionado a lo anterior, distintos autores han señalado que el RNA no codificante (ncRNA, del inglés *non-coding RNA*) participa como un andamio en la regulación de LLPS, ya que une múltiples proteínas de unión a RNA (RBP, del inglés *RNA-binding protein*) formando una red dinámica que da lugar a la formación de condensados (Pessina *et al.*, 2019 ; Lin *et al.*, 2015 ; Aumiller *et al.*, 2016). También otros estudios han revelado que los mRNA, mediante el apareamiento de bases complementarias, pueden formar condensados (Langdon *et al.*, 2018).

II. PLANTEAMIENTO DEL PROBLEMA.

El interior celular, no solo está constituido por el citosol y por distintos orgánulos encargados de llevar a cabo diversas funciones en regiones demarcadas por membranas para lograr especificidad y eficiencia (Mao YS, *et al.*, 2011), sino que también está conformado por compartimentos que no están delimitados por membranas, los cuales tienen propiedades que se asemejan a la de los líquidos y que se conocen como condensados biomoleculares (u orgánulos sin membrana) (Banani *et al.*, 2017, Shin y Brangwynne, 2017). Los condensados biomoleculares se forman a través del fenómeno de LLPS, en el cual se entiende como separación de fases a la aparición de una fase enriquecida en proteínas y una fase empobrecida en proteínas a partir de una solución uniformemente dispersa (Chong y Forman-Kay, 2016 ; Mitrea y Kriwacki, 2016), de las cuales la fase enriquecida a menudo asume la forma de gotas líquidas que contienen proteína altamente concentrada, similar a las gotas de aceite en una mezcla de agua y aceite.

A diferencia de los orgánulos clásicos, los condensados biomoleculares que subyacen del fenómeno de LLPS, resultan ser un poderoso mecanismo de regulación, porque su formación y disolución se controla rápida y dinámicamente mediante modificaciones postraduccionales, en respuesta a señales ambientales (Anderson y Kedersha, 2009) o al incremento en la concentración de proteínas citoplásmicas (Matsumoto *et al.*, 2012).

Los condensados biomoleculares están conformados por proteínas que impulsan su formación reversible a las cuales se les ha clasificado como “andamios” y por componentes macromoleculares que se

dividen preferentemente en los condensados que se han definido como “clientes” (Banani *et al.*, 2017 ; Banani *et al.*, 2016). Es importante destacar que Wang *et al.*, 2018, explican que una célula a través del fenómeno de LLPS es capaz de formar condensados biomoleculares de proteínas, discriminando aquellos clientes que se mantendrán dentro y fuera del compartimento (Banani *et al.*, 2016 ; Feric *et al.*, 2016 ; Zhang *et al.*, 2015).

A la fecha se sabe que un tipo de dominios proteicos que tienen la capacidad de experimentar el fenómeno de LLPS y, por lo tanto, formar condensados biomoleculares son las IDR, las cuales carecen de una estructura tridimensional bien definida (Romero *et al.*, 2001 ; Uversky *et al.*, 2000). Dunker *et al.*, 2013, explican que tanto los dominios estructurados y las regiones intrínsecamente desordenadas en proteomas eucariotas, son importantes para el repertorio de funciones que una proteína puede tener en una variedad de contextos celulares. Se ha reportado que la composición de aminoácidos, pero no la secuencia primaria *per se*, es importante para ejercer las funciones específicas de las IDR (Mao *et al.*, 2010 ; Marsh y Forman, 2010). Esto es consistente con el hecho de que los niveles de conservación de secuencia primaria en las IDR son muy bajos.

Los condensados biomoleculares representan una estrategia innovadora para responder de forma rápida y eficiente a condiciones adversas. Las células de diferentes organismos pueden estar sometidas a condiciones de estrés abiótico como la osmolaridad alta, cambios de pH, temperatura, estado redox, entre otros (He *et al.*, 2018 ; Miermont *et al.*, 2013). Algunos reportes en células humanas han mostrado que existen condensados biomoleculares que se forman inmediatamente después del aumento en la osmolaridad extracelular, lo cual induce un aumento en el amontonamiento macromolecular al interior de la célula (Jalihal, A. *et al.*, 2020 ; Yasuda *et al.*, 2020 ; Cai, D *et al.*, 2019). En dichos casos, se demostró que la formación de los condensados promueve la respuesta eficiente a las condiciones de estrés. Además, se mostró que las proteínas que facilitan la formación de los

condensados biomoleculares contienen IDR las cuales, en algunos casos, son necesarias y suficientes para su biogénesis. Los resultados descritos sugieren que este tipo de respuesta ante las condiciones de estrés podría estar conservada en otros organismos que están constantemente expuestos a cambios en su ambiente abiótico, como las plantas, sin embargo, a la fecha no existen reportes sobre proteínas que experimentan LLPS en respuesta a estrés abiótico en estos organismos.

III. HIPÓTESIS.

La composición de aminoácidos de las regiones intrínsecamente desordenadas de proteínas identificadas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico está conservada en las regiones intrínsecamente desordenadas de las proteínas ortólogas de la planta modelo *Arabidopsis thaliana*.

IV. OBJETIVO GENERAL.

Identificar proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico, determinar la composición de aminoácidos de sus regiones intrínsecamente desordenadas y compararla con las regiones intrínsecamente desordenadas ortólogas de *Arabidopsis thaliana*.

V. OBJETIVOS ESPECÍFICOS.

- Buscar en la literatura reportes de proteínas de cualquier organismo que experimentan separación de fases líquido-líquido durante condiciones de estrés osmótico.
- Caracterizar los niveles de desorden y composición de aminoácidos de las regiones intrínsecamente desordenadas de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.

- Determinar los aminoácidos abundantes y carentes en las regiones intrínsecamente desordenadas de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.
- Identificar las proteínas de *Arabidopsis thaliana* ortólogas a las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.
- Comparar la composición de aminoácidos de las regiones intrínsecamente desordenadas entre las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico y los ortólogos de *Arabidopsis thaliana*.
- Predecir la capacidad de experimentar separación de fases líquido-líquido en las proteínas ortólogas de *Arabidopsis thaliana* con base en la composición de aminoácidos y utilizando el predictor PSPredictor.
- Establecer un procedimiento de análisis de proteínas para predecir si experimentarán separación de fases líquido-líquido en respuesta a estrés osmótico en base a su composición de aminoácidos.

VI. MATERIALES Y MÉTODOS

VI.I Obtención de los grupos LLPS, IDR-LLPS y Delta-IDR-LLPS.

Las proteínas del grupo LLPS se obtuvieron de la búsqueda en la literatura de proteínas con la capacidad de experimentar separación de fases líquido-líquido en respuesta a estrés osmótico. La búsqueda se realizó en septiembre del 2020, utilizando la base de datos de PubMed con el criterio de búsqueda (*hyperosmotic stress*) AND (*phase separation*). Los artículos encontrados en la búsqueda fueron analizados individualmente para determinar la presencia de proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico. Además, se realizó una búsqueda manual de los artículos publicados en las revistas Nature, Science, Cell y PNAS entre los años 2019 a 2020. Una vez identificadas las proteínas candidatas a estudio, su secuencia de aminoácidos (grupo LLPS), así como sus regiones intrínsecamente desordenadas (grupo IDR-LLPS) se obtuvieron del servidor UniProt (<https://www.uniprot.org/>) y Pfam (<http://pfam.xfam.org/>). El grupo Delta-IDR-LLPS que corresponde a la secuencia de aminoácidos de la proteína, pero sin la región intrínsecamente desordenada, se obtuvo al eliminar la IDR previamente identificada en la secuencia.

VI.II Obtención del grupo AtHomólogos e IDR-AtHomólogos.

Los ortólogos en *Arabidopsis thaliana* (grupo AtHomólogos) a las proteínas del grupo LLPS se obtuvieron utilizando el servidor Blastp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) en conjunto con el servidor Phytozome (<https://phytozome-next.jgi.doe.gov/blast-search>). Los

ortólogos obtenidos se alinearon individualmente con su respectiva proteína del grupo LLPS en el servidor T-Coffee (<https://www.ebi.ac.uk/Tools/msa/tcoffee/>) para determinar su nivel de conservación e identificar la IDR en el ortólogo (grupo IDR-AtHomólogos).

VI.III Caracterización de grupos de estudio.

A los grupos obtenidos se les determinaron diferentes parámetros. Para obtener el porcentaje de desorden de cada proteína se utilizó la herramienta en línea *Predictor of Natural Disordered Regions* (PONDR) (<http://www.pondr.com/>) utilizando el algoritmo VLXT. Para determinar el porcentaje de estructura secundaria particular (alfa hélice, lamina beta o “random coil”) se utilizó la herramienta en línea Pole Rhone-Alpes de Bioinformatique (prabi) (<http://www.prabi.fr/>) utilizando el algoritmo MLRC. El porcentaje de cada aminoácido en la composición de la secuencia se obtuvo con la herramienta en línea ProtParam (<https://web.expasy.org/protparam/>).

VI.IV Integración de parámetros en una base de datos.

Los parámetros obtenidos para cada grupo de estudio se organizaron en un archivo de Excel, indicando en cada columna el parámetro correspondiente para cada proteína indicada en cada fila del archivo. Para facilitar el análisis gráfico del archivo, se utilizó el formato CSV para su manejo con apoyo de Rstudio.

VI.V Análisis de la base de datos utilizando Rstudio.

Para el análisis de la base de datos se utilizó el software Rstudio a través de la paquetería ggplot2. Se elaboraron dos códigos para el análisis gráfico de barras y el de puntos, mediante los cuales se estudió la composición particular de cada aminoácido de cada grupo de proteínas, así como de sus ortólogos.

VII. RESULTADOS

VII.I Búsqueda de proteínas que experimentan separación de fases líquido-líquido durante condiciones de estrés osmótico.

Se realizó una búsqueda en la literatura de proteínas con la capacidad de experimentar LLPS bajo condiciones de estrés osmótico (ver materiales y métodos). La base de datos de PubMed arrojó 16 resultados al momento de la búsqueda (septiembre del 2020). Tras analizar cada una de las publicaciones encontradas, se encontraron cuatro proteínas: YAP, DCP1A, RAD23B y G3BP1 (Tabla 1). Todas las proteínas encontradas fueron proteínas de humanos. Estas proteínas están conformadas por diferentes dominios proteicos, entre los que podemos destacar los dominios estructurados y las IDR. En la Tabla 1 se presentan las proteínas encontradas, su función molecular y el dominio responsable de la LLPS.

Tabla 1. Proteínas que experimentan LLPS bajo condiciones de estrés osmótico en células humanas.

Proteína	Función molecular	Condensados biomoleculares	Dominio IDR	Referencia
YAP	Transcripción de genes involucrados en la proliferación celular	Condensados nucleares	TAD	<i>Cai, D et al., 2019</i>
DCP1A	Formación de cuerpos P	Cuerpos P	CTD	Jalihal, A. et al., 2020
RAD23B	Modulación de la degradación en el proteasoma	Condensados de proteasoma	UBA1, UBA2	<i>Yasuda, S. et al., 2020</i>
G3BP1	Ensamblaje de gránulos de estrés	Gránulos de estrés	RG, Ácido	<i>Guillén, J. et al., 2020</i>

A continuación, se describen brevemente aspectos generales de cada una de las proteínas encontradas.

YAP (*Yes-associated Protein*).

YAP es un coactivador transcripcional encargado de promover la proliferación y supervivencia celular al unirse a un conjunto selectivo de potenciadores para la activación de genes diana; sin embargo, se desconoce cómo YAP coordina estas respuestas transcripcionales.

En estudios recientes, Cai, D *et al.*, 2019 demostraron que YAP tiene la capacidad de experimentar LLPS en el núcleo, formando condensados líquidos en dicha región en respuesta al hacinamiento macromolecular (Cai, D *et al.*, 2019).

Cabe destacar que, en su estudio, Cai, D *et al* 2019 demostraron que la eliminación del dominio intrínsecamente desordenado de activación de la transcripción, mejor conocido como TAD (del inglés *Transcription activation domain*), evitó la formación de condensados de YAP y disminuyó la señalización de YAP corriente abajo. En la Figura 3 se muestra la estructura de YAP.



Figura 3. Estructura de la proteína YAP. P (*Proline rich region*), TEAD (*transcriptional enhanced associate domain*), TAD (*transcription activation domain*). Modificado de Cai, D *et al.*, 2019.

DCP1A (*mRNA-decapping enzyme 1A*).

DCP1A es un componente específico de los cuerpos P, se une específicamente al 5' Cap de los mRNA y lo elimina, promoviendo así la degradación del mRNA (Aizer *et al.*, 2013).

En estudios recientes, Jalihal, A. *et al.*, 2020 observaron que DCP1A tiene la capacidad de experimentar rápidamente (dentro de ~ 10s) el fenómeno de LLPS en células de mamíferos durante condiciones de estrés hiperosmótico, siendo reversible tras el rescate isosmótico (más de ~ 100 s) con un impacto mínimo en la viabilidad celular, incluso después de múltiples ciclos de perturbación osmótica (Jalihal, A. *et al.*, 2020).

La capacidad de DCP1A para experimentar LLPS bajo condiciones de estrés osmótico se debe a su dominio de trimerización C-terminal (TrD, del inglés *Trimerization domain*). En la Figura 4 se muestra la estructura de DCP1A (Jalihal, A. *et al.*, 2020).

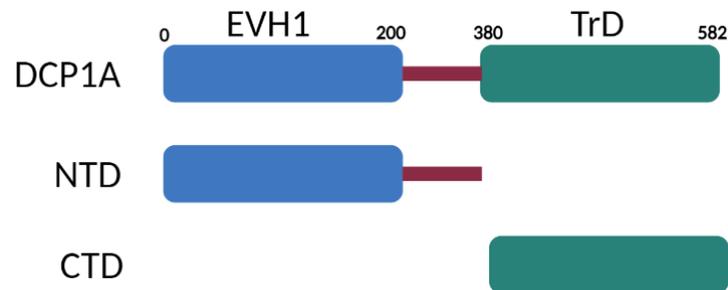


Figura 4. Estructura de la proteína DCP1A. EVH1 (*enabled/vasodilator-stimulated phosphoprotein homology domain 1*), TrD (*trimerization domain*), NTD (N-terminal domain), CTD (C-terminal domain). Modificado de Jalihal, A. *et al.*, 2020.

RAD23B (*UV excision repair protein RAD23 homolog B*).

El proteasoma es una máquina proteolítica importante que regula la proteostasis celular a través de la degradación selectiva de proteínas ubiquitinadas (Sun, D. *et al.*, 2018 ; Turakhiya, A. *et al.*, 2018).

En estudios recientes realizados por Yasuda *et al.*, 2020, revelan que bajo condiciones de estrés hiperosmótico agudo, se forman focos nucleares con propiedades de gotas líquidas que contienen proteasomas, proteínas ubiquitinadas, p97 (VCP, del inglés *Valosin-containing protein*) y múltiples proteínas que interactúan con el proteasoma, que colectivamente constituyen un centro proteolítico encargado de la degradación de proteínas ribosomales no ensambladas correctamente (Yasuda *et al.*, 2020). El ensamblaje de dichos cuerpos ocurre mediante el proceso de separación de fases líquido-líquido. La proteína RAD23B, la cual según Yasuda *et al.*, 2020, es indispensable en conjunto con proteínas ubiquitinadas para la formación de los focos nucleares gracias a interacciones multivalentes de dos dominios asociados a ubiquitina de RAD23B (UBA1, del inglés *Ubiquitin-associated domain 1* y UBA2, del inglés *Ubiquitin-associated domain 2*) y cadenas de ubiquitina que constan de cuatro o más moléculas de ubiquitina. Además, también se sabe que RAD23B participa en la reparación por escisión de nucleótidos del genoma global (GG-NER, del inglés *global genome nucleotide excision repair*) al actuar como componente del complejo XPC (del inglés *Xeroderma pigmentosum, complementation group C*). En cooperación con CETN2 (del inglés *Centrin-2*) parece estabilizar XPC. Puede proteger XPC de la degradación proteosomal (UniProt, s.f.). En la Figura 5 se muestra la estructura de RAD23B.

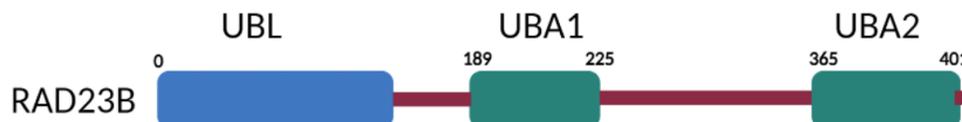


Figura 5. Estructura de la proteína RAD23B. UBL (*ubiquitin-like domain*), UBA1 (*ubiquitin-associated domain 1*), UBA2 (*ubiquitin-associated domain 2*). Modificado de Yasuda *et al.*, 2020.

G3BP1 (*Ras GTPase-activating protein-binding protein 1*).

En estudios recientes realizados por Guillén-Boixet *et al.*, 2020 mencionan que las células estresadas detienen la traducción, liberan mRNA de los polisomas y forman SG a través de una red de interacciones que involucran a la proteína G3BP1. Esta proteína está conformada por distintos dominios, de los cuales dos son IDR, una de ellas es la IDR-ácida que se encuentra unida a la IDR rica en glicina y arginina (IDR-RG) a través de interacciones intramoleculares electrostáticas. Sin embargo, tras el estrés, el mRNA liberado supera esta interacción intramolecular induciendo la expansión de G3BP1, liberando la IDR-RG (estado expandido) para participar en interacciones proteína-RNA, que permiten la unión cooperativa del RNA y el ensamblaje del condensado, siendo esta región la encargada de proporcionar a G3BP1 el control sobre las fuerzas impulsoras de la condensación (Figura 6). (Guillén-Boixet *et al.*, 2020).

Por lo tanto, G3BP1 aprovecha la separación de fases líquido-líquido para formar condensados líquidos en respuesta a cambios locales en la concentración de RNA, inhibiendo las interacciones RNA-RNA evitando su agregación (Guillén-Boixet *et al.*, 2020). En la Figura 6 se muestra la estructura de G3BP1.

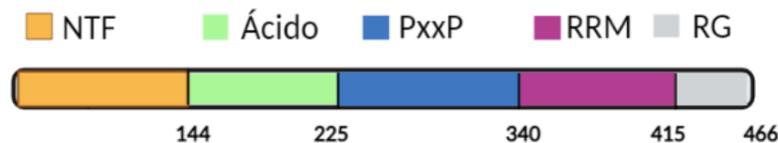


Figura 6. Estructura de la proteína G3BP1. NTF (*N-terminal nuclear transport factor-like domain*), Ácido (*central acidic IDR*), PxxP (*proline residues*), RRM (*RNA-recognition motif*), RG (*RG-rich region*). Modificado de Guillén-Boixet *et al.*, 2020.

VII.II Caracterización de los niveles de desorden y composición de aminoácidos de las regiones intrínsecamente desordenadas de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.

El siguiente paso fue caracterizar las propiedades particulares de las proteínas que experimentan LLPS en respuesta a estrés osmótico. Entre estas características se encuentran el porcentaje de desorden, los niveles de estructura secundaria y la composición de aminoácidos, mediante el uso de herramientas bioinformáticas como PONDR, Prabi y Protparam. La información de estas propiedades se recopiló en una base de datos para poder determinar cuáles son aquellos aminoácidos abundantes en la región intrínsecamente desordenada con respecto al resto de la secuencia de cada proteína, independientemente de si promueven o no el desorden estructural, bajo la hipótesis de que su abundancia es importante para la función. Así mismo, se determinaron los aminoácidos carentes en la región intrínsecamente desordenada. En la Tabla 2 se presenta cómo está organizada la base de datos, mostrándose que las proteínas que experimentan LLPS se clasificaron en tres grupos distintos. La primera fue caracterizando la totalidad de las proteínas, denominando a este grupo como “LLPS”, por otro lado, las otras dos formas de clasificar a estas proteínas fue separando sus IDR y el resto de la secuencia que no corresponde a la IDR, denominando a estos dos grupos como IDR-LLPS y Delta-IDR-LLPS, respectivamente.

Como se puede observar en la Tabla 2, en el caso de DCP1A, la región Delta-IDR-LLPS está 19.53 % más desordenada que la IDR-LLPS descrita por los autores, mientras que para el caso de YAP no difiere mucho el porcentaje de desorden de ambas regiones, lo cual se debe a la distribución de aminoácidos promotores de orden y desorden en las secuencias. Aun así, se consideró como región desordenada de interés al dominio descrito por los autores de los respectivos artículos como el responsable de dirigir el fenómeno

de LLPS en respuesta a estrés osmótico en ambas proteínas, incluso si el resto de la secuencia era más desordenada.

Este análisis se realizó debido a que en los artículos encontrados en la literatura se encontró que la IDR de cada proteína es necesaria y suficiente para experimentar LLPS.

Tabla 2. Proteínas que experimentan LLPS en células humanas y cuyas secuencias son parte de la base de datos de este trabajo.

Proteína	Grupo	% Desorden	% Estructura secundaria
YAP	LLPS	75.5	32.38
DCP1A	LLPS	59.62	34.71
RAD23B	LLPS	60.88	42.05
G3BP1	LLPS	53.22	34.54
IDR-YAP	IDR-LLPS	71.94	45.41
IDR-DCP1A	IDR-LLPS	48.28	42.85
IDR-RAD23B(UBA1)	IDR-LLPS	67.50	45
IDR-RAD23B(UBA2)	IDR-LLPS	76.74	41.86
IDR-G3BP1(Región ácida)	IDR-LLPS	100	39.02
IDR-G3BP1(RG)	IDR-LLPS	93.75	3.12
Delta-IDR-YAP	Delta-IDR-LLPS	72.65	24.49
Delta-IDR-DCP1A	Delta-IDR-LLPS	67.81	34.04
Delta-IDR-RAD23B	Delta-IDR-LLPS	53.68	42.03
Delta-IDR-G3BP1	Delta-IDR-LLPS	35.51	42.9

VII.III Determinación de los aminoácidos abundantes y carentes en las regiones intrínsecamente desordenadas de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.

Una vez construida la base de datos se comparó la composición de aminoácidos en el grupo IDR-LLPS versus el grupo Delta-IDR-LLPS para poder determinar cuáles eran aquellos aminoácidos abundantes y carentes en las IDR de cada proteína. En las Figuras 7 - 12 se muestran los resultados de la comparación. A continuación, se describirá el análisis realizado en cada una

de las IDR pertenecientes a las proteínas que experimentan LLPS en respuesta a estrés osmótico.

En la Figura 7 se observa que los aminoácidos Arg, Asn, Asp, Glu, Leu, Met y Ser son más abundantes en la IDR que en la región de la proteína YAP que no es IDR. A este tipo de aminoácidos se les denominó aminoácidos “abundantes”. Por otro lado, se observó que la IDR de la proteína YAP carece del aminoácido Cys. A este tipo de aminoácidos se les denominó aminoácidos “carentes”.

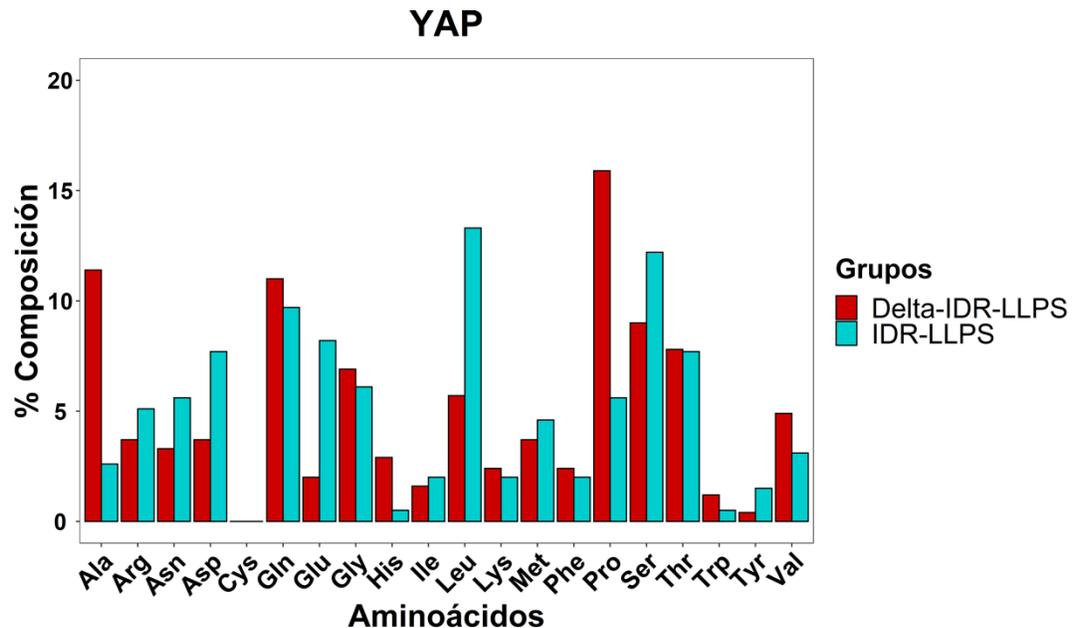


Figura 7. Composición de aminoácidos en la región IDR-LLPS (barras azules) y en la región Delta-IDR-LLPS (barras rojas) de la proteína YAP. Cada aminoácido se representa mediante el código de tres letras.

En la Figura 8 se observa que los aminoácidos abundantes en la IDR de la proteína DCP1A son Ala, Gln, Leu, Lys, Ser, Thr y Val. Por otro lado, los aminoácidos carentes en la IDR de la proteína DCP1A son Trp y Cys.

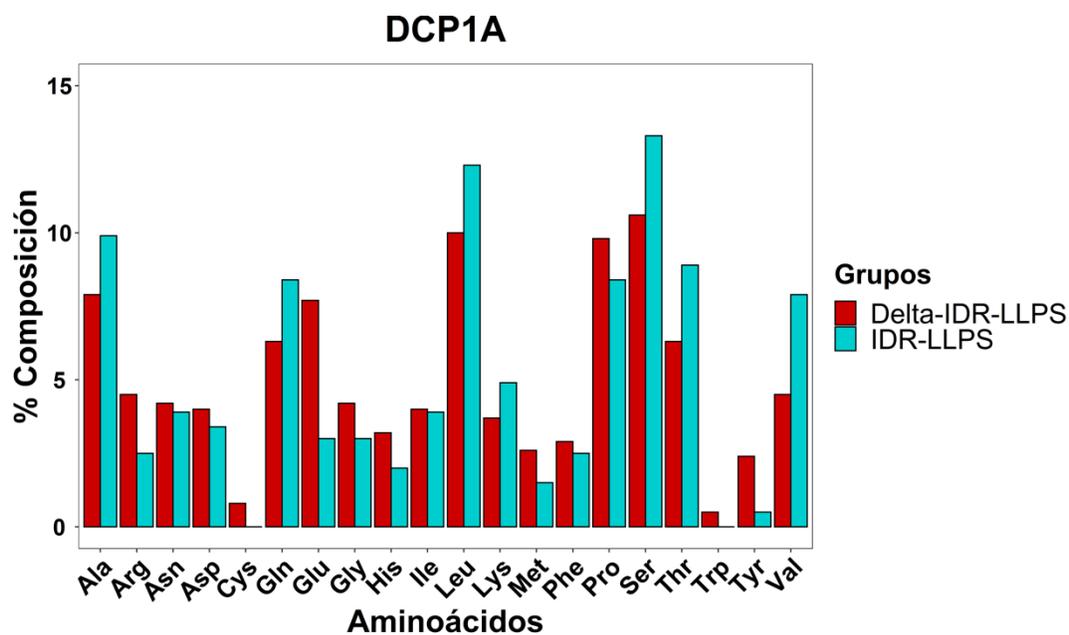


Figura 8. Composición de aminoácidos en la región IDR-LLPS (barras azules) y en la región Delta-IDR-LLPS (barras rojas) de la proteína DCP1A. Cada aminoácido se representa mediante el código de tres letras.

En la Figura 9 se observa que los aminoácidos abundantes en la IDR (UBA1) de la proteína RAD23B son Asn, Arg, Met, Ser y Thr. Por otro lado, los aminoácidos carentes en la IDR de la proteína RAD23B son Cys, His, Lys y Trp.

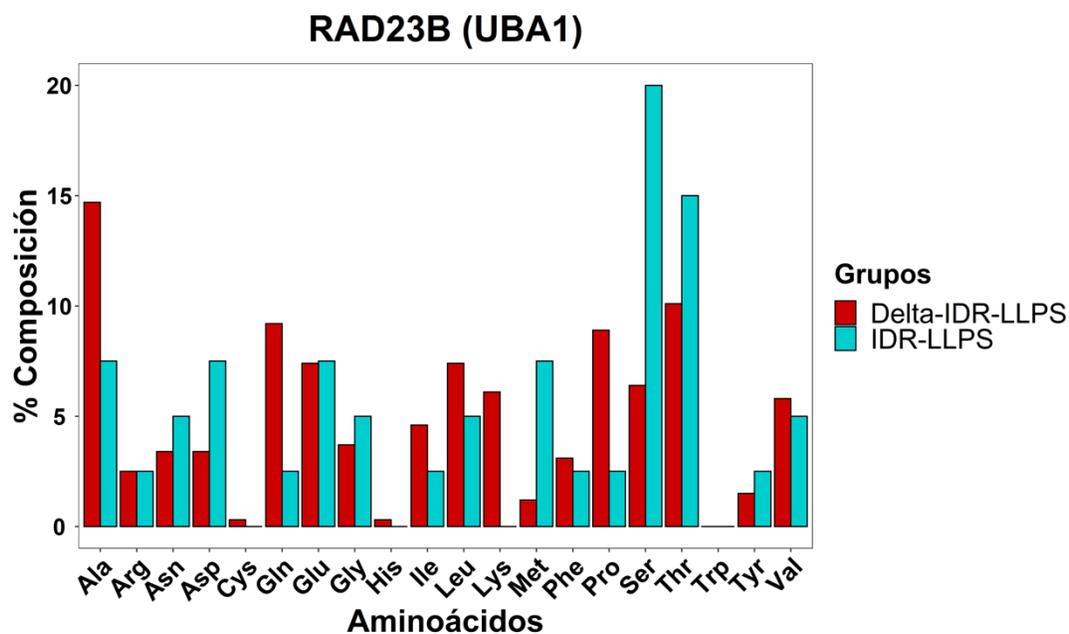


Figura 9. Composición de aminoácidos en la región IDR-LLPS (barras azules) y en la región Delta-IDR-LLPS (barras rojas) de la proteína RAD23B (UBA1). Cada aminoácido se representa mediante el código de tres letras.

En la Figura 10 se observa que los aminoácidos abundantes en la IDR (UBA2) de la proteína RAD23B son Asn, Gln, Glu, Gly, His, Ile y Met. Por otro lado, los aminoácidos carentes en la IDR de la proteína RAD23B son Arg, Asp, Cys, Lys, Thr, Trp.

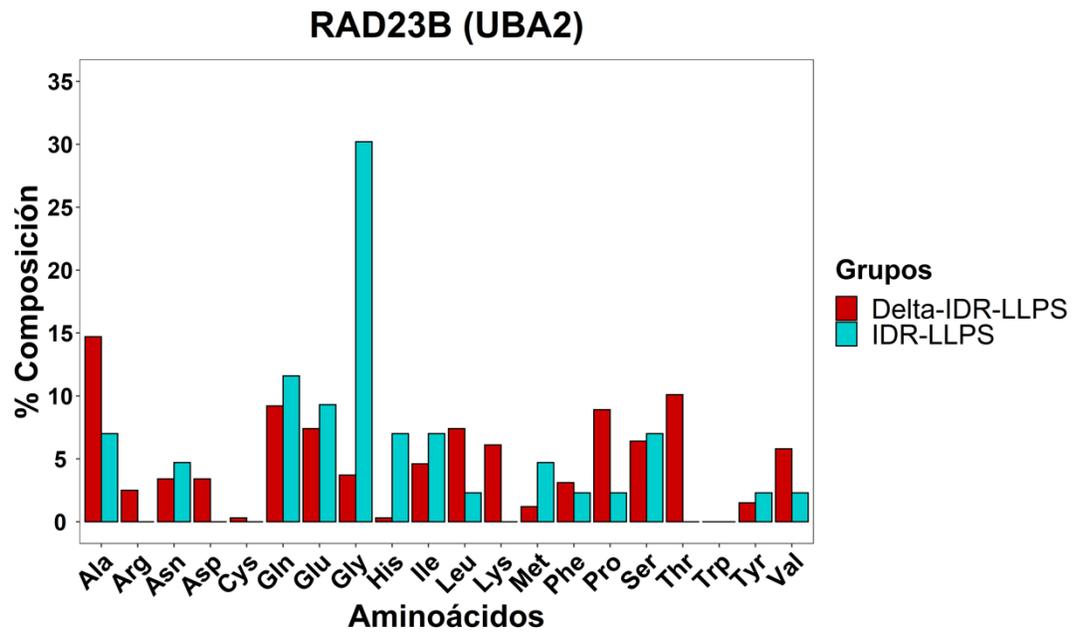


Figura 10. Composición de aminoácidos en la región IDR-LLPS (barras azules) y en la región Delta-IDR-LLPS (barras rojas) de la proteína RAD23B (UBA2). Cada aminoácido se representa mediante el código de tres letras.

En la Figura 11 se observa que los aminoácidos enriquecidos en la IDR (RG) de la proteína G3BP1 son Arg, Gly, Met y Pro. Por otro lado, los aminoácidos carentes en la IDR de la proteína G3BP1 son Ala, Asn, Asp, Cys, Glu, His, Ile, Ser, Thr, Trp y Tyr.

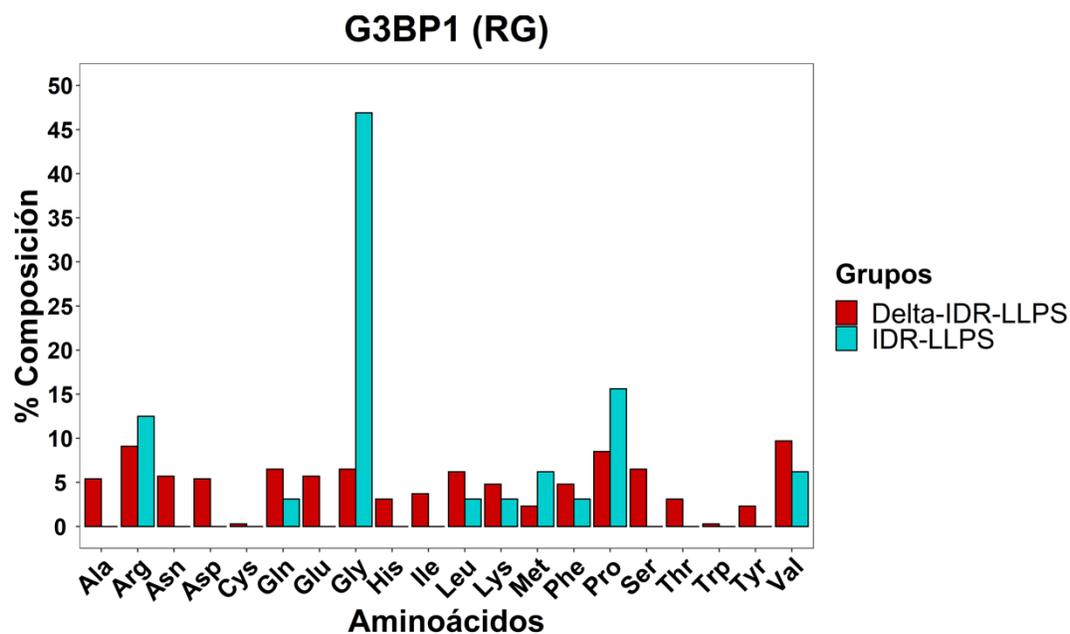


Figura 11. Composición de aminoácidos en la región IDR-LLPS (barras azules) y en la región Delta-IDR-LLPS (barras rojas) de la proteína G3BP1 (RG). Cada aminoácido se representa mediante el código de tres letras.

En la Figura 12 se observa que los aminoácidos abundantes en la IDR (región ácida) de la proteína G3BP1 son Glu y Pro. Por otro lado, los aminoácidos carentes en la IDR de la proteína G3BP1 son Cys y Trp.

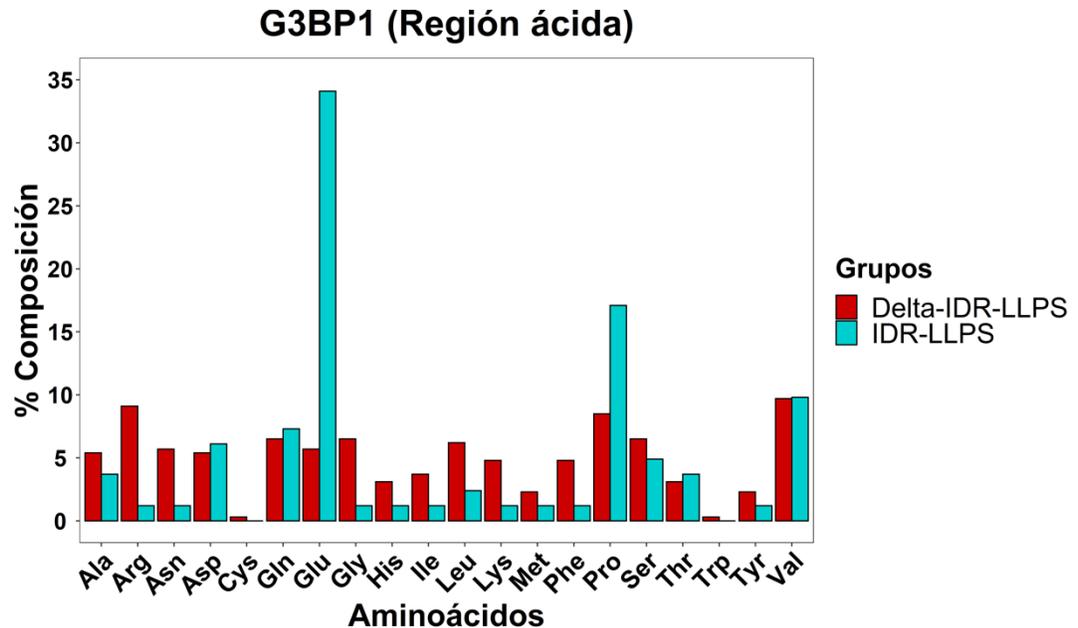


Figura 12. Composición de aminoácidos en la región IDR-LLPS (barras azules) y en la región Delta-IDR-LLPS (barras rojas) de la proteína G3BP1 (Región ácida). Cada aminoácido se representa mediante el código de tres letras.

Los análisis anteriores nos permitieron encontrar que aminoácidos son abundantes y carentes en las IDR. Se esperaba en un inicio que los aminoácidos abundantes y carentes fuesen similares en todas las IDR, pero fueron particulares para cada una.

Para poder determinar cuáles de estos aminoácidos eran significativamente abundantes en el grupo IDR-LLPS con respecto al grupo Delta-IDR-LLPS, fue necesario definir un parámetro nuevo. El parámetro, denominado como “enriquecimiento”, indica qué tan enriquecido está un

aminoácido en la región IDR-LLPS con respecto al Delta-IDR-LLPS. La ecuación para calcular este parámetro es la siguiente.

$$\text{Enriquecimiento} = 1 - \left(\frac{\%aa_{\Delta IDR}}{\%aa_{IDR}} \right)$$

En la Tabla 3 se muestran los resultados de enriquecimiento de algunos aminoácidos (Ala, Arg, Asn, Asp, Cys, Glu, Gln y Gly) obtenidos para cada IDR.

Tabla 3. Porcentajes de enriquecimiento para cada aminoácido en las IDR de las proteínas analizadas.

Proteína	Ala	Arg	Asn	Asp	Cys	Glu	Gln	Gly
YAP	-3.385	0.275	0.411	0.519	-	0.756	-0.134	-0.131
DCP1A	0.202	-0.800	-0.077	-0.176	-	-1.567	0.250	-0.400
RAD23B (UBA1)	-0.960	0.0	0.320	0.547	-	0.013	-2.680	0.260
RAD23B (UBA2)	-1.100	-	0.277	-	-	0.204	0.207	0.877
G3BP1(Región ácida)	-0.459	-6.583	-3.750	0.115	-	0.833	0.110	-4.417
G3BP1 (IDR-RG)	-	0.272	-	-	-	-	-1.097	0.861

Los guiones representan aquellos aminoácidos carentes para cada proteína.

Una vez obtenidos los valores de enriquecimiento para cada aminoácido, se definió el siguiente punto de corte: aquellos aminoácidos que tuviesen un valor mayor o igual a 0.4 de enriquecimiento se consideraron aminoácidos abundantes significativamente en la IDR. Un enriquecimiento de 0.4 indica que el aminoácido de interés se encuentra enriquecido 0.4 veces más en la región IDR-LLPS con respecto al Delta-IDR-LLPS, es decir, 40 % más. Asumimos que el que un aminoácido estuviese 0.5 veces más enriquecido en la región desordenada con respecto al resto de la secuencia tenía una razón evolutiva importante para la función de la proteína, sin embargo, elegimos el valor de corte de 0.4 para hacer más flexible el parámetro de enriquecimiento y poder tener un mayor número de aminoácidos candidatos a ser abundantes. Además del “enriquecimiento”, para considerar a un aminoácido como abundante de forma significativa en el grupo IDR-LLPS se consideró el “porcentaje de composición”, que se refiere al porcentaje de

cada aminoácido en cada secuencia. Para este porcentaje se estableció que un aminoácido con un valor mayor al 5% se considera como abundante en la IDR.

Los aminoácidos que cumplieron con el criterio de enriquecimiento se enlistaron y posteriormente se graficó su porcentaje de composición por pares. Las líneas punteadas en el gráfico indican el parámetro de corte de composición de 5 % para aminoácidos. Por otro lado, también se graficaron por pares los aminoácidos carentes, sin embargo, para estos no se estableció un valor de corte ya que su composición es del 0 %. Los gráficos correspondientes a las demás proteínas se encuentran en el Anexo I.

En la Figura 13 se muestran los aminoácidos Asp y Met, aminoácidos que cumplieron con los criterios de enriquecimiento y composición para la IDR (UBA1) de la proteína RAD23B como aminoácidos abundantes, observándose que ambos se encuentran en mayor composición en la IDR con respecto al resto de la secuencia.

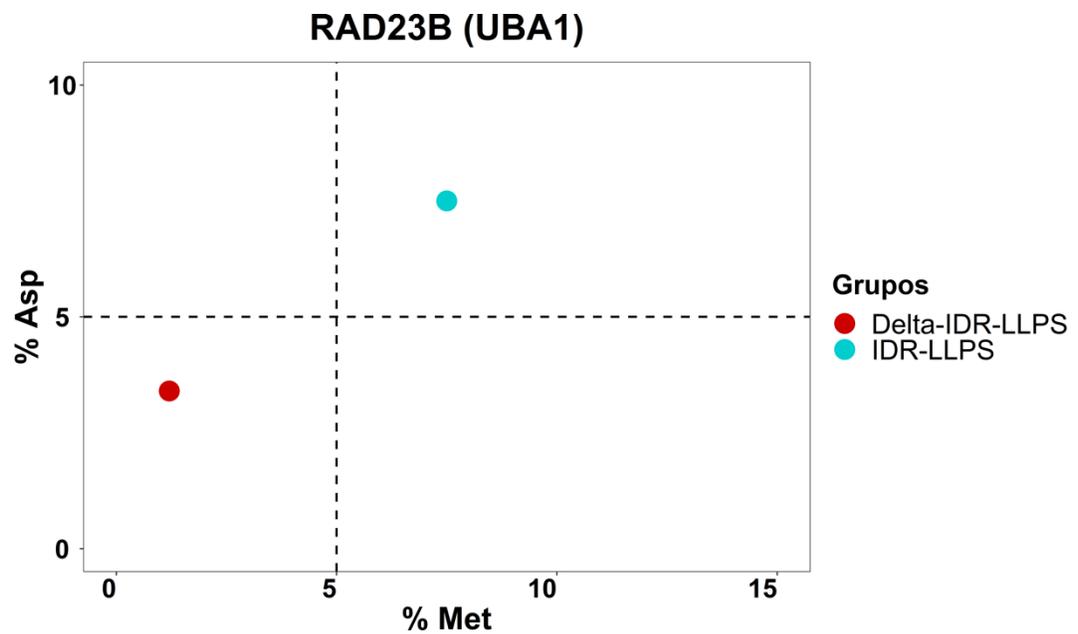


Figura 13. Porcentaje de Asp y porcentaje de Met, aminoácidos abundantes en la IDR UBA1 de RAD23B.

En la Figura 14 se muestran los aminoácidos His y Cys, cuya composición es del 0% en la IDR respecto al resto de la secuencia, definiéndolos por lo tanto como aminoácidos carentes en la región desordenada UBA1 de RAD23B.

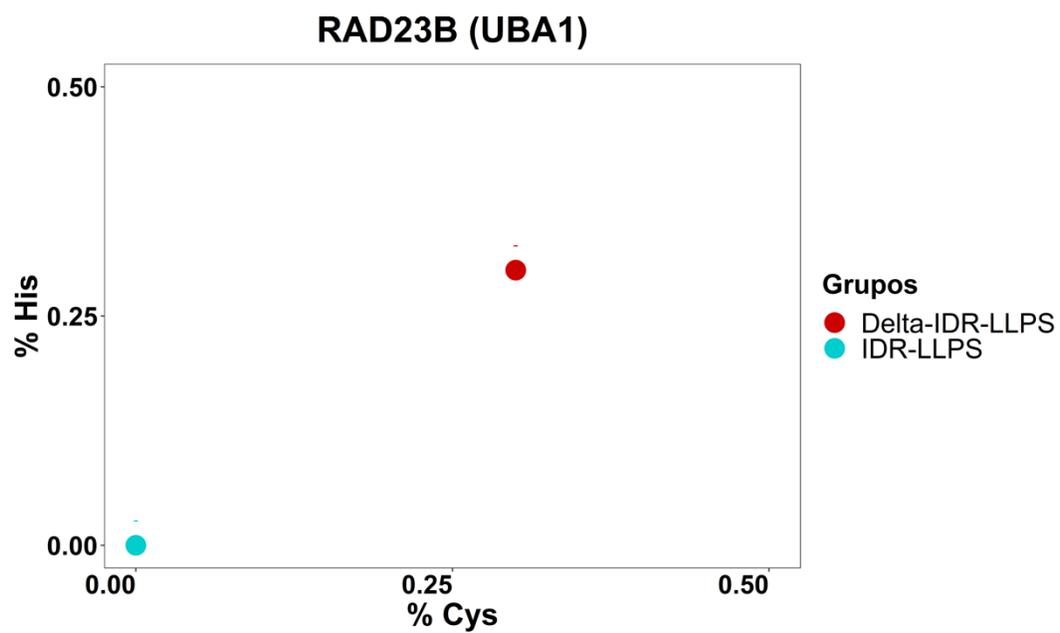


Figura 14. Porcentaje de His y porcentaje de Cys, aminoácidos carentes en la IDR UBA1 de RAD23B.

En la Figura 15 se muestran los aminoácidos Gly y His, aminoácidos que cumplieron con los criterios de enriquecimiento y composición para la IDR (UBA2) de la proteína RAD23B como aminoácidos abundantes, observándose que ambos se encuentran en mayor composición en la IDR con respecto al resto de la secuencia.

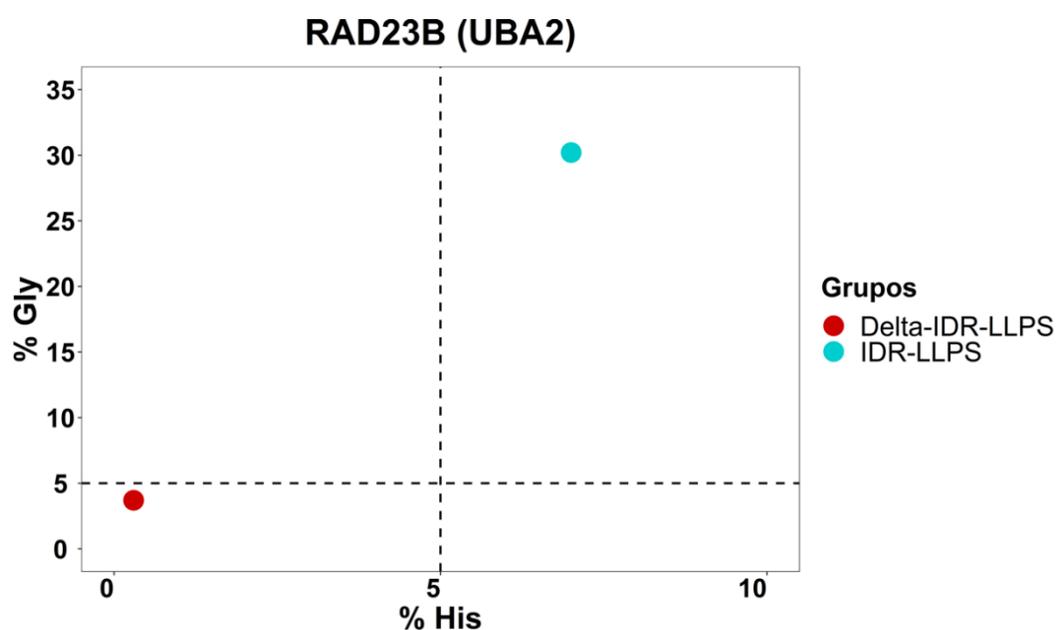


Figura 15. Porcentaje de Gly y porcentaje de His, aminoácidos abundantes en la IDR UBA2 de RAD23B.

En la Figura 16 se muestran los aminoácidos Thr y Cys, cuya composición es del 0% en la IDR respecto al resto de la secuencia, definiéndolos por lo tanto como aminoácidos carentes en la región desordenada UBA2 de RAD23B.

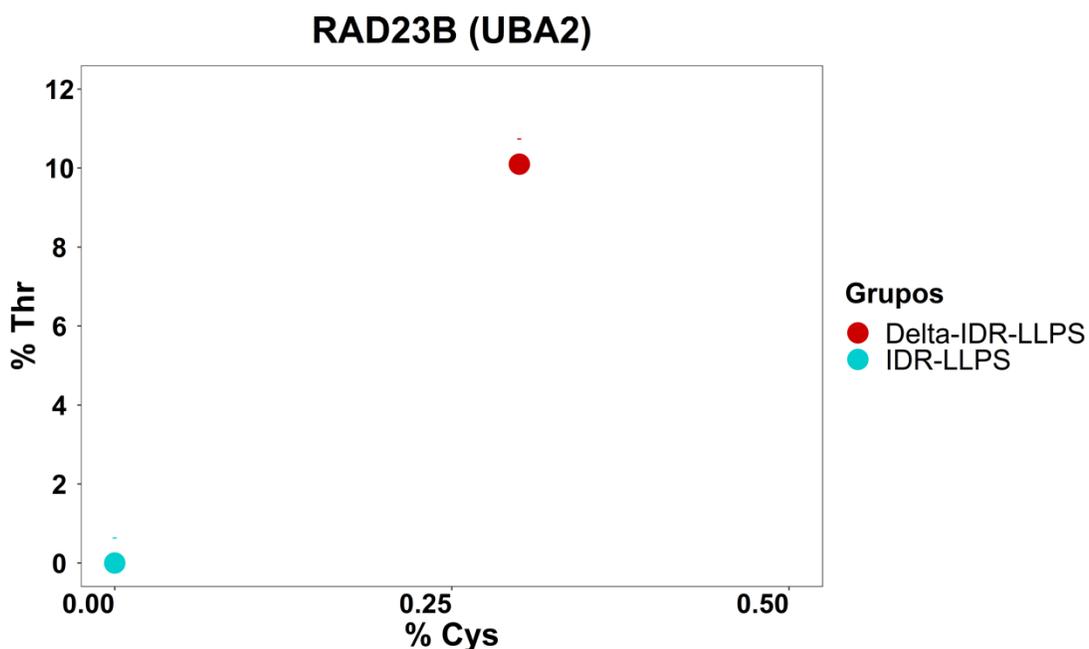


Figura 16. Porcentaje de Thr y porcentaje de Cys, aminoácidos carentes en la IDR UBA2 de RAD23B.

Como se observó en las Figuras 13 y 15 los aminoácidos abundantes encontrados para las IDR (UBA1 y UBA2) de la proteína RAD23B, sí cumplieron con los parámetros de enriquecimiento y composición, además, en las Figuras 14 y 16 se muestra la composición nula de los aminoácidos carentes en la IDR, previamente observada en los gráficos de barras. Por otro lado, para la proteína G3BP1, el grupo IDR-LLPS no cumplió con el porcentaje de composición, aun así, al verse una diferencia notoria en cuanto a composición con respecto al grupo Delta-IDR-LLPS, los aminoácidos

encontrados se consideraron como abundantes. En la Figura 17 y Figura 18 se ejemplifica lo descrito para la región ácida y la región RG de la proteína G3BP1, respectivamente. En las Figuras 19 y 20 se muestra la composición nula de los aminoácidos carentes en la IDR, previamente observada en los gráficos de barras.

En la Figura 17 se muestran los aminoácidos Glu y Pro, aminoácidos que se consideraron como abundantes en la región ácida de G3BP1 a pesar de que el grupo Delta-IDR-LLPS tiene una composición también arriba del 5 % para ambos aminoácidos, ya que, aun así, la diferencia de composición entre ambos grupos fue muy notoria.

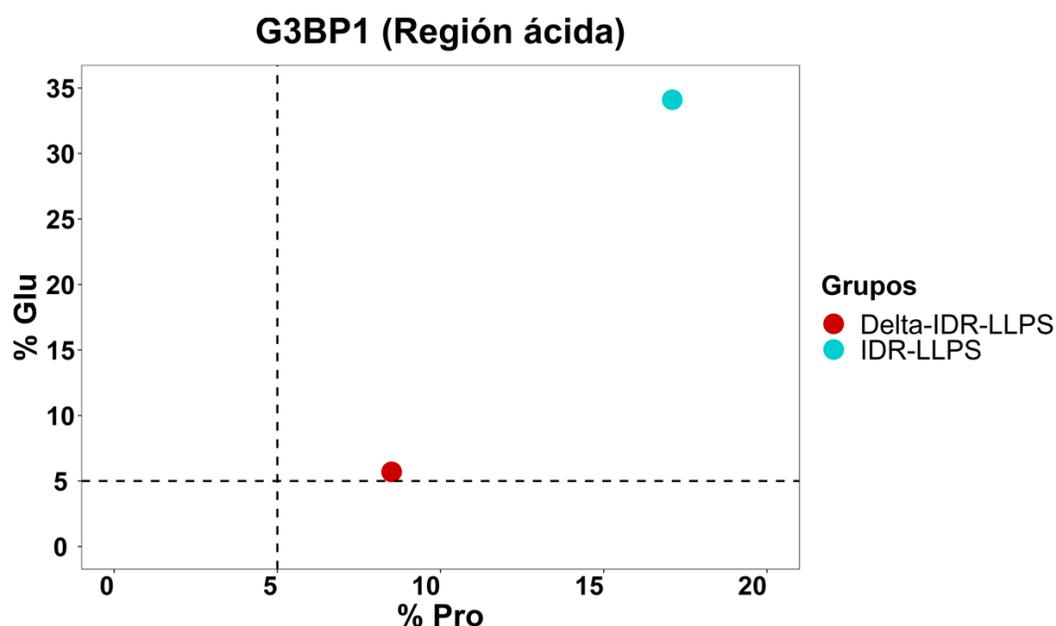


Figura 17. Porcentaje de Glu y porcentaje de Pro, aminoácidos abundantes en la región ácida de G3BP1.

En la Figura 18 se muestran los aminoácidos Gly y Arg, aminoácidos que se consideraron como abundantes en la IDR RG de G3BP1 a pesar de que el grupo Delta-IDR-LLPS tiene una composición también arriba del 5 % para ambos aminoácidos, ya que, aun así, la diferencia de composición entre ambos grupos fue muy notoria. Cabe señalar que el aminoácido Arg tuvo un enriquecimiento de 0.272, estando por debajo del punto de corte (0.4), sin embargo, al ser un aminoácido característico de esta región desordenada y tener una mayor composición en el grupo IDR-LLPS con respecto a Delta-IDR-LLPS, se considera también como abundante.

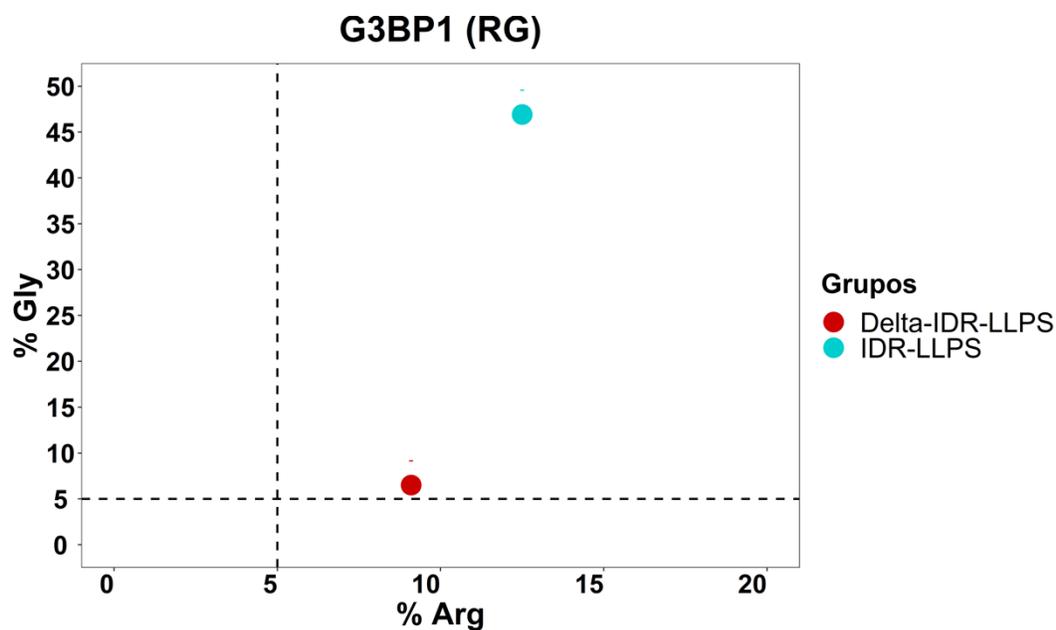


Figura 18. Porcentaje de Gly y porcentaje de Arg, aminoácidos abundantes en la IDR RG de G3BP1.

En la Figura 19 se muestran los aminoácidos Trp y Cys, cuya composición es del 0% en la IDR respecto al resto de la secuencia, definiéndolos por lo tanto como aminoácidos carentes en la región ácida de G3BP1.

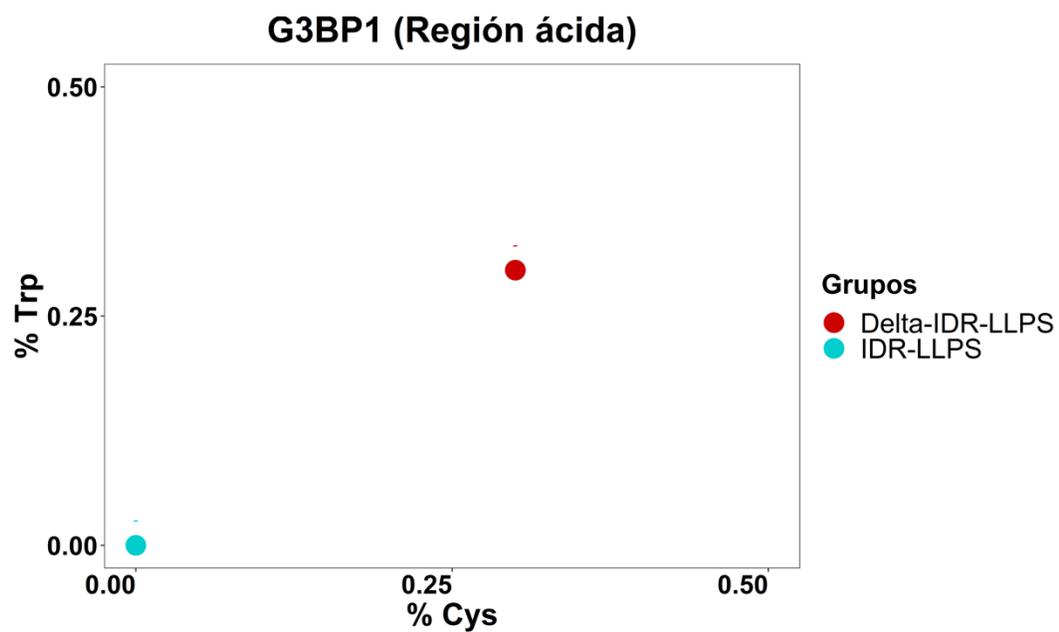


Figura 19. Porcentaje de Trp y porcentaje de Cys, aminoácidos carentes en la región ácida de G3BP1.

En la Figura 20 se muestran los aminoácidos Trp y Cys, cuya composición es del 0% en la IDR respecto al resto de la secuencia, definiéndolos por lo tanto como aminoácidos carentes en la IDR RG de G3BP1.

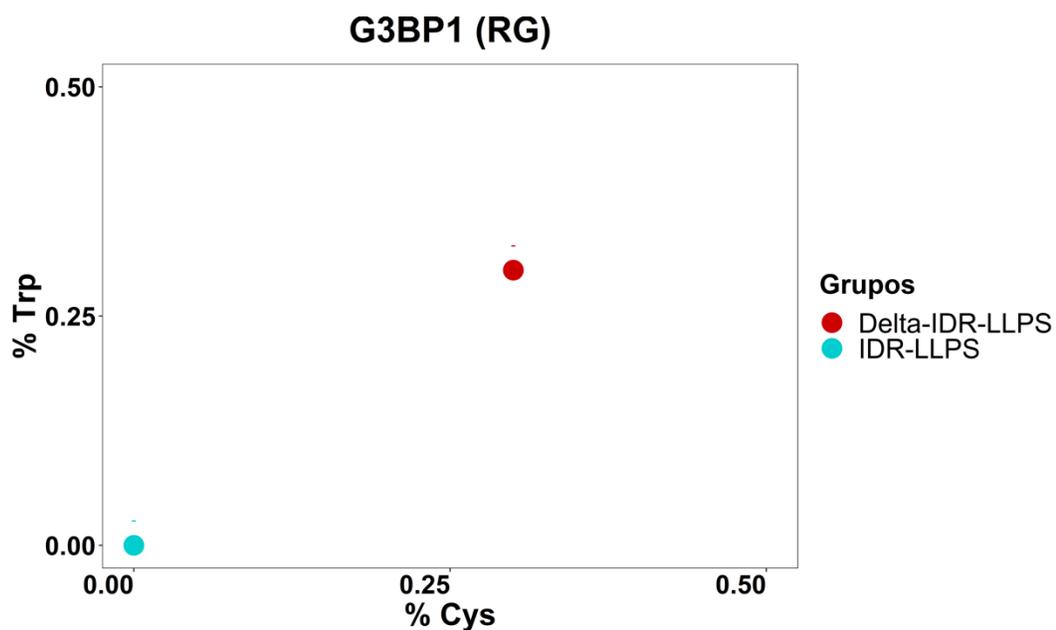


Figura 20. Porcentaje de Trp y porcentaje de Cys, aminoácidos carentes en la IDR RG de G3BP1.

Aplicando este mismo análisis a las proteínas restantes se obtuvieron aquellos aminoácidos abundantes y carentes para cada una. En la Tabla 4 se muestran los aminoácidos significativamente abundantes y los aminoácidos carentes encontrados en las IDR de cada proteína.

Tabla 4. Aminoácidos abundantes y carentes en las IDR de las proteínas que experimentan LLPS en respuesta a estrés osmótico en células humanas.

Proteína (IDR)	Aminoácidos abundantes en IDR	Aminoácidos carentes en IDR
YAP	Asn, Asp, Glu	Cys
DCP1A	Val	Trp, Cys
RAD23B (UBA1)	Asp, Met	Lys, Cys, His, Trp
RAD23B (UBA2)	Gly, His	Lys, Thr, Arg, Asp, Cys, Trp
G3BP1 (Región ácida)	Glu, Pro	Trp, Cys
G3BP1 (IDR-RG)	Gly, Pro, Arg, Met	Ala, Asn, Asp, Cys, Glu, Ser, His, Ile, Thr, Trp, Tyr

VII.IV Búsqueda de proteínas de *Arabidopsis thaliana* ortólogas a las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico.

El siguiente objetivo consistió en determinar si los aminoácidos abundantes y carentes en las IDR de las proteínas que experimentan LLPS en humanos se encuentran conservados en los ortólogos de la planta modelo *Arabidopsis thaliana*, con la hipótesis de que aquellos ortólogos que cumplan con lo anterior serían posibles candidatos a experimentar LLPS en respuesta a estrés osmótico.

Mediante el uso de herramientas como BLASTp y Phytozome, se buscaron proteínas ortólogas a las proteínas YAP, DCP1A, RAD23B y G3BP1 en la planta modelo *Arabidopsis thaliana*. Los ortólogos encontrados se caracterizaron por su porcentaje de desorden, estructura y composición de aminoácidos y se incluyeron en la base de datos, denominado a este nuevo grupo como “AtHomólogos”. Los ortólogos encontrados se resumen en la Tabla 5.

Tabla 5. Ortólogos en *A. thaliana* a proteínas que experimentan LLPS bajo condiciones de estrés osmótico.

Proteína LLPS	Ortólogos en <i>Arabidopsis thaliana</i> (AGI)
YAP	*
DCP1A	At1g08370
RAD23B	At1g16190.1 At3g02540.1 At5g16090 At5g38470 At1g79650
G3BP1	At2g03640 At1g13730 At1g69250 At5g60980 At3g25150 At5g48650 At5g43960

*No existen ortólogos para YAP.

Debido a que las proteínas globulares generalmente no experimentan LLPS (Zhou *et al.*, 2018), se incluyeron y caracterizaron un grupo aleatorio de veinte proteínas globulares de *Arabidopsis thaliana* a la base de datos, denominado a este nuevo grupo como “AtGlobulares”, con la hipótesis de que estas proteínas tendrán propiedades de desorden y estructura distintas al grupo AtHomólogos y al grupo LLPS, así como para los respectivos grupos IDR, distinguiendo a las proteínas que tienen la capacidad de experimentar el fenómeno de LLPS de las que no, funcionando por lo tanto como un grupo control. En la Tabla 6 se resumen las proteínas globulares mencionadas.

Tabla 6. Proteínas globulares de *Arabidopsis thaliana* seleccionadas para este estudio.

Proteína globular	AGI	Proteína globular	AGI
G6PD3	AT1G24280	ENO1	AT1G74030
HEX03	AT1G65590	PGK3	AT1G79550
CAO	AT1G44446	LPCAT2	AT1G63050
PP2AA1	AT1G25490	MGL	AT1G64660
LPCAT1	AT1G12640	PKP3	AT1G32440
APR2.	AT1G62180	PME18	AT1G11580
GLYK	AT1G80380	PMI2	AT1G67070
KAS2	AT1G74960	GLPK	AT1G80460
AGC1-7	AT1G79250	HT1	AT1G62400
GAPCP1	AT1G79530	GGAT1	AT1G23310

Respecto a lo anterior, se determinó a nivel global si efectivamente existía una diferencia en cuanto a desorden y estructura entre los grupos LLPS, AtHomólogos y AtGlobulares.

En la Figura 21 se representa el resultado de la comparación, en la cual se observó que el grupo AtHomólogos comparte características de desorden y estructura muy similares al grupo LLPS, separándose ambos grupos de las proteínas globulares. Esto sugiere que los ortólogos son posibles candidatos a experimentar LLPS bajo condiciones de estrés osmótico. Por otro lado, también se demostró con esto que el grupo AtGlobulares se comporta como grupo control de proteínas que probablemente no tienen la capacidad de experimentar LLPS, el cual se utilizó para el resto de la investigación.

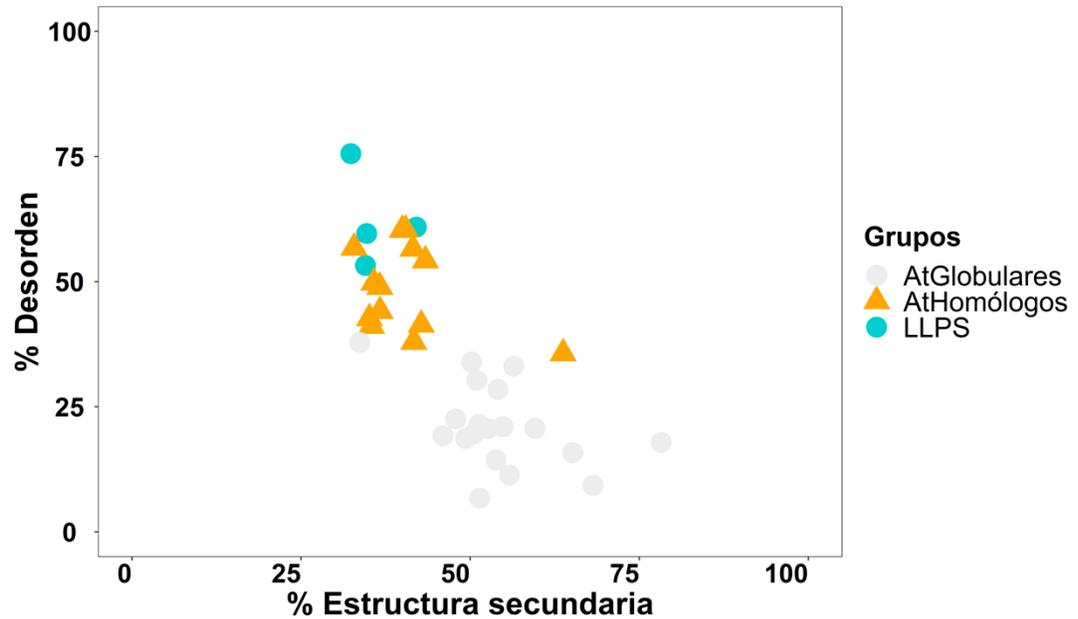


Figura 21. Comparación del contenido predicho de desorden y estructura secundaria entre el grupo LLPS, AtHomólogos y AtGlobulares.

Sabiendo que los ortólogos comparten propiedades de estructura y desorden similares al grupo LLPS, distinguiéndose del grupo control, el siguiente paso fue encontrar las regiones intrínsecamente desordenadas del grupo AtHomólogos, teniendo en cuenta que las IDR son indispensables para la formación de condensados biomoleculares del grupo LLPS. Utilizando la herramienta T-COFFEE, se alinearon y analizaron las secuencias de las proteínas del grupo LLPS con sus respectivos ortólogos de *A. thaliana*. A través de este análisis, se determinó el dominio correspondiente a la IDR en los ortólogos. En la Figura 22 se muestra el alineamiento entre RAD23B y uno de sus ortólogos, resaltándose en amarillo los aminoácidos que conforman ambas regiones intrínsecamente desordenadas de la proteína (UBA1 y UBA2). Como se puede observar en la Figura 22, existen niveles bajos de conservación de la secuencia primaria de las IDR de RAD23B y las IDR de los ortólogos. Los alineamientos correspondientes a las demás proteínas se encuentran en el Anexo II.

```

AT5G38470      -TASVPVPTSGTATAAAPATAASVQTDVYGQAASNLVAGTTLESTVQQIL
RAD23B        KPAETPVATSPTATDS--TSGDSSRSNLFEDATSALVTGQSYENMVTEIM
               .*.***.* * * * :  :. * : : : : : * : * * * : * . * : * :

AT5G38470      DMGGGSWDRDRTVVRALRAAFNNPERAVEYLYSGIPAQAEIPPVAQAPATG
RAD23B        SMG---YEREQVIAALRASFNNDRAVEYLLMGIPGDRESQAVVDPP---
               .**      : : * : * : * : * : * : * : * : * : * : *

AT5G38470      EQAANPLAQFQQAAPAAATGGPNANPLNLFQGMPAADAGAGAGNLDLFL
RAD23B        -QAASTGAPQSSAVAAAAAT-----TTATTTTTTSSGGHPLEFL
               ***. . * . . * . * . * * * * * * * * * * * * * * *

AT5G38470      RNSQQFQALRAMVQANPQILQPMLQELGKQNPQLVRLIQEHQADFLRLIN
RAD23B        RNQPFQFQMRQIIQQNPSLLPALLQQIGRENPLLQQISQHQEHFIQMLN
               ** . * * * : * : : * * * . * : * : * : * : * : * : *

AT5G38470      EPVE-----GEENVMEQLEAAMPQAVTVTPEEREAIERLEGMGF
RAD23B        EPVQEAGGQGGGGGGGGGGIAEAGS-GHMNYIQVTPQEKEAIERLKALGF
               ***:          * . : * . . : : * : * : * : * : * : *

AT5G38470      DRAMVLEVFFACNKNEELAANYLLDHMHFEFDQ
RAD23B        PEGLVIQAYFACEKNENLAANFLLQQNFD-E-D
               . : * : : : * : * : * : * : * : * : : : * :

```

Figura 22. Alineamiento de RAD23B y su ortólogo en *A. thaliana*. En amarillo se resaltan las regiones IDR de RAD23B.

Conociendo la secuencia de la IDR en los ortólogos, el siguiente paso fue caracterizar sus niveles de desorden, estructura y composición, denominando a este nuevo grupo como IDR-AtHomólogos.

Nuevamente comparamos el porcentaje de desorden y estructura, pero ahora entre los grupos IDR-LLPS, IDR-AtHomólogos y AtGlobulares. El resultado de esta comparación es consistente con el de la Figura 21, teniendo los primeros dos grupos características similares de desorden y estructura secundaria, diferenciándose notablemente del grupo de proteínas globulares. Esto sustenta la idea de que las proteínas ortólogas son posibles candidatos a experimentar LLPS en respuesta a estrés osmótico. La Figura 23 muestra el resultado obtenido en esta comparación.

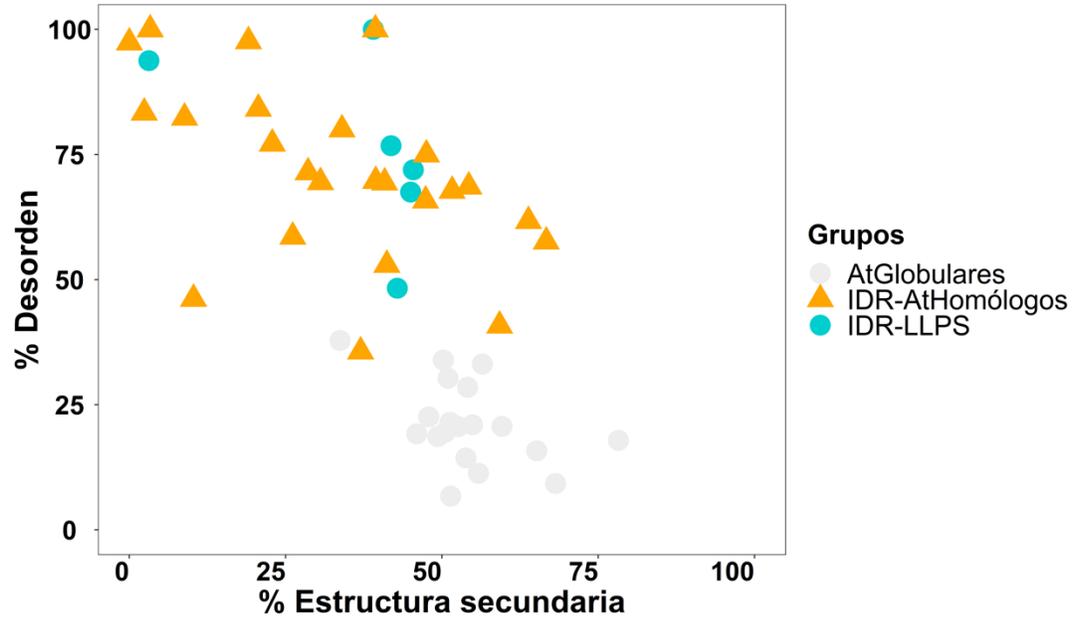


Figura 23. Comparación del contenido predicho de desorden y estructura secundaria entre el grupo IDR-LLPS, IDR-AtHomólogos y AtGlobulares.

VII.V Comparación de la composición de aminoácidos de las regiones intrínsecamente desordenadas entre las proteínas de humanos y los ortólogos de *Arabidopsis thaliana*.

Debido a que se ha reportado que la composición de aminoácidos es importante para la función de las IDR en diversos sistemas (Martin y Mittaj, 2018 ; Wang *et al.*, 2018; Lin *et al.*, 2017), el siguiente análisis consistió en determinar si la composición de aminoácidos resumidos en la Tabla 4 para el grupo IDR-LLPS se encontraba conservada en cada ortólogo en el grupo IDR-AtHomólogos. Para lograr esto, se graficaron por pares los aminoácidos de la Tabla 4 para cada IDR, utilizando al grupo AtGlobulares como control. Además, en las Tablas 7-10 se resume cuáles son aquellos aminoácidos conservados como abundantes y carentes para cada ortólogo y si se predicen según nuestro análisis de conservación como candidatos a promover LLPS. Para esto último, los ortólogos debían tener conservados la mitad de los

aminoácidos abundantes y obligatoriamente Cys y Trp conservados como carentes para poder considerarse como candidatos a experimentar LLPS.

A continuación, en las Figuras 24 - 31, se muestran los resultados obtenidos para RAD23B y G3BP1. Los gráficos correspondientes a las demás proteínas se encuentran en el Anexo III.

En la Figura 24 se muestran aquellos aminoácidos abundantes en la región intrínsecamente desordenada UBA1 de la proteína RAD23B. De los cuatro ortólogos, el aminoácido Asp se encuentra conservado en sólo uno, teniendo una composición por encima del 5 %, por otro lado, el aminoácido Met sólo está conservado en dos de los ortólogos teniendo ambos una composición mayor al 5 %.

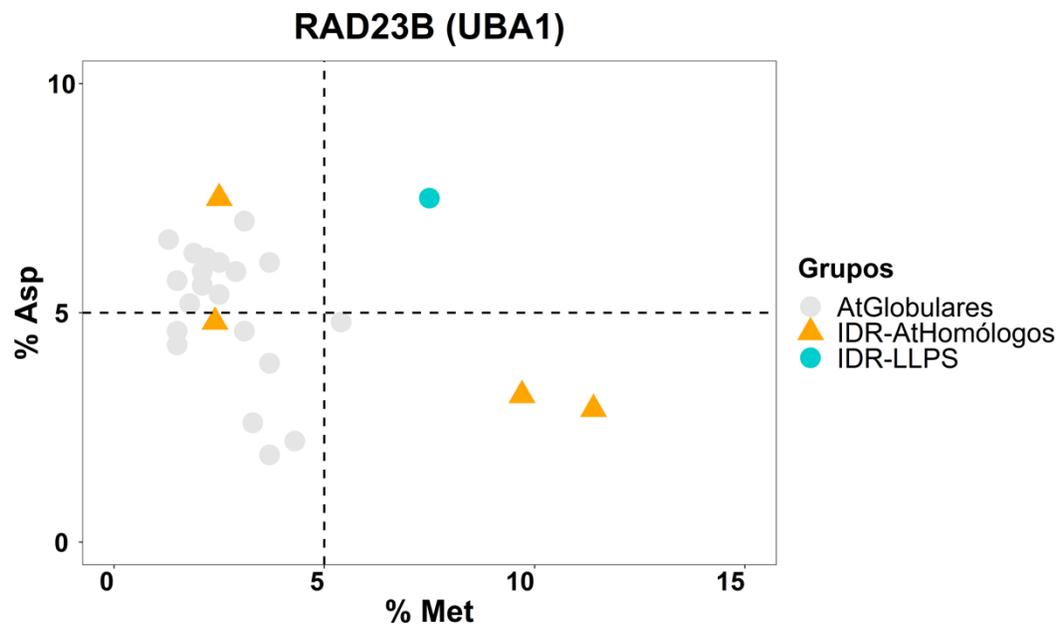


Figura 24. Porcentaje de Asp y porcentaje de Met, aminoácidos abundantes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína RAD23B (UBA1).

En la Figura 25 se muestran aquellos aminoácidos carentes en la región UBA1 de RAD23B, observándose que los ortólogos tienen una composición idéntica (0 %) al grupo IDR-LLPS, distinguiéndose ambos del grupo control. Por lo tanto, His y Cys se consideran como aminoácidos carentes del grupo IDR-LLPS y dicha carencia se conserva en el grupo IDR-AtHomólogos.

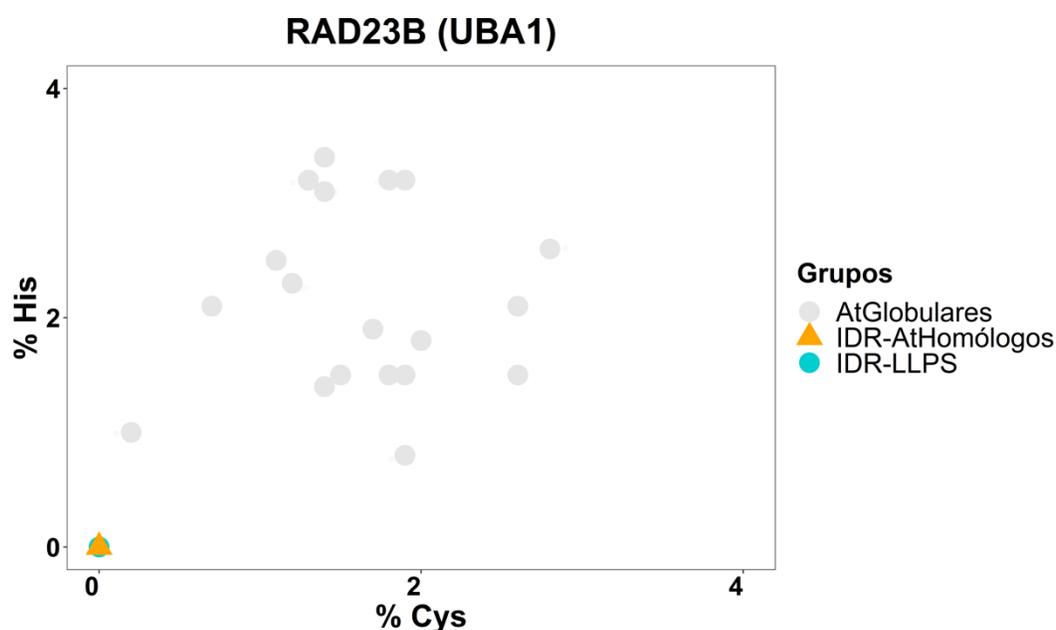


Figura 25. Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogos de la proteína RAD23B (UBA1).

La Tabla 7 se construyó con los resultados del análisis de las dos Figuras anteriores, en la cual se resume para cada ortólogo cuales son aquellos aminoácidos del grupo IDR-LLPS conservados en el grupo IDR-AtHomólogos. En la última columna se menciona si el ortólogo es candidato a promover LLPS dependiendo de la cantidad de aminoácidos conservados como abundantes y/o carentes de acuerdo con lo mencionado al inicio de esta sección. Este mismo análisis se hizo para el resto de las proteínas.

En la Tabla 7 se muestran aquellos aminoácidos abundantes y carentes del grupo IDR-LLPS conservados o no en el grupo IDR-AtHomólogos, de los cuales, el ortólogo At5g38470 no se considera como candidato a promover LLPS ya que ninguno de los aminoácidos abundantes esta conservado.

Tabla 7. Aminoácidos conservados como abundantes o carentes en los ortólogos de la proteína RAD23B (UBA1).

Ortólogo	Abundantes		Carentes				Candidato a promover LLPS
	Asp	Met	Lys	Cys	His	Trp	
At1g16190.1	NO	SI	SI	SI	SI	SI	SI
At3g02540.1	SI	NO	SI	SI	SI	SI	SI
At5g38470	NO	NO	SI	SI	SI	SI	NO
At1g79650	NO	SI	SI	SI	SI	SI	SI

En la Figura 26 se muestran aquellos aminoácidos abundantes en la región intrínsecamente desordenada UBA2 de la proteína RAD23B. De los cinco ortólogos, el aminoácido Gly se encuentra conservado en tres, teniendo una composición por encima del 5 %, por otro lado, el aminoácido His no está conservado en ninguno de los ortólogos, teniendo todos una composición menor al 5 %.

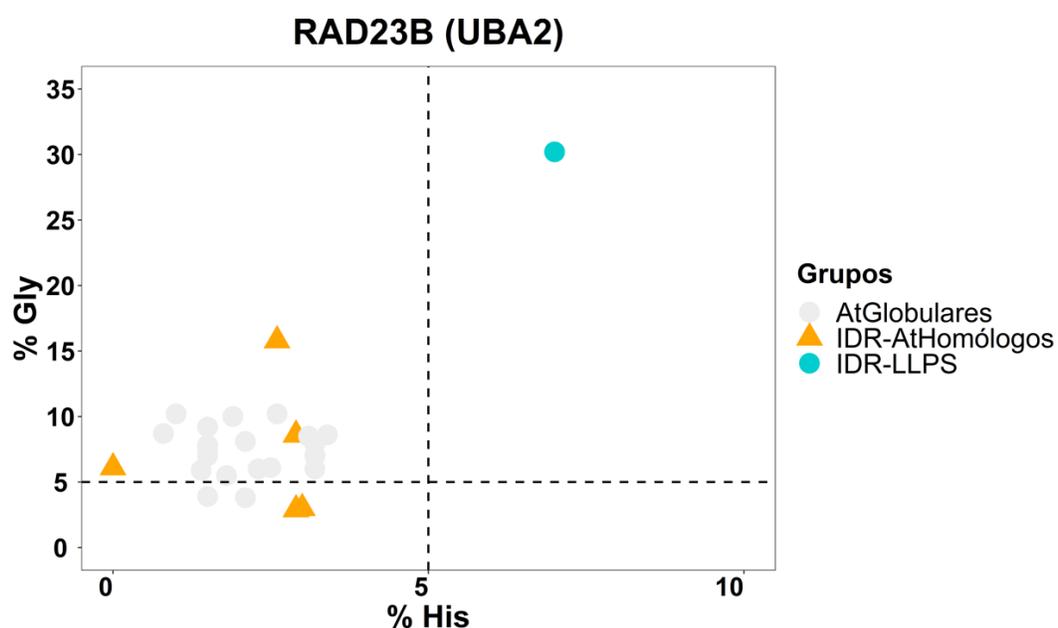


Figura 26. Determinación de la conservación de aminoácidos abundantes del grupo IDR-LLPS en el grupo IDR-AtHomólogos de la proteína RAD23B (UBA2).

En la Figura 27 se muestran aquellos aminoácidos carentes en la región UBA2 de RAD23B, observándose que los ortólogos tienen una composición idéntica (0 %) al grupo IDR-LLPS, distinguiéndose ambos del grupo control. Por lo tanto, Thr y Cys se consideran como aminoácidos del grupo IDR-LLPS conservados en el grupo IDR-AtHomólogos.

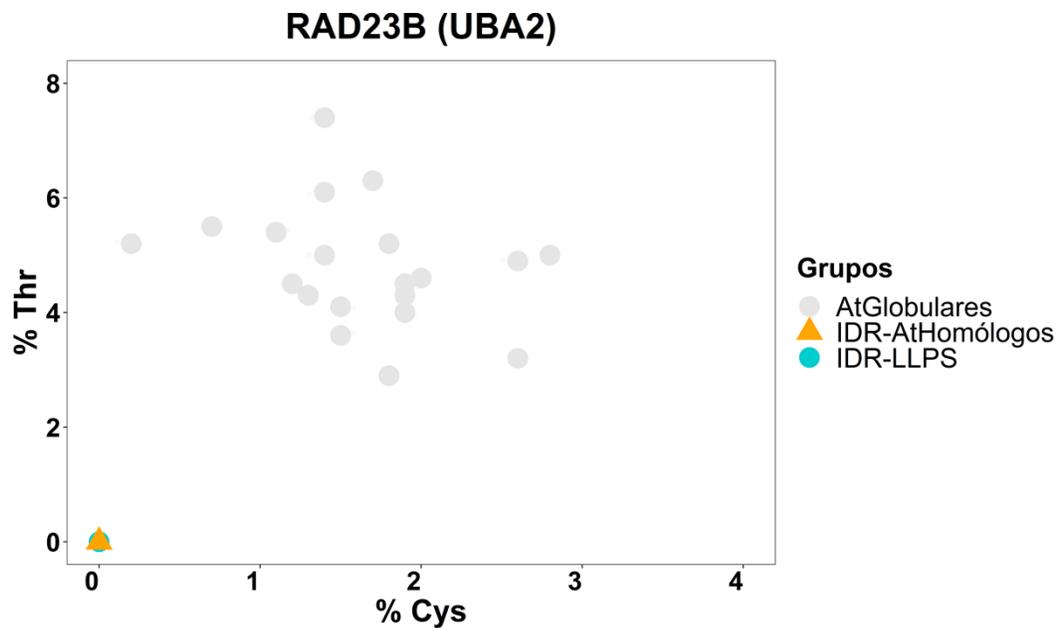


Figura 27. Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogos de la proteína RAD23B (UBA2).

En la Tabla 8 se muestran aquellos aminoácidos abundantes y carentes del grupo IDR-LLPS conservados o no en el grupo IDR-AtHomólogos, de los cuales, los ortólogos At5g16090 y At5g38470 no se consideran como candidatos a promover LLPS ya que ninguno de los aminoácidos abundantes está conservado.

Tabla 8. Aminoácidos conservados como abundantes o carentes en los ortólogos de la proteína RAD23B (UBA2).

Ortólogo	Abundantes		Carentes						Candidato a promover LLPS
	Gly	His	Lys	Thr	Arg	Asp	Cys	Trp	
At1g16190.1	SI	NO	SI	SI	SI	NO	SI	SI	SI
At3g02540.1	SI	NO	SI	SI	NO	NO	SI	SI	SI
At5g16090	NO	NO	NO	SI	NO	NO	SI	SI	NO
At5g38470	NO	NO	SI	SI	NO	NO	SI	SI	NO
At1g79650	SI	NO	SI	SI	SI	NO	SI	SI	SI

En la Figura 28 se muestran aquellos aminoácidos abundantes en la región intrínsecamente desordenada ácida de la proteína G3BP1, observándose que los ortólogos se distinguen del grupo de proteínas globulares al tener una mayor composición en Glu y Pro, por tal motivo ambos aminoácidos del grupo IDR-LLPS se consideran como conservados en el grupo IDR-AtHomólogos.

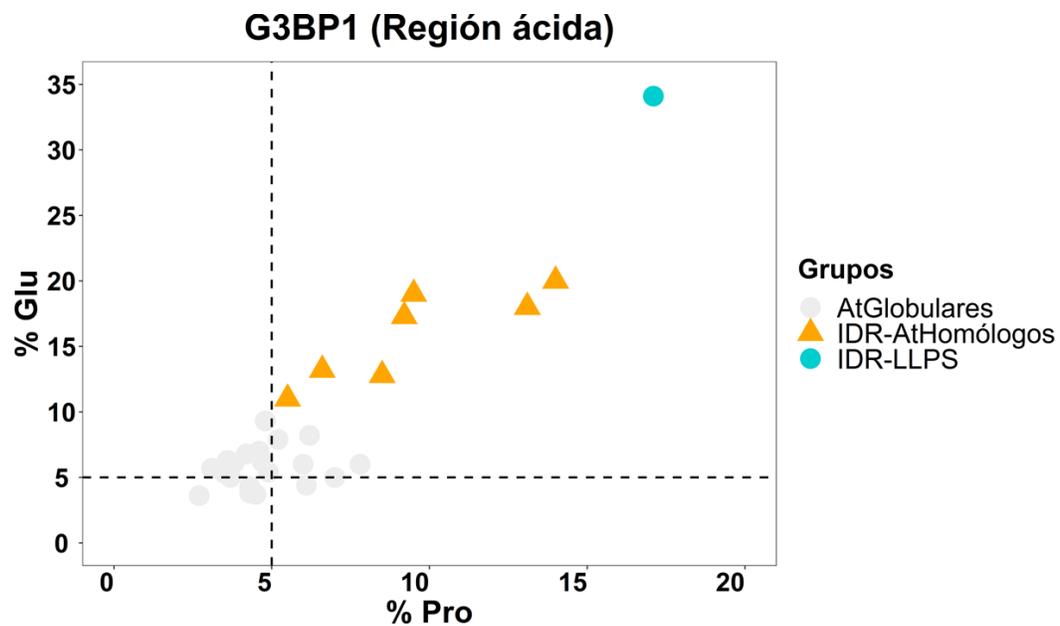


Figura 28. Determinación de la conservación de aminoácidos abundantes del grupo IDR-LLPS en el grupo IDR-AtHomólogos de la proteína G3BP1 (Región ácida).

En la Figura 29 se muestran aquellos aminoácidos carentes en la región ácida de G3BP1, observándose que los ortólogos tienen una composición idéntica (0 %) al grupo IDR-LLPS para Trp, sin embargo, para Cys los ortólogos At1g69250 y At5g48650 no cumplen.

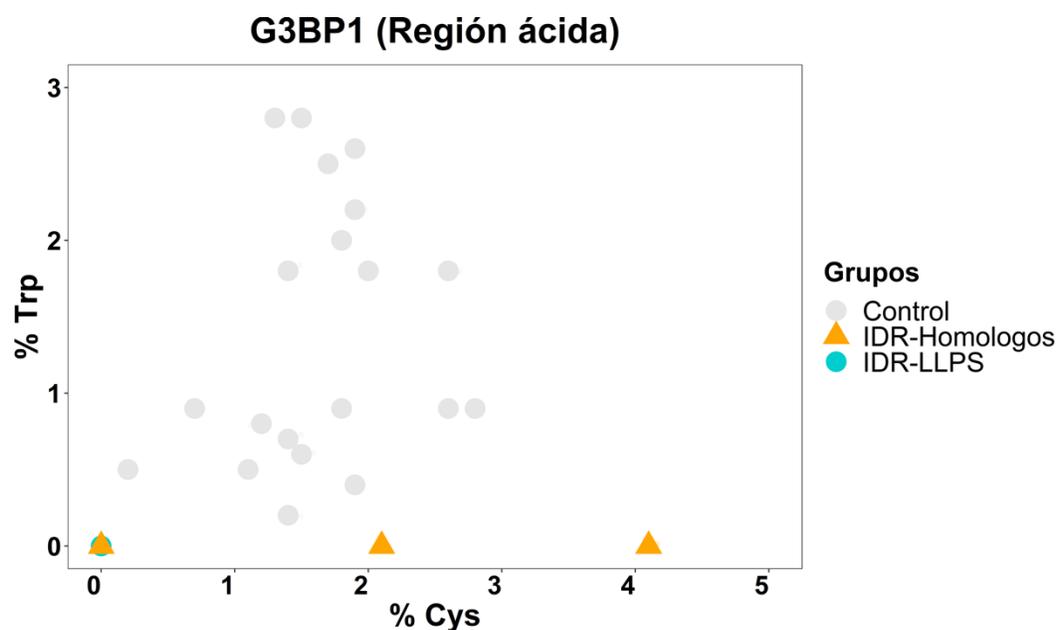


Figura 29. Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (Región ácida).

En la Tabla 9 se muestran aquellos aminoácidos abundantes y carentes del grupo IDR-LLPS conservados o no en el grupo IDR-AtHomólogos, de los cuales, los ortólogos At1g69250 y At5g48650 no se consideran como candidatos a promover LLPS ya que el aminoácido Cys no está conservado como carente en ambos.

Tabla 9. Aminoácidos conservados como abundantes y carentes en los ortólogos de la proteína G3BP1 (Región ácida).

Ortólogo	Abundantes		Carentes		Candidato a promover LLPS
	Glu	Pro	Cys	Trp	
At2g03640	SI	SI	SI	SI	SI
At1g13730	SI	SI	SI	SI	SI
At1g69250	SI	SI	NO	SI	NO
At5g60980	SI	SI	SI	SI	SI
At3g25150	SI	SI	SI	SI	SI
At5g48650	SI	SI	NO	SI	NO
At5g43960	SI	SI	SI	SI	SI

La Figura 30 muestra aquellos aminoácidos abundantes en la región intrínsecamente desordenada rica en arginina y glicina de la proteína G3BP1, observándose que todos los ortólogos a excepción de uno están igualmente enriquecidos en Arg y Gly, separándose notoriamente de las proteínas globulares. Por tal motivo ambos aminoácidos del grupo IDR-LLPS se consideran conservados en el grupo IDR-AtHomólogos.

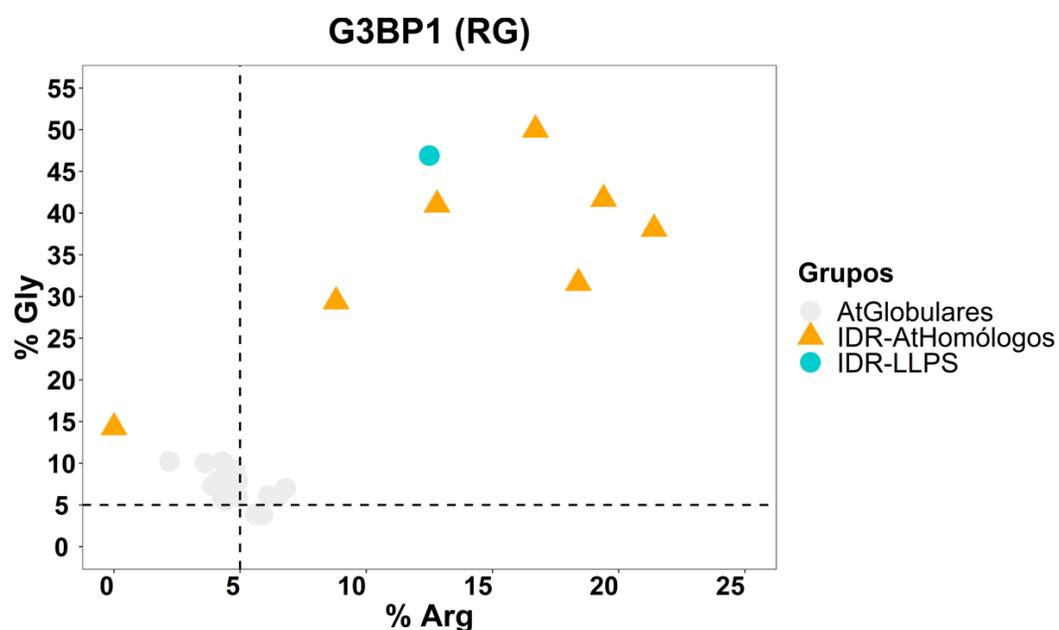


Figura 30. Determinación de la conservación de aminoácidos abundantes en el grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (RG).

En la Figura 31 se muestran aquellos aminoácidos carentes en la región RG de G3BP1, observándose que los ortólogos tienen una composición idéntica (0 %) al grupo IDR-LLPS, distinguiéndose ambos del grupo control. Por lo tanto, Trp y Cys se consideran como aminoácidos del grupo IDR-LLPS conservados en el grupo IDR-AtHomólogos.

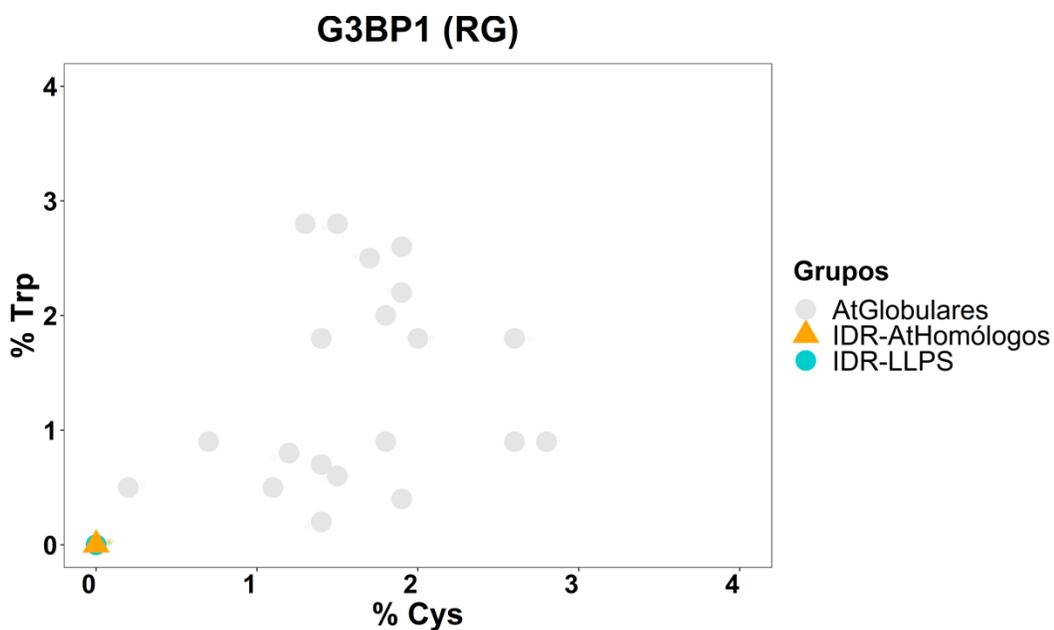


Figura 31. Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (RG).

En la Tabla 10 se muestran aquellos aminoácidos abundantes y carentes del grupo IDR-LLPS conservados o no en el grupo IDR-AtHomólogos, de los cuales, el ortólogo At1g69250 no se considera como candidato a promover LLPS, ya que, de los cuatro aminoácidos abundantes, no están conservados tres.

Tabla 10. Aminoácidos conservados como abundantes y carentes en los ortólogos de la proteína G3BP1 (RG).

Ortólogo	Abundantes				Carentes											Candidato a promover LLPS
	Gly	Pro	Arg	Met	Ala	Asn	Asp	Cys	Glu	Ser	His	Ile	Thr	Trp	Tyr	
At2g03640	SI	NO	SI	NO	NO	NO	NO	SI	NO	NO	NO	SI	NO	SI	NO	SI
At1g13730	SI	NO	SI	NO	NO	NO	NO	SI	NO	NO	NO	SI	SI	SI	NO	SI
At1g69250	SI	NO	NO	NO	NO	NO	NO	SI	NO	NO	SI	SI	NO	SI	SI	NO
At5g60980	SI	NO	SI	NO	NO	NO	SI	SI	NO	NO	SI	SI	SI	SI	NO	SI
At3g25150	SI	NO	SI	NO	NO	NO	NO	SI	SI	NO	SI	SI	NO	SI	NO	SI
At5g48650	SI	NO	SI	NO	SI	NO	NO	SI	SI	NO	SI	SI	SI	SI	NO	SI
At5g43960	SI	SI	SI	NO	NO	NO	NO	SI	NO	NO	SI	SI	NO	SI	NO	SI

Una vez obtenidos los resultados de predicción para las IDR de cada ortólogo como candidatas a promover LLPS según nuestro análisis, procedimos a determinar los niveles de predicción de experimentar LLPS utilizando un predictor conocido como PSPredictor y, de esta manera, determinar si la conservación en la composición de aminoácidos correlaciona con la predicción. A continuación, en las Tablas 11 – 13 se muestra la comparación entre los ortólogos candidatos a experimentar LLPS, así como su valor de PSP Score según el predictor.

En la Tabla 11 se muestra la comparación entre la predicción de nuestro análisis para el ortólogo de la proteína DCP1A contra PSPredictor. Como se puede observar, para el ortólogo At1g08370 existe una correlación entre el análisis de composición y el PSPredictor, sugiriendo que sí es candidato a experimentar separación de fases líquido-líquido al igual que DCP1A.

Tabla 11. Predicción de la capacidad de experimentar separación de fases líquido-líquido en respuesta a estrés osmótico para DCP1A.

Proteína	Ortólogo	Candidato LLPS	PSP Score (¿LLPS?)
DCP1A	At1g08370	SI	0.8741 (SI)

En la Tabla 12 se muestra la comparación entre la predicción de nuestro análisis para los ortólogos de la proteína RAD23B contra PSPredictor. Como se puede observar, para todos los ortólogos, excepto At5g38470, existe una correlación entre el análisis de composición y el PSPredictor, sugiriendo aquellos ortólogos que sí son candidatos a experimentar separación de fases líquido-líquido al igual que RAD23B.

Tabla 12. Predicción de la capacidad de experimentar separación de fases líquido-líquido en respuesta a estrés osmótico para RAD23B.

Proteína	UBA1	UBA2	Candidato	PSP Score (¿LLPS?)
	Ortólogo	Ortólogo		
RAD23B	At1g16190.1	At1g16190.1	SI	0.9588 (SI)
	At3g02540.1	At3g02540.1	SI	0.957 (SI)
	-	At5g16090	NO	0.007 (NO)
	At5g38470	At5g38470	NO	0.9666 (SI)
	At1g79650	At1g79650	SI	0.9495 (SI)

En la Tabla 13 se muestra la comparación entre la predicción de nuestro análisis para los ortólogos de la proteína G3BP1 contra PSPredictor. Como se puede observar, para los ortólogos, excepto At1g13730, At5g48650 y At5g43960, existe una correlación entre el análisis de composición y el PSPredictor, sugiriendo aquellos ortólogos que sí son candidatos a experimentar separación de fases líquido-líquido al igual que G3BP1.

Tabla 13. Predicción de la capacidad de experimentar separación de fases líquido-líquido en respuesta a estrés osmótico para G3BP1.

Proteínas	RG	Ácido	Candidato	PSP Score (¿LLPS?)
	Ortólogo	Ortólogo		
G3BP1	At2g03640	At2g03640	SI	0.9709
	At1g13730	At1g13730	SI	0.1058
	At1g69250	At1g69250	NO	0.0316
	At5g60980	At5g60980	SI	0.9774
	At3g25150	At3g25150	SI	0.8274
	At5g48650	At5g48650	NO	0.968
	At5g43960	At5g43960	SI	0.4959

VIII. DISCUSIÓN.

Muchas proteínas celulares tienen la capacidad de formar condensados moleculares ricos en biomoléculas (proteínas y RNA), que subyacen a la formación de compartimentos subcelulares a través del fenómeno de LLPS, que además de llevar a cabo diversas funciones fisiológicas, también están relacionados a su vez con enfermedades debido a errores en su ensamblaje (Bondos *et al.*, 2022 ; Kanekura y Kuroda, 2022 ; Shukla *et al.*, 2022 ; Guillén-Chable *et al.*, 2021 ; Borgia *et al.*, 2018), de tal forma que resulta muy importante el estudiar y comprender la base molecular de este fenómeno (Saar *et al.*, 2021).

Considerando las características particulares de secuencia (grado de desorden y estructura) y composición de aminoácidos propias de las proteínas que tienen la capacidad de experimentar separación de fases líquido-líquido, en este proyecto de tesis mediante el análisis *in silico* de las proteínas del grupo LLPS, se demostró que cumplen con aquellas propiedades típicas de composición y desorden para las proteínas con la capacidad de formar condensados biomoleculares en respuesta a estrés osmótico. Esto también nos permitió determinar específicamente aquellos aminoácidos abundantes y carentes en sus regiones intrínsecamente desordenadas, siendo posiblemente esa composición particular la responsable de conferirle a dichas regiones de complejidad baja la capacidad de dirigir la formación de los condensados para responder al estrés osmótico.

Es importante señalar que, en nuestro estudio, las proteínas del grupo LLPS compartían como característica particular de composición la ausencia de Trp y Cys, sin embargo, diferían en composición de aminoácidos

abundantes de las IDR de YAP y DCP1A respecto a las IDR de RAD23B y G3BP1, ya que las dos primeras comparten la presencia de un aminoácido promotor del orden (Leu). Resulta interesante como perspectiva averiguar mediante estudios de mutaciones, si dicho aminoácido es relevante en la función de ambas proteínas.

En otros estudios, investigadores han sido capaces de determinar mediante el uso de bases de datos de proteínas desordenadas (DisProt) la composición particular de todos los IDR reportados en dicha base, siendo Gln, Ser, Pro, Lys y Glu los aminoácidos que más promueven el desorden. Además, otros aminoácidos como Tyr, Asn, Asp, Arg, Gly, Thr y His son también abundantes en algunos IDR (Liu *et al.*, 2019 ; Sickmeier *et al.*, 2007 ; Vacic *et al.*, 2007).

En este estudio también investigamos si las características de secuencia y composición de aminoácidos (abundantes y carentes) del grupo LLPS se encontraban conservadas en las IDR de sus ortólogos en *Arabidopsis thaliana* mediante el uso de herramientas bioinformáticas como T-COFFE y un software para el análisis gráfico (Rstudio). Nuestro análisis mostró que los ortólogos conservan las IDR del grupo LLPS, un grado de desorden y estructura similares, así como una conservación en cuanto a composición para la mayoría de los aminoácidos de interés. Este hallazgo indica hasta ahora que los ortólogos de *Arabidopsis thaliana* son posibles candidatos a experimentar LLPS bajo condiciones de estrés osmótico.

En relación con lo anterior, se utilizó también un predictor de LLPS para proteínas (PSPredictor), cuyos valores de predicción arrojados para experimentar LLPS o no, correlacionaron con la mayoría de nuestros resultados. Sin embargo, la razón por la que nuestra predicción para algunos ortólogos a experimentar LLPS no coincidió con el predictor, posiblemente se deba a que nuestro análisis no considera la presencia de otros aminoácidos en la composición de los ortólogos de cada proteína, los cuales podrían ser

los responsables de estos resultados particulares que no correlacionan con los de PSPredictor.

Con respecto a PSPredictor, es una herramienta útil para predecir si una proteína tiene una alta probabilidad a experimentar LLPS, sin embargo, a diferencia de nuestro método de predicción, esta herramienta no considera ninguna condición de estrés, únicamente se basa en el análisis de la secuencia de aminoácidos de la proteína. Esto es una limitante para el predictor, porque el que una proteína tenga una alta probabilidad de experimentar LLPS no quiere decir que esta forme condensados biomoleculares bajo distintas condiciones de estrés (temperatura, osmótico, oxidativo, etc.). Por otro lado, la metodología desarrollada en este proyecto, si nos permite determinar posibles candidatos a experimentar LLPS bajo una condición particular de estrés (osmótico), lo cual podría aplicarse para encontrar otras proteínas candidatas, en *Arabidopsis thaliana* o en algún otro organismo, a separarse en fases bajo otra condición de estrés aplicando la misma metodología.

Considerando que hasta ahora un número limitado de IDR se han caracterizado en modelos experimentales (Hatos *et al.*, 2020), es importante señalar que los nuevos métodos computacionales desarrollados o que están en desarrollo para predecir estas regiones directamente a partir de secuencias de proteínas, han dado lugar a una nueva forma de identificar e investigar funcionalmente los IDR (Ahmed *et al.*, 2022 ; French *et al.*, 2022 ; McFadden y Yanowitz, 2022 ; Zhao y Kurgan, 2022 ; Shen *et al.*, 2021 ; Liu *et al.*, 2019 ; Meng *et al.*, 2017 ; Cozzetto y Jones, 2013), siendo aún más viable que los métodos experimentales tradicionales debido a que es más eficiente y económica (Piovesan *et al.*, 2022 ; Fischer, 2003).

En la actualidad, a pesar de que se han desarrollado alrededor de 70 predictores de desorden en proteínas (Ohue *et al.*, 2013), aún queda un extenso trabajo que hacer para lograr elaborar herramientas bioinformáticas predictivas más precisas y eficientes que superen los desafíos y limitaciones

que derivan del creciente conjunto de datos de proteínas, así como de las características de estas (Lindorff-Larsen y Kragelund, 2021).

La estrategia que optamos emplear en esta investigación para la determinación de proteínas candidatas a separarse en fases en *Arabidopsis thaliana*, usando como base proteínas de otro organismo (*Homo sapiens*) pero que ya están bien caracterizadas en experimentar LLPS bajo ciertas condiciones de estrés, resultó útil al grado de arrojar resultados de “predicción” que coinciden con los de un predictor más complejo. Por lo tanto, nuestro trabajo aporta una metodología útil para la búsqueda y/o predicción de posibles proteínas con la capacidad de formar condensados biomoleculares que les permitan a los organismos contender contra condiciones de estrés osmótico. Sin embargo, cabe señalar que podría mejorarse nuestra metodología utilizando un grupo LLPS y un grupo control más extensos que nos permitan tener resultados aún más confiables y a mayor escala para la determinación de aminoácidos abundantes y carentes en las regiones desordenadas.

IX. CONCLUSIONES.

- Se encontró en la literatura que las proteínas humanas YAP, DCP1A, RAD23B y G3BP1 experimentan separación de fases líquido-líquido durante condiciones de estrés osmótico.
- Parámetros como la composición de aminoácidos, así como el nivel de desorden de las regiones intrínsecamente desordenadas de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico en células humanas, difieren a los del resto de su secuencia.
- Los aminoácidos abundantes en las regiones intrínsecamente desordenadas de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico en células humanas son Asn, Asp, Glu, Val, Met, Gly, His, Pro y Arg. Además, se encontró que los aminoácidos carentes en las regiones intrínsecamente desordenadas en todas las proteínas analizadas son Cys y Trp.
- Se identificaron los ortólogos de las proteínas DCP1A, RAD23B y G3BP1 en *Arabidopsis thaliana*.
- Los ortólogos de *Arabidopsis thaliana* conservan la región intrínsecamente desordenada de las proteínas que experimentan separación de fases líquido-líquido en respuesta a estrés osmótico en células humanas, así como la composición para la mayoría de los aminoácidos abundantes y carentes.
- Se predijo la capacidad de experimentar separación de fases líquido-líquido en los ortólogos de *Arabidopsis thaliana* en base a la conservación de la composición de aminoácidos.

- Se desarrolló una metodología para encontrar proteínas candidatas a experimentar separación de fases líquido-líquido en respuesta a estrés osmótico, basada en el análisis de conservación de aminoácidos en las regiones intrínsecamente desordenadas.

X. PERSPECTIVAS.

Estudiar aquellas proteínas candidatas en ensayos de estrés osmótico *in vivo* para verificar si experimentan LLPS bajo condiciones de estrés, lo cual a su vez nos permitiría validar los resultados de predicción arrojados por nuestra metodología *in silico*.

Aplicar nuestra metodología, pero a una mayor escala como sería el proteoma de *Arabidopsis thaliana* para la búsqueda de más proteínas candidatas a experimentar LLPS.

Realizar este mismo estudio, en otros organismos distintos a *Arabidopsis thaliana* como las levaduras o algún otro organismo cuyo número de proteínas que experimenten LLPS sea reducido.

XI. BIBLIOGRAFÍA.

1. Ahmed, S. S., Rifat, Z. T., Lohia, R., Campbell, A. J., Dunker, A. K., Rahman, M. S., & Iqbal, S. (2022). Characterization of intrinsically disordered regions in proteins informed by human genetic diversity. *PLOS Computational Biology*, *18*(3), e1009911. <https://doi.org/10.1371/journal.pcbi.1009911>
2. Alshareedah, I., Kaur, T., Ngo, J., Seppala, H., Kounatse, L. A. D., Wang, W., Moosa, M. M., & Banerjee, P. R. (2019). Interplay between short-range attraction and long-range repulsion controls reentrant liquid condensation of ribonucleoprotein–RNA Complexes. *Journal of the American Chemical Society*, *141*(37), 14593–14602. <https://doi.org/10.1021/jacs.9b03689>
3. Anderson, P., & Kedersha, N. (2009). Stress granules. *Current Biology*, *19*(10), R397–R398. <https://doi.org/10.1016/j.cub.2009.03.013>
4. Arimoto, K., Fukuda, H., Imajoh-Ohmi, S., Saito, H., & Takekawa, M. (2008). Formation of stress granules inhibits apoptosis by suppressing stress-responsive MAPK pathways. *Nature Cell Biology*, *10*(11), 1324–1332. <https://doi.org/10.1038/ncb1791>
5. Aumiller, W. M., Pir Cakmak, F., Davis, B. W., & Keating, C. D. (2016). RNA-based coacervates as a model for membraneless organelles: formation, properties, and interfacial liposome Assembly. *Langmuir*, *32*(39), 10042–10053. <https://doi.org/10.1021/acs.langmuir.6b02499>
6. Babinchak, W. M., & Surewicz, W. K. (2020). Liquid–liquid phase separation and its mechanistic role in pathological protein aggregation. *Journal of Molecular Biology*, *432*(7), 1910–1925. <https://doi.org/10.1016/j.jmb.2020.03.004>
7. Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, *18*(5), 285–298. <https://doi.org/10.1038/nrm.2017.7>
8. Banani, S. F., Rice, A. M., Peeples, W. B., Lin, Y., Jain, S., Parker, R., & Rosen, M. K. (2016). Compositional control of phase-separated cellular bodies. *Cell*, *166*(3), 651–663. <https://doi.org/10.1016/j.cell.2016.06.010>
9. Bondos, S. E., Dunker, A. K., & Uversky, V. N. (2022). Intrinsically disordered proteins play diverse roles in cell signaling. *Cell Communication and Signaling*, *20*(1). <https://doi.org/10.1186/s12964-022-00821-7>
10. Borgia, A., Borgia, M. B., Bugge, K., Kissling, V. M., Heidarsson, P. O., Fernandes, C. B., Sottini, A., Soranno, A., Buholzer, K. J., Nettels, D., Kragelund, B. B., Best, R. B., & Schuler, B. (2018). Extreme disorder in an ultrahigh-affinity protein complex. *Nature*, *555*(7694), 61–66. <https://doi.org/10.1038/nature25762>
11. Broucker, C., Thompson, D. C., & Vasiliou, V. (2012). The role of hyperosmotic stress in inflammation and disease. *BioMolecular Concepts*, *3*(4), 345–364. <https://doi.org/10.1515/bmc-2012-0001>

12. Buchan, J. R. (2014). mRNP granules. *RNA Biology*, 11(8), 1019–1030. <https://doi.org/10.4161/15476286.2014.972208>
13. Burkewitz, K., Choe, K., & Strange, K. (2011). Hypertonic stress induces rapid and widespread protein damage in *C. elegans*. *American Journal of Physiology-Cell Physiology*, 301(3), C566–C576. <https://doi.org/10.1152/ajpcell.00030.2011>
14. Cai, D., Feliciano, D., Dong, P., Flores, E., Gruebele, M., Porat-Shliom, N., Sukenik, S., Liu, Z., & Lippincott-Schwartz, J. (2019). Phase separation of YAP reorganizes genome topology for long-term YAP target gene expression. *Nature Cell Biology*, 21(12), 1578–1589. <https://doi.org/10.1038/s41556-019-0433-z>
15. Chahar, H. S., Chen, S., & Manjunath, N. (2013). P-body components LSM1, GW182, DDX3, DDX6 and XRN1 are recruited to WNV replication sites and positively regulate viral replication. *Virology*, 436(1), 1–7. <https://doi.org/10.1016/j.virol.2012.09.041>
16. Chong, P. A., & Forman-Kay, J. D. (2016). Liquid–liquid phase separation in cellular signaling systems. *Current Opinion in Structural Biology*, 41, 180–186. <https://doi.org/10.1016/j.sbi.2016.08.001>
17. Cozzetto, D., & Jones, D. T. (2013). The contribution of intrinsic disorder prediction to the elucidation of protein function. *Current Opinion in Structural Biology*, 23(3), 467–472. <https://doi.org/10.1016/j.sbi.2013.02.001>
18. Cuevas-Velazquez, C. L., & Dinneny, J. R. (2018). Organization out of disorder: liquid–liquid phase separation in plants. *Current Opinion in Plant Biology*, 45, 68–74. <https://doi.org/10.1016/j.pbi.2018.05.005>
19. 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>
20. Cumberworth, A., Lamour, G., Babu, M., & Gsponer, J. (2013). Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochemical Journal*, 454(3), 361–369. <https://doi.org/10.1042/bj20130545>
21. Dunker, A. K., Babu, M. M., Barbar, E., Blackledge, M., Bondos, S. E., Dosztányi, Z., Dyson, H. J., Forman-Kay, J., Fuxreiter, M., Gsponer, J., Han, K. H., Jones, D. T., Longhi, S., Metallo, S. J., Nishikawa, K., Nussinov, R., Obradovic, Z., Pappu, R. V., Rost, B., . . . Uversky, V. N. (2013). What’s in a name? Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins*, 1(1), e24157. <https://doi.org/10.4161/idp.24157>
22. Emara, M. M., & Brinton, M. A. (2007). Interaction of TIA-1/TIAR with west nile and dengue virus products in infected cells interferes with stress granule formation and processing body assembly. *Proceedings of the National Academy of Sciences*, 104(21), 9041–9046. <https://doi.org/10.1073/pnas.0703348104>
23. Escoté, X., Zapater, M., Clotet, J., & Posas, F. (2004). Hog1 mediates cell-cycle arrest in G1 phase by the dual targeting of Sic1. *Nature Cell Biology*, 6(10), 997–1002. <https://doi.org/10.1038/ncb1174>
24. Eulalio, A., Behm-Ansmant, I., Schweizer, D., & Izaurralde, E. (2007). P-Body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Molecular and Cellular Biology*, 27(11), 3970–3981. <https://doi.org/10.1128/mcb.00128-07>
25. Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., Kriwacki, R. W., Pappu, R. V., & Brangwynne, C. P. (2016). Coexisting liquid phases underlie

- nucleolar subcompartments. *Cell*, 165(7), 1686–1697. <https://doi.org/10.1016/j.cell.2016.04.047>
26. French-Pacheco, L., Rosas-Bringas, O., Segovia, L., & Covarrubias, A. A. (2022). Intrinsically disordered signaling proteins: Essential hub players in the control of stress responses in *Saccharomyces cerevisiae*. *PLOS ONE*, 17(3), e0265422. <https://doi.org/10.1371/journal.pone.0265422>
 27. Galea, C. A., Wang, Y., Sivakolundu, S. G., & Kriwacki, R. W. (2008). Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits. *Biochemistry*, 47(29), 7598–7609. <https://doi.org/10.1021/bi8006803>
 28. Gilks, N., Kedersha, N., Ayodele, M., Shen, L., Stoecklin, G., Dember, L. M., & Anderson, P. (2004). Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Molecular Biology of the Cell*, 15(12), 5383–5398. <https://doi.org/10.1091/mbc.e04-08-0715>
 29. Gomes, E., & Shorter, J. (2019). The molecular language of membraneless organelles. *Journal of Biological Chemistry*, 294(18), 7115–7127. <https://doi.org/10.1074/jbc.tm118.001192>
 30. Grant, W. D. (2004). Life at low water activity. *Philosophical transactions of the royal society of london. Series B: Biological Sciences*, 359(1448), 1249–1267. <https://doi.org/10.1098/rstb.2004.1502>
 31. Guillén-Boixet, J., Kopach, A., Holehouse, A. S., Wittmann, S., Jahnel, M., Schlüßler, R., Kim, K., Trussina, I. R., Wang, J., Mateju, D., Poser, I., Maharana, S., Ruer-Gruß, M., Richter, D., Zhang, X., Chang, Y. T., Guck, J., Honigmann, A., Mahamid, J., . . . Franzmann, T. M. (2020). RNA-induced conformational switching and clustering of G3BP drive stress granule assembly by condensation. *Cell*, 181(2), 346–361.e17. <https://doi.org/10.1016/j.cell.2020.03.049>
 32. Guillen-Chable, F., Bayona, A., Rodríguez-Zapata, L. C., & Castano, E. (2021). Phase separation of intrinsically disordered nucleolar proteins relate to localization and function. *International Journal of Molecular Sciences*, 22(23), 13095. <https://doi.org/10.3390/ijms222313095>
 33. Guo, Q., Shi, X., & Wang, X. (2021). RNA and liquid-liquid phase separation. *Non-coding RNA research*, 6(2), 92–99. <https://doi.org/10.1016/j.ncrna.2021.04.003>
 34. Hatos, A., Hajdu-Soltész, B., Monzon, A. M., Palopoli, N., Álvarez, L., Aykac-Fas, B., Bassot, C., Benítez, G. I., Bevilacqua, M., Chasapi, A., Chemes, L., Davey, N. E., Davidović, R., Dunker, A., Elofsson, A., Gobeill, J., Foutel, N. S., Sudha, G., Guharoy, M., . . . Piovesan, D. (2019). DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Research*, 48(D1), D269–D276. <https://doi.org/10.1093/nar/gkz975>
 35. He, M., He, C. Q., & Ding, N. Z. (2018). Abiotic stresses: General defenses of land plants and chances for engineering multistress tolerance. *Frontiers in Plant Science*, 9. <https://doi.org/10.3389/fpls.2018.01771>
 36. Henninger, J. E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M. M., Andrews, J. O., Zamudio, A. V., Lazaris, C., Hannett, N. M., Lee, T. I., Sharp, P. A., Cissé, I. I., Chakraborty, A. K., & Young, R. A. (2021). RNA-mediated feedback Control of Transcriptional Condensates. *Cell*, 184(1), 207–225.e24. <https://doi.org/10.1016/j.cell.2020.11.030>
 37. Hohmann, S. (2002). Osmotic stress signaling and osmoadaptation in yeasts. *Microbiology and Molecular Biology Reviews*, 66(2), 300–372. <https://doi.org/10.1128/mubr.66.2.300-372.2002>

38. Hyman, A. A., Weber, C. A., & Jülicher, F. (2014). Liquid-liquid phase separation in biology. *Annual Review of Cell and Developmental Biology*, 30(1), 39–58. <https://doi.org/10.1146/annurev-cellbio-100913-013325>
39. Jalihal, A. P., Pitchiaya, S., Xiao, L., Bawa, P., Jiang, X., Bedi, K., Parolia, A., Cieslik, M., Ljungman, M., Chinnaiyan, A. M., & Walter, N. G. (2020). Multivalent proteins rapidly and reversibly phase-separate upon osmotic cell volume change. *Molecular Cell*, 79(6), 978–990.e5. <https://doi.org/10.1016/j.molcel.2020.08.004>
40. Jawerth, L. M., Ijavi, M., Ruer, M., Saha, S., Jahnel, M., Hyman, A. A., Jülicher, F., & Fischer-Friedrich, E. (2018). Salt-dependent rheology and surface tension of protein condensates using optical traps. *Physical Review Letters*, 121(25). <https://doi.org/10.1103/physrevlett.121.258101>
41. Kanekura, K., & Kuroda, M. (2022). How can we interpret the relationship between liquid-liquid phase separation and amyotrophic lateral sclerosis? *Laboratory Investigation*. <https://doi.org/10.1038/s41374-022-00791-x>
42. Kaur, T., Alshareedah, I., Wang, W., Ngo, J., Moosa, M., & Banerjee, P. (2019). Molecular crowding tunes material states of ribonucleoprotein condensates. *Biomolecules*, 9(2), 71. <https://doi.org/10.3390/biom9020071>
43. Kedersha, N., Ivanov, P., & Anderson, P. (2013). Stress granules and cell signaling: more than just a passing phase? *Trends in Biochemical Sciences*, 38(10), 494–506. <https://doi.org/10.1016/j.tibs.2013.07.004>
44. Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., Lykke-Andersen, J., Fritzler, M. J., Scheuner, D., Kaufman, R. J., Golan, D. E., & Anderson, P. (2005). Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *Journal of Cell Biology*, 169(6), 871–884. <https://doi.org/10.1083/jcb.200502088>
45. Kültz, D., & Chakravarty, D. (2001). Maintenance of genomic integrity in mammalian kidney cells exposed to hyperosmotic stress. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 130(3), 421–428. [https://doi.org/10.1016/s1095-6433\(01\)00440-8](https://doi.org/10.1016/s1095-6433(01)00440-8)
46. Lachke, S. A., Alkuraya, F. S., Kneeland, S. C., Ohn, T., Aboukhalil, A., Howell, G. R., Saadi, I., Cavallesco, R., Yue, Y., Tsai, A. C. H., Nair, K. S., Cosma, M. I., Smith, R. S., Hodges, E., AlFadhli, S. M., Al-Hajeri, A., Shamseldin, H. E., Behbehani, A., Hannon, G. J., . . . Maas, R. L. (2011). Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. *Science*, 331(6024), 1571–1576. <https://doi.org/10.1126/science.1195970>
47. LANG, F., BUSCH, G. L., RITTER, M., VÖLKL, H., WALDEGGER, S., GULBINS, E., & HÄUSSINGER, D. (1998). Functional significance of cell volume regulatory mechanisms. *Physiological Reviews*, 78(1), 247–306. <https://doi.org/10.1152/physrev.1998.78.1.247>
48. Langdon, E. M., Qiu, Y., Ghanbari Niaki, A., McLaughlin, G. A., Weidmann, C. A., Gerbich, T. M., Smith, J. A., Crutchley, J. M., Termini, C. M., Weeks, K. M., Myong, S., & Gladfelter, A. S. (2018). mRNA structure determines specificity of a polyQ-driven phase separation. *Science*, 360(6391), 922–927. <https://doi.org/10.1126/science.aar7432>
49. Lin, Y., Currie, S. L., & Rosen, M. K. (2017). Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *Journal of Biological Chemistry*, 292(46), 19110–19120. <https://doi.org/10.1074/jbc.m117.800466>

50. Lin, Y., Protter, D., Rosen, M., & Parker, R. (2015). Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Molecular Cell*, *60*(2), 208–219. <https://doi.org/10.1016/j.molcel.2015.08.018>
51. Lindorff-Larsen, K., & Kragelund, B. B. (2021). On the potential of machine learning to examine the relationship between sequence, structure, dynamics and function of intrinsically disordered proteins. *Journal of Molecular Biology*, *433*(20), 167196. <https://doi.org/10.1016/j.jmb.2021.167196>
52. Liu, L., Weiss, E., Panas, M. D., Götte, B., Sellberg, S., Thaa, B., & McInerney, G. M. (2019). RNA processing bodies are disassembled during old world alphavirus infection. *Journal of General Virology*, *100*(10), 1375–1389. <https://doi.org/10.1099/jgv.0.001310>
53. Liu, Y., Wang, X., & Liu, B. (2019). A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Briefings in Bioinformatics*, *20*(1), 330–346. <https://doi.org/10.1093/bib/bbx126>
54. Malinowska, L., Kroschwald, S., & Alberti, S. (2013). Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1834*(5), 918–931. <https://doi.org/10.1016/j.bbapap.2013.01.003>
55. 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*, *107*(18), 8183–8188. <https://doi.org/10.1073/pnas.0911107107>
56. Mao, Y. S., Zhang, B., & Spector, D. L. (2011). Biogenesis and function of nuclear bodies. *Trends in Genetics*, *27*(8), 295–306. <https://doi.org/10.1016/j.tig.2011.05.006>
57. Marsh, J. A., & Forman-Kay, J. D. (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>
58. Martin, E. W., & Mittag, T. (2018). Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry*, *57*(17), 2478–2487. <https://doi.org/10.1021/acs.biochem.8b00008>
59. Matsumoto, K., Nakayama, H., Yoshimura, M., Masuda, A., Dohmae, N., Matsumoto, S., & Tsujimoto, M. (2012). PRMT1 is required for RAP55 to localize to processing bodies. *RNA Biology*, *9*(5), 610–623. <https://doi.org/10.4161/rna.19527>
60. McFadden, W. M., & Yanowitz, J. L. (2022). idpr: A package for profiling and analyzing Intrinsically Disordered Proteins in R. *PLOS ONE*, *17*(4), e0266929. <https://doi.org/10.1371/journal.pone.0266929>
61. Meng, F., Uversky, V. N., & Kurgan, L. (2017). Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cellular and Molecular Life Sciences*, *74*(17), 3069–3090. <https://doi.org/10.1007/s00018-017-2555-4>
62. Miermont, A., Waharte, F., Hu, S., McClean, M. N., Bottani, S., Leon, 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*, *110*(14), 5725–5730. <https://doi.org/10.1073/pnas.1215367110>
63. Mitrea, D. M., & Kriwacki, R. W. (2016). Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling*, *14*(1). <https://doi.org/10.1186/s12964-015-0125-7>
64. Mollet, S., Cougot, N., Wilczynska, A., Dautry, F., Kress, M., Bertrand, E., & Weil, D. (2008). Translationally repressed mRNA transiently cycles through stress granules

- during stress. *Molecular Biology of the Cell*, 19(10), 4469–4479. <https://doi.org/10.1091/mbc.e08-05-0499>
65. Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., Dosztányi, Z., Uversky, V. N., Obradovic, Z., Kurgan, L., Dunker, A. K., & Gough, J. (2012). D2P2: database of disordered protein predictions. *Nucleic Acids Research*, 41(D1), D508–D516. <https://doi.org/10.1093/nar/gks1226>
 66. Oeser, M. L., Amen, T., Nadel, C. M., Bradley, A. I., Reed, B. J., Jones, R. D., Gopalan, J., Kaganovich, D., & Gardner, R. G. (2016). Dynamic sumoylation of a conserved transcription corepressor prevents persistent inclusion formation during hyperosmotic stress. *PLoS Genetics*, 12(1). <https://doi.org/10.1371/journal.pgen.1005809>
 67. Ohue, M., Matsuzaki, Y., Shimoda, T., Ishida, T., & Akiyama, Y. (2013). Highly precise protein-protein interaction prediction based on consensus between template-based and de novo docking methods. *BMC Proceedings*, 7(S6). <https://doi.org/10.1186/1753-6561-7-s7-s6>
 68. Onuchic, P. L., Milin, A. N., Alshareedah, I., Deniz, A. A., & Banerjee, P. R. (2019). Divalent cations can control a switch-like behavior in heterotypic and homotypic RNA coacervates. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-48457-x>
 69. Panas, M. D., Ivanov, P., & Anderson, P. (2016). Mechanistic insights into mammalian stress granule dynamics. *Journal of Cell Biology*, 215(3), 313–323. <https://doi.org/10.1083/jcb.201609081>
 70. Pederson, T. (2011). The nucleolus. *Cold Spring Harbor Perspectives in Biology*, 3(3), a000638. <https://doi.org/10.1101/cshperspect.a000638>
 71. Pessina, F., Giavazzi, F., Yin, Y., Gioia, U., Vitelli, V., Galbiati, A., Barozzi, S., Garre, M., Oldani, A., Flaus, A., Cerbino, R., Parazzoli, D., Rothenberg, E., & D’Adda Di Fagagna, F. (2019). Functional transcription promoters at DNA double-strand breaks mediate RNA-driven phase separation of damage-response factors. *Nature Cell Biology*, 21(10), 1286–1299. <https://doi.org/10.1038/s41556-019-0392-4>
 72. Petelenz-Kurdziel, E., Eriksson, E., Smedh, M., Beck, C., Hohmann, S., & Goksör, M. (2011). Quantification of cell volume changes upon hyperosmotic stress in *Saccharomyces cerevisiae*. *Integrative Biology*, 3(11), 1120–1126. <https://doi.org/10.1039/c1ib00027f>
 73. Piovesan, D., Monzon, A. M., Quaglia, F., & Tosatto, S. C. E. (2022). Databases for intrinsically disordered proteins. *Acta Crystallographica Section D Structural Biology*, 78(2), 144–151. <https://doi.org/10.1107/s2059798321012109>
 74. Pothof, J., Verkaik, N. S., Hoeijmakers, J. H., & van Gent, D. C. (2009). MicroRNA responses and stress granule formation modulate the DNA damage response. *Cell Cycle*, 8(21), 3462–3468. <https://doi.org/10.4161/cc.8.21.9835>
 75. Riggs, C. L., Kedersha, N., Ivanov, P., & Anderson, P. (2020). Mammalian stress granules and P bodies at a glance. *Journal of Cell Science*, 133(16). <https://doi.org/10.1242/jcs.242487>
 76. Rogers, J. M., Steward, A., & Clarke, J. (2013). Folding and binding of an intrinsically disordered protein: Fast, but not ‘Diffusion-Limited’. *Journal of the American Chemical Society*, 135(4), 1415–1422. <https://doi.org/10.1021/ja309527h>
 77. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Structure, Function, and Genetics*, 42(1), 38–48.

78. Saar, K. L., Morgunov, A. S., Qi, R., Arter, W. E., Krainer, G., Lee, A. A., & Knowles, T. P. J. (2021). Learning the molecular grammar of protein condensates from sequence determinants and embeddings. *Proceedings of the National Academy of Sciences*, *118*(15), e2019053118. <https://doi.org/10.1073/pnas.2019053118>
79. Shen, B., Chen, Z., Yu, C., Chen, T., Shi, M., & Li, T. (2021). Computational screening of phase-separating proteins. *Genomics, Proteomics & Bioinformatics*, *19*(1), 13–24. <https://doi.org/10.1016/j.gpb.2020.11.003>
80. Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science*, *357*(6357). <https://doi.org/10.1126/science.aaf4382>
81. Shukla, S., & Parker, R. (2016). Hypo- and hyper-assembly diseases of RNA–protein complexes. *Trends in Molecular Medicine*, *22*(7), 615–628. <https://doi.org/10.1016/j.molmed.2016.05.005>
82. Shukla, S., Agarwal, P., & Kumar, A. (2022). Disordered regions tune order in chromatin organization and function. *Biophysical Chemistry*, *281*, 106716. <https://doi.org/10.1016/j.bpc.2021.106716>
83. Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z., & Dunker, A. K. (2007). DisProt: the database of disordered proteins. *Nucleic Acids Research*, *35*(Database), D786–D793. <https://doi.org/10.1093/nar/gkl893>
84. Silva, R. D., Sotoca, R., Johansson, B., Ludovico, P., Sansonetti, F., Silva, M. T., Peinado, J. M., & Côrte-Real, M. (2005). Hyperosmotic stress induces metacaspase- and mitochondria-dependent apoptosis in *Saccharomyces cerevisiae*. *Molecular Microbiology*, *58*(3), 824–834. <https://doi.org/10.1111/j.1365-2958.2005.04868.x>
85. Simon, M., & Hancock, J. M. (2009). Tandem and cryptic amino acid repeats accumulate in disordered regions of proteins. *Genome Biology*, *10*(6), R59. <https://doi.org/10.1186/gb-2009-10-6-r59>
86. Slaninová, I., Šesták, S., Svoboda, A., & Farkaš, V. (2000). Cell wall and cytoskeleton reorganization as the response to hyperosmotic shock in *Saccharomyces cerevisiae*. *Archives of Microbiology*, *173*(4), 245–252. <https://doi.org/10.1007/s002030000136>
87. Thomas, M. G., Loschi, M., Desbats, M. A., & Boccaccio, G. L. (2011). RNA granules: The good, the bad and the ugly. *Cellular Signalling*, *23*(2), 324–334. <https://doi.org/10.1016/j.cellsig.2010.08.011>
88. 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)
89. 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>
90. UniProt (s.f.). UV excision repair protein RAD23 homolog B. <https://www.uniprot.org/uniprot/P54727>
91. Uversky, V. N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current Opinion in Structural Biology*, *44*, 18–30. <https://doi.org/10.1016/j.sbi.2016.10.015>
92. Uversky, V. N. (2019). Intrinsically disordered proteins and their “mysterious” (Meta)Physics. *Frontiers in Physics*, *7*. <https://doi.org/10.3389/fphy.2019.00010>
93. Uversky, V. N., & Dunker, A. K. (2010). Understanding protein non-folding. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1804*(6), 1231–1264. <https://doi.org/10.1016/j.bbapap.2010.01.017>

94. Uversky, V. N., Gillespie, J. R., & Fink, A. L. (2000). Why are? natively unfolded? proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Genetics*, *41*(3), 415–427.
95. Vacic, V., Uversky, V. N., Dunker, A. K., & Lonardi, S. (2007). Composition profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics*, *8*(1). <https://doi.org/10.1186/1471-2105-8-211>
96. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., & Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. *Chemical Reviews*, *114*(13), 6589–6631. <https://doi.org/10.1021/cr400525m>
97. Viegas, J. O., & Meshorer, E. (2019a). The princess and the P: Pluripotent stem cells and P-Bodies. *Cell Stem Cell*, *25*(5), 589–591. <https://doi.org/10.1016/j.stem.2019.10.008>
98. Walters, R. W., & Parker, R. (2015). Coupling of ribostasis and proteostasis: Hsp70 proteins in mRNA metabolism. *Trends in Biochemical Sciences*, *40*(10), 552–559. <https://doi.org/10.1016/j.tibs.2015.08.004>
99. Wang, J., Choi, J. M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaître, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R. V., Alberti, S., & Hyman, A. A. (2018). A molecular grammar governing the driving forces for phase separation of Prion-like RNA binding proteins. *Cell*, *174*(3), 688–699.e16. <https://doi.org/10.1016/j.cell.2018.06.006>
100. Wright, P. E., & Dyson, H. J. (2009). Linking folding and binding. *Current Opinion in Structural Biology*, *19*(1), 31–38. <https://doi.org/10.1016/j.sbi.2008.12.003>
101. Xie, W., & Denman, R. B. (2011). Protein methylation and stress granules: Posttranslational remodeler or innocent bystander? *Molecular Biology International*, *2011*, 1–14. <https://doi.org/10.4061/2011/137459>
102. Yasuda, S., Tsuchiya, H., Kaiho, A., Guo, Q., Ikeuchi, K., Endo, A., Arai, N., Ohtake, F., Murata, S., Inada, T., Baumeister, W., Fernández-Busnadiego, R., Tanaka, K., & Saeki, Y. (2020). Stress- and ubiquitylation-dependent phase separation of the proteasome. *Nature*, *578*(7794), 296–300. <https://doi.org/10.1038/s41586-020-1982-9>
103. Zemła, J., Danilkiewicz, J., Orzechowska, B., Pabijan, J., Seweryn, S., & Lekka, M. (2018). Atomic force microscopy as a tool for assessing the cellular elasticity and adhesiveness to identify cancer cells and tissues. *Seminars in Cell & Developmental Biology*, *73*, 115–124. <https://doi.org/10.1016/j.semcd.2017.06.029>
104. Zhang, H., Elbaum-Garfinkle, S., Langdon, E. M., Taylor, N., Occhipinti, P., Bridges, A., Brangwynne, C., & Gladfelter, A. (2015). RNA controls PolyQ protein phase transitions. *Molecular Cell*, *60*(2), 220–230. <https://doi.org/10.1016/j.molcel.2015.09.017>
105. Zhang, J., Okabe, K., Tani, T., & Funatsu, T. (2011). Dynamic association–dissociation and harboring of endogenous mRNAs in stress granules. *Journal of Cell Science*, *124*(23), 4087–4095. <https://doi.org/10.1242/jcs.090951>
106. Zhao, B., & Kurgan, L. (2022). Deep learning in prediction of intrinsic disorder in proteins. *Computational and Structural Biotechnology Journal*, *20*, 1286–1294. <https://doi.org/10.1016/j.csbj.2022.03.003>

107. Zhou, H. X., Nguemaha, V., Mazarakos, K., & Qin, S. (2018). Why do disordered and structured proteins behave differently in Phase Separation? *Trends in Biochemical Sciences*, 43(7), 499–516. <https://doi.org/10.1016/j.tibs.2018.03.007>

ANEXOS.

Anexo I. Aminoácidos abundantes y carentes del grupo IDR-LLPS respecto al grupo Delta-IDR-LLPS.

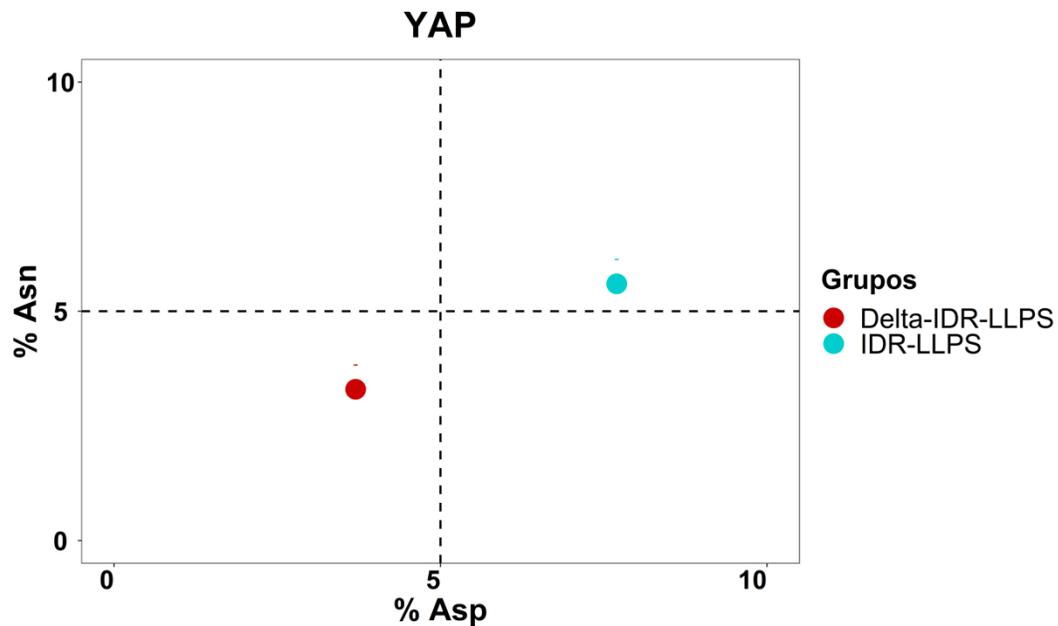


Figura A.1.1. Porcentaje de Asn y porcentaje de Asp, aminoácidos abundantes en la IDR de YAP.

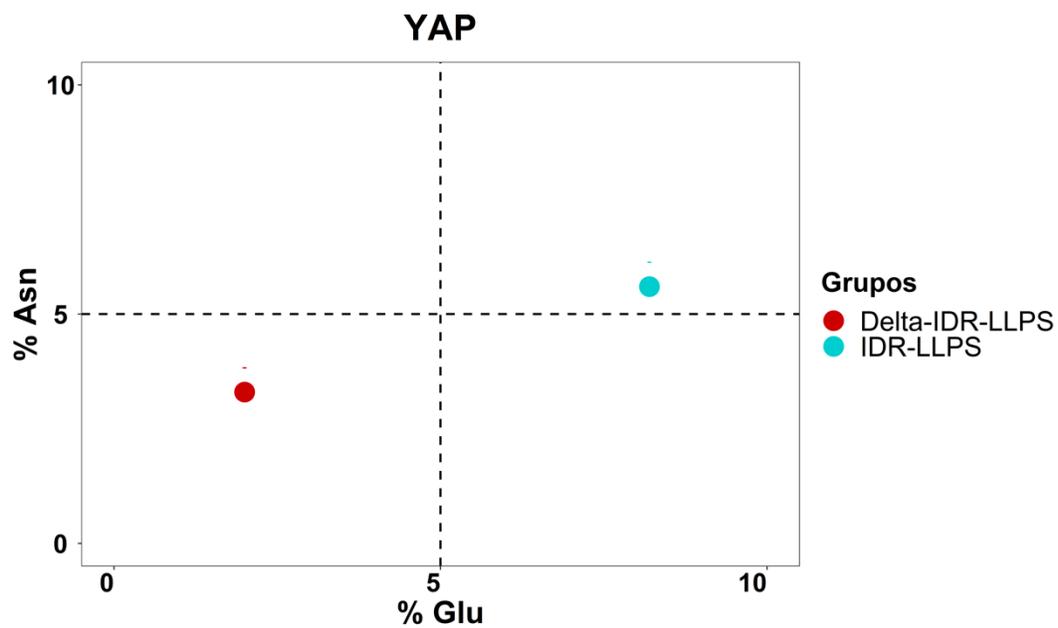


Figura A.1.2. Porcentaje de Asn y porcentaje de Glu, aminoácidos abundantes en la IDR de YAP.

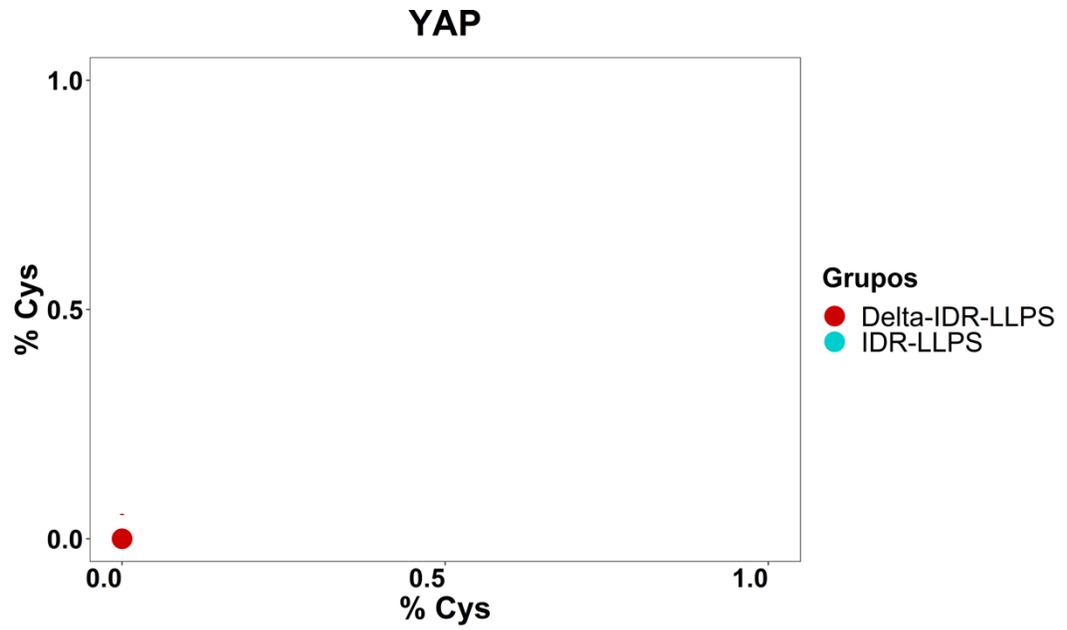


Figura A.1.3. Porcentaje de Cys, aminoácido carente en la IDR de YAP. (Ambos grupos están sobrelapados).

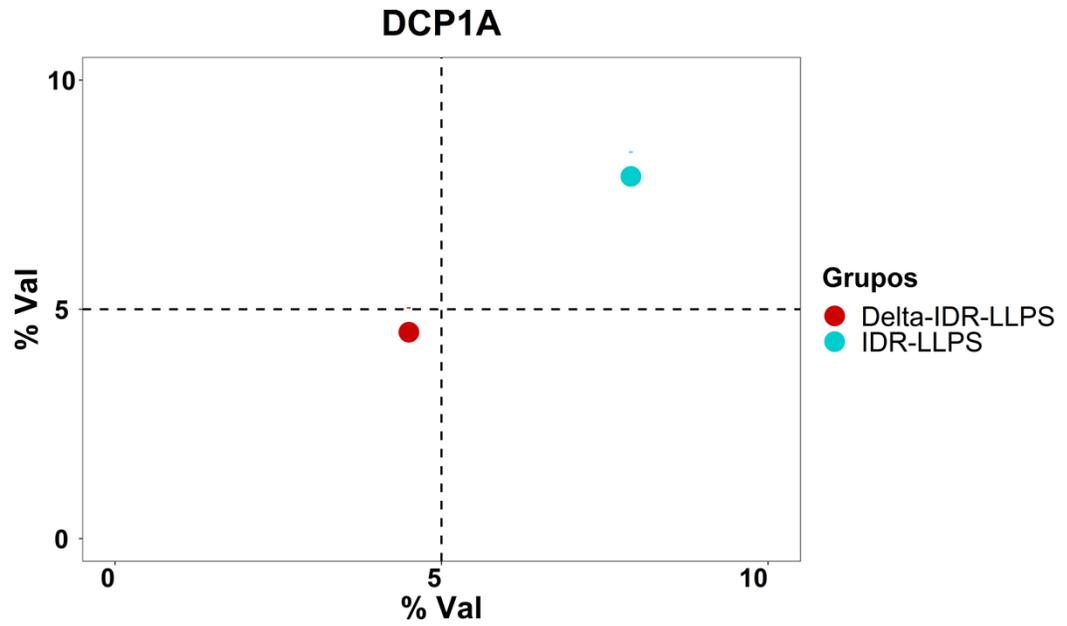


Figura A.1.4. Porcentaje de Val, aminoácido abundante en la IDR de DCP1A.

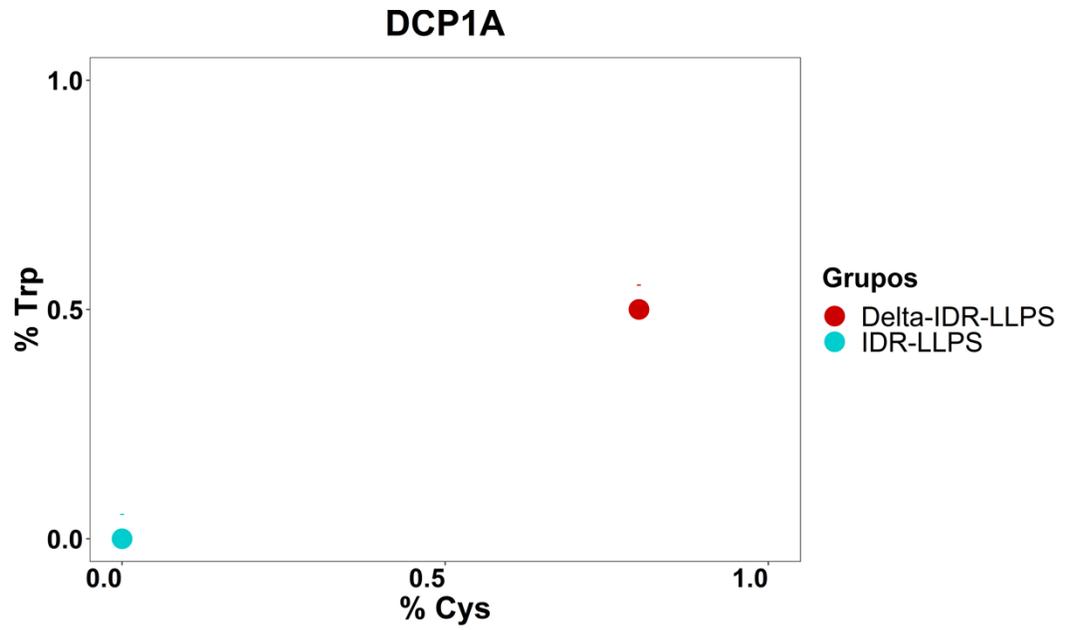


Figura A.1.5. Porcentaje de Trp y porcentaje de Cys, aminoácidos carentes en la IDR de DCP1A.

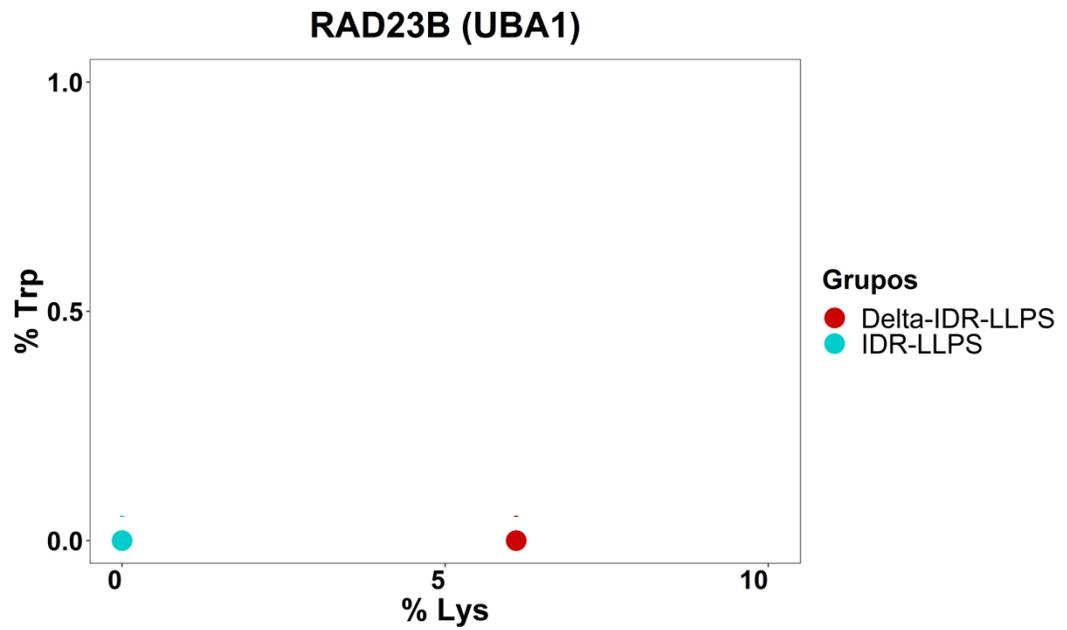


Figura A.1.6. Porcentaje de Trp y porcentaje de Lys, aminoácidos carentes en la IDR de DCP1A.

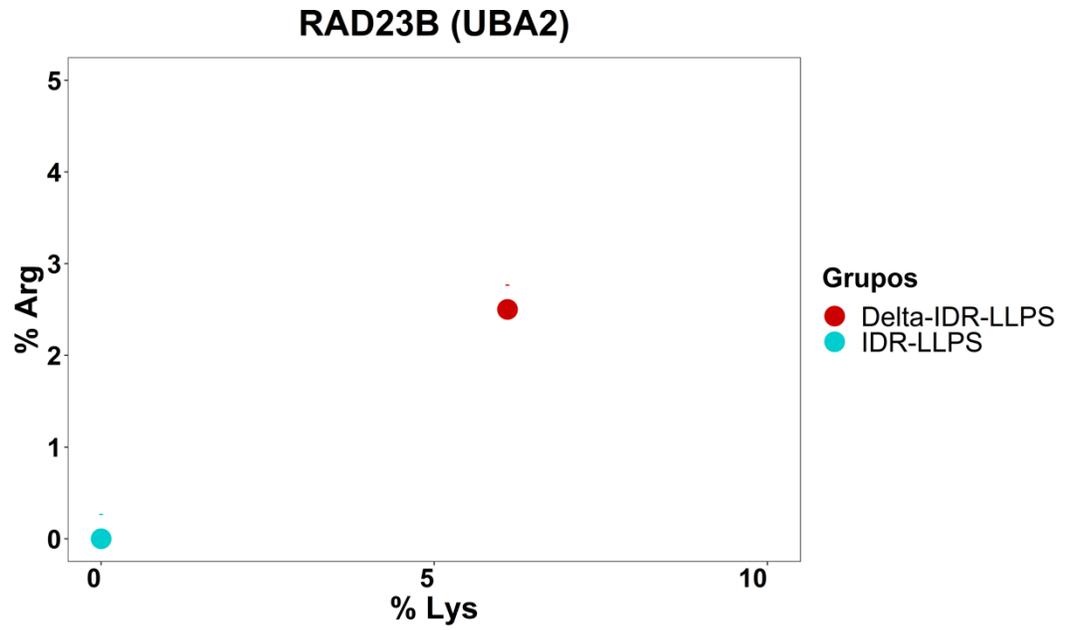


Figura A.1.7. Porcentaje de Arg y porcentaje de Lys, aminoácidos carentes en la IDR UBA2 de RAD23B.

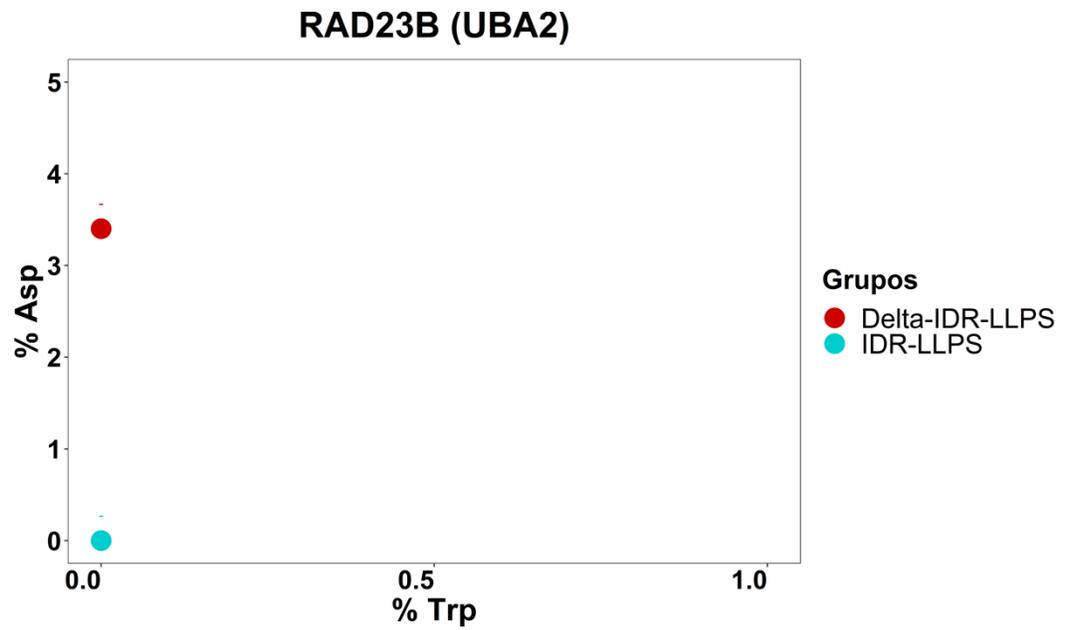


Figura A.1.8. Porcentaje de Asp y porcentaje de Trp, aminoácidos carentes en la IDR UBA2 de RAD23B.

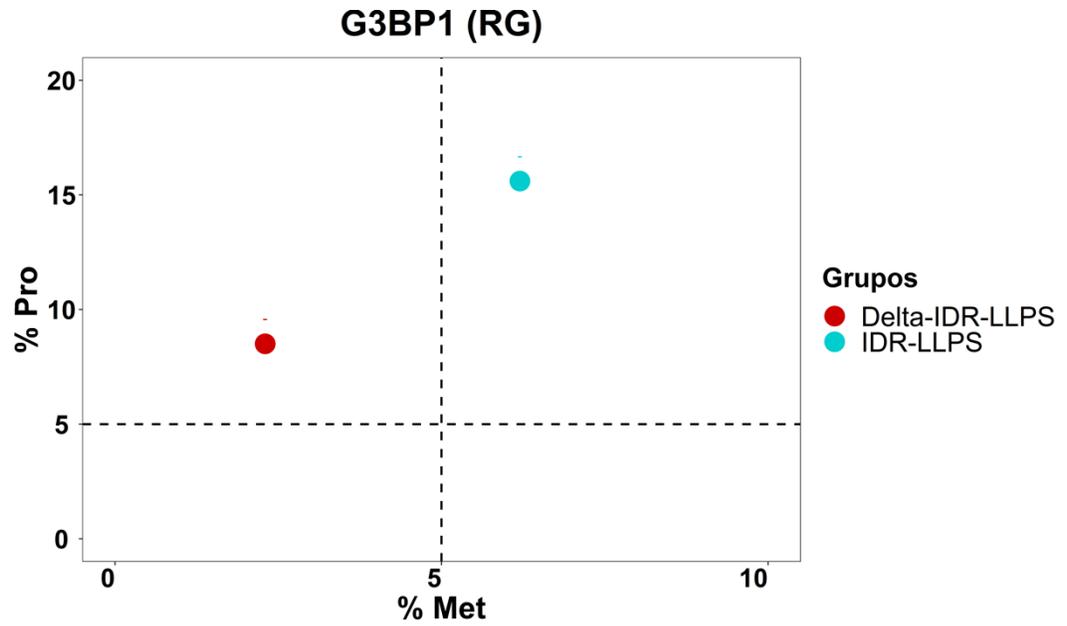


Figura A.1.9. Porcentaje de Pro y porcentaje de Met, aminoácidos abundantes en la IDR RG de G3BP1.

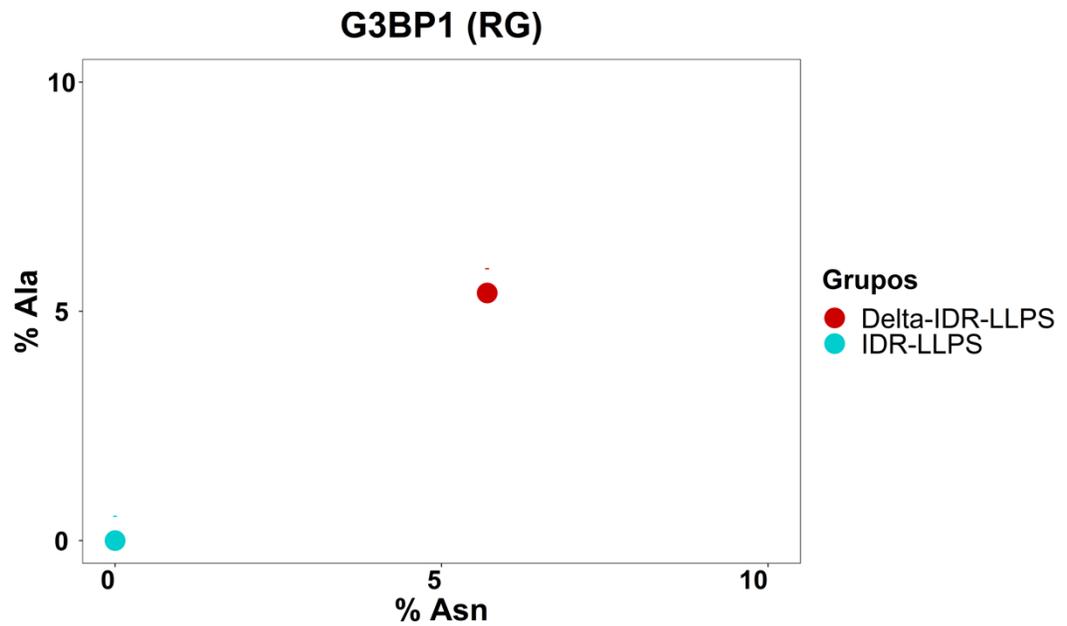


Figura A.1.10. Porcentaje de Ala y porcentaje de Asn, aminoácidos carentes en la IDR RG de G3BP1.

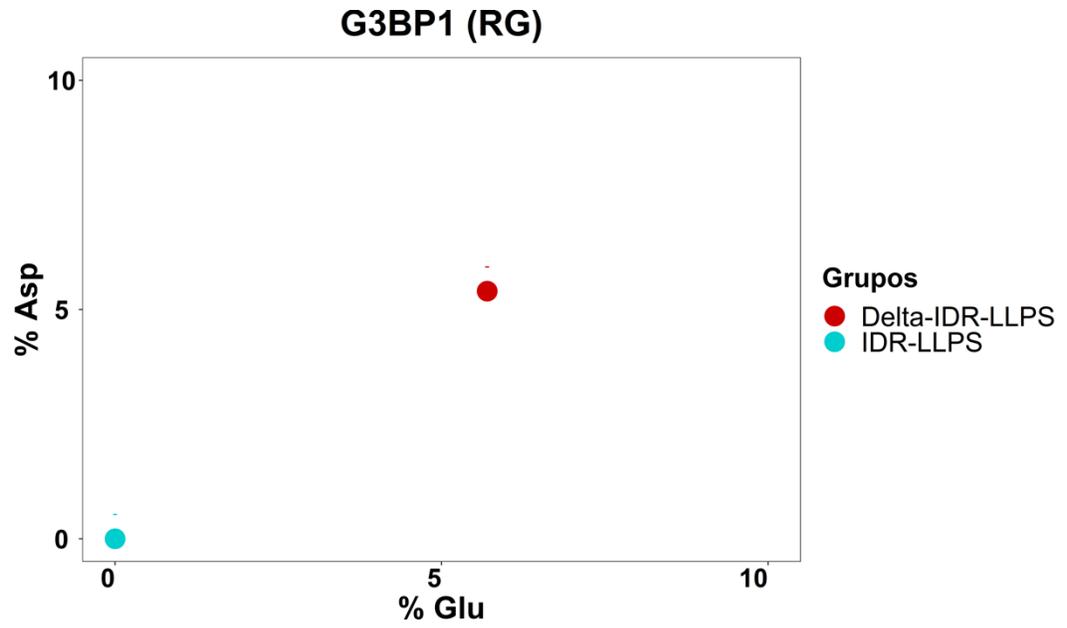


Figura A.1.11. Porcentaje de Asp y porcentaje de Glu, aminoácidos carentes en la IDR RG de G3BP1.

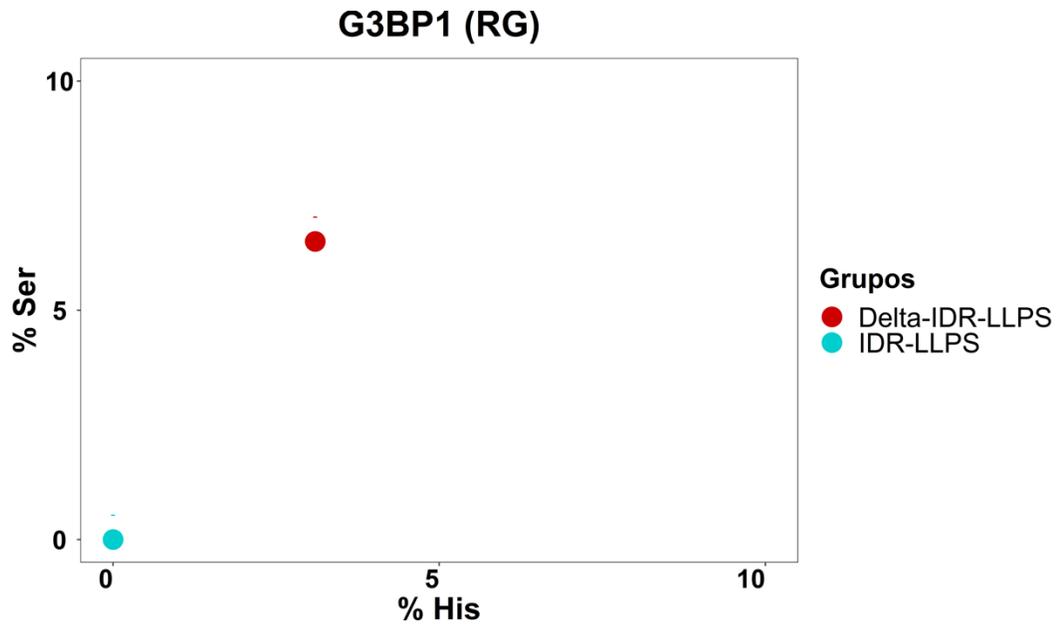


Figura A.1.12. Porcentaje de Ser y porcentaje de His, aminoácidos carentes en la IDR RG de G3BP1.

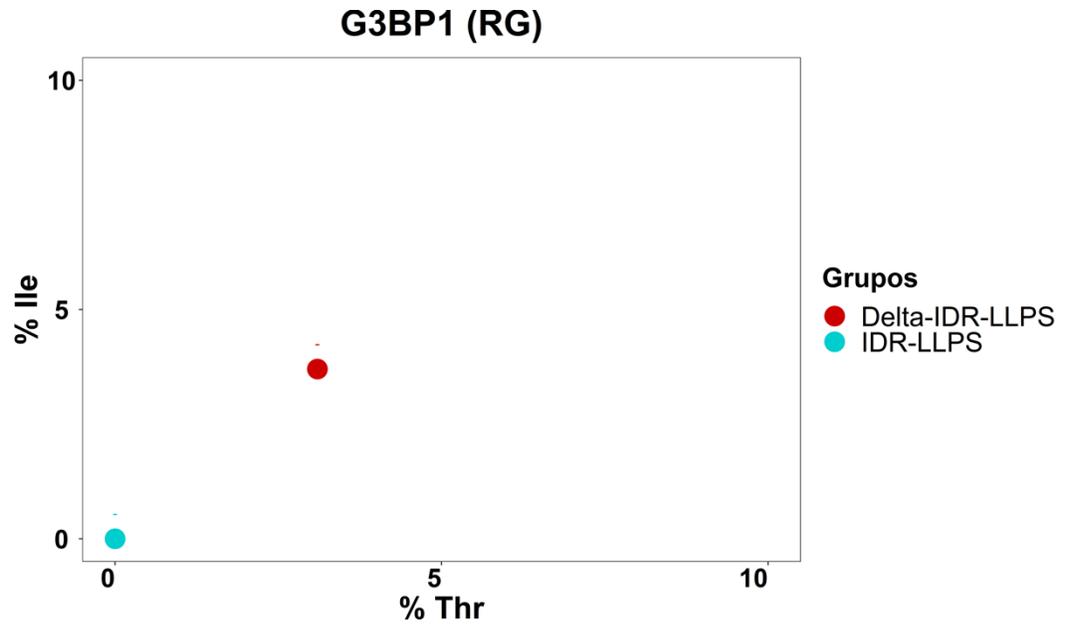


Figura A.1.13. Porcentaje de Ile y porcentaje de Thr, aminoácidos carentes en la IDR RG de G3BP1.

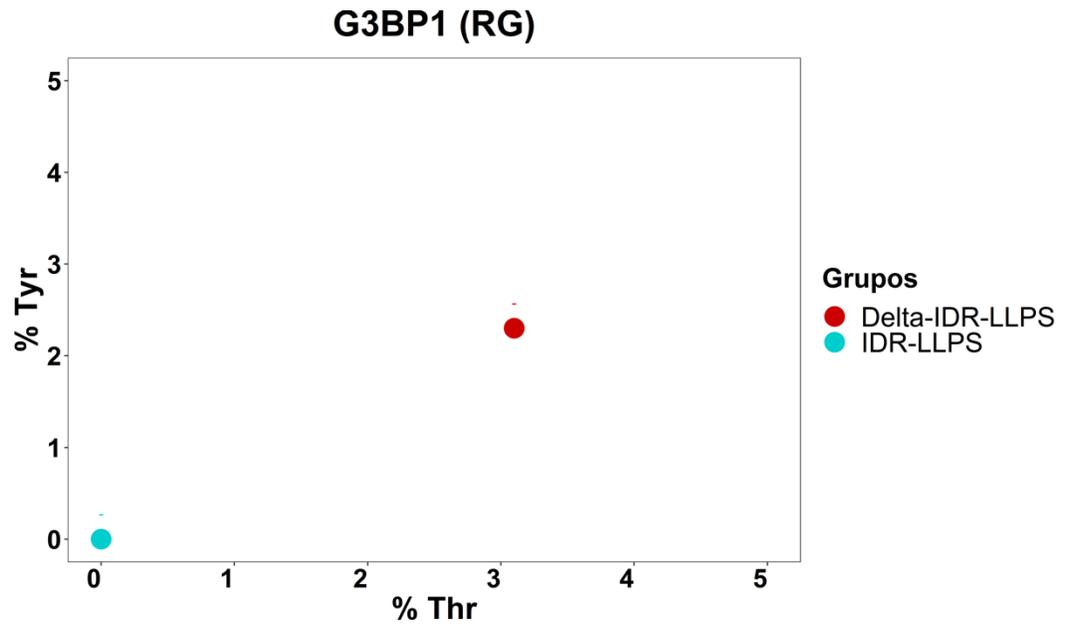


Figura A.1.14. Porcentaje de Tyr y porcentaje de Thr, aminoácidos carentes en la IDR RG de G3BP1.

2: DCP1A 29.58 100.00

Ortólogo AT1G16190.1 de RAD23B en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```
AT1G16190.1 MKLTVKTLKGSHEIRVLPDITMAVKKNIEDSQSKDNYPCGQQLLIHNG
RAD23B MQVTLKTLQQQTFKIDIDPEETVKALKEIESEKGDAPFPVAGQKLIYAG
*:*:*:*:* . *:* : * :*: *:*:*:*:*:*:*:* * . * *:* : *

AT1G16190.1 KVLKDETTLVENKVTEEGFLVVMLSKSKTASSAGPSSTQ---PTSTTTST
RAD23B KILNDDTALKEYKIDKKNFVVMVTKPKAVSTPAPATTQOSAPASTTAVT
*:*:*:*:* * * : * :*: *:*:*:*:*:*:*:*:*:*:* * :*:*:* : *

AT1G16190.1 ISSTTLAA-----PSTTQSIIVPAS-----
RAD23B SSTTTVAQAPTVPALAPTSTPASITPASATASSEPPAPASAAKQEKPAE
*:* * . * * * * * * * * * * * * * * * * * * * * * * *

AT1G16190.1 --NSTPVQEQPTAQSDTY-----GQAASTLVSGSSIEQMVQQIMEM
RAD23B KPAETPVATSPTATDSTSGDSSRSNLFEDATSALVTGQSYENMVTEIMSM
.*** .*** ..* :*:*:*:*:*:* *:*:*:*:*:*

AT1G16190.1 GGSWDKETVTRALRAAYNNPERAVDYLKSGIPETVTIPATNLSGVGSGR
RAD23B G---YEREQVIAALRASFNNDRAVEYLLMGIPGDRESQAV-----
* :*: * * * * * :*:*:*:*:* * * * * * *

AT1G16190.1 ELTAPP---SGGNSSPLDLFPQE---AVSDAAGGDLGTLEFLRGNDQ
RAD23B --VDPPQAASTGAPQSSAVAAAAATTTATTTTTSSGG--HPLEFLRNQFPQ
. * * :*:*:*:*:* . :* :* * * * * * * * * *

AT1G16190.1 FQQLRSMVNSNPQILQPMQLQELGKQNPQLLRLIQENQAEFLQLLNEPYEG
RAD23B FQQMRQIIQQNPSLLPALLQQIGRENQQLLQQISQHQEHFIQMLNEPVQE
***:*:*:*:*:*:* * :*:*:*:*:*:*:* * :*:* * :*:*:*:* * :

AT1G16190.1 SD-----GDVDIFDQPDQEMPHSVNVTPEEQESIERLEAMGFDRIV
RAD23B AGGQGGGGGGGGGGIAEAGSGHM-NYIQVTPQEKEAIERLKALGFPEGLV
:. * . * : . * : :*:*:*:*:*:*:*:* * * * * *

AT1G16190.1 IEAFLSCDRNEELAANYLLEHSADFED
RAD23B IQAYFACEKNENLAANFLLQQNFD-ED
*:*:*:*:*:*:*:*:*:*:* * * *
```

Percent Identity Matrix - created by Clustal2.1

```
#
#
1: AT1G16190.1 100.00 44.86
2: RAD23B 44.86 100.00
```


Ortólogo AT5G16090 de RAD23B en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

AT5G16090      MKIIVKTLKGI RFEIEV KPNDSVAEVKKN IETVMGASEYPAAQQILIHKR
RAD23B         MQVTLKTLQQQTFKIDIDPEETVKALKEKIESEKGDAPFPVAGQKLIYAG
                *:: :***:  *:*::*:::*  :*::**:*  * . :*. * * **

AT5G16090      EKLRDETTMEANKVFDKSVIAIIITKGC-----
RAD23B         KILNDDTALKEYKIDEKNFVVMVTKPKAVSTPAPATTQQSAPASTAVT
                : *.*:*:::  * : *..:..:.*

AT5G16090      -----
RAD23B         SSTTTTVAQAPTVPALAPTSTPASITPASATASSE PAPANASAAKQEKPAE

AT5G16090      -----
RAD23B         KPAETPVATSPTATDSTSGDSSRSNLFEDATSALVTGQSYENMVTEIMSM

AT5G16090      -----
RAD23B         GYEREQVIAALRASFNNDRAVEYLLMGI PGDRESQAVVDPPQAASTGAP

AT5G16090      -----
RAD23B         QSSAVAAAAATTTATTTTTSSGGHPLEFLRNQPQFQMRQIIQQNPSLLP

AT5G16090      --LEEMEKQNPPLFQMIRHNSAGFV PVLNKESFERDNE-----LA
RAD23B         ALLQQIGRENQQLQQISQHQEHFIQMLNEPVQEAGGQGGGGGGSGGIA
                *::: :** *:* * :.  * : **:*  * ..  :*

AT5G16090      QP-EEDLLQLQVTAVDDEAINRLEAMGFERRVVLEVFLACNKNEQLAANF
RAD23B         EAGSGHMNYIQVTPQEKEAIERLKALGFPEGLVIQAYFACEKNENLAANF
                :. . . : :***. :.*::*:*:* * . :*:::.*::**:*::**

AT5G16090      LLDHIH---
RAD23B         LLQQNFDED
                **:: .
    
```

Percent Identity Matrix - created by Clustal2.1

#

1: AT5G16090	100.00	34.50
2: RAD23B	34.50	100.00

Ortólogo AT5G38470 de RAD23B en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

AT5G38470      MKIFVKTLSGSNFEIEVKPADKVS DVKTAIETVKGAE-YPAAKQMLIHQG
RAD23B         MQVTLKTLQQQTFKIDIDPEETVKALKEKIESEKGD AFPVAGQKLIYAG
                *:: :***. ..*:*::.* :.* :* ** : ** : :*. * * ** : *

AT5G38470      KVLKDETTLEENNVVENSFIVIMLSKTKAS--PSGASTA-SAPAPSATQP
RAD23B         KILNDDTALKEYKIDKKNFVVMVTKPKAVSTPAPATTQQSAPASTAVT
                *:*:*:*:*:* : : *..*:*:*:*:* ** * : *:* * **** : : : .

AT5G38470      QTVATPQVSAP-----
RAD23B         SSTTTTVAQAPTVPALAPTSTPASITPASATASSE PAPANASAAKQEKPAE
                ..:* . **

AT5G38470      -TASVPVPTSGTATAAAPATAASVQTDVYGQAASNLVAGTTLESTVQQIL
RAD23B         KPAETPVATSPTATDS--TSGDSSRSNLFEDATSALVTGQSYENMVTEIM
                *..**.* ** * : : : * : : : : : * * * : * . * : * :

AT5G38470      DMGGGSWDRD TVVRALRAAFNNPERAVEYLYSGIPAQAEIPPVAQAPATG
RAD23B         SMG---YEREQVIAALRASFNNDRAVEYLLMGI PGDRESQAVVDPP---
                .* * : : : * : * * * : * * * * * * * * * * * * * * * * * *
                . * . . . *

AT5G38470      EQAANPLAQPQAAAAPAAATGGPNANPLNLFPPQMPAADAGAGAGNLDLFL
RAD23B         -QAASTGAPQSSAVAAAAAT-----TTATTTTSSGGHPLEFL
                ***.. * ..*.*.*** * : : : * . * : * *

AT5G38470      RNSQQFQALRAMVQANPQILQPMLQELGKQNPQLVRLIQEHQADFLRLIN
RAD23B         RNQPQFQQMRQIIQQNP SLLPALLQQIGREN PQLLQI SQHQEHFIQMLN
                ** . *** : * : * * * . * . * * : * * * * * * * * * * * * *

AT5G38470      EPVE-----GEENVMEQLEAAMPQAVTVTPPEEREAIERLEGMGF
RAD23B         EPVQEAGGQGGGGGGSGGIAEAGS-GHMNYIQVTPQEKEAIERL KALGF
                ***: * . : * . . : : * * * * * * * * * * * * * * *

AT5G38470      DRAMVLEVFFACNKNEELAANYLLDHMHFEFDQ
RAD23B         PEGLVIQAYFACEKNENLAANFLLQQNFD-E-D
                ..*:::***:*:*:*:*:*:*: : * :
    
```

Percent Identity Matrix - created by Clustal2.1

#

1: AT5G38470	100.00	42.37
2: RAD23B	42.37	100.00

Ortólogo AT1G79650 de RAD23B en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

AT1G79650      MKLTVKTLKGSFHEIRVLPSTIMAVKKNIEDSQGKDNYPGQQLLIHNG
RAD23B         MQVTLKTLQQQTFKIDIDPEETVKALKEKIESEKGDAPVAGQKLIYAG
                *:*:*:*:* . *:* : *:*: *:*:*:*:*:*:*:* * . * **:*
                *

AT1G79650      KVLKDETSLVENKVTEEGFLVVMLS KSKSGGSAGQASVQTSSVSQPVSAT
RAD23B         KILNDDTALKEYKIDEKNFVVVMVTKPKAVSTPAPATTQQSAPASTTAVT
                *:*:*:*:* * * : *:*:*:*:*:*:*:* . . . . *:*.* * : : . . . . *

AT1G79650      TSSTKPA-----AP-----STTQSSPVPASPIPAQEQPA
RAD23B         SSTTTTVAQAPTVPALAPTSTPASITPASATASSEPA PASA-AKQEKPA
                :*:*...          **                :*:*.*.*.*.*. . **:*

AT1G79650      VYAFVFSFAGLAFCLYGFYGFPKV-----SMAQTDTYGQAAS TLVSGSSLE
RAD23B         EKP-----AETPVATSPTATDSTSGDSSRSNLFEDATSALVTGQSYE
                .          * * : *..          . : : : : : *:*:*:*:*.* *

AT1G79650      QMVQQIMEMGGGSWDKETVTRALRAAYNNPERAVDYL YSGIPIQTAEVAVP
RAD23B         NMVTEIMSMG---YEREQVIAALRASFNNDRAVEYLLMGI PGDRESQAV
                :* *:*.* * : : * * * * * : * * * * * * * * * * * *

AT1G79650      VPEAQIAGSGAAPVAPASGGPNSSPLDLFPQETVAAAGSGDL--GTLEFL
RAD23B         VDPPQAAS-----TGAPQSSAVAAAAATTTATTTTSSGGHPLEFL
                * . * * .          :*:*:*:*:* . * . * : : . . *.*.*

AT1G79650      RNNDQFQQLRTMVHSNPQILQPLQELGKQNPQLLRLIQENQAEFLQLVN
RAD23B         RNQPQFQQMRQIIQQNPSLLPALLQQIGRENPLLQQISQHQEHFIQMLN
                ** : * * * * : * : : * * . * . * * : * * * * * * * * * *

AT1G79650      EPYEGSDGE-----GDMFDQPEQEMPHAINVTPAEQEAIQRLEAMGF
RAD23B         EPVQEAGGQGGGGGGGGGGIAEAGSGHM-NYIQVTPQEKEAIERL KALGF
                ** : : * : * : : . * : * : * * * * * * * * * * * *

AT1G79650      DRALVIEAFLACDRNEELAANYLLENSGDFED
RAD23B         PGLVIQAYFACEKNENLAANFLLQQNFD-ED
                ..**:*:*:*:*:*:*:*:*:*:*:* * **
    
```

Percent Identity Matrix - created by Clustal2.1

#

1: AT1G79650	100.00	42.47
2: RAD23B	42.47	100.00

Ortólogo At2g03640 de G3BP1 en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

At2g03640      MTPESNAPSVDPQFVGNQVQVEYYNHLYDSTSEVHKFYLEDSEMSIRPGLD
G3BP1          MVMEKP----SPLLVGREFVRRYYTLLNQAPDMLHRFYGKNSSYVHGGLD
                * . * .      . * : ** . ** : : ** . * : : . . : * : ** : : * :   : **
                .

At2g03640      GEIV---TIKSLKGINDQIMSIDYKSSRIEILTADSQSTLKNQVVTLVTG
G3BP1          SNGKPADAVYQKEIHRKVMSQNFNTCHTKIRHVDAAHATLNDGTVVQVMG
                . :           : : . * * : : ** : : . . . : : * . * : : ** : : ** . * *
                .

At2g03640      LVIGNDGGRRKFSQSFFLVSR---NGSYFVLNDTFRYVSDEFV----EP-
G3BP1          LLSNNNQALRRFMQTFVLAPEGSVANKFYVHNDIFRYQDEVFGGFVTEPQ
                * : . * : . * : * * . * . . . . . . . . : : * * * * * . : *   * *
                .

At2g03640      ----EATKEVEESQSTNAITAEPANESVEAVIVPTEAKTTVTKPASAI PN
G3BP1          EEEEEVEEPEERQQTPEVVPDDSGTFYDQAVVSNMEEHLEEPVAEPEP
                * . : * * * * . * : : . . : : . : * . . : :   : : * . :
                .

At2g03640      GHAKVPEEKVVNE-----NS--SLPKAAEAKLQEEV
G3BP1          DPEPEPEQEPVSEIQEEKPEPVLEETAPEDAQKSSSPAPADIAQTVQEDL
                .   * * : : * . *           : * . *   : : * : :
                .

At2g03640      PKKSFALIVQSLAQSAAGTLQVKASPVK-----
G3BP1          RTFSWASVTSKNLPPSGAVPVTGIPPHVVKVPASQPRPESKPESQIPPQR
                . * : * : . . . .   . : * : : * . . * :
                .

At2g03640      ----RKPVEKPVAAPERKAPSPIRKQASAESIKPQA----QGSSIFVAN
G3BP1          PQRDQRVREQRINIPQQRGPRPIREAGEQGDIERRMVRHPDQSHQLFIGN
                : :   * : :   * : : . * * * : . .   . * : * :   . .   . : * : . *
                .

At2g03640      LPMDATIEQLYETFKSFGAIRKDGIVRSYPEKKNKIGFVAFENGEAVKN
G3BP1          LPHEVDKSELKDFQSYGNVVE--LRINSGG-KLPNFGFVVFDSEFPVQK
                ** : .   . : * : * : * : * * : :   : : . *   *   : * * * . * : . * : :
                .

At2g03640      VFQAHRESPIRIGNRRASIEEKRGGNQNGNRVSTRNNSGYKNEEDGFRRD
G3BP1          VLSN-RPIMF-RGEVRLNVEEKKTRAAREGDRRDNRRLRG-----
                * : .   *   :   * : * . : * * * :   : : * * . . * . .
                .

At2g03640      GYKPRGSGVNGGRGYGRRNSENGDGKAYQNNHGNTAKN---
G3BP1          -----PGGPRGGLGGMRGPPRG---MVQKPGFVGRGLAPRQ
                . * . * * * * . .   . *   * : * . * . .
                .
    
```

Percent Identity Matrix - created by Clustal2.1

#

1: At2g03640	100.00	28.10
2: G3BP1	28.10	100.00

Ortólogo At1g13730 de G3BP1 en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

At1g13730      MALESNAPVVDPNNTIGNSFVEKYYNLLYKSPSQVHQFYLDSSV--LGRPG
G3BP1          VMMEKPSPL----LVGREFVRQYYTLLNQAPDMLHRFYGKNSSYVHGGLD
               *.:*. :*:      :*..**.:**.* ** :*: .*:** :.*  * .

At1g13730      SDGEMV-SVKSLKAINQIMSFYIEISKIQILTADSQASYMNGVVTLVTG
G3BP1          SNGKPADAVYGQKEIHRKVMSONFTNCHTKIRHVDAAHATLNDGVVVQVMG
               *:*: . :* . * *.:** ** :. :.* .*:** :***. * *

At1g13730      LLTVKEGQRMRFQSFFLVPLN---GSYFVLNDVFRYVADEI-----VEP
G3BP1          LLSNNNQALRRFMQTFVLAPEGSVANKFYVHNDIFRYQ-DEVFGGFVTEP
               **: :.  ** *.*.*. . .:.* **:* ** ** :.  **

At1g13730      EA-N-----KKEVEEVIP-----QVVQP---TEQV-DEVAEPV
G3BP1          QEESEEEVEEPEERQQTPEVVPDDSGTFYDQAVVSNDMEEHLEEPVAEPE
               : .      :.:. **:*      *.* .      *:* : ****

At1g13730      TIPTQQPEAKQTTENTVKKPERAVANGHPKTQEDNV---VNDKSNQVDAP
G3BP1          PDPEPEPEQEPVSEIQEEKPEPVLEETAPEDAQKSSSPADIAQTVQED
               . *  :** : .:*  :*** .: : * : ..      * : : *

At1g13730      KKSFA--HIV-QDLAQNGAT-----FNAKASPAKPKSKPVTKPSAA
G3BP1          LRTFSWASVTSKNLPPSGAVPVTGIPPHVVKVPASQRPESKPEIQIPPQ
               :*:  . :.*. **.      . . ** .:*** ** :. .

At1g13730      R-----ESKAPAPVSEHSSAATIDQ----QAEGYTIFVA
G3BP1          RPQRDQRVREQRINIPPQRGPRPIREAGEQGDIEPRMVRHPDSHQLFIG
               *      .:. * * : * . . . * :      :.:. : **:

At1g13730      NLLMDATPEQLNETFKGFGAITKDGIVRSYRLKGNCFGFVTFASAEAVK
G3BP1          NLPHEVDKSELKDFQSYGNVVELRINS--G-GKLPNFGFVVFDSEPVQ
               ** :.  .:*** :*:** :. : * :      * ***** * .:*. :

At1g13730      LVLQAHKESAIRIGNRRVSIIEKRGNNDNGRPAMRNGGYRNDNGYRNDGY
G3BP1          KVLNSR--PIMFRGEVRLNVEEKKTR-----AAR-EGDRRDNRRLRGGG
               **. : . : * : *.:***: .      * * * *.* ** * . *

At1g13730      RPRNGSNGGRGYGRNGSERRGESRNGEAHNGDGKVVHQNQTVEASR
G3BP1          -PR-----GGLGGMRGPPRGGMVQKPGFGVGRGLAPRQ-----
               **      ** * * .*. * * :.      * * . :
    
```

Percent Identity Matrix - created by Clustal2.1

#

1: At1g13730	100.00	31.41
2: G3BP1	31.41	100.00

Ortólogo At1g69250 de G3BP1 en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

At1g69250      MATEGVVPSAQDIAAEFVRQYYHVLGQLPHEARRLYVDASV--VSRPDVT
G3BP1          MVMEK--PSPLLVGREFVRQYYTLLNQAPDMLHRFYGKNSSYVHGGLDSN
               * . *   ** .   . . ***** :*. * .   :*. * . *   . * .

At1g69250      GTMM-SFTSVEAINKHILSCDFENTKFEVLSVDSQNSLEDGIFIMVIGFM
G3BP1          GKPADAVYGQKEIHRKVMSQNFNTCHTKIRHVDAHATLNDGVVQVMGLL
               * .   :. . . : *:::.* * * : :.  **:: :*:***.:. :*::

At1g69250      TGKDNQRRKFSQMFYLARQN---TLVVLNDMLRYVDQEDSSTTETPCEP
G3BP1          SNNNQALRRFMQTFVLAPEGSVANKFYVHNDIFRYQDEVFGGFVTEPQEE
               :.:::  ** * * * ** :.   .: * **::** * :   . .   * *

At1g69250      VT-EIVRPADGLKK-----AE
G3BP1          SEEEVEEPEERQQTPEVVPDDSGTFYDQAVVSNDMEEHLEEPVAEPEPDP
               * : . * :   :.

At1g69250      KTELKQKNVASVEKSVNAAVEKNAAPLDNGKMKQSEKAVITQKVTEPDAA
G3BP1          EPEPEQEPVSEIQEEKPEPVLEETAPEDAQKSSSPAPADIAQTVQED---
               :.* :*: *:::..   . * ::** * * ... * *:*.* *

At1g69250      PQQDGAKRSFADIVGSM--AKNAAPFQVK--SPVQAPVQKPKYVQGPRAA
G3BP1          ----LRTFSWASVTSKNLPPSGAVPVTGIPPHVVKVPASQPRPESKPESQ
               . *:*.....   ...*.*   *:*.....:  :*:.

At1g69250      AAPQKPAYVSK-----SIKK---NDQKVIEVP-GTS
G3BP1          IPPQRQORDQRVREQRINIPQRGPRPIREAGEQGDIERRMVRHPDSHQ
               .**.*   .:   : : : : : . * . .

At1g69250      IFVANLPLNAMPPQLFELFKDFGPIKENGIQVRSSRGANPVCFGFISFE
G3BP1          LFIGNLPHEVDKSELKDFQSYGNVVE---LRINSGGKLP-NFGFVVFD
               :*:.*** :.   .:* :*:.* : *   :* . * . * ***: *

At1g69250      TVASVQSVLQAAKNTPFML-ADRKLRVKEKEVD--YDGSKPSGKTKG---
G3BP1          DSEPVQKVL---SNRPIMFRGEVRLNVEEKKTRAAREGDRRDNRRLRGGG
               .**.*   . * *:* :. : *.****.   :*:. . : : *

At1g69250      ---GSNKTQNGSADSSKTENGSAEDSADSKTNGSAEDGEKEFKQVKSRRNRKK
G3BP1          PRGGLGGGMRGPPR-----GGMVQKPGFGVGRGLAPRQ-
               * .   .**   ..   : *   :.   * :

At1g69250      SEAAH
G3BP1          -----

```

Percent Identity Matrix - created by Clustal2.1

#

1: At1g69250	100.00	26.55
2: G3BP1	26.55	100.00

Ortólogo At5g60980 de G3BP1 en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

At5g60980      MAQQEASPSPGAIEVVGRAFVEQYYHILHQSPGLVHRFYQD-SSFLT-RPD
G3BP1          MVMEKPSP----LLVGREFVRQYYTLLNQAPDMLHRFYGKNSSYVHGGLD
* . : : . * *      : * * * * * * * * * * * * * * * * * * * * * * *
At5g60980      VTGA-VTTVTTMQAINDKILSLKYEDYTAEIETADAQESHERGVIVLVTG
G3BP1          SNGKPADAVYGQKEIHRKVMSONFTNCHTKIRHVDAHATLNDGVVVQVMG
. * . : * : * : * * * * : : : : * . * * * : : * * * * *
At5g60980      RLTGNDNVRKKFSQSFFLAPQDK---GYFVLNDVFRFLEEKEVTAQARSV
G3BP1          LLSNNNQALRRFMQTFVLAPEGSVANKFYVHNDIFRYQDE--VFGGFVTE
* : : * : : * * * * * * * * * * * * * * * * * * * * * *
At5g60980      PINGTTRDVQAPIEPER-----VVV-----
G3BP1          PQEESSEEEVEEPEERQQTPEVVPDSDSGTFYDQAVVSNDMEEHLEEFVAEP
* : : : * * * * * : : : : * * * * * * * * * * * * * * *
At5g60980      --SHEPEVEPEPVASIEEEDLDNVAEVYDPSDKDEGVVVDVEPIEPPTQI
G3BP1          EPDPEPEPEQEPVSEIQEEKPEPVLEETAPEDAQKSSSP-----
. * * * * * * * * * * * * * * * * * * * * * * * * * * *
At5g60980      SHNEILSVPQGDAPKHSYASIL-----KQMK
G3BP1          APADIAQTVQEDLRTFSWASVTSKNLPPSGAVPVTGIPPHVVKVPASQPR
: : * . . * * . . * * * * : : * * * * * * * * * * *
At5g60980      SSPAPTHVARNKPR-PAP-VNQKLTAPPAEPAARPEASAHE--NVPNSS
G3BP1          PESKPESQIIPPQRPQRDQVRVREQRINIPPQ-RGPRPIREAGEQGDIENR
. . . * : : : : * * : * : : * * . . * * . * * * : :
At5g60980      HVDVEDDGHSIYVRNLPFDSTPTQLEEVFKNFGAIKHEGIQVRSNKQQ-G
G3BP1          MV-RHPDShQLFIGNLPHEVDKSELKDFQSYGNVVE----LRINSGGKL
* . * . * . : : * * * * : : * * : : * * * * * * * * * *
At5g60980      FCFGFVEFETSSGKQSALEASPVTI-GDRQAVVEEKKTNSRGGGNNGGSR
G3BP1          PNFGFVVFDDSEPVQKVLNSRNPIMFRGEVRLNVEEKKTRAAREGDRRDNR
* * * * * * * * * * * * * * * * * * * * * * * * * * *
At5g60980      GRYFSGRGSFRNESFKGGRGGGRGGYGRGGGEFSGRPKSSNPRNGGEGY
G3BP1          LRG-----
*
At5g60980      QRVPQNGGGGRGGREGGGRGGARGGSS-----
G3BP1          -----PGGPRGGLGGGMRRGPPRGGMVQKPGFGVGRGLAPRQ
* * * * * * * * * * * * * * * * * * * * * * * * * *

```

Percent Identity Matrix - created by Clustal2.1

#

1: At5g60980	100.00	33.07
2: G3BP1	33.07	100.00

Ortólogo At3g25150 de G3BP1 en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

At3g25150      MNFSVLLYINSKATFFSSSSIFLFSRRSLFFSRFSCCKTLTSAIAKEAFLR
G3BP1          MVM-----
* :

At3g25150      SSLRNFPVNRQLRFLQFLDRFFITSICAYLLRESSSHLAFELDTDHTAM
G3BP1          -----

At3g25150      AMLGAQQVPAAACTPDMVGNFVQPYYHILHQSPHVHRFYQEISKL--G
G3BP1          -----EKPSPLLVGREFVRYTLLNQAPDMLHRFYGKNSSVYHG
                :* :***. ** *** :*:*: :***** : * . *

At3g25150      RPEENGLMS-ITSTLQAIIDKKIMALGYGVISAEIATVDTQESHGGGYIVL
G3BP1          GLDSNGKPADAVYGQKEIHRKVMSQNFTNCHTKIRHVDAHATLNDGVVVQ
                :.* * : . : *:*:*: : : ::* **:: : ..* :*

At3g25150      VTGYLTGKDSVRRTFSQTFFLAPQET---GYFVLNDMFRFIDE---GTVV
G3BP1          VMGLLSNNNQALRRFMQTFVLAPEGSVANKFYVHNDIFRYQDEVFGGFVT
* * * : : : . . * * * * . * * * : : : : * * * : * * * * *

At3g25150      HGNQIPVNNVQAPVNTYQDTAAAKEIPDDFVQEKYVQENHAVKQTEVLSK
G3BP1          EPQEESEEEVEEPEERQQT---PEVVPDDS-GTFYDQAVVSNDMEEHLEE
                . : . : : * : * . : : * * * * * * * : . * * :

At3g25150      SINEPEKVFTPS-EDEQVSAEEALVT-----ETVNEAPIEVQ
G3BP1          PVAEPEPDPEPEPEQEPVSEIQEEKPEPVLEETAPEDAQSSSPAPADIA
                . : * * * * * . * * * * * : * : : . * * :

At3g25150      KVGESDSRT-----GEIPKRSYA-SIVKVMKENAAPMSASR
G3BP1          QTVQEDLRTFSWASVTSKNLPPSGAVPVTGIPPHVVKVPASQ--PRPESK
                . : . * * * * * * : * . . : * * * . : * . * :

At3g25150      TPTKVEPKK-----QEDQAIHIPL---PTPLSEKSDSGANVAVNENNQE
G3BP1          PESQIPPQRPQRDQRVREQRINIPQRGPRPIREAGEQGDIEP----RRM
                . : : * : : . : * * : * * * * * * * : . . :

At3g25150      NERALGPSIYLKGLPLDAPALLENFQKFGILRTNGIQVRSQKGF CFGF
G3BP1          VRHPDSHQLFIGNLPHEVDKSELKDFQSYGNVVELRINSG-GKLPNFGF
                . : . . : : : . * * . : : * : : * * . * : * : * * *

At3g25150      VEFESASSMQSAIEASPVMNLNGH-KVVVEEKRSTARGNYRGRSTFGVNTG
G3BP1          VVFDDSEPVQKVLNRPIMFRGEVRLNVEEKKTRAA-----REG
* * : : : : * . . . . * : * : * . : : * * * : : * . *

At3g25150      YRNEGGRGRGSFGGGRGGYGRDTFNNGYGNRRGNRRGGYANRA-NGDGGGF
G3BP1          DRRD-NRLRGP-GGPRGGLG-----GGMRGPPRGGMVQKPGFVGRG-
                * . : . * * * . * * * * * * * . * * * * * . : : . * * *

At3g25150      PRANGNNGRVRGGGNDANRATKPVDDAPRVSVAA
G3BP1          -----LAPRQ-----
                ***
    
```

Percent Identity Matrix - created by Clustal2.1

1: At3g25150 100.00 31.82
2: G3BP1 31.82 100.00

Ortólogo At5g48650 de G3BP1 en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```
At5g48650      MDSTAATKRVDPLTVGSAFVNQYYYIFCNMPEHLPRFYQEISRVRGVGQ
G3BP1          MVME-----KPSPLLVGREFVRQYYTLLNQAPDMLHRFYGKNSSVHGGL
                *          .** **  **.*** : : * : * ** : * : *

At5g48650      DGVMRDFSTFQGISEE---LKRLTYGDCNSAEITSYDTQESHNGGFLLFV
G3BP1          DSNKGPADAVYGQKEIHRKVMSQNFNCH-TKIRHVDAAHATLNDGVVVQV
                * . : . . . * . * : . : ** : : * * : : * . * : : *

At5g48650      TGYFTLNERSRRKFTQTFFLAPQE---KGFFVLNDILRFVNDDAKDNVPE
G3BP1          MGLLSNNNQALRRFMQTFLAPEGSVANKFYVHNDIFRYQDEVFGGFVTE
                * : : * : : * : * **.*** : : : * : * ** : * : : . * . *

At5g48650      TI---DGEVVS-GINSTPTIINGMKGSEQAACVSVNPVCKEVSKPL---
G3BP1          PQEESSEEEVEEPEERQQTPEVVPDDSGTFYDQAVVSNMEEHLEEPVAEP
                . : * * . . . ** : : . * : . * * : : : : * :

At5g48650      --DNENAKDNVLVPEIANEVARTEITCKEV-ADDSQKNYDPDDG-L----
G3BP1          EPDPEPEPEQEPEVSEIQEE--KPEPVLEETAPEDAQKSSSPAPADIAQTV
                * * : : * . * * : * : * . : * . : * : * . * . :

At5g48650      -ADAPKKSYSVAVLKVTKDKF-GVPAVSL-----PSPKKIPKDQE
G3BP1          QEDLRTFSWASVTSKNLPPSGAVPVTGIPPHVVKVPASQPRPESKPES--
                * . * : * * . . . * . * . : * * * . * : .

At5g48650      HQAPSDPSTGQIL-----KDQGGQASSDPSQVIESDTVS
G3BP1          QIPPQRPQRDQRVREQRINIPPQRGPRPIREAGEQGDIEPRMVRH----
                : . * . * . * : : : * : * . : * : : .

At5g48650      ESDVASENGHNQEAVAEGTSIYVRHLPFNANIDMLEAEFKQFGAITNGGI
G3BP1          -----PDSHQLFIGNLPHEVDKSELKDFQSYGNVVEL--
                . : . . : : * * . : : . * : * : : * : :

At5g48650      QVINQRGLGYPYGFVEFEEADAHRAIEASPVKI-GGLRAFVEEKLSTS-
G3BP1          -RINSGGKLPNFGFVVFDDSEPVQKVLNRPIMFRGEVRLNVEEKKTRAA
                ** . * : * * * : : : : : : . * : : * : * * * : :

At5g48650      -RG-KRGNNGVGYGNRN--VGVMRGRGSYGYDYRRGG--RGPG-GGG
G3BP1          REGDRRDNRRLRPGGPRGGLGGMRGPP-----RGGMVQKPGFGVG
                . * : * . * * . . : * * * * * * * * : * * * *

At5g48650      RSFNRRGNEYVASINSY
G3BP1          RGLAPRQ-----
                * . : *
```

Percent Identity Matrix - created by Clustal2.1

#

1: At5g48650 100.00 28.99
2: G3BP1 28.99 100.00

Ortólogo At5g43960 de G3BP1 en *Arabidopsis thaliana*.

CLUSTAL W (1.83) multiple sequence alignment

```

AT5G43960      MATPYPGATQVGSYFVGQYYQVLQQQPDLIHQFYSEPSRAIRID----GD
G3BP1          MVMEKPSPLLVGREFVRQYYTLLNQAPDMLHRFYGKNSSYVHGGLDSNGK
                * . * . . ** * * * * * : * : * * * : * : * . * .

AT5G43960      STETANSLHLHIHNMVMSLNFTAIE--VKTINSVESWEGGVLVVVSGSVKT
G3BP1          PADAVYGQKEIHRKVMSQNFTNCHTKIRHVDAHATLNDGVVVQVMGLLSN
                . : : . . . * . * * * * * . : : : : : : : * * * * * * . .

AT5G43960      KEFSNRRSFVQTFFLAPQE---KGYFVLSDVFLFVDEGT-VYYHQPSYLS
G3BP1          NNQALR-RFMQTFVLAPEGSVANKEFYVHNDIFRYQDEVFGGFVTEPQEEES
                : : : * * : * * * . * * * : : : * . * * : * * : : * . *

AT5G43960      EIKHEAQLNPPTRHPDPQVSDYVLEEEASDYVNAVQIKDDLVDKYSLQED
G3BP1          E---EEVEEPEERQQTPEV---VPDDSGTFYDQAV-VSNDMEEH--LEEF
                * * * : * * : * * * * : : . : * : * * : : * * : : * *

AT5G43960      QHQPQHEDYEDEVAIEETPREEVAVDVVHEHRAAPVEE-PVGE-----
G3BP1          VA---EP---EPDPEPEPEQE-PVSEIQEEKPEPVLEETAPEDAQKSSS
                * * * * * * * : * : * . : : * : . * * * . . *

AT5G43960      -----K-SKMSYASIL-KVAKEAATVPVAATQPSYNKSSQ---
G3BP1          PAPADIAQTVQEDLRTFSWASVTSKNLPPSGAVPVTGIPPHVVKVPASQP
                . : * : * * : * : : * * : . * * * .

AT5G43960      -----DINEW-----DQPMRTPSPQLAAPLAPIQQSNS
G3BP1          RPESKPESQIPPQRQDRQVRVREQRINIPPQRGPRPI-----
                * . * * * *

AT5G43960      STYVSDYGAAEDGSGFEDFEFKSVYVRNLPSDISASEIEEEFKNFGTIK
G3BP1          --REAGEQGDIEPRRMVRHPDSDLFIGNLPHEVDKSELKDFQSYGNVV
                : . : : * . . . : : : : * * * : . * * : : * : * :

AT5G43960      PDGVFLRTRKDVMGVCYAFVEFEDMTSVENAIKASPIYLGGR-QVYIEER
G3BP1          E----LRINSGGKLPNFGFVVFDSEPVQKVLNRPIMFRGEVRLNVEEK
                * * . . . : * * * : * . * : : . * : : * * :

AT5G43960      RPNPAGVRGARRGG-GRGRGGYPTEAPRGRFGGRGSGRGNQDGGDYRPRG
G3BP1          KTRAA-REGDRRDNRRLRGP-----GGPRGGLGG--GMRGPPRGGMVQKPG
                : . . * . * * * . * * . * * * : * * . * * * * : *

AT5G43960      NGYYRGGGR--
G3BP1          FGVGRGLAPRQ
                * * * .
    
```

Percent Identity Matrix - created by Clustal2.1
#

1: AT5G43960	100.00	30.37
2: G3BP1	30.37	100.00

Anexo III. Aminoácidos conservados como abundantes y carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogos.

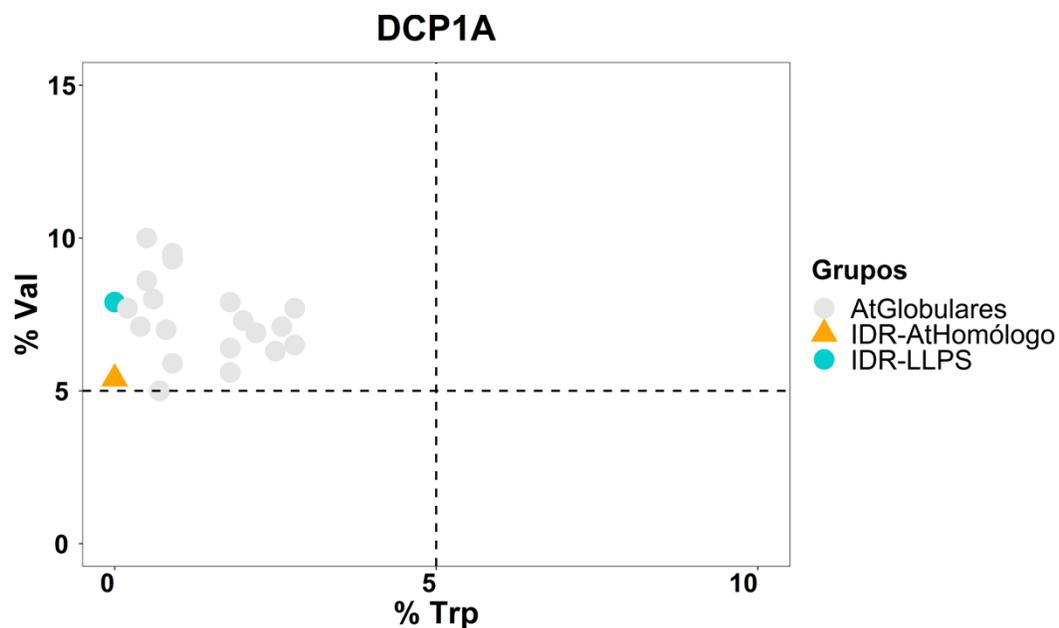


Figura A.3.1. Determinación de la conservación de aminoácidos abundantes en el grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína DCP1A.

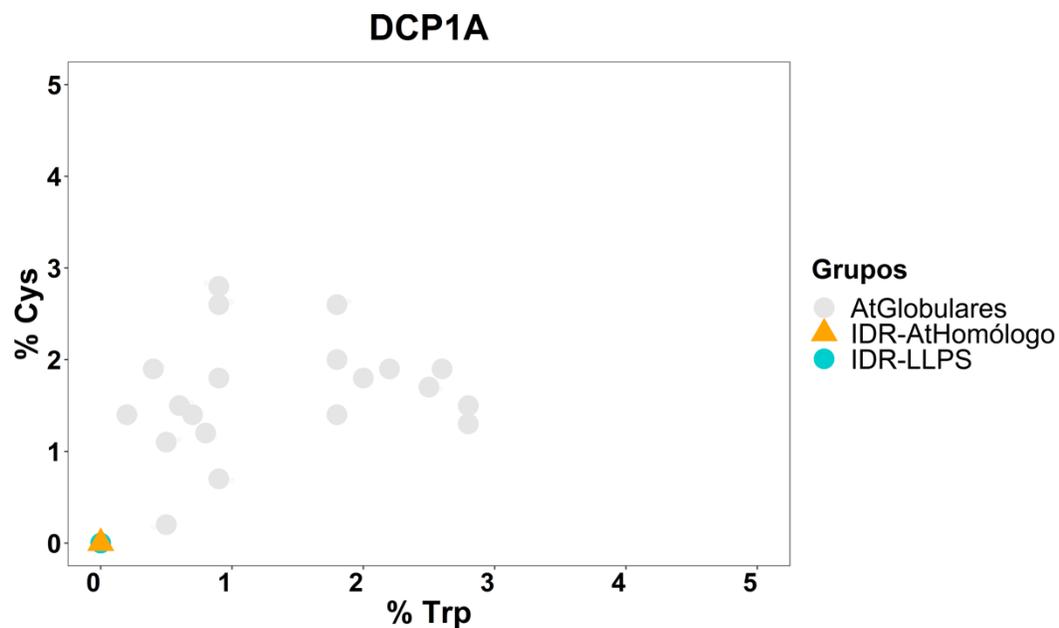


Figura A.3.2. Determinación de la conservación de aminoácidos carentes en el grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína DCP1A.

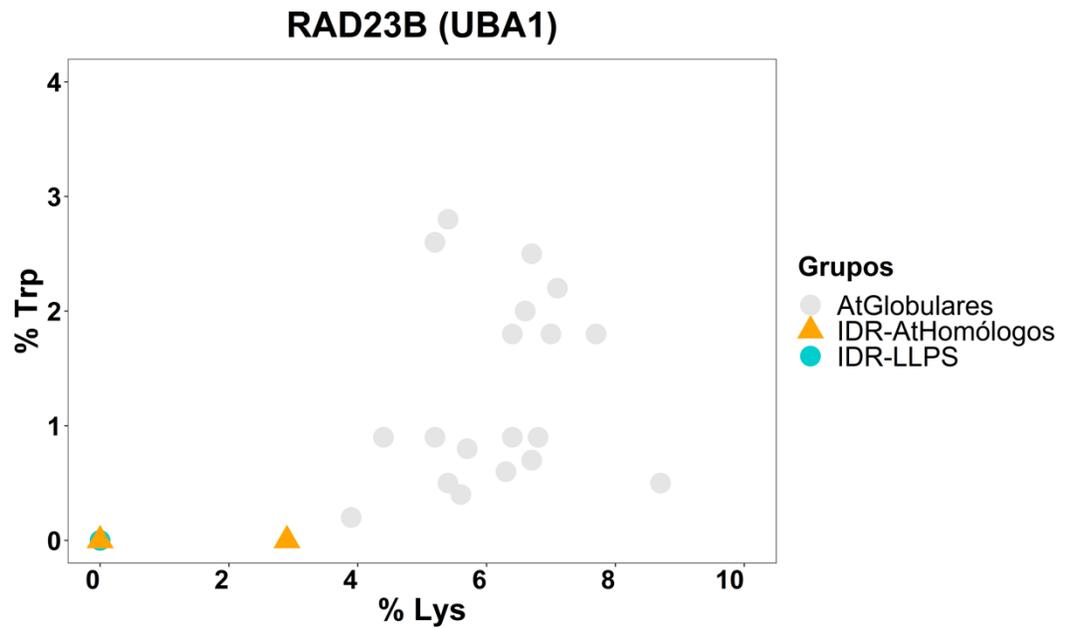


Figura A.3.3 Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína RAD23B (UBA1).

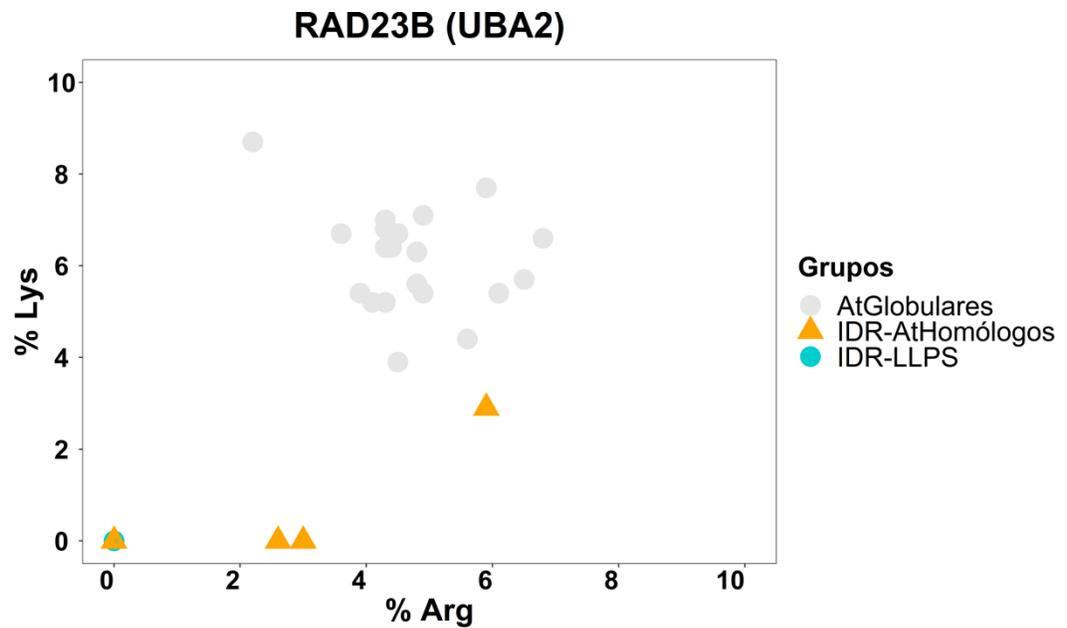


Figura A.3.4 Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína RAD23B (UBA2).

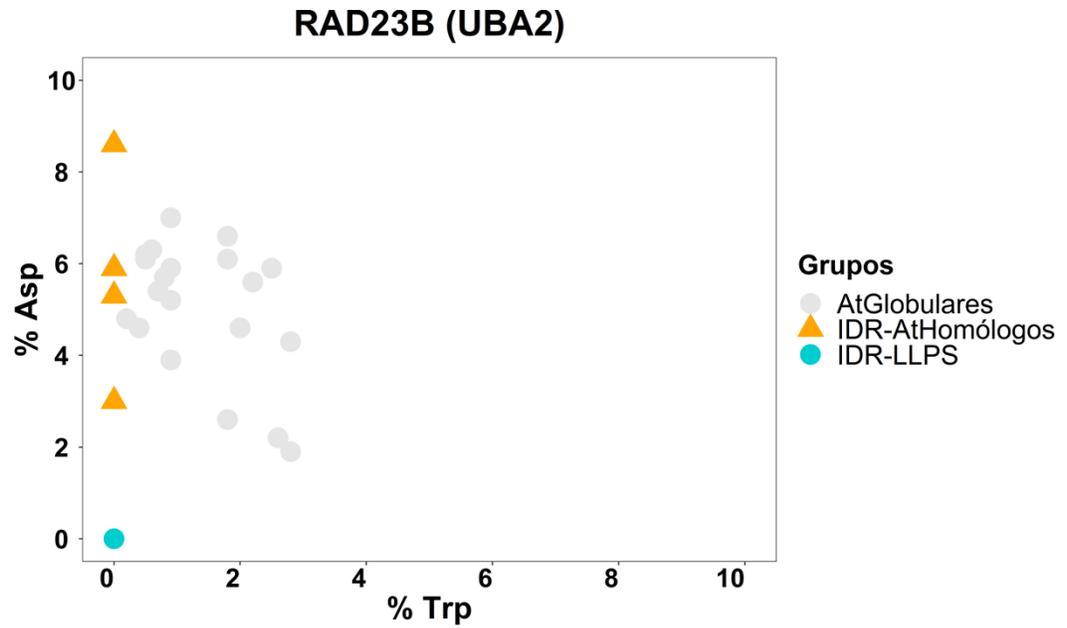


Figura A.3.5. Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína RAD23B (UBA2).

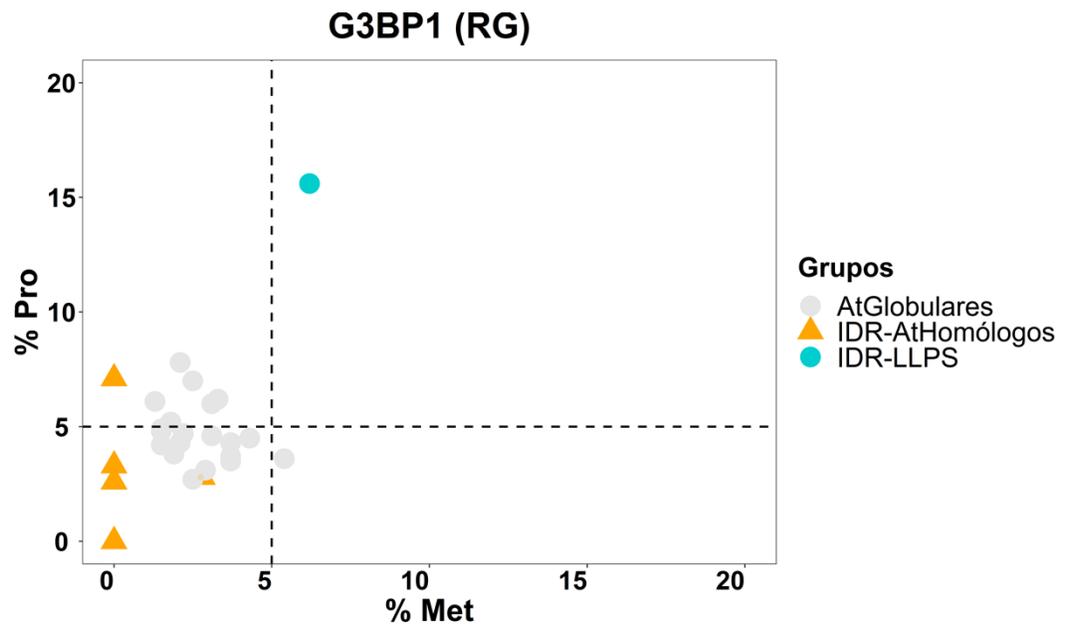


Figura A.3.6 Determinación de la conservación de aminoácidos abundantes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (RG).

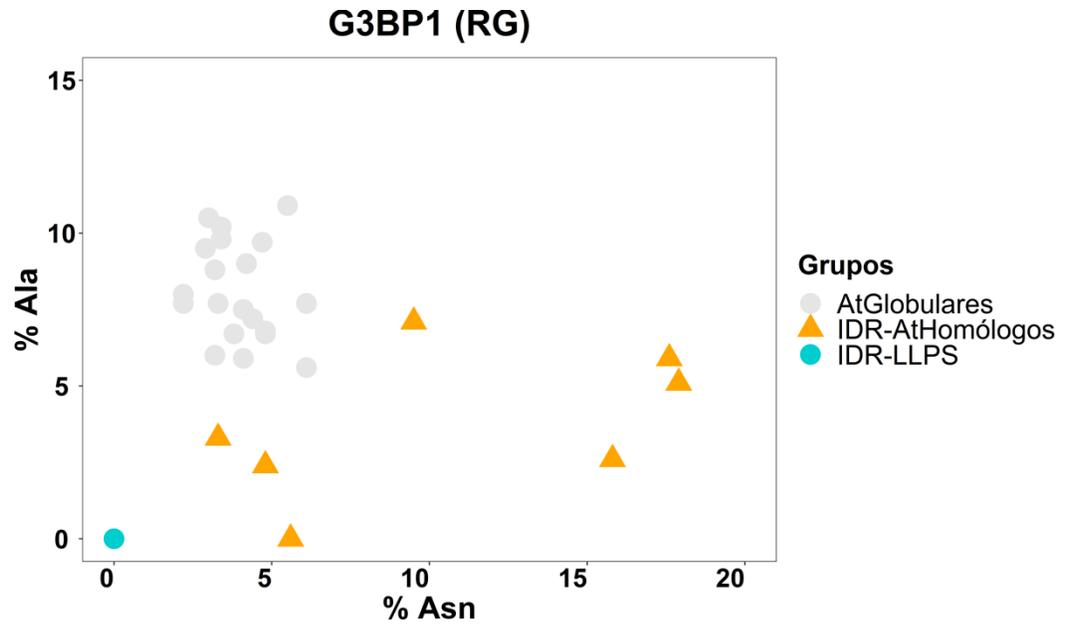


Figura A.3.7 Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (RG).

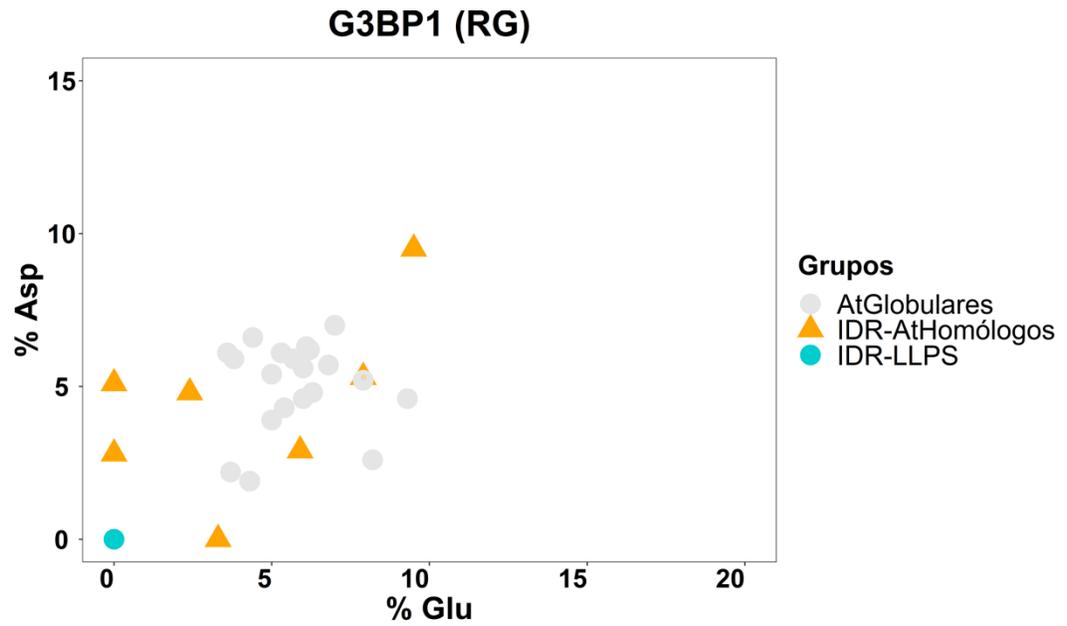


Figura A.3.8 Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (RG).

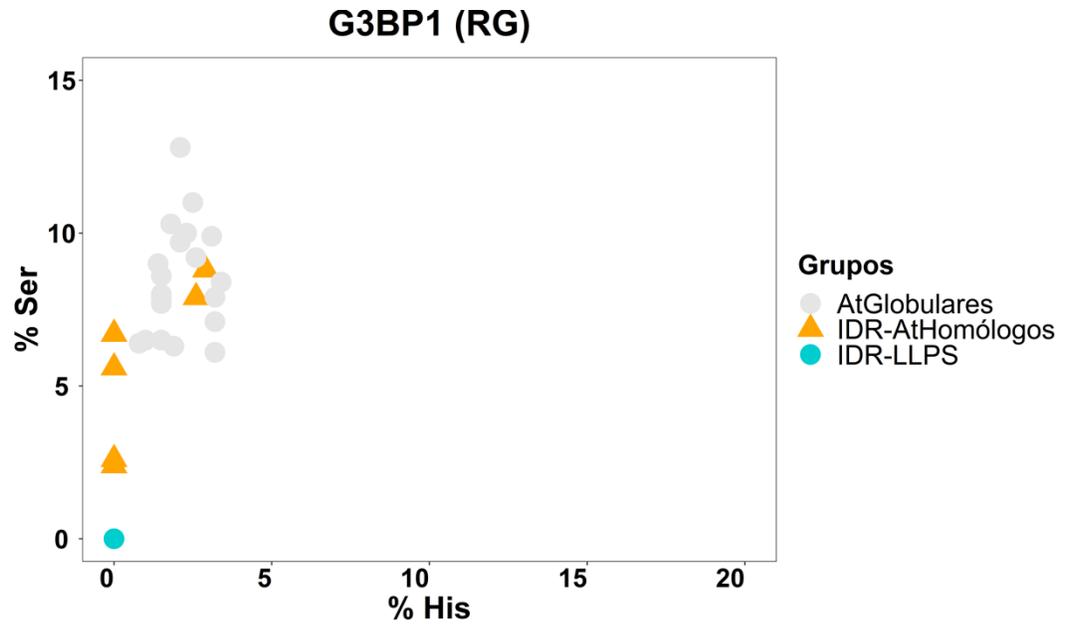


Figura A.3.9 Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (RG).

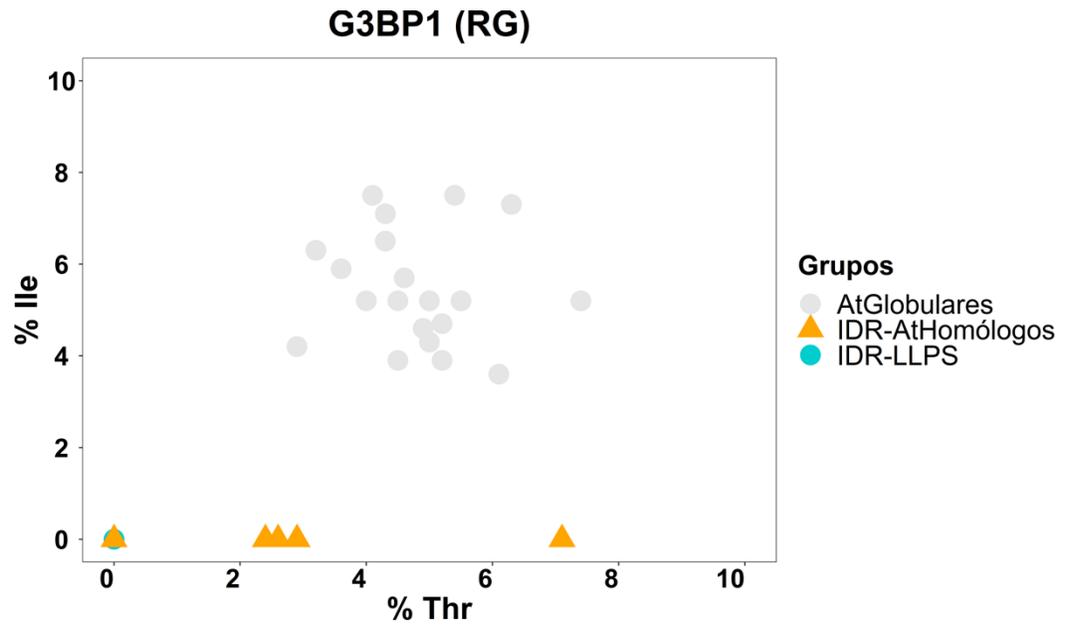


Figura A.3.10 Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (RG).

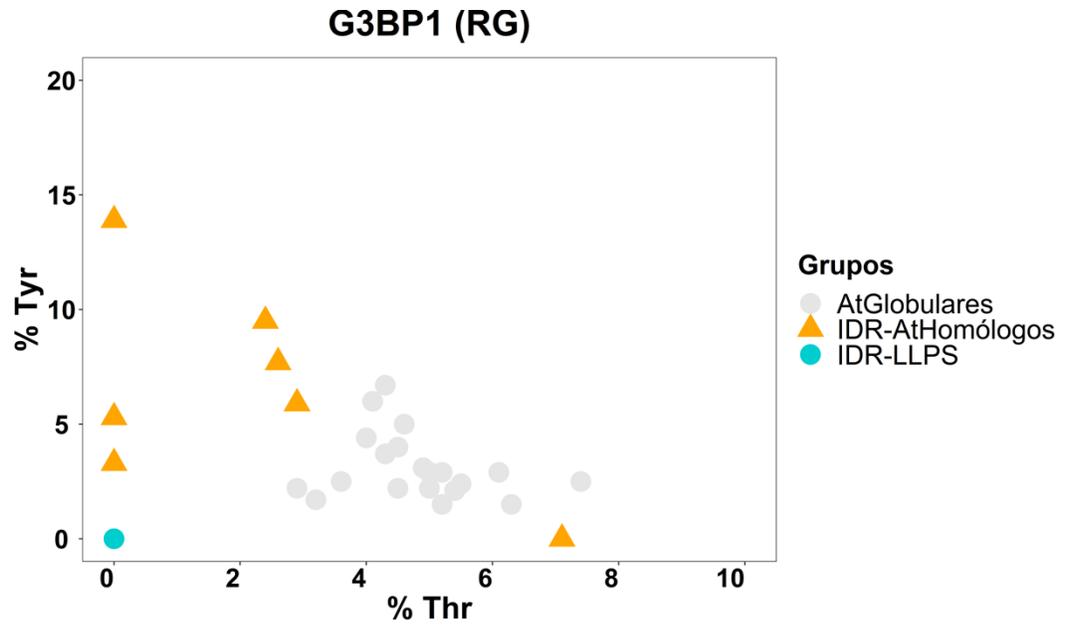


Figura A.3.11. Determinación de la conservación de aminoácidos carentes del grupo IDR-LLPS en el grupo IDR-AtHomólogo de la proteína G3BP1 (RG).