

CENTRO DE INVESTIGACION Y DE ESTUDIOS AVANZADOS DEL
INSTITUTO POLITÉCNICO NACIONAL



UNIDAD ZACATENCO

DEPARTAMENTO DE BIOQUÍMICA

Caracterización parcial del dominio KH-QUA2 de la
subunidad grande del Factor Auxiliar snRNP U2 (U2AF84) de
Entamoeba histolytica

Tesis que presenta

Ricardo Martínez Baltazar

Para obtener el grado de

Maestro en Ciencias

En la especialidad de

BIOQUÍMICA

Director de tesis:

Dr. Jesús Valdés Flores

Ciudad de México

Septiembre, 2016.

El presente trabajo fue realizado en el laboratorio 11 del departamento de Bioquímica bajo la dirección del Dr. Jesús Valdés Flores. Durante el desarrollo del trabajo, el autor fue becario CONACYT (No. 336869).

Agradecimientos:

Al Dr. Jesús Valdés Flores

I.B.Q Sandra Lorena Palacio Molina

José Martin Velázquez Lozano

Dr. Carlos Gómez Lojero

Dr. Nicolás Villegas Sepúlveda

Dr. Carlos Ortuño Pineda

Raúl Bonilla Moreno

Dra. María Sarahi Mendoza Figueroa

Dra. Mónica de Nova Ocampo

Q.B.P Adán Arizmendi Izazaga

M.C José Manuel Jáuregui Wáde

Dedicatoria

A mi familia: Yeyri Sofia Martinez Solís (hija), Yenny Solís de la O (esposa).

Familia materna: Josefina Baltazar Gonzales (madre), Roque Maldonado Luna+ (padre), América Sarabia Baltazar, Rodrigo Martinez Baltazar, María de Jesús Sarabia Baltazar, Felipe de Jesús Sarabia Baltazar, José Alberto Sarabia Baltazar.

Amigos y personas especiales: Sandra Lorena Palacio Molina y Francisco Sandoval, José Manuel Jáuregui Wáde.

“SOLO HAY UNA COSA QUE NO EXISTE.... AQUELLA QUE NO ESTÁ NI SI QUIERA EN TU IMAGINACIÓN”.

INDICE

LISTA DE ABREVIATURAS	8
RESUMEN	9
1. INTRODUCCIÓN	11
1.1 Mecanismo de <i>splicing</i> : ensamblaje del spliceosoma.....	11
1.2 El <i>splicing</i> alternativo.....	13
1.3 <i>Entamoeba histolytica</i>	15
1.4 <i>Splicing</i> en <i>E. histolytica</i>	16
1.5 Las proteínas de unión al sitio de ramificación.....	17
1.5.1 Las proteínas de unión a BP son sitio específicas para la unión al RNA.....	18
1.6 Interacción entre MSL5-MUD2: funciones en el reconocimiento del BP.....	19
1.7 Bases estructurales para el reconocimiento del sitio de ramificación por las proteínas de unión a BP.....	21
1.8 Reconocimiento de los BP en secuencias no conservadas.....	25
2. ANTECEDENTES DIRECTOS Y JUSTIFICACIÓN	26
3. HIPOTESIS	26
4. OBJETIVOS	27
5. ESTRATEGIA EXPERIMENTAL	28
6. MATERIALES Y METODOS	30
7. RESULTADOS	37
8. DISCUSIÓN	52
9. CONCLUSIÓN	56

10. PERSPECTIVAS.....	57
11. BIBLIOGRAFIA.....	58
12. ANEXOS.....	61

LISTA DE ABREVIATURAS

BP	Branch point
CDC	Centro de control y prevención de enfermedades
CLIP	Inmunoprecipitación de complejos tempranos de <i>splicing</i> entrecruzados con UV
CLPD	<i>Cross linking-Pull Down</i>
<i>E. histolytica</i>	<i>Entamoeba histolytica</i>
EhSF1	Factor de <i>splicing</i> -1 de <i>E. histolytica</i>
EMSA	Esayo de cambio de la movilidad electroforética
GST	Etiqueta de glutation S-transferasa
HIS	Etiqueta de hisitidinas
IR	Retención intronica
KH-QUA2	Dominio de unión a BP
nt.	Nucleótido
pEhExHA	Plásmido de expresion en <i>E. histolytica</i> con etiqueta de hemaglutinina
pre-mRNA	Transcrito no procesado
RNA	Acido ribonucleico
RRM	Motivo de reconocimiento a RNA
SF1	Factor de <i>splicing</i> -1 humano
U2 snRNA	Secuencia de RNA de U2 que contacta con el BP
U2AF35	Subunidad pequeña del factor auxiliar U2
U2AF65	Subunidad grande del factor auxiliar U2
U2AF84	Subunidad grande del factor auxiliar U2 de <i>E. histolytica</i>
UHM	Motivo de homología a U2AF
ULM	Motivo de unión a UHM
3'ss	Sitio de <i>splicing</i> 3'
5'ss	Sitio de <i>splicing</i> 5'

RESUMEN

Los intrones de *Entamoeba histolytica* (*E. histolytica*) no poseen una secuencia de ramificación (BP) conservada, por ello SF1 (Splicing Factor 1) y los factores auxiliares de U2 (U2AF65 y U2AF35) son fundamentales para su reconocimiento así como del tracto de pirimidinas y el sitio de splicing 3'. La inmunoprecipitación de complejos tempranos de *splicing* entrecruzados con UV (CLIP) acoplados a espectrometría de masas identificaron 32 factores de *splicing* amebianos, incluyendo U2AF84 (el posible ortólogo de U2AF65), SF1 y U2AF33 (posible ortólogo de U2AF35). Previa e inesperadamente se observó que el silenciamiento de U2AF84 incrementa el *splicing* de algunos transcritos amebianos, sugiriendo que está implicado en la retención intrónica (IR) mediada por inhibición del *splicing*. A diferencia de U2AF65 y Mud2 en levadura, U2AF84 contiene una región extra de 226 a.a. dentro de los cuales 136 a.a. son muy similares al dominio KH-QUA2 de los SF1 reportados. Esto nos hizo preguntarnos si el dominio KH-QUA2 de U2AF84 (KQ2-84) y el dominio KH-QUA2 de SF1 (KQ2-SF1) compiten por la unión al BP o interaccionan entre ellos y si esto conlleva a un incremento en el *splicing* o una inhibición del mismo. Para explorar estas interrogantes, sobreexpresamos el cDNA de U2AF84 y U2AF84 Δ C (deletadas del dominio KH-QUA2) en trofozoítos amebianos lo que resultó en una acumulación o disminución del pre-mRNA de algunos transcritos, respectivamente, indicando que este dominio es el responsable de la inhibición del *splicing* en *E. histolytica*. Para explorar cómo se lleva a cabo esta inhibición realizamos ensayos de retardo, *cross linking* con UV, con formaldehído, o con ambos y ensayos de CLIP. En solución, KQ2-84 bloquea la unión de KQ2-SF1 al BP mediante una interacción independiente de RNA. Las interacciones entre KQ2-84 y KQ2-SF1 se comprobaron por medio de far-wb y además KQ2-84 estabiliza las interacciones previamente formadas entre el KQ2-SF1 y el BP bloqueando parcialmente el intercambio de SF1 por el U2 snRNA en el BP. En conclusión, nuestros datos sugieren que el dominio 84-KQ2 no es funcionalmente equivalente al de SF1, sin embargo las interacciones entre ellos inhiben parcial del *splicing* que resultan en la IR.

ABSTRACT

The introns of *E. histolytica* possess conserved 5' (GUUUGU) and 3' (UAG) splice sites, but the branch point sequences (BS) lack such conservation. Therefore the polypyrimidine tract, SF1 (*Splicing Factor 1*) and the U2 snRNP auxiliary factors are essential for 3' splice site recognition. Recently, the snRNP protein U1A was tag-cloned and used in UV cross-linking immunoprecipitation (CLIP) assays coupled to tandem mass spectrometry analyses identified 32 amoeba splicing-specific factors including U2AF84 (a possible ortholog of the human U2AF65), SF1, and U2AF33 (possible ortholog of U2AF35). We previously showed that silencing of U2AF84 increased the splicing variants of some amoeba transcripts, suggesting that it is involved in intron retention via splicing inhibition. Unlike U2AF65 and Mud2, U2AF84 contains 226 extra amino acids of which 136 are very similar to the KH-QUA2 domain of the human SF1. We asked whether the KH-QUA2 domain of U2AF84 (KQ2-84) and the KH-QUA2 domain of SF1 (KQ2-SF1) compete for BP binding, or interact with each other enhancing or inhibiting splicing. To explore these possibilities, in vivo splicing assays were carried out with the wild type and U2AF84 Δ C (without KH-QUA2 domain) amoeba transformants and found that the pre-mRNA of some transcripts accumulated or decreased, respectively, indicating that the KQ2-84 domain causes IR. To shed some light in the mechanism involved, EMSA, UV/FH-CL and CLIP assays carried out with recombinant KQ2-84 and KQ2-SF1. These experiments suggested that, in solution, KQ2-84 prevents binding of KQ2-SF1 to the BP, through an RNA-independent interaction, and is able to enhance KQ2-SF1 binding to the BP; KQ2-84-KQ2-SF1 protein-protein interactions were confirmed by far-wb. Furthermore, KQ2-84 stabilizes the preformed interactions between KQ2-SF1 and the BP blocking SF1 to U2 snRNA interchange. We conclude that 84-KQ2 is not equivalent functionally to KQ2-SF1, however this domain is responsible for the splicing Inhibitory function that leads to intron retention.

1. INTRODUCCIÓN

1.1 Mecanismo de *splicing*: ensamblaje del spliceosoma

Durante el proceso de expresión génica, un paso muy importante es la remoción de las secuencias no codificantes (intrones) y la unión de las secuencias codificantes (exones); este proceso es conocido como *splicing*. La maquinaria de *splicing* no solo reconoce y remueve los intrones de los pre-mRNA para hacer el mensaje correcto que será posteriormente traducido a proteína, sino que también puede generar isoformas las cuales pueden ser funcionalmente diversas. A este último mecanismo de regulación se le conoce como *splicing* alternativo (Brett et al., 2001; Ast, 2004). La reacción de *splicing* se lleva a cabo por el spliceosoma, una partícula ribonucleoproteica altamente dinámica de 60 S (Staley & Guthrie, 1998). El spliceosoma consiste de cinco complejos de ribonucleoproteínas pequeñas nucleares U1, U2, U4, U5 y U6 y centenares de otras proteínas. El spliceosoma lleva a cabo su función a través de una gran cantidad de interacciones RNA-RNA, RNA-proteínas y proteína-proteína, esto con la finalidad de escindir cada intrón y unir los exones en el orden correcto.

El *splicing* se lleva a cabo a través de dos reacciones de transesterificación de tipo SN2 (sustitución nucleofílica dos). Estas reacciones químicas involucran grupos funcionales de 3 regiones reactivas en el pre-mRNA: durante la primera reacción el hidroxilo en la posición 2' de la secuencia BP ataca nucleofílicamente al fosfato 5' del dinucleótido GU en el sitio donador de *splicing* ó 5'ss y esto da lugar a la formación de un intrón lariat. El rearrreglo del RNA por la maquinaria de *splicing* permite el segundo ataque nucleofílico, el cual es ahora dirigido por el hidroxilo 3' disponible en el primer exón (o exón 5'), el cual ataca al fosfato del nucleótido G del sitio aceptor de *splicing* o sitio 3'ss (Fig. 1) (Gordon et al., 2000; Lewin, 2002; Turner et al., 2004).

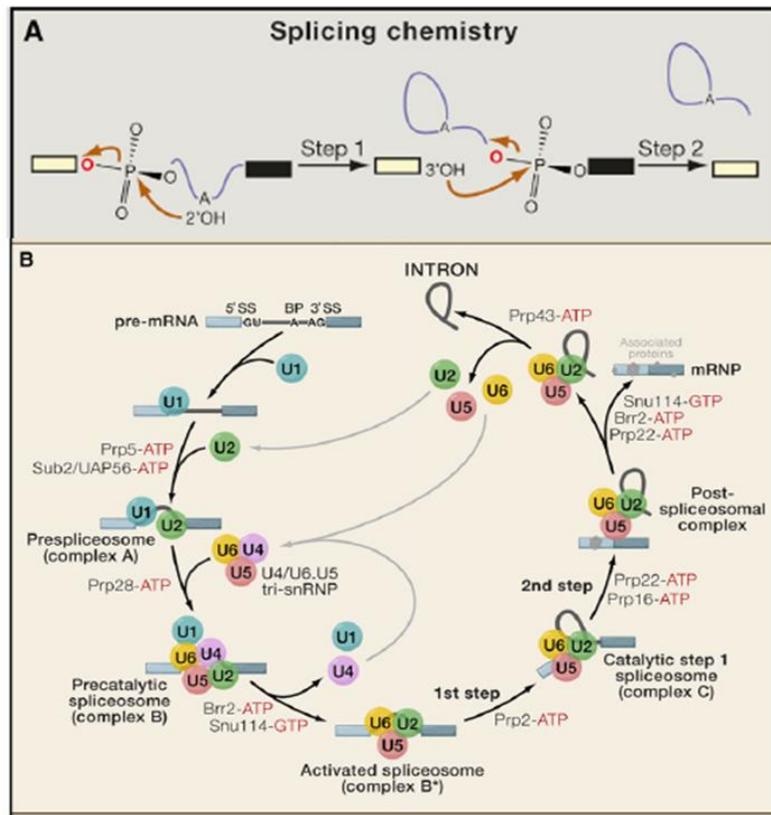


Fig 1. Ensamblaje del spliceosoma. Los hidroxilos responsables de las dos reacciones de trans esterificación se muestran en el panel A. En el panel B podemos apreciar la dinámica del spliceosoma durante la remoción de los intrones. Se muestran algunas helicasas requeridas en el ensamblaje del spliceosoma así como la corrección del splicing. Los pasos que requieren ATP son indicados en el dibujo. Adaptada de Markus et al., 2009.

Un requisito esencial para un *splicing* eficiente son las secuencias en cis por lo que el pre-mRNA debe contar con un sitio donador conservado, un BP seguido por un tracto rico en pirimidinas y un sitio aceptor también conservado (Fig. 2 panel superior). Uno de los primeros pasos en el ensamblaje del spliceosoma sobre el RNA es la formación del complejo E el cual involucra el reconocimiento del sitio donador por U1 snRNP (ribonucleoproteína nuclear U1), y la asociación de factores no snRNP tales como proteínas SR, U2AF35, U2AF65 (del inglés U2 snRNP Auxiliary Factor), y sobre todo la unión de SF1 al BP. En seguida el U2

snRNP se aparea con el BP lo cual involucra un rearrreglo del spliceosoma y la disociación de SF1 del BP, lo que da lugar a la formación del complejo A (Fig. 2 panel inferior). La subsecuente asociación del tri-snRNP U4/U6.U5 con el pre-mRNA resulta en la formación del complejo B, y finalmente se forma el complejo C remodelando las interacciones RNA-RNA y RNA-proteínas para dar lugar al spliceosoma catalíticamente competente (Fig. 1B) (Hastings et al., 2001; Lewin, 2004).

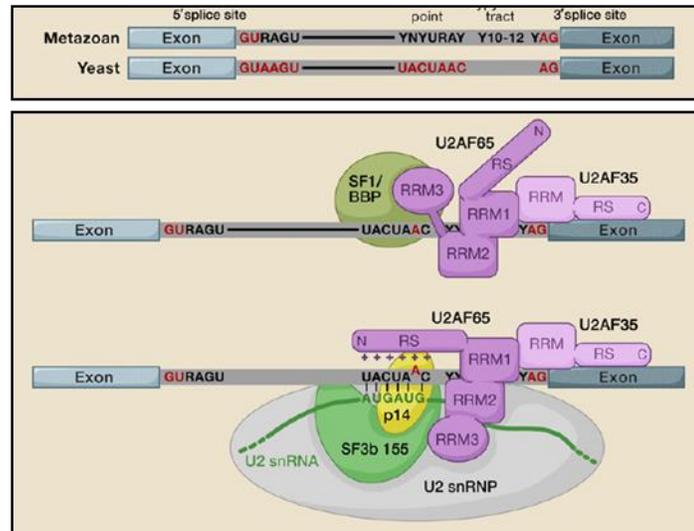


Fig 2. Transición del complejo E-A. En el panel superior se señalan algunos elementos esenciales para el inicio del *splicing* para metazoos y levaduras. El panel inferior refiere a la transición del complejo E-A donde se aprecia la salida de SF1 al entrar U2 a través del apareamiento de bases entre U2 y el pre-mRNA. Tomada de Markus et al., 2009.

1.2 El *splicing* alternativo

El *splicing* alternativo se refiere a la inclusión o exclusión diferencial de exones en el mRNA de modo que un solo gen puede resultar en múltiples proteínas (Rothrock et al., 2005). Entonces el inesperado número de genes en humanos (~26 000) que no concuerda con el número de proteínas (~90 000) ha llevado en los últimos años a proponer que el *splicing* alternativo es un mecanismo que genera diversidad de proteínas a partir de los genes con los que se cuenta. Hay 5 formas principales de *splicing* alternativos (Fig. 3): exón *skipping* que abarca el 38% de todos los eventos, sitios donador y aceptor alternativos que abarcan 8% y

18%, respectivamente; IR que comprende menos del 3% y finalmente un 33% para eventos más complejos de *splicing* alternativo (Ast, 2004 y Hastings et al., 2001).

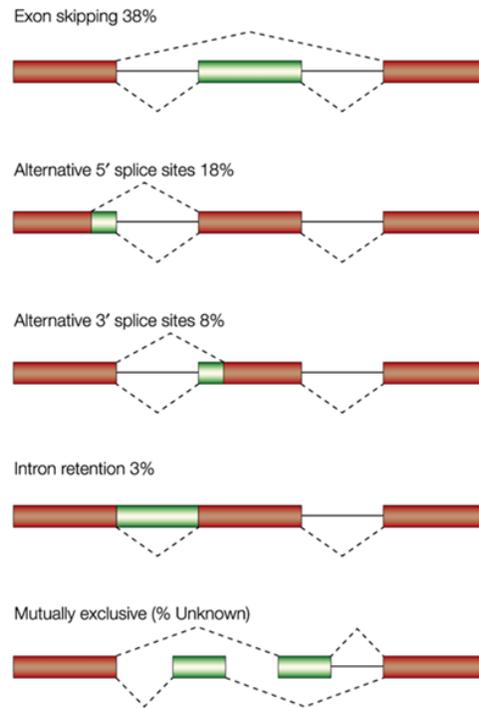


Fig 3. Tipos de *splicing* alternativo. En rojo se muestran los exones constitutivos y en verde los exones alternativos, el porcentaje que abarca cada uno de estos eventos es mostrado arriba de cada uno. Tomada de Ast, 2004.

1.3 *Entamoeba histolytica*

En el presente trabajo se utilizó como modelo de estudio a *E. histolytica* el cual es un parásito eucariótico causante de gran morbilidad y mortalidad a nivel mundial. Se estiman unos 50 millones de casos de amebiasis invasiva al año con un estimado de 40,000 a 110,000 muertes cada año (Petri et al., 2000). Las enfermedades más comunes de la infección por *E. histolytica* es la disentería, pero también ocasiona abscesos hepáticos, pulmonares y en cerebro (Carol A et al., 2001). De acuerdo a la CDC el ciclo biológico de la ameba comienza con la ingesta de la forma infectiva (quiste) a través de alimentos o agua contaminada. Dentro del huésped ocurre el desenquistamiento que libera a 8 trofozoitos (forma activamente móvil). Posteriormente sigue la multiplicación de los mismos y el enquistamiento de algunos otros que son excretados en heces (Fig. 4). Una vez dentro del huésped pueden llegar al hígado a través del torrente sanguíneo y esto podría resultar en un absceso hepático. Para causar la enfermedad, *Entamoeba histolytica* cuenta con un arsenal de factores de virulencia y patogenicidad como son las proteasas, proteínas de movilidad, receptores de unión a células del huésped, entre otras.

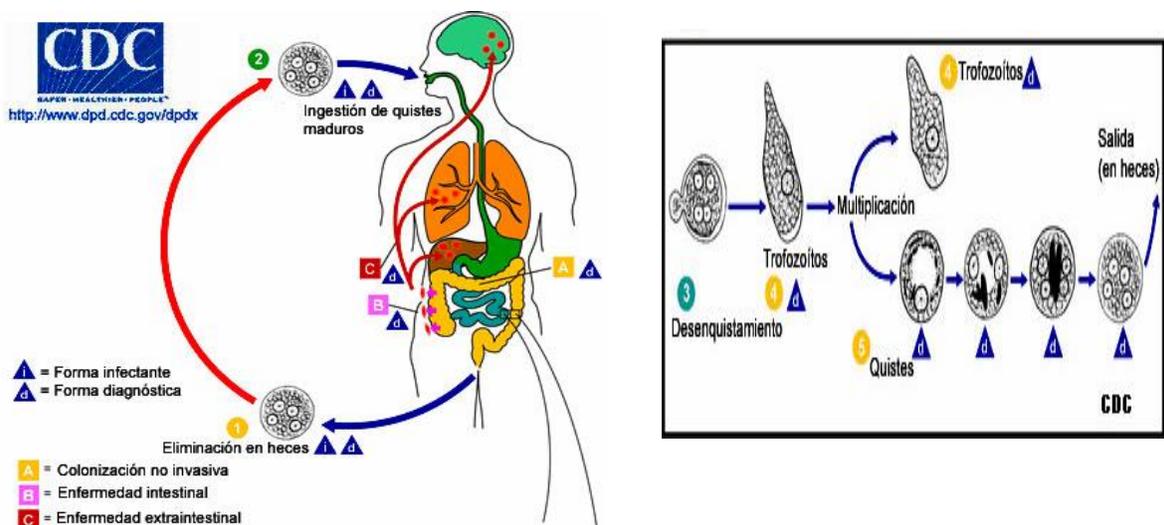


Fig 4. Ciclo biológico de *E. histolytica*. Se muestra cada una de las etapas de la infección y las formas infectantes y móviles del parásito. Además de cada una de las patologías que ocasiona cuando se localiza en algún órgano particular.

1.4 *Splicing* en *E. histolytica*

La gran mayoría de los eucariotas de ramificación temprana tienen pocos o ningún intrón, sin embargo alrededor de 3 mil intrones han sido identificados entre los 9938 genes reportados de *E. histolytica* (Loftus *et al.*, 2005). Los intrones de *E. histolytica* poseen 5´ss (GUUUGU) y 3´ss (UAG) conservados, pero la secuencia BP carece de tal grado de conservación (Wilihoeft *et al.*, 2001). Recientemente Valdés y colaboradores (2014) realizó la identificación del spliceosoma de *E. histolytica* mediante ensayos de inmunoprecipitación de complejos tempranos de *splicing* entrecruzados con UV (CLIP) acoplados a espectrometría de masas. De esta manera se identificaron 32 factores de *splicing* amebianos, incluyendo U2AF84, SF1, U2AF33 y las DExH/D helicasas requeridas para todos los pasos de *splicing*. Estos factores se suman a otras proteínas relacionadas en *splicing* identificadas previamente como la helicasa DEAD1 y Prp6 (1-3).

Muchos de los pre-mRNAs son mono intronicos y mediante análisis bioinformáticos se ha predicho que la IR es la ruta principal de *splicing* alternativo (Davis *et al.*, 2007; McGuire *et al.*, 2008). Estos datos en conjunto con la validación de intrones mediante ensayos de RT-PCR (Davis *et al.*, 2007) sugieren que el *splicing* no sólo participa en la expansión del proteoma amebiano, sino que puede constituir otro nivel de regulación de la expresión de genes. Otros trabajos han identificado al U2, U4, U5 y U6 snRNAs amebianos (Miranda *et al.*, 1996; Davis *et al.*, 2007). Además Morales Tovar (2015), observó que el silenciamiento de U2AF84 incrementa el *splicing* de algunos transcritos amebianos, sugiriendo que U2AF84 está implicado en la IR.

1.5 Las proteínas de unión al sitio de ramificación

En *Saccharomyces cerevisiae* la proteína de unión al sitio de ramificación (BBP) o MSL5 es un factor de procesamiento de pre-mRNA con características de una proteína tipo STAR/GSG, las cuales incluyen un sitio de unión a RNA central el cual está compuesto por un dominio tipo KH con un motivo QUA2 adyacente. En seguida del dominio KH-QUA2 tienen dos pequeños dedos de zinc del tipo CCHC y un dominio de interacción carboxilo terminal el cual es rico en prolinas (Fig 5). Otros miembros de la familia STAR/GSG, además de los dominios mencionados, contienen un motivo de homodimerización QUA1 localizado hacia el extremo amino. En levadura el dominio QUA1 es ausente y es sustituido por un sitio de unión a Mud2/U2AF65 denominado ULM. Como su nombre lo indica estas proteínas reconocen secuencias de RNA llamadas BP las cuales son conservadas en levadura (UACUAAC) y dicho motivo se encuentra cercano al 3' de los intrones. Dentro de esta secuencia se encuentra la Adenosina catalítica (subrayada) la cual dirige la primera transesterificación durante el *splicing* (Rymond, 2010). Cuando la BBP es reclutada en el sitio blanco, comienza el reclutamiento de una serie de subunidades spliceosomales y eventos de rearrreglos que termina con la configuración del sitio activo de la enzima para dar lugar a las trans esterificaciones ya mencionadas (Wahl MC et al., 2009). El homólogo de MSL5 en mamíferos ZFM1/ZNF162/D11S636/SF1 fue encontrado por primera vez mediante una búsqueda de genes asociados con la neoplasia endocrina múltiple tipo 1 (Toda T et al., 1994), posteriormente fue demostrado que funcionaba de manera similar a MSL5 en el *splicing* de mamíferos (Kramer A et al., 1992; Rain JC et al., 1998). Tanto MSL5 y SF1 son importantes para la viabilidad de los organismos desde levaduras, *Caenorhabditis elegans* hasta ratones. Los ratones heterocigotos SF1⁻/SF1⁺ mostraron ser más susceptibles a la tumorigénesis de colon inducida por azoxymetano (Shistashige M, et al., 2007).

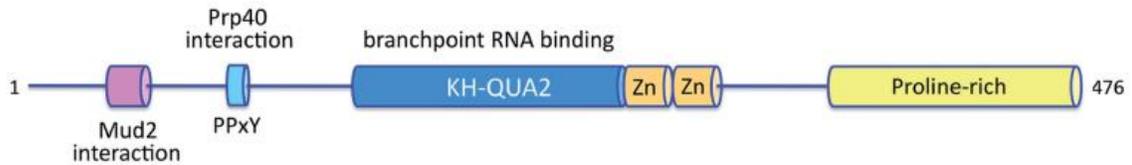


Fig 5. Organización del dominio y estructura primaria de MSL5. Esquema de los 476 aminoácidos componentes del polipéptido MSL5. Se indican el dominio que reconoce al BP, los dedos de Zn, el tracto rico en prolina y los sitios de interacción con Prp40 y con Mud2 (ULM).

1.5.1 Las proteínas de unión a BP son sitio específicas para la unión al RNA

El gen de *MSL5* fue uno de los siete mutantes letales sintéticos identificados por el laboratorio Rosbash. En estos ensayos se encontró que las cepas de levadura perdían viabilidad cuando se combinaban con una mutación no letal de *mud2* (Abovich N, et al., 1997), el ortólogo de U2AF65. *MSL5* es esencial en levaduras dado que las mutantes de *msl5* presentan anomalías durante el *splicing* principalmente en pre-mRNA que no tienen una secuencia consenso de BP (Rutz B, et al., 2000; Wang Q, et al., 2008). Los extractos depletados de MSL5 fallan durante el ensamblaje del complejo CC2 (complejo E en mamíferos), esta es la primera interacción estable entre proteínas y el BP del pre-mRNA y el snRNP unido al 5' ss (Abovich N, et al., 1997; Seraphin B, et al., 1989). MSL5 y SF1 se unen directa y específicamente al motivo BP del intrón (Berglund JA, et al., 1998; Garrey SM, et al., 2006). De igual manera también los estudios *Systematic evolution of ligands by exponential enrichment (SELEX)* encontraron como mejor candidato prevalente natural de levaduras a la secuencia UACUAAC como blanco de MSL5 (Garrey SM, et al., 2006). Además la estructura secundaria del RNA puede incrementar o disminuir la eficiencia del *splicing* (Buratti E, et al., 2004) y esto va en concordancia con que un orquilla proximal al BP, aunque no sea una característica de los intrones de levadura, puede influir en la selección de algunos pre-mRNAs a través de la asociación estabilizada por MSL5. De manera general el sitio de unión a RNA de SF1/MSL5 presenta una topología $\beta\alpha\alpha\beta\alpha$ lo cual es típico de un dominio KH tipo 1 (Keller et al., 1984). Las mutaciones en algunos de los aa presentes en este dominio reducen la unión del SF1/MSL5 al BP. Por

ejemplo, mutaciones en los dedos de zinc solo incrementan ligeramente la unión de SF1/MSL5 al BP. SF1 se une al RNA con una fuerza de 200 veces menor y con menor especificidad que MSL5, pero mediante mutaciones puntuales en R240K y K241R (ubicadas en la 4 hélice α del QUA2) puede aumentarse la fuerza de unión al RNA (Garrey et al., 2006; Garrey et al., 2008; Berglund et al., 1997). Otras mutaciones de Msl5/SF1 se discutirán más adelante.

La inespecificidad de SF1 en humanos es comprensible puesto que en humanos existe una variabilidad natural en los motivos BP cuyo consenso es YNCURAY (Keller et al., 1984). La unión débil de SF1 favorece la competencia alrededor de los sitios de *splicing* subóptimos y favorece el *splicing* alternativo, proceso que es fundamental durante la expresión génica de mamíferos (Wahl et al., 2009; Wang et al., 2008). En contraste en levaduras solo el 5% de los genes tienen intrones y solo de estos contienen un intrón desregulado (con secuencias no consenso). Por lo anterior la elevada fuerza de unión y especificidad de MSL5 favorece la eficiencia en el *splicing* en levaduras.

1.6 Interacción entre MSL5-MUD2: funciones en el reconocimiento del BP

La incompatibilidad genética de ciertas mutaciones en *msl5* y *mud2* reflejan una interacción física y funcional entre las proteínas codificadas por estos genes (Abovich et al., 1997; Wang et al., 2008; Fromont-Racine et al., 1997; Berglund et al., 1998). Además de estas pruebas genéticas la interacción física fue demostrada porque ambas proteínas copurifican en extractos de levadura y el heterodímero no necesita de un RNA para su unión (Wang et al., 2008). El extremo amino terminal de MSL5 contiene un sitio de unión para Mud2, específicamente los a.a. 41-141 son suficientes para la asociación estable con Mud2 (Rain et al., 1998) pero la delección de los a.a. 2-56 evita la recuperación del heterodímero (Wang et al., 2008). Un motivo típico de reconocimiento a RNA (RRM) llamado motivo de homología a U2AF (UHM) en el carboxilo terminal de Mud2 sirve como punto de contacto con MSL5 (Rain et al., 1998; Berglund et al.,

1998). Cooperativamente U2AF65 aumenta la afinidad de MSL5 por el RNA unas 20 veces.

U2AF65 se une al tracto de polipirimidinas el cual se encuentra entre el BP y el 3' ss de los intrones de mamíferos. Esta interacción es mediada por dos RRM canónicos posicionado río arriba del UHM de U2AF65 y esto es requerido para el reclutamiento del U2 snRNP durante el ensamblaje del spliceosoma (Zamore PD et al., 1992; Jenkins et al., 2008; Gaur et al., 1995; Banerjee et al., 2003; Valcarcel et al., 1996). Si comparamos las interacciones entre MSL5-MUD2, SF1-U2AF65 ambas revelan una asociación de subunidades conservadas. Por ejemplo, similar a MSL5 el N terminal de SF1 se une a un pocket hidrofóbico entre hélices A y B del dominio UHM de U2AF65 (Selenko et al., 2003). El motivo de ligando a UHM (ULM) incluye los residuos 15KKRKRSRW22. Cualquier mutación de la serina y el triptófano conservados, o la fosforilación de esta serina o la reversión de las cargas de los a.a. básicos del motivo resultan en la inhibición de la interacción entre SF1 y U2AF65 (Wang et al., 1999; Corsini et al., 2007). En mamíferos U2AF65 se encuentra asociada con la subunidad pequeña de U2AF, U2AF35 (Kielkops et al., 2004; Zamore et al., 1989). Esta interacción ocurre a través de un motivo ULM de U2AF65 y UHM de U2SF35. U2AF35 puede ser entrecruzado al 3' ss y esta interacción estabiliza la unión de U2AF65 al tracto de polipirimidinas (Wu et al., 1999; Zorio et al., 1999; Merendino et al., 1999).

El factor de splicing U2AF es central en el reconocimiento del 3' ss en eucariotas superiores, y participa importantemente en la regulación de la expresión génica. Oliver A. Kent et al., en el 2003 encontraron que U2AF65 solo es capaz de unirse al 3' ss y doblar el tracto de polipirimidinas para aproximar el 3' ss con el BP. Para refinar el modelo de interacción U2AF-RNA se considero la forma en que los RRM se unen con el tracto de polipirimidinas por medio de cristalografía y rayos X: específicamente la estructura de complejos de RNA con las proteínas Sxl y HuD que compiten con U2AF65 por trectos de pirimidinas en el splicing de algunos transcritos de *Drosophila* y *C. elegans*, respectivamente (Fig. 6A). En el modelo de

la figura 6B se aprecia la forma de como el U2AF65 estructura el 3´ss cuando se une a los trectos de polipirimidinas y además las interacciones con SF1 y U2AF35.

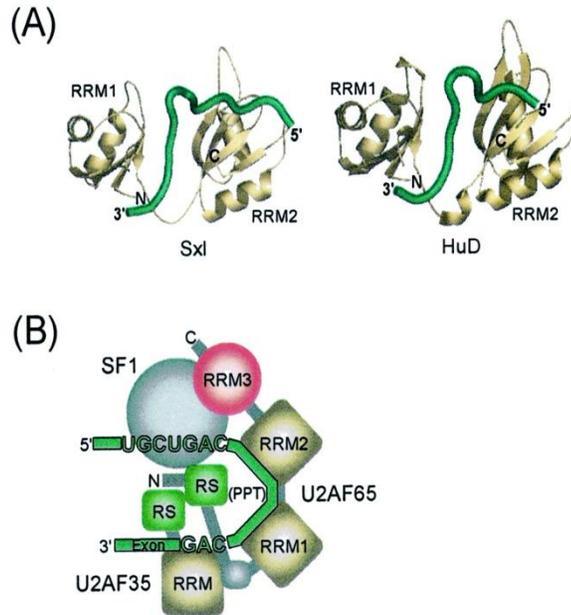


Fig 6. Estructuración del 3´ss por la unión del factor U2AF65 al tracto de polipirimidinas. La estructura de alta resolución de los complejos Sxl-RNA y HuD-RNA (Handa et al., 1999; Wang et al., 2001). Tanto la flexión del tracto de polipirimidinas y la orientación del C y N terminal de la proteína con la orientación 5´al 3´ del RNA empalman con la interacción observada de U2AF65 con el 3´ss. B, orientación de U2AF65 al 3´ss lo cual estructura el tracto de polipirimidinas hacia el BP, y el 3´ss posiciona el dominio SR de U2AF65 hacia la proximidad del BP las interacciones de U2AF65 con SF1 al BP y U2AF35 al 3´ss en un complejo multiproteico lo cual da lugar a la formación del complejo E.

1.7 Bases estructurales para el reconocimiento del sitio de ramificación por las proteínas de unión a BP

El complejo KH-QUA2/BPS RNA es estabilizado por una combinación de interacciones hidrofóbicas, puentes de hidrogeno y contactos electrostáticos (Fig. 7A). Consistente con la conservación de la A del BP, A8 es específicamente reconocida por el dominio KH. Con mayor precisión el N6 y N1 de la A8 forman

puentes de hidrogeno con el esqueleto amida y el oxigeno del carbonil de I177, la cual simula el grupo funcional tipo Watson-Crick de una uridina. Otros puentes de hidrogeno son encontrados entre el 2' hidroxilo de la A7 y el N7 de la A8. En combinación estos puentes de hidrogeno unen específicamente a una A en esta posición. Esto es consistente puesto que la A8 es estrictamente conservada en los BP y las mutaciones por otra base evitan el reconocimiento por parte de SF1 (Berglund et al., 1997). La A8 es rodeada por una cadena alifática conservada (I157, L164, I175, I177, V183) del dominio KH (Fig. 7B). La interacción π -catión (Gallivan et al., 1999) entre la A7 y la cadena de la K184 (Fig. 7C) es consistente con la preferencia por una purina en esta posición de los BP tanto en mamíferos como en *S. cerevisiae*. Una mutación K184A evita la unión al RNA (Berglund et al., 1997) lo cual demuestra la importancia de esta interacción. Los fosfatos de A7 y A8 quedan muy próximos a la cadena del péptido ¹⁵⁸GPRG¹⁶¹ el cual se estructura como un loop. Consecuentemente, las mutaciones G161D Y R160 (que contacta con el 3' fosfato de la A8) eliminan la unión al RNA (Rain et al., 1998) (Fig. 7B). El N6 de la A7 interactúa con la cadena del E149 conservado, el cual es posicionado vía unión de puente de hidrogeno con el esqueleto de la amida de I150 (Fig. 7C). En la estructura también se puede explicar la conservación del uracilo 6 (U6) en el BP. U6 está localizado entre la interface de los dominios KH y QUA2 (Fig. 7D). La base del U6 interactúa con L244 y L247 en la hélice QUA2 (α 4) y aproxima muy estrechamente a la cadena principal del péptido de la hélice α 1 lo cual es consistente con la conservación de la G154. El protón del imino de U6 es posicionado para formar un puente de hidrogeno con el esqueleto de la amida de la L155. El azúcar de U6 se une con la P159 en el loop GPRG y es estabilizado por un puente de hidrogeno entre su 2'OH y el esqueleto amida de R160 (Fig. 7D). Los residuos en la región QUA2 (L244, L247, T253, L254, R255) interactúan con la A2, A4 y C5. La descripción de la estructura del complejo SF1-RNA provee información valiosa que nos ayuda a entender un poco más acerca de los primeros pasos del ensamblaje del spliceosoma temprano (Liu et al., 2001).

De igual forma la estructura de la proteína MSL5 ya fue cristalizada, mostrando características de unión iguales a SF1, donde el BP queda unido en un pocket

hidrofóbico abarcando principalmente el dominio KH (UAAC), el dominio QUA2 interactúa con el 5' final del BP (ACU) (Fig. 8A).

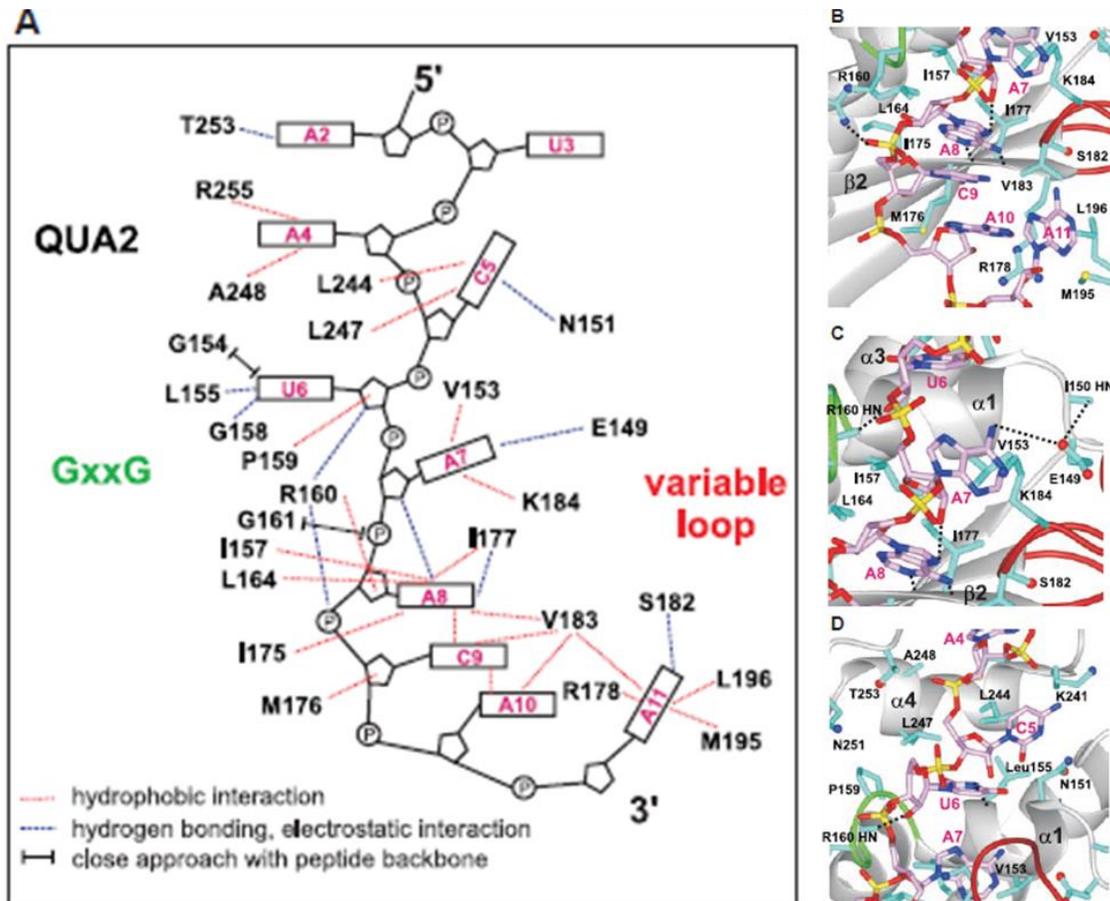


Fig 7. Bases estructurales del reconocimiento del RNA. (A) vista esquemática de la interacción proteína-RNA. Los nucleótidos de RNA se muestran en código de una letra y el esqueleto de la proteína en código de 3 letras. Se muestran las interacciones hidrofóbicas (en líneas punteadas color rojo), los puentes de hidrogeno (en azul). (B) reconocimiento de la Adenosina del BP (A8) y C9-A10-A11. Los puentes de hidrogeno son indicados por líneas discontinuas negras. (C) reconocimiento de la Adenosina 7 (A7) por el dominio KH. (D) reconocimiento del uracilo 6 y citocina 5 por el loop ¹⁵⁸GPRG¹⁶¹ cercano a la interface KH/QUA2, y por la hélice QUA2 (α 4). Tomada de Liu et al., 2001.

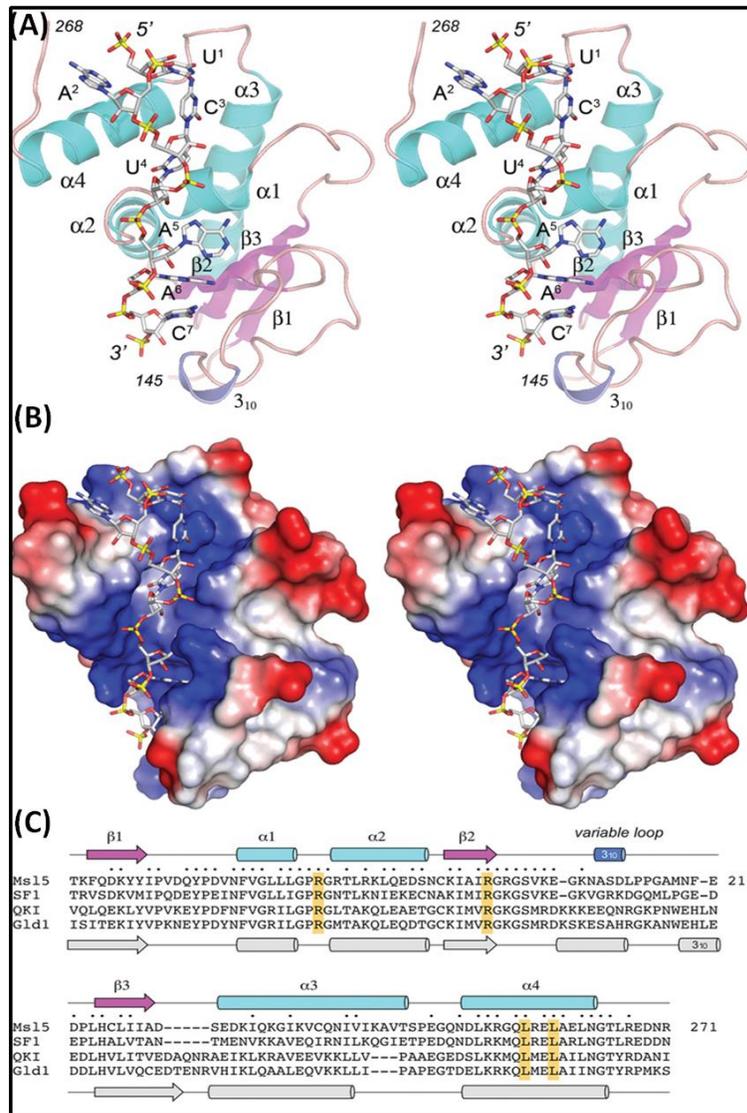


Fig 8. Estructura de MSL5-(KQ2) en complejo con el RNA. (A) vista estereoscópica de la estructura terciaria de MSL5 es mostrada en modelo de superficie, con las hojas β coloreadas en magenta las hélices α aparecen en color cyan y una α_{310} hélix es mostrada en color azul. El N y C terminal (a.a. 148 y 268) son indicados. El RNA se muestra en modelo de varillas. (B) una vista estereoscópica de la superficie electrostática del modelo del KQ2-MSL5, el cual fue generado en Pymol, se muestra en modelo de varillas. (C) la estructura secundaria de los elementos de la del KQ2-MSL5 de levadura (coloreada como en A) en el cual se muestra la secuencia de a.a. de MSL5, las hojas β aparecen como flechas gruesas y las hélices como cilindros. La estructura primaria de MSL5 se alinea con su ortólogo en humanos (SF1), QKI de humano, GLD-1 en *C. elegans* los *gaps* aparecen como líneas discontinuas, las cajas amarillas muestran los aminoácidos conservados que son importantes para MSL5 (tomada de Jacewicz et al., 2015).

1.8 Reconocimiento de los BP en secuencias no conservadas

Cuando se hace una comparación entre las secuencias primarias de los KH-QUA2 encontrados en diferentes especies (humano, *C. elegans* y levadura) podemos apreciar que aunque existe una variabilidad en las secuencias de a.a, los residuos importantes son conservados, por ejemplo el GPRG y las dos leucinas de la hélice $\alpha 4$. Sin embargo estas variabilidades pueden resultar en cambios estructurales que ayudan a estos dominios para interactuar con BPs no conservados como en humanos (Jacewicz et al., 2015). Un alineamiento de estructuras primarias con la secuencia de a.a. de de los 4 KQ2 muestran gran similitud entre ellas (Fig. 8C). El alineamiento de la estructura terciaria de MSL5, QKI, GLD1 y SF1 unidos a BP se muestran en la Figura 9. Las características sobresalientes de la diferencia entre MSL5 y QKI/GLD1 reside en el “loop variable” entre las hojas $\beta 2$ y $\beta 3$. Las secuencias de a.a. son virtualmente idénticas en QKI y GLD1, pero poco conservadas en MSL5 (Fig. 8C). En tanto que el loop de MSL5 tiene un segmento de solo 3 a.a. de estructura secundaria ($\alpha 3_{10}$ helix), el loop de QKI/GLD-1 comprende una α hélice de 7 a.a. (Figs. 8C y 9). En contraste la estructura del componente QUA2 de SF1 no tiene una estructura ordenada (Fig. 9). Estas diferencias estructurales pueden explicar los hallazgos reportados anteriormente que arguyen que el dominio QUA2 es el que dicta la alta especificidad y afinidad por el reconocimiento del BP en MSL5 comparado con SF1 de humanos (Garrey et al., 2008).

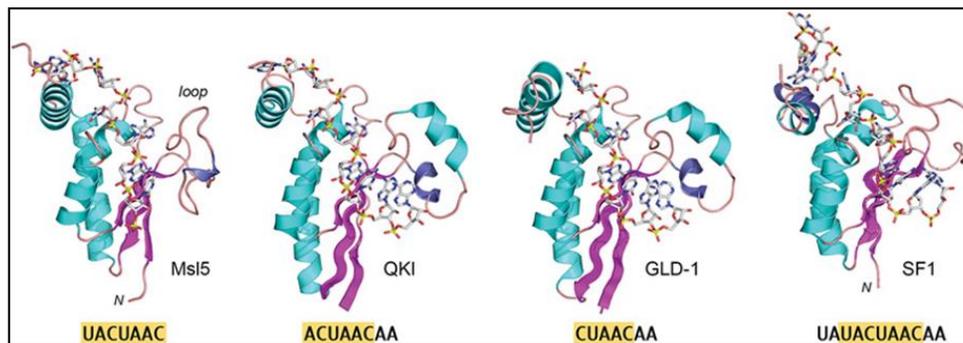


Fig 9. Comparación estructural entre distintos KQ2. La estructura terciaria de los dominios KQ2 de MSL5, QKI (pdb JVH), GLD-1 (pdb 4JVY) y SF1 (pdb 1K1G) fueron alineadas y se selecciono la vista horizontal. El RNA aparece en modelo de varilla. La secuencia de las bases que forman el RNA blanco para cada caso aparece en la parte inferior de las estructuras. En color oro se muestra las posiciones idénticas con respecto al consenso de levadura.

2. ANTECEDENTES DIRECTOS Y JUSTIFICACIÓN

U2AF84 y Ehsf1 son dos de los 36 factores identificados en los complejos tempranos de *splicing* inmunoprecipitados tras entrecruzamiento con UV. U2AF84 posee todos los elementos que conforman al U2AF65 de otras especies, pero también contiene una región extra de 225 a.a. que es muy similar a KH-QUA2. En tanto que U2AF65 favorece el *splicing*, U2AF84 está involucrada en la inhibición del mismo que resulta en retención intrónica. Durante el ensamblaje del complejo temprano del *splicing* el ambiente del 3'ss es definido por la interacción de SF1 con el BP mediante su dominio KH-QUA2, interacción estabilizada mediante la asociación con U2AF65 y de forma distal con la asociación de U2AF35 que reconoce el dinucleótido del 3'ss. Todas estas interacciones son fundamentales para la entrada de U2, el cual disocia a SF1 al interactuar directamente con el BP mediante un apareamiento de bases a través del U2 snRNA. Aparentemente, U2AF84 podría estar realizando las mismas funciones que U2AF65, sin embargo no queda claro cómo es que U2AF84 inhibe el *splicing*. Por todo esto es decidimos estudiar cual es la dinámica molecular que resulta en dicha inhibición.

3. HIPÓTESIS

Debido a que a diferencia de U2AF65 humano y Mud2 de levadura y de otro posible ortólogo de U2AF65 (EhU2AF45), el factor U2AF84 contiene un dominio de 136 a.a. muy similar al dominio de unión al BP, KH-QUA2, de los SF1 reportados (incluido el Ehsf1 putativo), es posible que el dominio KH-QUA2 de U2AF84 sea funcionalmente equivalente a los presentes en SF1 y que éste reconozca BP distintos a los reconocidos por SF1, compitan por la unión al BP o interaccionen entre ellos y esto conlleve a un incremento o una inhibición en el *splicing*.

Para responder a ésta hipótesis nos planteamos los siguientes

4. OBJETIVOS

OBJETIVO GENERAL

Nos propusimos investigar si la región extra de U2AF84 es la responsable de la inhibición del *splicing* y si esto es así, cuál es el mecanismo molecular que media dicha inhibición, particularmente si dicho mecanismo involucra una interacción con los componentes del complejo de *splicing* E, por ejemplo EhSF1 y el BP.

OBJETIVOS PARTICULARES

Para evaluar la función del dominio KQ2-84 durante el *splicing*, realizamos ensayos de *splicing* in vivo en *E. histolytica* transfectantes de U2AF84 y una versión deletada del KQ2-84.

Para poder determinar el mecanismo molecular por el cual U2AF84 participa en la inhibición del *splicing* se obtuvieron proteínas recombinantes rKH-QUA2 de U2AF84, EhSF1 y msl5.

La capacidad de unión de los KH-QUA2 a BP se realizó mediante EMSA de sondas BPs con secuencias de Rab13 y el consensus de levadura.

Para evaluar el papel de los KQ2-84 y KQ2-SF1 se realizaron ensayos de complementación entre los dos KH-QUA2 mencionados, usando las dos secuencias BP.

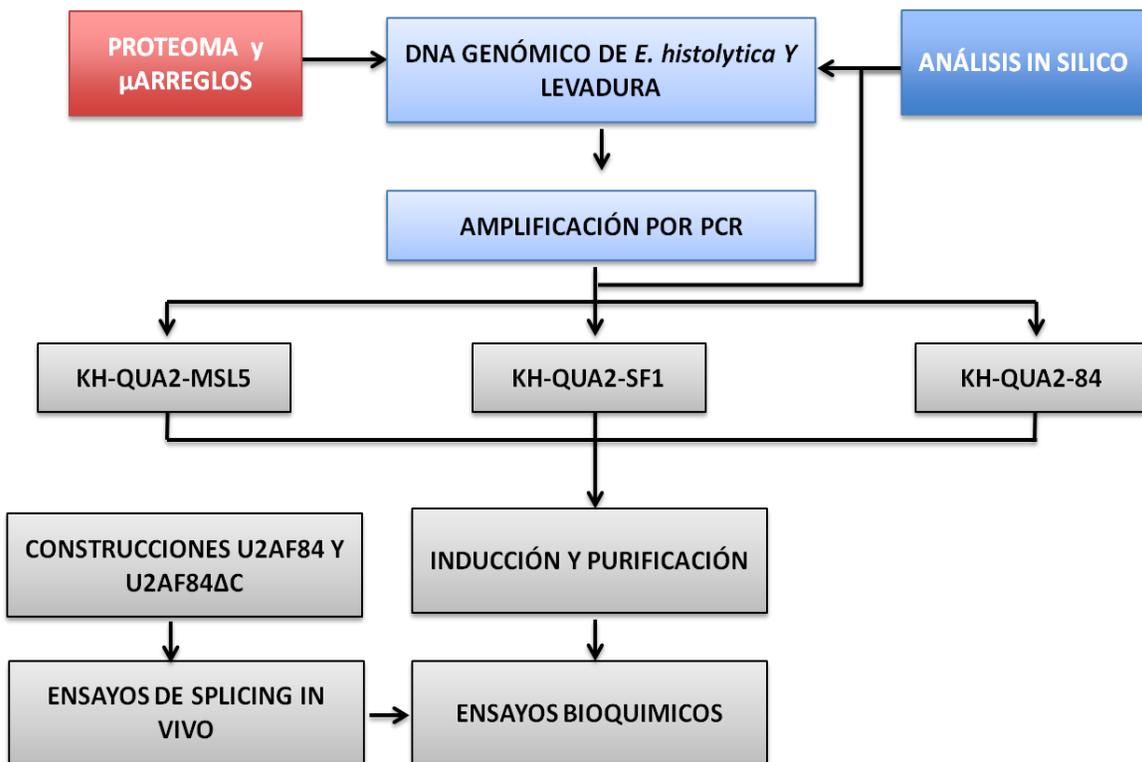
Para determinar si existe una interacción entre KQ2-84 y KQ2-SF1 se realizaron ensayos de CL-PD usando etiquetas diferentes para cada proteína, con esta misma finalidad se diseñaron ensayos de far-wb los cuales dan una mayor certeza de los resultados con solo versiones recombinantes.

Debido a que las proteínas U2AF84 y SF1 están involucradas en la formación del complejo E, nosotros evaluamos la inhibición de transición de complejo E-A por U2AF84 usando las secuencias las secuencias de BP de RabX13 y levadura.

5. ESTRATEGIA EXPERIMENTAL

Para poder cumplir con los objetivos planteados, primero realizamos la evaluación del efecto en el *splicing* por parte de la región KH-QUA2 de U2AF84 por medio de ensayos de *splicing* in vivo. En breve, se realizaron ensayos de RT-PCR utilizando RNA total de amebas transfectadas con plásmidos que codifican la proteína de fusión HA-U2AF84 y una versión deletada del dominio KH-QUA2 de U2AF84 (HA-U2AF84 Δ C). Una vez probada la participación del dominio KH-QUA2 de U2AF84 en la inhibición del *splicing* y para probar cómo se lleva éste a cabo, se procedió a realizar EMSA y *crosslinking* con UV para evaluar las características de las interacciones de los KH-QUA2 con BP de RabX13 y el consenso para levadura. En seguida procedimos a los ensayos de complementación y *crosslinking* con formaldehído para determinar la dinámica de los KH-QUA2 de EhSF1 y U2AF84. Los ensayos de pulldown y far-wb nos sirvieron para determinar la interacción entre las rKH-QUA2 y su dependencia a secuencias de RNA. Por último, durante la transición del complejo E-A del *splicing* el U2 snRNA hibrida con la secuencia del BP con la concomitante disociación de SF1, quedando presente U2AF65 unido al tracto de polipirimidinas. Nosotros simulamos la transición del *splicing in vitro*, usando la sonda U2 snRNA como competidor en un experimento de mezclas y lo evaluamos contra los las secuencias BP de RabX13 y el consenso de levadura.

ESQUEMA DE LA ESTRATEGIA EXPERIMENTAL



6. MATERIALES Y METODOS

Extracción de RNA

Se extrajo RNA total utilizando el reactivo TRIzol (Invitrogen, No. de catálogo 15596-018), las muestras se agitaron y se centrifugaron a 12000 xg durante 10 minutos a 4°C, la fase acuosa fue transferida a un tubo nuevo y se agregaron 200 µL de cloroformo, se agitó nuevamente durante 15 segundos y se incubó a temperatura ambiente durante 3 minutos entonces las muestras fueron centrifugadas a 12000 xg durante 15 minutos a 4°C, se eliminó el sobrenadante y el precipitado fue resuspendido en 225 µL de buffer de extracción SDS 1X [SDS 0.5%, EDTA 1 mM (pH 8), Tris-HCl 20 mM (pH 7.5)], luego se agregó acetato de sodio pH 5.3 a una concentración final de 0.3 M, dos volúmenes de fenol-cloroformo-isoamílico (25:24:1) y se mezcló en vortex durante un minuto. La muestra fue centrifugada a 12000 xg durante 3 minutos a 4°C y se recuperó la fase acuosa, a esta se le agregó nuevamente acetato de sodio a una concentración final de 0.3 M, 2.5 volúmenes de etanol absoluto, se mezcló. El RNA se dejó precipitando a -70°C durante 16 horas. Posteriormente se centrifugó a 12000 xg durante 15 minutos a 4°C, se eliminó el sobrenadante y el precipitado se lavó con 300 µL de etanol 70% y nuevamente se centrifugó a 7500 xg durante 15 minutos a 4°C. El precipitado se dejó secar a temperatura ambiente y se resuspendió en 75 µL de agua miliQ estéril. Para una disolución completa del RNA, las muestras se incubaron a 55°C durante 5 minutos. La integridad del RNA fue verificada en un gel de agarosa 2% y se cuantificó espectrofotométricamente. Posteriormente las muestras de RNA se trataron con DNasa I (Sigma-Aldrich No. De catálogo 4716728001), incubando durante 20 minutos a 25 °C, luego las muestras se llevaron a un volumen final de 200 µL con agua miliQ estéril y se agregó un volumen de fenol-cloroformo-isoamílico (25:24:1). Las muestras fueron agitadas en vortex durante 30 segundos y se centrifugaron a 13000 rpm durante 3 minutos, posteriormente se recuperó la fase acuosa y a esta se le agregó un volumen de cloroformo-isoamílico (25:1), se repitió la agitación en vortex y la centrifugación. Se recuperó la fase acuosa y se agregó a cada muestra 1 µL de

glicógeno (5 mg/mL) y un volumen de isopropanol frío. El RNA se precipitó a -70°C durante 16 horas. Posteriormente las muestras se centrifugaron a 13000 rpm durante 15 minutos a 4°C se eliminó el sobrenadante y el precipitado se lavó con 300 μL de Etanol 70 %, se centrifugó nuevamente bajo las mismas condiciones. El precipitado se dejó secar a temperatura ambiente hasta que se evaporó el etanol después las muestras se re suspendieron en 50 μL de agua miliQ y se cuantificaron espectrofotométricamente.

RT-PCR

Las reacciones de retro transcripción se llevaron a cabo utilizando la enzima M-MLV (Invitrogen, No. de catálogo 28025-021) y siguiendo el protocolo descrito por el proveedor. Como molde se utilizaron 1.5 μg de RNA total. El RNA se mezcló con 1 μL de hexámeros, 1 μL dNTP's 10 mM y $\text{H}_2\text{O}_{\text{mQ}}$ estéril llegando a un volumen final de 13 μL . Las reacciones se incubaron a 65°C , durante 5 min y se colocaron en hielo por 2 min más. En seguida, se agregaron 4 μL del 5X First Strand Buffer (Invitrogen, No. de catálogo 28025-021) y 2 μL de DTT 0.1 M, la mezcla se agitó suavemente y se incubó a 37°C por 2 min. Posteriormente, se agregaron 200 unidades (1 μL) de la enzima M-MLV, la mezcla se agitó suavemente, y se incubó una vez más a 37°C durante 1 hora. Finalmente, las muestras se incubaron durante 10 min a 70°C , y el cDNA obtenido se usó inmediatamente o fue almacenado a -20°C hasta su uso. Las reacciones de PCR se llevaron a cabo de la siguiente manera: Se utilizó como molde el 10% de la reacción respectiva de RT, MgCl_2 de acuerdo a la Tabla 1 del anexo 3, 1X del buffer de PCR 10X (Invitrogen, No. de catálogo 11615-010), 0.2 μM de los oligonucleótidos sentido y antisentido, 1.25 unidades de la Taq DNA Polimerasa (Invitrogen, No. de catálogo 11615-010) y $\text{H}_2\text{O}_{\text{mQ}}$ estéril a un volumen final de 25 μL . Como control de carga se realizó bajo las mismas condiciones la RT-PCR para la RNA Polimerasa II. Como control negativo, en cada uno de los experimentos se realizaron reacciones de PCR utilizando como molde al RNA sin reacción previa de RT.

Construcción de los plásmidos de sobreexpresión

Se realizó la amplificación por PCR del fragmento que codifica para los primeros 486 nt de U2AF84, usando como molde la construcción HA-U2AF84 donada por el Dr. Jesús Valdés. Los oligos incluyen los sitios de corte para la enzima SmaI (sentido), un codón de paro seguido por un sitio de corte para XhoI (antisentido). El vector de expresión y los amplicones se cortaron usando las enzimas de restricción SmaI y XhoI (NEB No. de catálogo #R0141L y #R0146L), luego se realizó la ligación usando una relación vector: inserto 1:3 en 4 µL de Buffer de Ligación 5x y una unidad de Ligasa T4 (Invitrogen No. de catálogo 15224-017) en un volumen final de 20 µL, la reacción se dejó incubando toda la noche a 16 °C. La ligación se empleó para transformar bacterias competentes *Escherichia. coli* cepa TOP10. La selección de las clonas fue realizada en medio Luria-Bertani con ampicilina (100 µg/mL). Las clonas seleccionadas se sembraron en tubos con 3 mL de medio Luria-Bertani con ampicilina (100 µg/mL) y se incubaron en agitación toda la noche a 37 °C. Se extrajo el DNA plasmídico por miniprep acorde a la técnica descrita por Del Sal *et al* (1998), los plásmidos obtenidos fueron restringidos con las enzimas de restricción SmaI y XhoI y se monitoreo la liberación de los fragmentos. Se seleccionó una clona positiva de cada construcción para extraer DNA plasmídico usando el *Plasmid Midi Kit* (QIAGEN No. de catálogo 12143), una alícuota del plásmido fue usado para la secuenciación y usados en las transfecciones de amebas de acuerdo al protocolo descrito abajo y en colaboración con la IBQ. Sandra Lorena Palacio Molina.

Transfección de los plásmidos de sobreexpresión en trofozoitos de *E. histolytica*

La transfección de la cepa HM1:IMSS de *E. histolytica* se realizó en una placa de 12 pozos, en donde se sembraron ~750,000 amebas por pozo en los cuáles se adicionó medio TYI-S-33 tibio y se incubaron a 37 °C en presencia de CO₂ durante 30 minutos para que se adhirieran a la superficie, durante este lapso se realizaron en esterilidad las mezclas de transfección, en la primera se agregaron 5 µg de los plásmidos control (pEhExHA vacío) y de sobreexpresión (pEhExHA-U2AF84 y

pEhExHA-U2AF84 Δ C), 10 μ L de PLUS reagent (Invitrogen No. de catálogo 11668027) y se llevó a un volumen de 50 μ L añadiendo el medio de transfección (Opti-MEM suplementado con 1 mg/mL ácido ascórbico y 5 mg/mL L-cisteína pH 6.8), en la segunda mezcla se agregaron 20 μ L de Lipofectamina 2000 (Invitrogen No. de catálogo 11668027) y 30 μ L de medio de transfección, ambas mezclas se incubaron por separado 15 minutos a temperatura ambiente, posteriormente se colocaron en un solo tubo de reacción y se incubaron nuevamente bajo las mismas condiciones mencionadas anteriormente, luego se agregaron 400 μ L más de medio de transfección para tener un volumen final de 500 μ L. Transcurridos los tiempos de incubación se retiró el medio TYI-S-33 y se colocó la mezcla final de transfección adicionando lentamente gota a gota sobre la monocapa de amebas de cada pozo, la placa se incubó nuevamente ahora durante 5 horas. Finalmente, la placa se colocó en un baño de hielo-agua durante 15 minutos y se colectó el medio con las amebas, las cuales se sembraron en tubos de 7 mL con medio TYI-S-33 y se incubaron nuevamente en las mismas condiciones ahora durante 24 horas. Posteriormente se inició la selección con el antibiótico G418 (*Life technologies* No. de catálogo 11811098) a una concentración de 1 μ g/mL, la concentración fue aumentando conforme los tubos alcanzaron una confluencia de 80%, hasta llegar a una concentración de 10 μ g/mL de antibiótico en las transfecciones estables. Una vez que se establecieron los cultivos a las concentraciones de antibiótico deseadas, se expandieron en cajas de 25 cm² y se recolectaron las muestras después de 48 horas de incubación para la posterior extracción de RNA y proteínas.

Western blot

Las proteínas se extrajeron de acuerdo al protocolo de extracción de proteínas con TRIzol (Invitrogen No. de catálogo 15596-026). Para los ensayos de WB, se emplearon 20 µg proteínas las cuales fueron separadas en un gel de poliacrilamida SDS-PAGE al 8%, se realizó la transferencia a una membrana de nitrocelulosa de 0.45 µm de diámetro (Trans-Blot Transfer Medium, de BIORAD) bajo las siguientes condiciones: 400 V, 350 mA, 1 hora en buffer de transferencia (Tris-glicina). Posteriormente se verificó la transferencia de la membrana con rojo de Ponceau. En seguida se procedió a la realización del bloqueo de la membrana con PBS 1x-leche al 5%, durante 1 hora en agitación suave a 25 °C. Se retiró el exceso de leche con PBS 1x. Inmediatamente después, la membrana fue incubada con la preparación del anticuerpo primario (anti-HA Covance No. de catálogo MMS-101R dilución 1:25000, en PBS 1x) durante 3 horas. La membrana se lavó 3 veces durante 10 min con PBS 1x-Tween20 0.05% y se realizó la incubación con el anticuerpo secundario (goat anti-mouse IgG-HRP SC2768 de *Santa Cruz Biotechnology Inc.* dilución 1:10000 en PBS 1x-Tween20 0.05%) durante 1 hora a temperatura ambiente en agitación suave. Una vez que el proceso anterior fue realizado se realizaron tres lavados con PBS 1x-Tween20 0.05%. La membrana se lavó con PBS 1x 3 veces durante 1 min y se procedió al revelado. El proceso de revelado fue realizado por quimioluminiscencia usando el kit "*Western lightning Plus-ECL*" (PerkinElmer, Inc. No. de catálogo NEL103001EA). Como control de carga se empleó el anticuerpo primario anti-Histona H3 de conejo (Merckmillipore No. de catálogo #97275), a una dilución 1:1000 y el anticuerpo secundario IgG de cabra anti-conejo (Zymax No. de catálogo AQ132P) a una dilución 1:10000, empleando los mismos tiempos de bloqueo incubación y lavado.

Purificación de KH-QUA2 de EhSF1 y U2AF84

Un fragmento de DNA de 360 nt. que codifica para el dominio KH-QUA2 de EhSF1, uno de 384 de U2AF84 (ver alineamientos) y uno de 384 nt que corresponde al KH-QUA2 de MSL5 de levadura fueron amplificados por PCR a

partir de DNA genómico con primers que introducen el sitio BamHI, y un sitio XhoI inmediatamente después del codón de paro. El producto de PCR fue digerido con las enzimas BamHI y XhoI e insertado entre los sitios BamHI y XhoI del vector Pet28a (o pGex 6p1 según se indique). Esto generó un plásmido de expresión que codifica los polipeptidos de los KH-QUA2 fusionados a un N-terminal de Histidinas (6His), los plásmidos fueron utilizados para transformar la cepa *Escherichia coli* top101. Las clonas se evaluaron mediante la resistencia a antibiótico y algunas fueron seleccionadas para llevar a cabo la extracción de plásmidos por minipreps. Los plásmidos obtenidos fueron digeridos con BamHI y XhoI para evaluar la liberación del fragmento. Los plásmidos fueron secuenciados en ambos sentidos, se verificó el marco de lectura. Los plásmidos positivos se usaron para transformar a la cepa BL21 (DE3). Una colonia fue amplificada en un volumen de 500 mL con 25 µg/mL de kanamicina o 100 µg/mL de ampicilina (según el vector) en medio *TB broth*. Cuando se alcanzó una D.O de 0.7 a 600 nm fueron agregados 0.5 mM de IPTG e incubados por 4 horas a 37 °C en agitación constante. Los pasos siguientes fueron hechos a 4°C. Las células fueron cosechadas por centrifugación (8000 RPM durante 5 minutos), y resuspendidas en 5 mL de buffer de lisis de imidazol (50 mM Tris-HCl, pH: 7.5, 500 mM de KCl, 15 mM de imidazol), con 2x de un cocktail de inhibidor de proteasas (Roche). Las células fueron lisadas por sonicación 6X ciclos de 30 s a 75% de amplitud en hielo, y el material insoluble fue removido por sonicación a 12000 RPM durante 20 minutos. El sobrenadante fue mezclado con 1 mL de NiNTA-QIAGEN (perlas de agarosa acopladas con Nickel), durante 1 hora en agitación constante. La resina se recuperó en una columna de purificación previamente equilibrada con buffer de lisis, se hizo un lavado con un volumen de 10 mL de buffer de lavado (50 mM Tris-HCl, pH: 7.5, 500 mM de KCl, 20 mM de imidazol). Posteriormente se hicieron 5 eluciones con volúmenes de 1 mL de buffer de elución (50 mM Tris-HCl, pH: 7.5, 500 mM de KCl, 300 mM de imidazol), las muestras fueron monitoreadas en SDS-PAGE 12% y teñidas con azul de Coomassie. Las fracciones que contenían a las proteínas recombinantes fueron dializadas toda la noche en un volumen de 2L de buffer de diálisis (20 mM de Tris HCl pH: 7.5, 100 mM de KCl y glicerol al 10%) con membrana de corte de

3.5 kDa. la proteína dializada fue cuantificada en Qiubit y almacenada a -20 °C o a -70 según el tiempo de uso.

Ensayos de cambio de la movilidad electroforética (EMSA) o de retardamiento

La base de un EMSA es el cambio de la movilidad electroforética de un ácido nucleico cuando se une a una proteína u otra molécula. Las proteínas purificadas fueron incubadas (según se indica la cantidad) durante 15 min a temperatura ambiente con buffer BB5X (50 mM de Hepes K-OH, pH 7.6; 40 mM de MgCl₂, 20 mM de espermidina; 0.1 mM de DTT, 0.5 mM de EDTA, 1.5 mM de ATP y 50% de glicerol), 2.5 µg de tRNA, en un volumen final de reacción de 20 µL. La mezcla se puso a interactuar con el RNA marcado con γ -ATP (reacción catalizada por la *T4 Polynucleotide Kinase* (10 U/µL) (Cat. EK0031, *thermo scientific*). La mezcla fue incubada 15 min a 4 °C (a menos que se indique otro tiempo) y en seguida los complejos fueron separados en un gel de acrilamida nativo al 10% (200 V, 40 min, a 4 °C). El gel fue secado al Vacío y luego se reveló la marca en una placa autorradio gráfica. En los ensayos de mezcla, se siguió el mismo protocolo, excepto que se agregó la cantidad de proteína indicada durante la primera incubación. Para el caso de la competencia con el U2 snRNA, la sonda fría fue agregada a la relación molar indicada, durante 1 hora a 4 °C previo a la adición de las proteínas, tal como se describe en los resultados.

Ensayo de *cross-linking* con UV y formaldehído

Una vez concluido el tiempo de formación de complejos (descrito en EMSA), la reacción fue irradiada con luz UV durante 20 minutos a una longitud de onda de 254 nm en cama de hielo, y entonces tratada con buffer de carga de proteínas al 2X y hervida durante 5 minutos. Las muestras fueron separadas en SDS-PAGE 12% e inmediatamente fueron transferidas a una membrana de nitrocelulosa para después revelar la radioactividad en una película autorradiográfica. La membrana fue recuperada y rehidratada para realizar la técnica de WB con el anti cuerpo que se indica en cada resultado. Para el *cross-linking* con formaldehído, los complejos

fueron tratados con formaldehido al 2% durante 15 minutos, y entonces la reacción se detuvo con glicina 0.25M en PBS. La reacción fue desnaturalizada con buffer de carga 2X y hervida durante 5 min.

Ensayos de pull down

Después de la formación de los complejos (y la irradiación con UV según se indique), se adicionaron 20 μ L de perlas de GSH, y se llevo la reacción a un volumen de 60 μ L con Buffer de unión, la reacción se incubo en agitación durante 2 horas a 4°C y entonces las perlas fueron colectadas por centrifugación a 1000 rcf durante 5 minutos y entonces lavadas 3 veces con PBS 1X. El precipitado fue tratado con buffer de carga de proteínas 2X y hervido durante 5 minutos y así cargado a en un gel SDS-PAGE al 12% para su análisis. La radiactividad fue revelada y después la membrana fue rehidratada y usada para ensayo de WB.

Ensayo de far-wb

20 μ g de extractos proteicos de las transfectantes de *E. histolytica* fueron separados en un gel de poliacrilamida SDS-PAGE al 12%, transferidas en a una membrana de nitrocelulosa y se verifico la transferencia mediante rojo de Ponceau (5 min con Rojo de Ponceau al 1%). La membrana fue bloqueada (PBS-Tween20 al 0.02% con leche al 5%) durante 1 hora a temperatura ambiente en agitación constante. La membrana fue incubada durante toda la noche a 4°C en agitación constante, con la proteína recombinante indicada a una concentración de 10 nM en el mismo buffer de bloqueo. Después de la interacción entre las proteínas, se realizaron 3 lavados con PBS y se procedió a realizar el ensayo de WB como se describió arriba con el anticuerpo correspondiente a la etiqueta.

7. RESULTADOS

Análisis bioinformático de U2AF84 y EhSF1

Por medio de análisis bioinformáticos se analizaron los dominios que conforman a U2AF84 (Fig. 10A). Igual que U2AF65 y U2AF45 amebiana, el extremo amino de U2AF84 de 712 a.a. contiene una región rica en serinas y argininas (SR), seguido de 3 motivos de reconocimiento a RNA (RRM). A diferencia de U2AF65, U2AF84 contiene una región extra de 241 a.a., de los cuales 137 a.a. guardan cercana homología con el dominio KH-QUA2 presente en las proteínas Quaking de diferentes especies, incluyendo SF1 humana y la EhSF1 putativa (Fig. 10B).

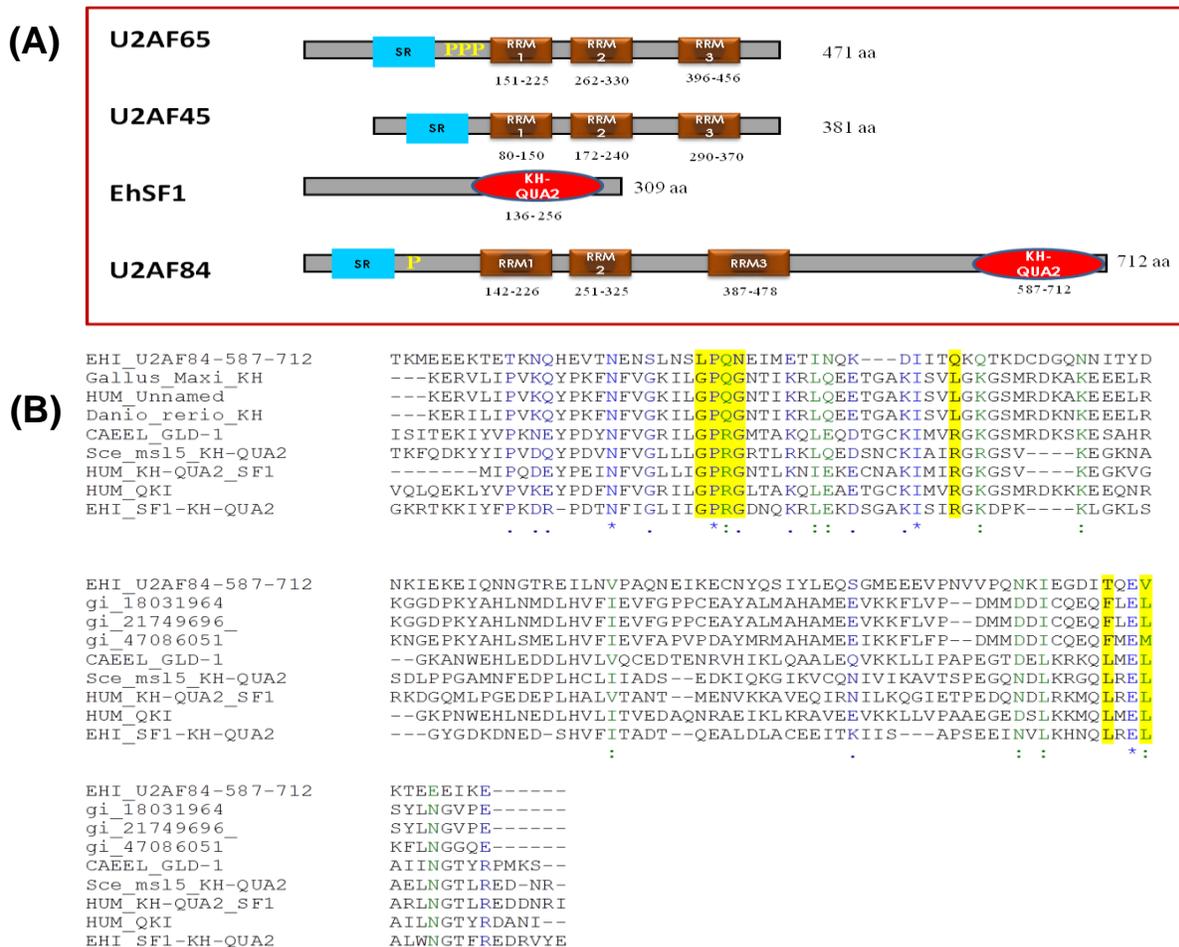


Fig 10. U2AF84 contiene un dominio KH-QUA2. A) Se utilizó el programa pfam para predecir los dominios de U2AF65, U2AF45, EhSF1 y U2AF84. Se indican el número de aminoácidos y la región de cada dominio: motivo de reconocimiento a RNA (RRM), región serinas-argininas (SR), sitios de fosforilación región rica en Prolinas (P), KH-QUA2 (dominios KH-QUA2). **B)** Alineamiento de 9 dominios KH-QUA2 usados como referencia para identificar el KH-QUA2 de U2AF84. CAEEL_GLD1: Female germline-specific tumor supresor gld1 de *C. elegans*; Sce_msl5 (levadura), HUM (humano), EHI (Entamoeba histolytica, los códigos g1 pertenecen a proteínas putativas de humano).

El dominio carboxilo extra de U2AF84 es el responsable de la inhibición de *splicing* in vivo

Ya que U2AF84 es una proteína similar a U2AF65, Morales-Tovar (2015) silenció U2AF84 en *E. histolytica* cepa G3 y encontró que a mayor concentración de agente selectivo, es decir a mayor número de copias del plasmido silenciador, la concentración del mRNA incrementó y el pre-mRNA disminuyó, lo que sugiere que U2AF84 es un inhibidor de *splicing*. Para confirmar que U2AF84 inhibe el *splicing* sobre expresé U2AF84 etiquetada con hemaglutinina (HA-U2AF84) en *E. histolytica* cepa HM1-IMSS. En concordancia con datos de Morales Tovar, los ensayos de *splicing* in vivo para los transcritos RabX13, Cdc2, ClcB1 y Sam50 mostraron una acumulación del pre-mRNA en las transfectantes de HA-U2AF84 (84) respecto a las transfectantes con el vector vacío (V).

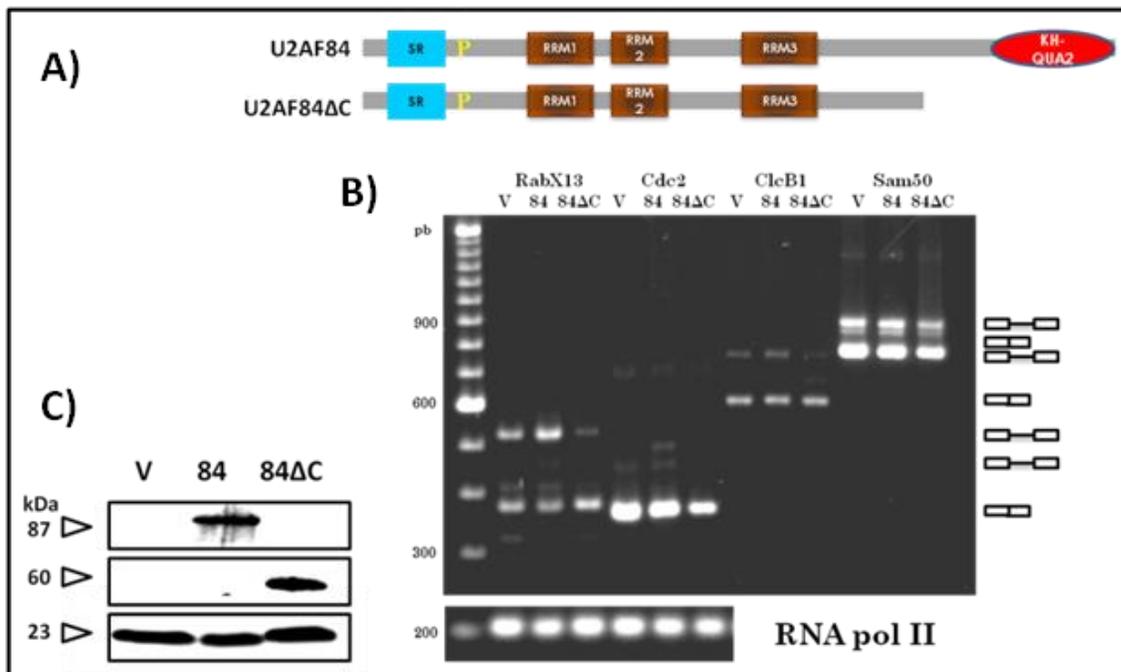


Fig 11. La región KH-QUA2 de U2AF84 inhibe el *splicing* en *E. histolytica*. A) Esquema de las proteínas de fusión U2AF84 y U2AF84ΔC (sin etiqueta HA) expresadas en trofozoítos. B) Ensayos de RT-PCR para los transcritos RabX13, Cdc2, ClcB1 y Sam50 fueron realizados para el vector vacío (V), U2AF84 (84) y U2AF84ΔC (84ΔC). A la derecha aparecen los símbolos para el mRNA y pre-mRNA; el transcrito de la RNA pol II fue utilizado como control de carga. C) La expresión de las proteínas expresadas en las transfectantes se monitorizó en western blots revelados con anti-HA. Anitcuerpo anti-Histonas fue utilizado como control de carga. Se indican los pesos moleculares en kDa.

Debido a que el extremo carboxilo terminal de U2AF84 presenta una región extra respecto a U2AF65 nosotros decidimos analizar si esta región está involucrada en la inhibición del *splicing*. Para tal fin, se realizó la construcción y transfección de U2AF84 Δ C (1-471 a.a.). Los ensayos de *splicing* in vivo mostraron una acumulación de mRNA en las transfectantes con la delección y una disminución del pre-mRNA respecto a las transfectantes con U2AF84 silvestre y con el vector vacío (Fig. 11B). Por medio de ensayos de western blot se verificó la expresión de las proteínas respectivas en extractos protéicos totales de las distintas transfectantes utilizando anticuerpos anti-HA (Fig. 11C). Estos datos sugieren que el dominio extra de U2AF84 inhibe el *splicing* en forma de retención intrónica.

Interacción de los dominios KQ2-SF1 y KQ2-84 con los BP de RabX13 y Levadura

Como se mostró anteriormente, el dominio KH-QUA2 de U2AF84 (KQ2-84) es el responsable de la inhibición de *splicing*. Además, el dominio UHM en el extremo carboxilo terminal de U2AF65 interactúa con el dominio ULM del extremo amino de SF1 y el dominio KH-QUA2 de SF1 (KQ2-SF1) interactúa con el BP lo que resulta en la formación del complejo E de *splicing*, paso esencial para que el *splicing* sea llevado a cabo (Raymon et al., 2010). Por tal motivo decidimos estudiar in vitro el mecanismo molecular por el cual este dominio puede conllevar a la inhibición en el *splicing* ya sea por una competencia por el sitio de unión al BP entre KQ2-84 y KQ2-SF1 o por alguna interacción proteína-proteína o proteína-RNA-proteína entre los dominios KQ2-84 y KQ2-SF1. Particularmente analizamos la naturaleza de las posibles interacciones entre los respectivos dominios KH-QUA2 mediante ensayos de cambio de la movilidad electroforética (EMSA) y entrecruzamiento con UV y formaldehído, utilizando distintas sondas de RNA correspondientes a BP, silvestres y mutantes, tanto de *E. histolytica* como de levadura, usado como control positivo. Para este fin nosotros obtuvimos proteínas

recombinantes de los dominios KH-QUA2 de EhSF1 (KQ2-SF1), U2AF84 y de MSL5 (KQ2-MSL5) de levadura, utilizado como control positivo en los EMSAS. Las proteínas fueron purificadas por afinidad (Fig. 12, panel inferior).

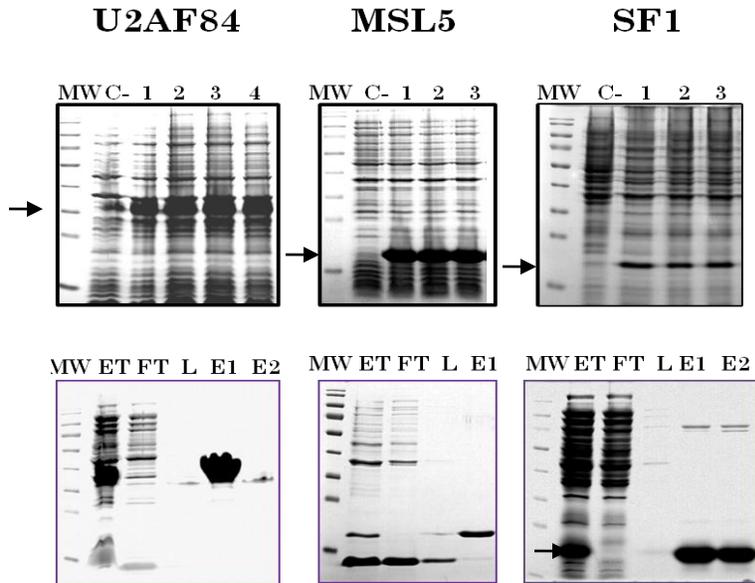


Fig 12. Purificación de los dominios KH-QUA2. El proceso de inducción de las proteínas fue monitoreado en un SDS-PAGE al 12% y teñidas con azul de Comassie, se muestran los KQ2-84, KQ2-MSL5, KQ-SF1 panel superior. En el panel inferior se muestra una alícuota de las proteínas purificadas. MW (marcador de peso), C- control negativo (vector vacío), ET (extracto total), FT (flujo a través de la columna), L (lavado), E (eluciones). Las flechas indican la altura a la cual se encuentra la proteína sobreexpresada.

La unión de los dominios KQ2 analizados con el BP fue llevado a cabo con cantidades crecientes de los dominios KH-QUA2 y con 80 fmolas de RNA de la secuencia del BP de RabX13 ³²P-AACUUUUUUUUU (P+) o con la secuencia consenso del BP de levadura ³²P-UAUACUAACAA (C+). Como se aprecia en la figura 13A, el KQ2-SF1 y KQ2-MSL5 se unen a la sonda C+ a una concentración de 5 y 10 microgramos respectivamente, en tanto que no se aprecia la formación evidente de un complejo de KQ2-84 incluso a 50 microgramos. Para el caso de la sonda P+ la unión de los KQ2-SF1 y KQ2-MSL5 es evidente a 50 microgramos y en el caso de KQ2-84 no apreciamos la formación del complejo incluso a 200 microgramos (panel inferior). En trabajos anteriores ya ha sido reportado que la interacción de los dominios KH-QUA2 es secuencia específica con los BP, por ello evaluamos la especificidad de unión de estos dominios y utilizando sondas mutantes C- y P- a las cuales les fueron sustituidos algunos nucleótidos importantes (indicados en rojo) como la adenosina responsable de primer ataque nucleofílico durante el *splicing*. Es evidente que con las sondas C- mutantes la formación de los complejos es ineficiente. Para el caso de P- el dominio KQ2-SF1 es más específico respecto al control positivo KQ2-MSL5. Nuevamente en KQ2-

84 no encontramos efecto alguno (Fig. 13B), por tal motivo nos preguntamos si la unión de KQ2-84 presenta una cinética de unión que dependa del tiempo y para ello realizamos la interacción sonda-proteína desde 4 min hasta 16 min. De manera interesante el KQ2-84 pareciera retardar las sondas en mayor proporción a los 4 min respecto a los 16 min. En el caso de los KQ2-84 y KQ2-MSL5 no existe una diferencia entre los 3 tiempos analizados (Fig. 13).

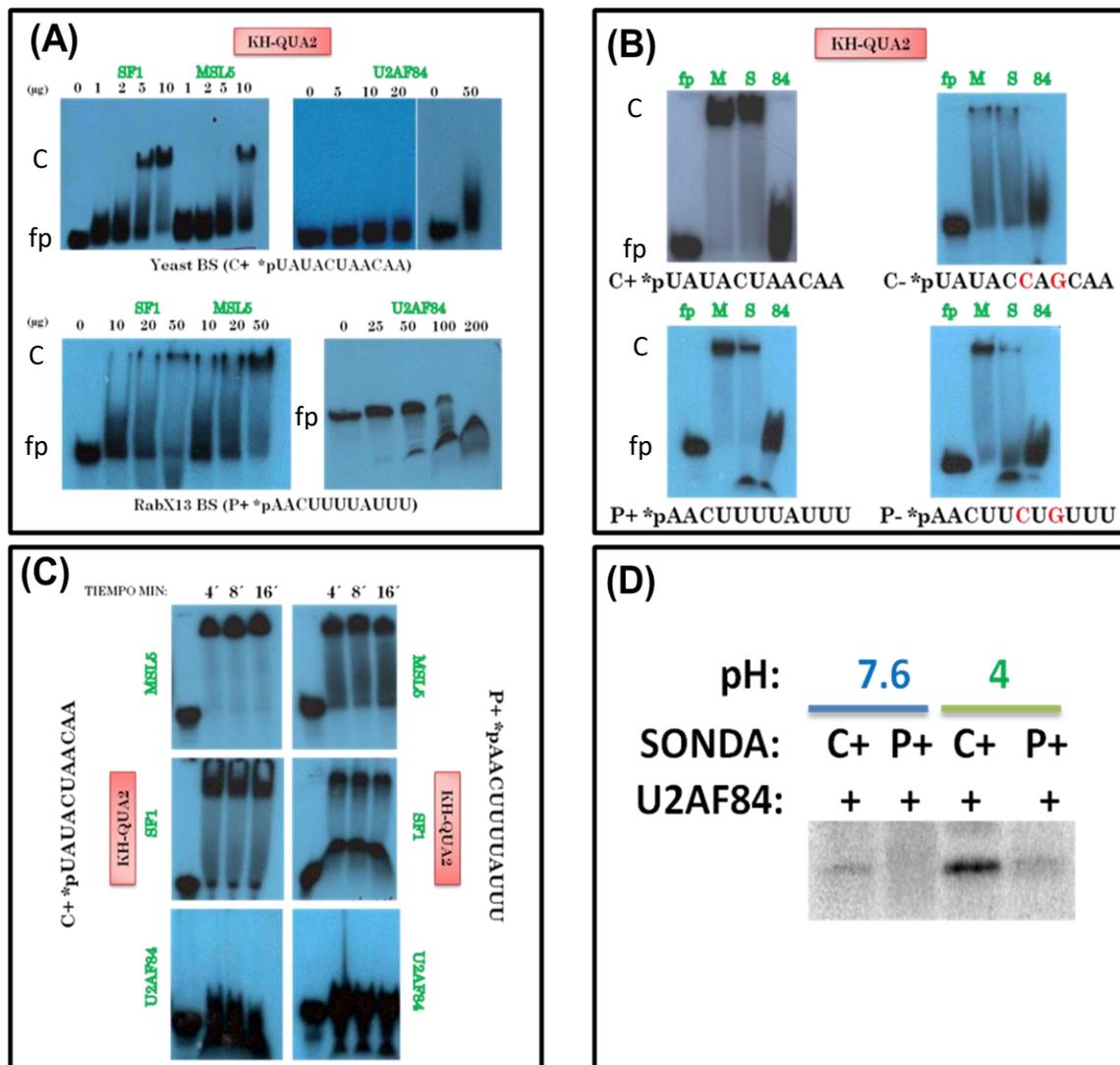


Fig 13. Los dominios KH-QUA2 interactúan con los BP de levadura y Rabx13. A) EMSA de cinética de concentraciones (en µg) de las proteínas recombinantes KQ2-84, KQ2-SF1 y KQ2-MSL5. En cada caso se indica la secuencia del BP utilizado. **B)** La especificidad del reconocimiento de los KH-QUA2 por su BP se probó utilizando BP mutantes (sustituciones de nt. señaladas en rojo). **C)** Cinética de tiempo (min) realizada con 5 y 50 µg de KQ2-SF1 y KQ2-84, respectivamente, para C+, y 50 µg de cada proteína para P+. **D)** Entrecruzamiento con UV, utilizando 50 µg de KQ2-84 a pH diferentes con ambos BP monitoreado en gel SDS-PAGE 10%. Los EMSAs fueron analizados en gels nativos al 10 %.

De acuerdo al análisis bioinformático y a los modelos solicitados en I-TASSER la región analizada de KQ2-84 si presenta un *pocket* de unión para el BP, pero su punto isoeléctrico se estimó en 4.02 (prot-param). Consistente con los datos obtenidos hasta ahora, los ensayos de unión se realizaron a pH 7.6, por los que se espera que esta proteína tenga carga negativa imposibilitándola para unirse al RNA. Para investigar la influencia del pH en la unión KQ2-84-BP, tanto la diálisis de la proteína purificada como la unión de la misma al BP se realizaron en amortiguadores a pH 4.0 y tal unión se evidenció por entrecruzamiento con UV. Bajo estas condiciones se puso de manifiesto una interacción entre el KQ2-84 y los BP aunque es más evidente en el caso de C+ (Fig. 13D). En conclusión las proteínas analizadas efectivamente son dominios KH-QUA2 con capacidad de unión a secuencias BP.

Formación de un súper complejo entre KQ2-SF1-KQ2-84 y BP

Para investigar si en condiciones fisiológicas existe una competencia por la unión al BP, existe una interacción entre KQ2-SF1 y KQ2-84 o si se favorece la unión de alguna de ellas al BP realizamos ensayos de mezcla utilizando cantidades crecientes de una proteína y manteniendo constante la otra y viceversa utilizando ambas sondas por separado. En presencia de P+ al incrementar la concentración de KQ2-84 a 0.1 y .5 veces la concentración inicial de KQ2-SF1 se observó una potenciación en la señal del complejo (carril 3 y 4 panel izquierdo); cuando KQ2-84 se mantuvo constante observamos que a medida que KQ2-SF1 se incrementó 1 y 10 veces respecto KQ2-84 se observó un súper complejo (carril 5 y 6 panel derecho). Para el caso de C+ cuando KQ2-SF1 se mantiene constante, se aprecia un potenciamiento en la señal en la señal del complejo una mezcla de 1:1 (carril 5 panel izquierdo); Cuando KQ2-84 se mantuvo constante observamos que a medida que KQ2-SF1 se incrementó 0.1, .5, 1 y 10 respecto a KQ2-84 se observó un súper complejo aun más evidente. De manera interesante en algunos casos (carril 5 y 6 de P+ panel izquierdo, carril 4 de C+ panel izquierdo) existe una disminución en la intensidad del complejo lo cual sugiere un papel inhibitorio de KQ2-84 en estas condiciones (Fig. 14). Para averiguar si en efecto KQ2-84 puede

tener un papel como inhibidor de la formación del complejo KQ-SF1-BP realizamos un ensayo de entrecruzamiento con UV utilizando mezclas de los distintos dominios KH-QUA2: el KQ2-SF1 y el KQ2-84. Observamos que En efecto KQ2-84 puede inhibir la interacción de KQ2-SF1 al BP tal como se muestra en el carril 3 y 6 de la figura 14 (panel inferior), panel izquierdo (P+) y panel derecho (C+). En conclusión el dominio KQ2-84 puede formar un súper complejo con el KQ2-SF1 y el BP, pero a demás puede funcionar como un inhibidor de la formación del complejo KQ2-SF1-BP.

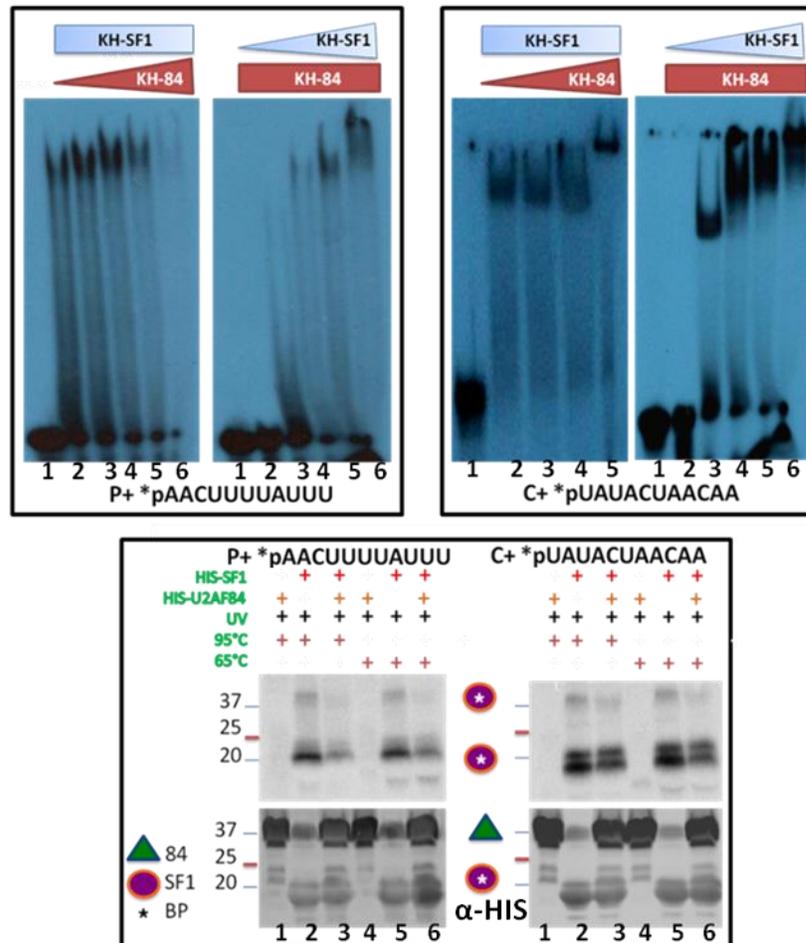


Fig 14. KQ2-84 forma un súper complejo con KQ2-SF1 y BP. En el panel superior se muestra el experimento de mezclas de las proteínas KQ2-84 y KQ2-SF1, donde la concentración inicial fue 5 µg para KQ2-SF1 y 50 µg para KQ2-84. Carril 1 sonda libre; carril 2 concentración inicial de la proteína constante (rectángulo), carriles 3, 4, 5 y 6 cantidades de 0.1, 0.5, 1, 10 veces de proteína creciente (triángulo) respecto a la proteína constante. En el panel inferior se muestra en ensayo de CL-UV con la(s) proteína(s) indicada(s) y utilizando las sondas P+ y C+; además se indica la temperatura a la cual fueron calentados los complejos después de 15 minutos de interacción. Hacia la izquierda de cada panel se indica el peso molecular en kDa; His, etiqueta de 6 Histidinas que se usó para la inmunodetección (α-HIS). Para mayor claridad se indican los RNA marcados (*), KQ2-84 (triángulo verde), KQ2-SF1 (óvalo morado) y KQ2-SF1 entrecruzado con el BP (* dentro del óvalo morado).

El súper complejo es formado mediante una interacción entre KQ2-SF1 y KQ2-U2AF84

En la introducción se describió que existe una interacción entre U2AF65 (UHM) y el ULM de Hum_SF1, lo cual estabiliza y estructura al 3'ss; esto sugiere que la interacción entre las proteínas juega un papel importante durante la formación del complejo temprano de *splicing*. En el resultado anterior mostramos la formación de un súper complejo entonces ahora nos preguntamos si este súper complejo es resultado de una interacción proteína-proteína, proteína-RNA-proteína, proteína-proteína-RNA o una combinación de las 3. Primero se procedió a realizar ensayos bioquímicos mediante entrecruzamiento con UV (para la formación de enlaces covalentes entre RNA-proteína) y formaldehído (enlaces covalentes proteína-proteína). Como se muestra en el carril 3 de la figura 15A cuando solo se irradia con UV podemos apreciar la formación del súper complejo el cual desaparece cuando se agrega el formaldehído primero y después se procede a la irradiación con UV (carril 2 panel A y B) indicando que en solución el dominio KQ2-84 bloquea la interacción del KQ2-SF1 con el BP; en contraste cuando se cambia el orden del entrecruzamiento (carril 6 de panel A y B) se encontró la formación del súper complejo en todos los casos, lo cual sugiere que el KQ2-84 interacciona con el complejo KQ2-SF1-BP. Cuando solo se trato con UV también podemos apreciar al súper complejo (carril 7 de panel A y B), puesto que el calentamiento en todos los casos fue a 65 °C podemos decir que la interacción entre las proteínas presenta una alta estabilidad o que existe una interacción entre proteína-RNA-proteína. Según los WB (panel inferior de 6A y 6B), cuando se utilizó formaldehído se encontró un doblete a la altura del súper complejo lo cual sugiere que existe una interacción proteína-proteína aun en ausencia de RNA tal como lo muestra el carril 8 de ambos WB. Siguiendo esta idea se realizó un análisis más detallado (panel C) en el cual podemos resumir que cuando se agrega el KQ2-SF1 y BP seguido de irradiación con UV y luego se adiciona el KQ2-84 y entonces formaldehído (carriles 1, 2 y 3 de C+ y P+), la señal de la radiación indica la existencia del súper complejo y en el WB se aprecia nuevamente el doblete que correspondería a las dos proteínas juntas (~53 kDa) aun con calentamiento a 95

°C. En seguida cuando se hace el mismo tratamiento pero colocando primero KQ2-84 (carriles 5, 6 y 7 de C+ y P+), podemos apreciar que la intensidad de la señal del súper complejo disminuye (según la señal del BP radioactivo) pero la interacción entre las proteínas se mantiene según el WB (panel inferior), en los carriles 4 y 8 de cada sonda en los cuales no se utilizo formaldehído nuevamente se disminuye la señal del doblete. En conclusión tal como lo muestra el modelo de la figura 15D; existe una interacción entre los KQ2-84 y KQ2-SF1, el KQ2-84 inte-

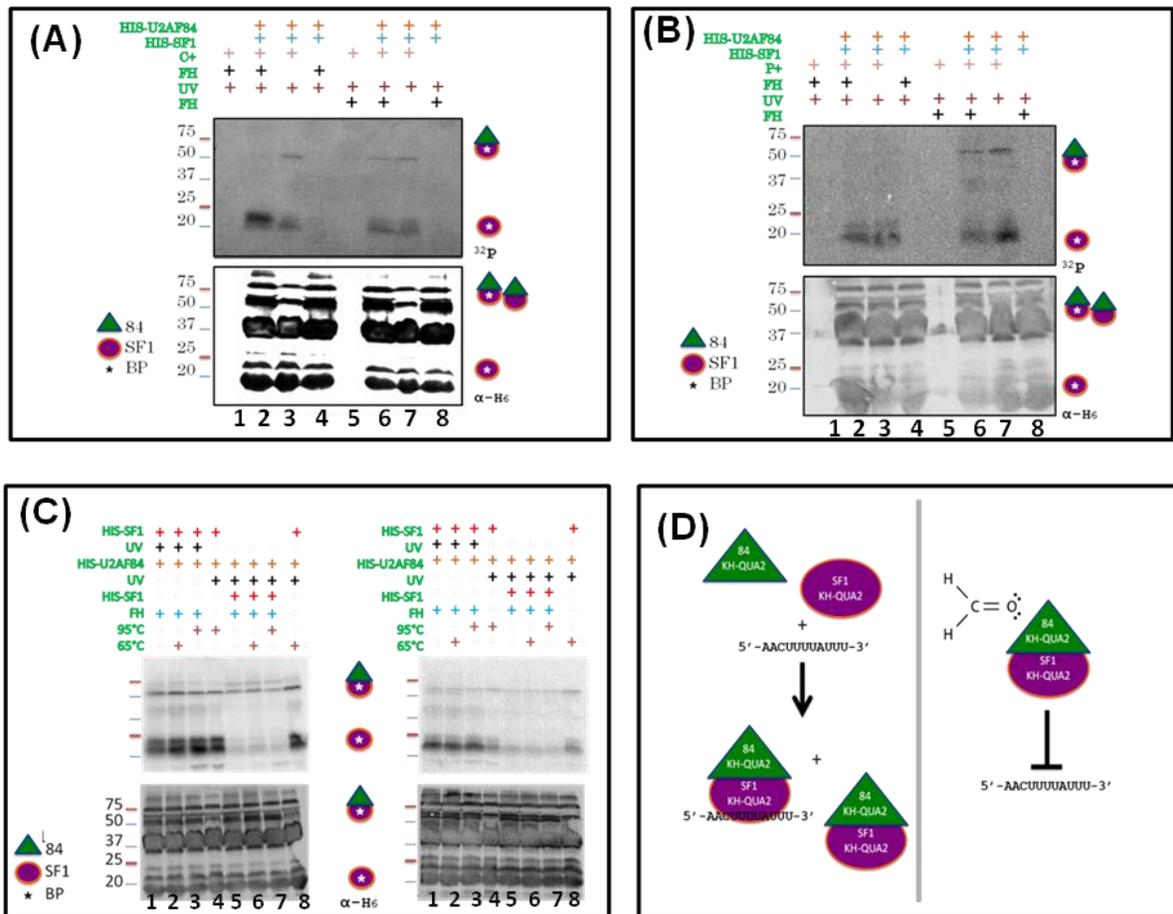


Fig 15. El KQ2-84 interacciona con KQ2-SF1 y con KQ2-SF1-BP. Los ensayos de CL con UV y FH fueron realizados con una concentración de 50 µg de ambas proteínas, con la sondas C+ (A) y P+ (B) marcadas y complementados en el orden que se indica (+); los complejos se disociaron a 65°C ó a 95°C según se indica. C) Igual que en A y B excepto que se agregó una proteína a la vez antes del entrecruzamiento; panel izquierdo P+, panel derecho C+. Los WB de los paneles inferiores en A, B y C fueron realizados con anti-Histidinas. (D) Resumen gráfico de los hallazgos en estos experimentos. Nomenclatura como en la Figura 14.

racciona con el complejo KQ2-SF1-BP y si se fija la interacción entre los KQ2-84 y KQ-SF1 la unión KQ2-SF1 y BP es bloqueada.

La interacción entre KQ2-84 y KQ2-SF1 es independiente de RNA.

Las evidencias anteriores muestran una interacción entre los dominios KQ2-84 y KQ2-SF1 la cual sucede aun en ausencia de RNA. De igual forma en reportes anteriores se ha descrito que las proteínas U2AF65 y SF1 pueden formar un heterodimero en ausencia de RNA. Para corroborar que esto es independiente de RNA se procedió a realizar ensayos de *Cross Linking Pull Down* (CLPD). Para lo cual se realizo la construcción de GST-KQ2-SF1 (GST-SF1) una vez purificada esta proteína se utilizo en un EMSA para comprobar su eficiencia y funcionalidad. En la figura 16A se muestra que la proteína interacciona con C+ desde los 2 µg de proteína siendo más evidente el complejo a concentraciones de 5, 10 y 20 µg. tal como se aprecia la etiqueta GST no retarda la sonda, una alícuota de la proteína His-KQ2-SF1 fue utilizada como control positivo del ensayo. De igual manera la formación del súper complejo fue confirmada utilizando la proteína GST-SF1 y KQ2-84, como se muestra en la figura 16B cuando se mezclan las dos proteínas a una relación 1:1, se puede apreciar un súper complejo para el caso de ambas sondas (C+ y P+).

Una vez calibrada la concentraciones de proteína necesarias para la formación del súper complejo (formado por KQ2-84, GST-SF1 y el BP), procedimos a realizar los ensayos de CLPD en presencia o ausencia de RNA, con o sin UV y con GST-SF1 o solo GST, pasado el tiempo de interacción, se agregaron las perlas de glutatión y se recuperaron por centrifugación. Las muestras fueron desnaturalizadas en agua hirviendo por 10 minutos con buffer de carga de proteínas 2X, posteriormente se analizaron en un SDS-PAGE al 12% y se realizo la transferencia a una membrana de nitrocelulosa tal como lo muestra la figura 16C. La señal del BP fue recuperado por autorradiografía y se encontró en el peso molecular de GST-SF1 cuando fue irradiado con UV carril 3 en ambos casos

sondas P+ y C+, después de haber capturado la señal de la radiación, la membrana fue rehidratada y se procedió a la tinción con rojo de Ponceau el cual muestra la presencia de las proteínas y la eficiencia de la precipitación con las perlas de glutatión (PR), la identidad de estas proteínas fue determinada mediante WB anti-Histidinas (para identificar a KQ2-84) y anti-GST (para identificar a GST-SF1) tal como se observa en la figura 16D. En efecto podemos apreciar la interacción de KQ2-84 con GST-SF1 (carriles 1, 2, 3 y 4) lo cual sugiere que la interacción de KQ2-84 y GST-SF1 ocurre independiente de la presencia de RNA y de la radiación con UV, los controles donde solo se uso la etiqueta GST (carriles 5, 6, 7 y 8) muestran que la interacción no es mediada por la etiqueta. Para comprobar la eficiencia de los anticuerpos, una alícuota de una u otra proteína fue cargada al gel (KQ2-84 como control de anti-Histidinas, GST-SF1 para el caso de anti-GST). En conclusión nuestros ensayos bioquímicos y de PD sugieren que existe una interacción proteína-proteína entre los dominios KQ2-84 y KQ2-SF1 y dicha interacción es independiente de RNA.

Los ensayos realizados confirman que existe una interacción entre KQ2-SF1 y KQ2-84, sin embargo estos experimentos fueron realizados in vitro con versiones recombinantes de ambas proteínas, entonces ahora nos preguntamos si nuestras proteínas recombinantes son capaces de interactuar con las proteínas de origen amebiano. Para lo cual se llevo a cabo el ensayo de far-wb (Fig. 17), en el cual se usaron extractos proteicos de amebas transfectadas con vector vacío, 84 y 84 Δ C (mismas usadas en los ensayos de RT-PCR). La interacción se llevo a cabo con 10 nM de GST-SF1 para la identificación de U2AF84 y con His-KQ2-84 para identificar a SF1. De acuerdo a lo encontrado en este experimento, en el panel izquierdo encontró una proteína a la altura de ~90 kDa la cual es más abundante en el carril que corresponde a los extractos de las amebas transfectadas con 84 (ver los carriles V y 84), en el caso del vector vacío esta señal es menos intensa y en el caso de 84 Δ C, no se aprecia señal lo cual pudiera ser resultado de una baja producción de U2AF84 debido a la sobreexpresión de la versión deletada. Como es de esperar el control donde no se uso proteína recombinante (S/P), no muestra señal en ningún caso lo cual nos garantiza que los anticuerpos usados en estos

ensayos, no reconocen proteínas endógenas de la ameba, si no que realmente nuestras recombinantes son las que interaccionan con las proteínas amebianas. Cuando se utilizo como bandera a His-KQ2-84, además de la señal esperada en una altura de 34 kDa la cual correspondería a SF1, se encontraron otras proteínas

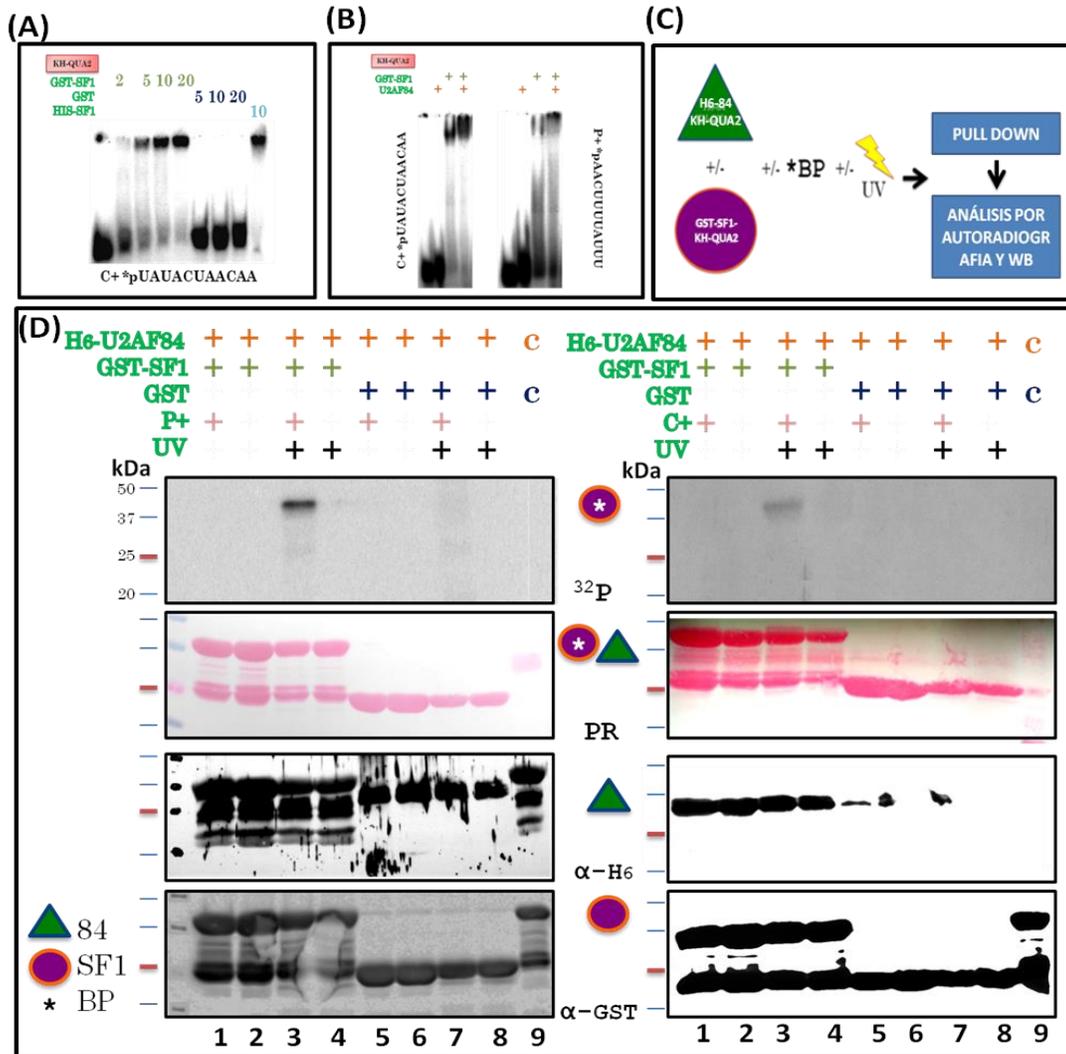


Fig 16. Interacción proteína-proteína entre los dominios KQ2-84 y KQ2-SF1 independiente de RNA. A) El dominio KQ2-SF1 fue etiquetado con GST y se usó en un EMSA para calibrar la concentración (en µg, indicadas para cada proteína) necesaria para la formación del complejo con la sonda C+; la etiqueta de GST fue utilizada para demostrar la especificidad de interacción del dominio KQ2-SF1. B) El súper complejo fue observado utilizando 50 µg de cada proteína, y para ambas sondas marcadas radioactivamente (indicado en cada panel). C) Estrategia experimental para el *crosslinking-pull down* realizado en D) Después de la interacción y la fijación de la misma (UV donde se indica), las muestras fueron hervidas en buffer de carga 2x y separadas en SDS-PAGE al 12 %. El panel superior muestra la marca radioactiva que indica la presencia del complejo GST-SF1-BP; el segundo panel hacia abajo muestra la tinción de la misma membrana con Rojo de Ponceau (PR); los paneles inferiores muestran consecutivamente los WB anti-Histidinas y anti-GST, respectivamente. Nomenclatura como en la Figura 14.

que pudiera ser el resultado de un mecanismo que la ameba lleva a cabo para bloquear el exceso de U2AF84 resultado de la sobreexpresión, o pudiera ser resultado de isoformas de cortas de SF1. En conclusión la interacción entre los dominios KH-QUA2 también sucede con las versiones endógenas de amebas, biológicamente esta interacción implicaría una inhibición del *splicing* ya que el KQ2-84 bloquea la unión de SF1 al BP y esto evita el ensamblaje del complejo E.

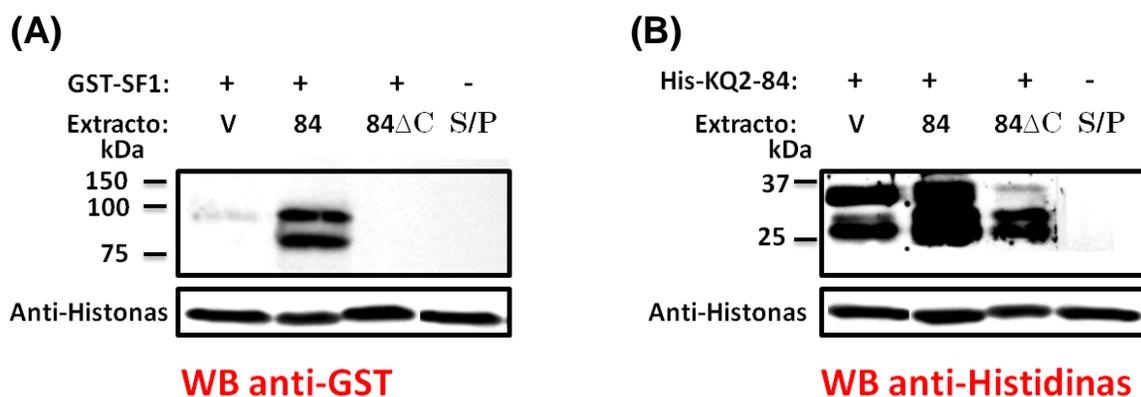


Fig 17. Interacción proteína-proteína entre los dominios KH-QUA2 de U2AF84 y SF1 en extractos amebianos. Para el far-wb 30 microgramos de extractos amebianos de transfectantes con vector vacío (V), HA-U2AF84 (84) y HA-U2AF84 Δ C (84 Δ C) fueron transferidos a nitrocelulosa e incubados con 10 nM de las proteínas recombinantes etiquetadas GST-KQ2-SF1 (A) e His-KQ2-84 (B). Después de la interacción se realizó el WB con el anticuerpo correspondiente de acuerdo a la etiqueta. Como control de carga el WB se reveló con anti-Histonas. Como control, extractos del vector se incubaron sin la proteína recombinante (S/P) y se revelaron con el anticuerpo correspondiente. A la izquierda aparecen los marcadores de peso molecular (kDa).

El dominio KQ2-84 inhibe la transición del complejo E-A

Durante el *splicing* el paso de la formación del complejo E-A involucra la disociación de SF1 del BP, debido a una interacción por complementariedad de bases por parte del snRNP U2 y el BP. Debido a que el KQ2-84 estabiliza la interacción del KQ2-SF1 al BP, decidimos investigar si la secuencia del RNA del snRNP U2 (U2 snRNA) disociaba el complejo formado por KQ2-SF1 y el BP en presencia de KQ2-84. Los ensayos de competencia entre KQ2-SF1 y U2 snRNA por el BP, muestran que en efecto el U2 snRNA es capaz de disociar el complejo KQ2-SF1-BP a una relación molar de 1:1, de manera interesante a una

concentración mayor no observamos efecto (Fig. 18A), este resultado fue observado para los dos BP (P+ y C+) usados en el ensayo, en resumen el U2 snRNA si es capaz de disociar al complejo lo cual representaría el paso del complejo E-A (panel superior Fig. 18C). De manera interesante cuando se realizo el experimento en presencia de KQ2-84 el U2 snRNA no fue capaz de disociar el complejo en ninguna de las dos sondas (Fig. 18B). En conjunto nuestros ensayos in vitro sugieren que el dominio KQ2-84 estabiliza la interacción formada entre el KQ2-SF1 y el BP, evitando así la entrada del U2 snRNA resultando en la inhibición del *splicing* (panel inferior Fig. 18C).

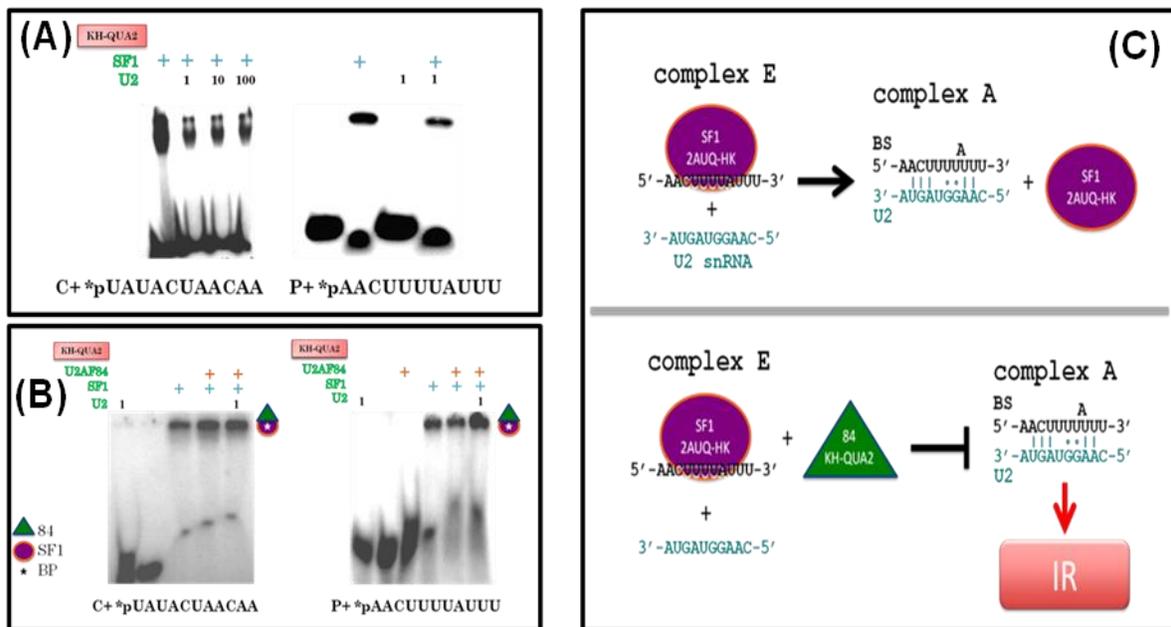


Fig 18. El dominio KQ2-84 inhibe la transición del complejo E-A. A) EMSA realizado con 50 μ g de KQ2-SF1 (óvalo morado) competido con 1, 10 y 100 veces de exceso molar de U2 snRNA frío (secuencia en celeste), para cada una de las sondas marcadas (200 fmolas; secuencia en negro). B) La inhibición de la competencia por KQ2-84 (triángulo verde) se realizó igual que en A) excepto que 50 μ g de KQ2-84 se añadieron a la reacción donde se indica. C) Resumen grafico del resultado obtenido, donde podemos apreciar la secuencia del U2 snRNA y la dinámica de los KQ2-84 y KQ2-SF1, a demás de la representación del paso E-A del *splicing*, en la parte inferior podemos apreciar el mecanismo inhibitorio propuesto para el KQ2-84, lo cual llevaría a retención intrónica (IR).

8. DISCUSIÓN

Mediante ensayos bioquímicos de CLIP-UV se identificaron los complejos tempranos de *splicing* de *E. histolytica* en donde se reportaron las proteínas EhSF1 y U2AF84 (Valdes et al., 2014) mismas que fueron analizadas en el presente trabajo. En tanto que en humanos SF1 y su homólogo en levadura MSL5 se unen al BP mediante un dominio KH-QUA2 (Garrey et al., 2006, Berglund et al., 1998), nuestro análisis bioinformático revela que U2AF84 presenta todos los elementos de U2AF65, pero además una región extra que presenta gran similitud con los dominios KH-QUA2 de los SF1 reportados (ver Fig. 10). Sin embargo este dominio KQ2-84 no une a la sonda BP en condiciones fisiológicas. Esto es de esperar pues el punto isoeléctrico de esta región es 4.02 (ver en anexo 1) por tanto a pH=7.5 al cual fue dializada y puesta a interactuar, la su carga neta es negativa lo cual lo imposibilita para unir al BP (Fig. 13 A y B), sin embargo esta proteína si posee un pocket en donde el RNA podría ser acoplado (anexo 2) tal como lo muestra el ensayo de *cross linking* realizado a pH=4.0 donde la carga de la proteína es positiva (Fig. 13D). Ya ha sido reportado que *E. histolytica* utiliza la retención intrónica como ruta principal de *splicing* alternativo (McGuire et al., 2008), sin embargo no se ha descrito el mecanismo por el cual puede darse dicho proceso. De acuerdo a los datos obtenidos en nuestro grupo de trabajo, U2AF84 está implicado en la retención intrónica, por ello procedimos a confirmar esto mediante experimentos más finos. La elucidación a nivel genético de la función del dominio KQ2-84 como inhibidor de *splicing* a nivel de pre-mRNA es un dato relevante, pues a diferencia de U2AF65, U2AF84 solo contiene la región KH-QUA2 extra y por tal motivo centramos nuestra atención en evaluar este dominio. Según la figura 11, el KQ2-84 es el responsable de la inhibición de *splicing*. En trabajos anteriores se encontró que las secuencias en cis en el 3' ss son las que regulan el *splicing* alternativo en particular exón *skipping* e IR. Sin embargo nuestros 4 transcritos con secuencias 3' ss distintas evaluados por ensayos de *splicing* in vivo presentaron IR mediada por el KH-QUA2 de U2AF84, lo cual sugiere que en *E. histolytica* es este dominio el responsable de la IR. Esto es más parecido al mecanismo encontrado en levaduras en donde las mutaciones puntuales L169A,

I189A, T265A y R267A dentro de la región del KH-QUA2 de MSL5 afectan la unión al BP y por ende el *splicing* es inhibido (Jacewicz et al., 2015). En otras palabras, el mecanismo de inhibición de U2AF84 se debe a una regulación por interacciones entre proteínas que unen RNA y no a competencia por el sitio de unión al RNA.

Durante el experimento de complementación usando los KQ2 de 84 y SF1, con ambos BP, encontramos la formación de un súper complejo el cual sugiere una estabilización del complejo. Dicho súper complejo resultó ser producto de una interacción entre las proteínas tal como lo revelan los ensayos bioquímicos como el *cross linking* con formaldehído (Fig. 15 A-C), ensayo de *pull down* (Fig. 16 D) y el far-wb (Fig. 17). Ya es conocido que MSL5/SF1 y MUD2/U2AF65 interactúan para estabilizar al complejo E, sin embargo esto involucra a los dominios UHM y RRM (Rain JC et al., 1998; Berglund JA et al., 1998) que no están presentes en las proteínas analizadas, de esta manera se trata de una interacción novedosa entre dos dominios KH-QUA2. Esta interacción fue realizada en condiciones fisiológicas donde el pH 7.5 mantiene a KQ2-SF cargado positivamente, mientras que KQ2-84 mantiene carga negativa, lo cual favorece aun más la interacción. Nosotros sugerimos que esta interacción es mediada por los dominios KH tal como se demostró con las variantes de Vg1RBP en experimentos de *cross linking* con dimetil de suberimidato, en donde todas las variantes de estos dominios KH fueron capaces de formar multimeros entre ellos (Anna Git & Nancy Standart 2002). De igual manera en el factor de *splicing* Nova-1 se demuestra un alineamiento anti-paralelo de los dos KH presentes en esta proteína, mismos que interaccionan entre sí (Teplova et al., 2002). A diferencia de estos reportes, nosotros detectamos una interacción novedosa entre dominios KH-QUA2 de distintas proteínas. Debido a que los dominios QUA2 confieren especificidad de unión al RNA (ref) y a que los dominios KH son capaces tanto de unirse al RNA como de interactuar con otros dominios KH, sugerimos que la interacción entre KQ2-84 y KQ2-SF1 este mediada por sus respectivos dominios KH.

Un dato que llamo nuestra atención durante los experimentos de mezcla, es que en ciertas relaciones en masa de los KQ2-84 y KQ2-SF1 encontramos una

disminución del complejo (carril 5 y 6 de P+ panel izquierdo, carril 4 de C+ panel izquierdo de la Fig.14), lo cual sugiere que el KQ2-84 funciona como inhibidor y estabilizador del complejo KQ-SF1-BP. La capacidad de inhibición de KQ2-84 puede ser explicada con nuestro resultado de *cross linking* con FH (Fig. 15 A y B), en el cual se muestra claramente como KQ2-84 interactúa con KQ2-SF1 y lo imposibilita para unirse al RNA. Pero si previamente se establece la unión de KQ2-SF1 con el BP, el KQ2-84 si puede interactuar con KQ2-SF1 dando estabilidad al complejo proteína RNA. En resumen el KQ2-84 puede interactuar con KQ2-SF1 independientemente del RNA imposibilitando a SF1 para contactar con el BP, pero además también puede interactuar con complejos KQ2-SF1-BP estabilizando dicho complejo e imposibilitándolo para el recambio de SF1 por el U2 snRNA, tal como se muestra en el experimento de inhibición de la transición del complejo E-A (Fig. 18). Es claro que el U2 snRNA no puede disociar a KQ2-SF1 de los BP mientras esté presente KQ2-84, esto podría ser resultado de la interacción fuerte entre los KH-QUA2, lo cual cambiaría la estructura del complejo KQ2-SF-BP imposibilitando al U2 snRNA para disociar el complejo o quizá la formación de un canal en el cual el BP quede atrapado entre los dos KH-QUA2.

Por último la implicación biológica de la función dual de KQ2-84 es la inhibición del *splicing* tal como lo revelan todas las pruebas presentadas y lo hace mediante un contacto proteína-proteína con el KQ2-SF1, sea antes de la interacción de SF1 con el BP o después de la misma. En los experimentos de far-wb se encontró una sobre expresión de SF1, en los carriles con sobreexpresión de U2AF84 respecto al vacío pero aun más evidente respecto al carril donde U2AF84 Δ C es sobreexpresada. Para este resultado nosotros proponemos que en abundancia de U2AF84 *E. histolytica* produce una mayor cantidad de SF1 para compensar la inhibición del *splicing* causada por U2AF84. Ahora bien cuando se le facilita el *splicing* mediante la versión deletada del KH-QUA2 (prácticamente una U2AF65), la producción de SF1 disminuye quizá debido a que la inhibición del *splicing* es prácticamente nula. De acuerdo a lo encontrado nosotros sugerimos que U2AF84 regula el *splicing* de un gran número de transcritos, dando lugar a IR, lo cual ocasiona una mayor diversidad en el transcriptoma de *E. histolytica* y con ello

podríamos sugerir a U2AF84 como regulador de la virulencia. Esta sugerencia está sustentada en el hecho de que la extensión de U2AF84 que contiene el dominio KH-QUA2 sólo está presente en las amibas virulentas del género (*E. histolytica*, *E. Nuttalli*, *E. invadens*) y ausente en las avirulentas (*E. dispar* y *E. moshkovskii*).

9. CONCLUSIÓN

De acuerdo a los resultados obtenidos durante los experimentos podemos decir que el dominio KQ2-84 está implicado en la inhibición del splicing en forma de IR mediante dos mecanismos: como inhibidor, mediante la interacción en solución con KQ2-SF1 evitando así la interacción entre KQ2-SF1 y el BP, y además puede funcionar como estabilizador del complejo KQ2-SF1-BP lo cual evita la entrada del U2 snRNA dando lugar a un bloqueo de la transición del complejo E-A y ocasionando la IR.

10. PERSPECTIVAS

Los ensayos de mutagénesis sitio dirigida sobre la región KH de la proteína KQ2-84 nos darían información valiosa acerca de los puntos de contactos que rigen la interacción con SF1 y con ello poder entender mejor como se regula la formación el *splicing* en *E. histolytica* por parte de U2AF84.

La evaluación de un mayor número de transcritos para saber cuáles son regulados por la interacción entre U2AF84 y SF1, esto podría realizarse mediante ensayos de CLIP para obtener el RNA y después hacer secuenciación masiva. Y hacer una comparación entre distintas cepas de Entamoeba.

Por último evaluar si U2AF84 está implicado de alguna forma en la regulación del ciclo biológico de *E. histolytica*, sobre todo poniendo énfasis en ensayos de patogenicidad y virulencia in vivo usando modelos animales como hámster.

11. BIBLIOGRAFIA

Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 2009; 136(4):701-718.

Toda T, Iida A, Miwa T et al. Isolation and characterization of a novel gene encoding nuclear protein at a locus (D11S636) tightly linked to multiple endocrine neoplasia type 1 (MEN1). *Hum Mol Genet* 1994; 3(3):465-470.

Kramer A. purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a pre-splicing complex. *Mol Cell Biol* 1992; 12(10):4545-4552.

Rain JC, Rafi Z, Rhani Z et al. conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* 1998; 4(5):551-565.

Shitashige M, Satow R, Honda K et al. Increased susceptibility of Sf1(+/-) mice to azoxymethane-induced colon tumorigenesis. *Cancer Sci* 2007; 98(12):1862-1867.

Abovich N, Rosbash M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 1997; 89(3):403-412.

Zamore PD, Patton JG, Green MR. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 1992; 355(6361):609-614.

Rutz B, Seraphin B. A dual role for BBP/ScSF1 in nuclear pre-mRNA retention and splicing. *EBMO J* 2000; 19(8):1873-1886.

Wang Q, Zhang L, Lynn B et al. A BBP-Mud2 heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast. *Nucleic Acids Res* 2008; 36(8):2787-2798.

Seraphin B, Rosbash M. identification of functional U1 snRNA-pre-mRNA complex committed to spliceosome assembly and splicing. *Cell* 1989; 59(2): 349-358.

Berglund JA, Fleming ML, Rosbash M. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA* 1998; 4(8):998-1006.

Garrey SM, Voelker R, Berglund JA. An extended RNA binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in RNA binding. *J Biol Chem* 2006; 281(37):27443-27453.

Buratti E, Baralle FE. Influence of RNA secondary structure on pre-mRNA splicing process. *Mol Cell Biol* 2004; 24(24):10505-10514.

Garrey SM, Cass DM, Wandler AM et al. Transposition of two amino acids changes a promiscuous RNA binding protein into a sequence-specific RNA binding protein. *RNA* 2008; 14(1):78-88.

Berglund JA, Chua K, Abovich N et al. the splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 1997; 89(5):781-787.

Keller EB, Noon WA. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. *Proc Natl Acad Sci USA* 1984; 81(23):7417-7420.

Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 2008; 14(5):802-813.

Fromont-Racine M, Rain JC, Legrain P. Toward a functional analysis of the yeast genome through exhaustive two- hybrid screens [see comments]. *Nat Genet* 1997; 16(3):277-282.

Berglund JA, Abovich N, Rosbash M. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* 1998; 12(6):858-867.

Kielkopf CL, Lucke S, Green MR. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* 2004; 18(13):1513-1526.

Jenkins JL, Shen H, Green MR et al. Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF65. *J Biol Chem* 2008; 283(48):33641-33649.

Gaur RK, Valcarcel J, Green MR. Sequential recognition of the pre-mRNA branch point by U2AF65 and a novel spliceosome-associated 28-kDa protein. *RNA* 1995; 1(4):407-417.

Garrey SM, Cass DM, Wandler AM, Scanlan MS, Berglund JA. 2008. Transposition of two amino acids changes a promiscuous RNA protein into a sequence-specific RNA binding protein. *RNA* 14: 78–88.

Banerjee H, Rahn A, Davis W et al. Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA* 2003; 9(1):88-99.

Valcarcel J, Gaur RK, Singh R et al. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected] [published erratum appears in *Science* 1996; 274(5284):21]. *Science* 1996; 273(5282):1706-1709.

Selenko P, Gregorovic G, Sprangers R et al. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell* 2003; 11(4):965-976.

Wang X, Bruderer S, Rafi Z et al. phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. *EMBO J* 1999; 18(16):4549-4559.

Corsini L, Bonnal S, Basquin J et al. U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat Struct Mol Biol* 2007; 14(7):620-629.

Zamore PD, Green MR. Identification, purification and biochemical characterization of U2 small ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci USA* 1989; 86(23):9243-9247.

Gordon PM, Sontheimer EJ and JD Piccirilli (2000). Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: Extending the parallels between the spliceosoma and group II introns. *RNA* 6:199-205.

Lewin B (2004). *Genes VIII*. Pearson, 697-725.

Turner IA, Norman CM, Cchurcher MJ and AJ Newman (2004). Roles of the U5 snRNP in spliceosoma dynamics and catalysis. *Biochemical society transactions* 32:928-931.

Hastings ML and Krainer (2001). Pre-mRNA splicing in the new millennium. *Cell Biol* 13:302-309

Rothrock CR, House A and KW Lynchc, (2005). HnRNP L represser exon splicing via a regulated exonic splicing silencer. *EMBO J* 24:2792-2802.

Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature* 2005;433:865–8.

Wilihoeft U, Campos-Gongora E, Touzni S, Bruchhaus I, Tannich E. Introns of *Entamoeba histolytica* and *Entamoeba dispar*. *Protist* 2001;152:149–56.

Davis CA, Brown MP, Singh U. Functional characterization of spliceosomal introns and identification of U2, U4, and U5 snRNAs in the deep-branching eukaryote *Entamoeba histolytica*. *Eukaryot Cell* 2007;6:940–8.

McGuire AM, Pearson MD, Neafsey DE, Galagan JE. Cross-kingdom patterns of alternative splicing and splicerecognition. *Genome Biol* 2008;9:R50.

Miranda R, Salgado LM, Sanchez-Lopez R, Alagon A, Lizardi PM. Identification and analysis of the u6 small nuclear RNA gene from *Entamoeba histolytica*. *Gene* 1996;180:37–42.

J. A. Berglund, K. Chua, N. Abovich, R. Reed, M. Rosbash, *Cell* **89**, 781 (1997).

J. C. Rain, Z. RaP, Z. Rhani, P. Legrain, A. Kra-mer, *RNA* **4**, 551 (1998).

P. Gallivan, D. A. Dougherty, Proc. Natl. Acad. Sci. U.S.A. **96**, 9459 (1999).

Petri, W. A., Jr., Vines, R. R., Lyerly, D., and Haque, R. (2000) *Parasitol. Today* **16**, 320–321.

Daubner, G.M., Brümmer, A., Tocchini, C., Gerhardy, S., Ciosk, R., Zavolan, M., Allain, F.H.-T., 2014. Structural and functional implications of the QUA2 domain on RNA recognition by GLD-1. *Nucleic Acids Res.* **42**, 8092–8105. doi:10.1093/nar/gku445.

Git, A., Standart, N., 2002. The KH domains of *Xenopus* Vg1RBP mediate RNA binding and self-association. *RNA* **8**, 1319–1333.

Jacewicz, A., Chico, L., Smith, P., Schwer, B., Shuman, S., 2015. Structural basis for recognition of intron branchpoint RNA by yeast Msl5 and selective effects of interfacial mutations on splicing of yeast pre-mRNAs. *RNA* **21**, 401–414. doi:10.1261/rna.048942.114.

Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Krämer, A., Sattler, M., 2001. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098–1102. doi:10.1126/science.1064719.

Rymond, B.C., 2010. The branchpoint binding protein: in and out of the spliceosome cycle. *Adv. Exp. Med. Biol.* **693**, 123–141.

Teplova, M., Malinina, L., Darnell, J.C., Song, J., Lu, M., Abagyan, R., Musunuru, K., Teplov, A., Burley, S.K., Darnell, R.B., Patel, D.J., 2011. Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure* **19**, 930–944. doi:10.1016/j.str.2011.05.002

Hernandez-Rivas, R., Ramirez, C., Guillen, N. and Vargas, M. (2000) DNA cloning of the *Entamoeba histolytica* PRP6 gene: a putative U4/U6 small nuclear ribonucleoprotein particle (snRNP). *Arch Med Res*, **31**, S294-295.

Marchat, L.A., Orozco, E., Guillen, N., Weber, C. and Lopez-Camarillo, C. (2008) Putative DEAD and DEXH-box RNA helicases families in *Entamoeba histolytica*. *Gene*, **424**, 1-10.

Lopez-Camarillo, C., de la Luz Garcia-Hernandez, M., Marchat, L.A., Luna-Arias, J.P., Hernandez de la Cruz, O., Mendoza, L. and Orozco, E. (2008) *Entamoeba histolytica* EhDEAD1 is a conserved DEAD-box RNA helicase with ATPase and ATP-dependent RNA unwinding activities. *Gene*, **414**, 19-31.

11. ANEXOS

Anexo 1: parametros bioquímicos de KQ2-84

ProtParam

User-provided sequence:

```
      10      20      30      40      50      60
IEEDINPQNN IQETKMEEEK TETKNQHEVT NENSLNSLPQ NEIMETINQK DIITQKQTKD

      70      80      90     100     110     120
CDGQNNITYD NKIEKEIQNN GTREILNVPA QNEIKECNYQ SIYLEQSGME EEVPNVVPQN

     130
KIEGDITQEV KTEEEIKE
```

Number of amino acids: 138

Molecular weight: 16051.4

Theoretical pI: 4.27

Amino acid composition:

Ala (A)	1	0.7%
Arg (R)	1	0.7%
Asn (N)	19	13.8%
Asp (D)	6	4.3%
Cys (C)	2	1.4%
Gln (Q)	14	10.1%
Glu (E)	26	18.8%
Gly (G)	4	2.9%
His (H)	1	0.7%
Ile (I)	16	11.6%
Leu (L)	4	2.9%
Lys (K)	12	8.7%
Met (M)	3	2.2%
Phe (F)	0	0.0%
Pro (P)	5	3.6%
Ser (S)	4	2.9%
Thr (T)	11	8.0%
Trp (W)	0	0.0%
Tyr (Y)	3	2.2%
Val (V)	6	4.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 32

Total number of positively charged residues (Arg + Lys): 13

Atomic composition:

Carbon	C	674
Hydrogen	H	1090
Nitrogen	N	188
Oxygen	O	254
Sulfur	S	5

Formula: C₆₇₄H₁₀₉₀N₁₈₈O₂₅₄S₅

Total number of atoms: 2211

Extinction coefficients:

This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient 4595
Abs 0.1% (=1 g/l) 0.286, assuming all pairs of Cys residues form
cystines

Ext. coefficient 4470
Abs 0.1% (=1 g/l) 0.278, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is I (Ile).

The estimated half-life is: 20 hours (mammalian reticulocytes, in vitro).
30 min (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

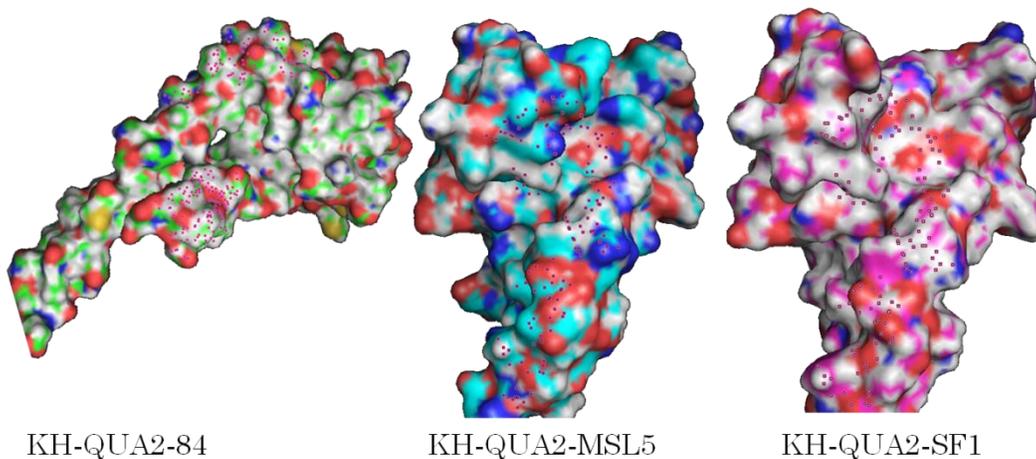
The instability index (II) is computed to be 82.00
This classifies the protein as unstable.

Aliphatic index: 69.86

Grand average of hydropathicity (GRAVY): -1.315

Anexo 2

Modelo de superficie de los dominios KH-QUA2



Anexo 2. Modelo de superficie de los dominios KH-QUA2 de las proteínas U2AF84, MSL5 y SF1. Los puntos muestran los sitios de contactos del RNA, de acuerdo a lo reportado (Liu et al., 2001; Jacewicz et al., 2015).

ANEXO 3

TABLA 1. SECUENCIA DE OLIGONUCLEOTIDOS, CONDICIONES DE Mg Y Tm PARA LOS TRANSCRITOS EVALUADOS.

Transcrito	Nombre	Secuencia de oligonucleótidos (5' → 3')	MgCl ₂ (mM)	Producto Amplificado (pb)	Tm (°C)
Cdc2	Cdc2 F	CAATTAGGAGAAGGAACATATGG	1.5	440-361	54
	Cdc2 R	GTGGTTTCATATCTCTGTGAAG			
CIC-B	Chlo F	ATGGAACAAAATTTACCCTC	2.5	652-596	55
	ChIRE2	CTATTGGAAGTGGATTATGGTG			
RabX13	Rab2 F	CGTTGTTGGAGACTCTTCAGTTGG	1.2	510-374	62
	Rab2R	GACCCATTTTCAGTTGAAACAGTTC			
Sam50	FSam50	GCAATGACAACAAGAATGCAA	2.5	835-763	64
	RSam50	AAGAAACCCCAACTCCACAA			
POL II	PolIIIF	GATCCAACATATCCTAAAACAACA	2	210	60
	PolIIIR	TCAATTATTTTCTGACCCGTCTTC			

Nota: El nombre del transcrito analizado se indica en la tabla, en seguida el código de los oligos usados (control interno), la secuencia de los oligonucleótidos, la concentración de Mg requerido, el tamaño esperado del producto amplificado (con intrón, sin intrón respectivamente) y por último la temperatura de alineamiento.