INSTITUTO POLITÉCNICO NACIONAL
Escuela Nacional de Medicina y Homeopatía
Sección de Estudios de Posgrado e Investigación

“Análisis funcional y molecular del factor de corte EhCFIm25 de *Entamoeba histolytica* como blanco terapéutico en la amibiasis humana”

TESIS
QUE PARA OBTENER EL GRADO DE
Doctor en Ciencias en Biotecnología

PRESENTA
M. en C. Juan David Ospina Villa

Directores de Tesis:
D. en C. Laurence Annie Marchat Marchau
D. en C. Absalom Zamorano Carrillo

México DF, octubre de 2017

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México, D.F. el día _05_ del mes de septiembre del año 2017, el que suscribe Juan David Ospina Villa_ alumno del Programa de _Doctorado en Ciencias en Biotecnología, con número de registro_A140220_, adscrito a la ENMyH_, manifiesta que es el autor intelectual del presente trabajo de Tesis bajo la dirección de la Dra. Laurence A. Marchat Marchau y el Dr. Absalom Zamorano Carrillo y cede los derechos del trabajo titulado _Análisis molecular y funcional del factor de corte EhCFIm25 como blanco terapéutico en la amibiasis, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del (de la) autor(a) y/o director(es) del trabajo. Este puede ser obtenido escribiendo a las siguientes direcciones__juando12358@gmail.com y lmarchat@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Juan David

Juan David Ospina Villa

SIP-13-BIS

INSTITUTO POLITÉCNICO NACIONAL
SECRETARIA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REGISTRO DE TEMA DE TESIS
Y DESIGNACIÓN DE DIRECTORES DE TESIS

México, D.F. a 28 de mayo del 2014

El Colegio de Profesores de Estudios de Posgrado e Investigación de la ENMyH en su sesión ordinaria No. 5 celebrada el día 28 del mes de Mayo conoció la solicitud presentada por el(la) alumno(a):

Ospina	Villa	Juan David
Apellido paterno	Apellido materno	Nombre (s)

Con registro: A 1 4 0 2 2 0

Aspirante de: Doctorado en Ciencias en Biotecnología

1.- Se designa al aspirante el tema de tesis titulado:
Análisis molecular y funcional del factor de corte EhCFIm25 como blanco terapéutico en la amibiasis

De manera general el tema abarcará los siguientes aspectos:
Análisis estructural de la proteína silvestre y mutantes tanto *in silico* como *in vitro*
Identificación del sitio de unión al RNAm de la proteína EhCFIm25
Evaluación del efecto de la inhibición de la proteína sobre la maquinaria de poli(A) *in vitro* e *in vivo*

2.- Se designan como Directores de Tesis a los Profesores:
Dra. Laurence A. Marchat Marchau , Dr. Absalom Zamorano Carrillo

3.- El trabajo de investigación base para el desarrollo de la tesina será elaborado por el alumno en:
Escuela Nacional de Medicina y Homeopatía, lab. 2 de Biomedicina Molecular

que cuenta con los recursos e infraestructura necesarios.

4.- El interesado deberá asistir a los seminarios desarrollados en el área de adscripción del trabajo desde la fecha en que se suscribe la presente hasta la aceptación de la tesis por la Comisión Revisora correspondiente:

Directores de Tesis

Dra. Laurence Annie Marchat Marchau

Dr. Absalom Zamorano Carrillo

Dr. César Augusto Sandino Reyes López
(Presidente del colegio)

SECRETARIA DE EDUCACION PUBLICA
ESCUELA NACIONAL DE MEDICINA Y HOMEOPATIA
SECCION DE ESTUDIOS DE POSGRADO

INSTITUTO POLITECNICO NACIONAL
SECRETARIA DE INVESTIGACION Y POSGRADO
30 SEP 2014
RECIBIDO
DIRECCION DE POSGRADO

SIP-14-BIS

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México siendo las 12:00 horas del día 05 del mes de Septiembre del 2017 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de La E. N. M. y H. para examinar la tesis titulada:

"Análisis molecular y funcional del factor de corte EhCFIm25 como blanco terapéutico en la amibiasis"

Presentada por el alumno:

Ospina	Villa	Juan David
Apellido paterno	Apellido materno	Nombre (s)

Con registro:

A	1	4	0	2	2	0

aspirante de:

Doctorado en Ciencias en Biotecnología

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***APROBAR LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de tesis

Dra. Laurence Annie Marchat Marchau

Dr. Absalom Zamorano Carrillo

Dra. Claudia Guadalupe Benítez Cardoza

Dra. María Esther Ramírez Moreno

Dr. Jorge Luis Rosas Trigueros

PRESIDENTE DEL COLEGIO DE PROFESORES

Dra. Mónica Ascención De Nova Ocampo

SECRETARÍA DE EDUCACIÓN PÚBLICA
INSTITUTO POLITECNICO NACIONAL
ESCUELA NACIONAL DE MEDICINA Y HOMEOPATIA
SECCIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

COMITÉ TUTORAL

Directores de Tesis:

D. en C. Laurence Annie Marchat Marchau

D. en C. Absalom Zamorano Carrillo

Asesores:

D. en C. Claudia Guadalupe Benítez-Cardoza

D. en C. Jorge Luis Rosas Trigueros

D. en C. María Esther Ramírez Moreno

Agradecimientos

Gracias infinitas a México por acogerme como un ciudadano más, fue y será siempre mi segundo hogar. He llegado a opinar igual que Juan O'Gorman *"México, para mí, representa el amor, la paz y todo aquello que es magnífico y maravilloso en el mundo".*

Les dedico este trabajo especialmente a mi esposa, y a mi hijo que viene en camino, son el motor de mi vida.

Quiero agradecer especialmente a la Dra. Laurence Marchat que más que mi asesora de tesis fue como una madre para mí. Me enseño muchísimas cosas no solo a nivel profesional sino también a nivel personal. La admiro y le agradezco infinitamente.

A mis compañeros del laboratorio de Biomedicina Molecular II, gracias por estar siempre ahí, pasamos grandes momentos juntos. Raúl García y Daniel Guzmán, gracias por ser más que compañeros, amigos para toda la vida.

Este trabajo fue realizado en el laboratorio de Biomedicina Molecular 2 y Laboratorio de Bioquímica y Biofísica Computacional de la Sección de Posgrado e Investigación (SEPI) de la Escuela Nacional de Medicina y Homeopatía del Instituto Politécnico Nacional (ENMyH-IPN) bajo la dirección de la Dra. en C. Laurence A. Marchat y del Dr. en C. Absalom Zamorano-Carrillo.

Este proyecto fue apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) con el número de proyecto 178550 y por la Secretaría de Investigación y Posgrado del Instituto Politécnico Nacional (SIP-IPN) con los números de proyecto 20150720, 20160801 y 20170969.

Agradezco el apoyo recibido por los programas de Becas de Doctorado del Consejo Nacional de Ciencia y Tecnología (CONACYT) con el número de becario A140220, así como del Programa Institucional para la Formación de Investigadores (BEIFI) por los proyectos 20150720, 20160801 y 20170969.

Igualmente, agradezco al programa ECOS (249554; M14S02) apoyado en el marco del Acuerdo México-Francia Relativo a la Formación Doctoral para la Investigación Científica, Desarrollo Tecnológico e Innovación, Programa ECOS, suscrito entre la SEP, el CONACYT, el Ministerio de Asuntos Extranjeros de la República Francesa y el ANUIES (SEP-CONACyT-ANUIES-ECOS NORD Francia), durante el periodo 2015-2018 por permitirme realizar mi estancia en el Instituto Pasteur en París – Francia.

Contenido

ABREVIATURAS

3D: tridimensional

a.a: aminoácidos

ADN: Ácido desoxirribonucleico

ARN: Ácido Ribonucleico

ARNm: ARN mensajero

CDC: *Centers for Disease Control and Prevention*

CF I y IIm: Factor de corte I y II m

CPSF: Factor específico de corte y poliadenilación

CstF: Factor estimulante del corte

CTD: Dominio carboxilo terminal

EMSA: Ensayo de retardamiento de la movilidad electroforética

MREs: Elementos de respuesta a microARNs

OMS: Organización Mundial de la Salud

PABP: Proteína de unión al tracto de poliA

PAP: PoliA polimerasa

PCR: Reacción en cadena de la polimerasa

PDB: *Protein Data Bank*

RBPs: Proteínas reguladoras

REMSA: Ensayo de retardamiento de la movilidad electroforética del ARN

RER: Retículo Endoplasmático Rugoso

RNAPII: ARN polimerasa II

RT-PCR: Transcripción reversa Reacción en cadena de la polimerasa

SELEX: *Systematic Evolution of Ligands by Exponential Enrichment*

UTR: Unstranslated Region - Región no codificante

LISTA DE FIGURAS

RESUMEN

La poliadenilación de los 3'UTR de los ARN mensajeros es un importante regulador de la expresión génica en células eucariotas. En nuestra búsqueda de nuevos mecanismos para el tratamiento de la amibiasis humana causada por el protozoario *Entamoeba histolytica*, analizamos si la alteración de la poliadenilación representa una estrategia eficiente para el control del parasito, usando herramientas de biología molecular y computacional. Primeramente, realizamos el silenciamiento del factor de corte EhCFIm25 mediante el cultivo de los trofozoítos en presencia de RNA de doble cadena dirigidos contra el gen *EhCFIm25*. Los resultados evidenciaron que la ausencia de EhCFIm25 provoca la muerte de los parásitos y la disminución de sus propiedades de virulencia; además los trofozoítos aparecieron más grandes y multinucleados. Estos cambios fueron asociados a cambios en la selección del sitio proximal de corte/poliadenilación lo que probablemente repercutió en la expresión génica. Por otra parte, siguiendo la estrategia SELEX, identificamos dos aptámeros de ARN de cadena sencilla (C4 y C5) que reconocen específicamente a la proteína EhCFIm25. Ensayos de interacción RNA-proteína mostraron que EhCFIm25 se une al motivo GUUG *in vitro,* el cual difiere del motivo UGUA reconocido por la proteína homóloga humana. Consiguientemente, los experimentos de docking y simulaciones de dinámica molecular confirmaron que la interacción con GUUG estabiliza la proteína EhCFIm25, pero no la mutante EhCFIm25*L135T que no es reconocida por los aptámeros. Además, sugieren que la cadena lateral de Leu135 es esencial para el equilibrio estructural, mientras que el grupo hidroxilo del Tyr236 participa directamente en la interacción con el RNA. De manera interesante, el secuestro de EhCFIm25 por los aptámeros resultó en la muerte rápida de los trofozoítos. En conjunto, nuestros datos indican que la alteración de la poliadenilación y más específicamente del factor de corte EhCFIm25 representa una vía efectiva para el control de *E. histolytica*. Además, el presente estudio es el primero que remarca el valor potencial de los aptámeros de ARN para controlar este patógeno humano.

ABSTRACT

Polyadenylation of 3'UTR of messenger RNAs is an important regulation mechanism of gene expression in eukaryotic cells. In our search for new alternative for the treatment of human amebiasis caused by the protozoan *Entamoeba histolytica*, we analyzed whether the alteration of polyadenylation represents an efficient strategy for parasite control using molecular and computational biology tools. First, we performed the silencing of the EhCFIm25 by the culture of the trophozoites in the presence of double-stranded RNA directed against the EhCFIm25 gene. The results evidenced that the absence of EhCFIm25 causes the death of the parasites and the diminution of their virulence properties. In addition, the trophozoites appeared larger and multinucleated. These changes were associated with changes in proximal cleavage/ polyadenylation site selection which probably had repercussions on gene expression. On the other hand, following the SELEX strategy, we identified two single chain RNA (C4 and C5) aptamers that specifically recognize the EhCFIm25 protein. RNA-protein interaction assays showed that EhCFIm25 binds to the GUUG motif in vitro, which differs from the UGUA motif recognized by the human homologous protein. Consequently, docking experiments and molecular dynamics simulations confirmed that the interaction with GUUG stabilizes the EhCFIm25 protein, but not the mutant EhCFIm25*L135T that is not recognized by the aptamers. In addition, they suggest that the Leu135 side chain is essential for structural equilibrium, while the hydroxyl group of Tyr236 is directly involved in the interaction with RNA. Interestingly, sequestration of EhCFIm25 by aptamers resulted in rapid death of trophozoites. Taken together, our data indicate that alteration of polyadenylation and more specifically of the EhCFIm25 represents an effective pathway for the control of *E. histolytica*. In addition, the present study is the first to highlight the potential value of RNA aptamers to control this human pathogen.

INTRODUCCIÓN

La expresión génica es un proceso biológico esencial para todos los organismos vivos ya que permite la transformación de la información codificada en el ADN (genes) en las proteínas necesarias para su desarrollo y funcionamiento. Este proceso involucra dos etapas principales, la transcripción que permite la síntesis del ARN mensajeros (ARNm) en el núcleo y la traducción que conlleva a la producción de las proteínas correspondientes en el citoplasma. Particularmente, la transcripción puede ser regulada en diferentes pasos como son la modulación de la estructura de la cromatina, la interacción de proteínas reguladoras (factores de transcripción) con motivos específicos en el promotor (ADN) y el procesamiento de los pre-ARNm.

Al mismo tiempo que se lleva a cabo el proceso de transcripción por la ARN polimerasa II (RNAP II) sucede el procesamiento de los transcritos nacientes. En Eucariotas, ese procesamiento se lleva a cabo en el núcleo y comprende tres procesos muy importantes: el *capping* que consiste en la adición de la estructura denominada caperuza o casquete en el extremo 5´, el *splicing* o eliminación de intrones, y la poliadenilación del extremo 3´. Estas reacciones son necesarias para que los ARNm puedan madurar, ser estables y ser exportados al citoplasma. En todos estos procesos, la subunidad mayor del dominio carboxilo terminal (CTD) de la RNAP II se ha visto fuertemente implicado y al parecer es la conexión entre la transcripción y las demás reacciones de procesamiento del RNA (Brown & Gilmartin 2003; Venkataraman *et al.*, 2005, Yang *et al.,* 2013, Mandel *et al.,* 2006, Ryan *et al.,* 2004). Muchos factores que participan en el procesamiento de los ARNm ya han sido identificados y caracterizados, pero todavía no es muy claro como la transcripción se relaciona con cada una de las maquinarias de procesamiento (Adamson *et al*., 2005).

1.1. Relevancia de la poliadenilación

En Eucariotas, alteraciones en la poliadenilación de los transcritos, ya sea por pérdida o ganancia de función, ocasionan defectos letales en organismos unicelulares como *Schizosaccharomyces pombe* (Wang *et al.,* 2005) o en células humanas, causando varias enfermedades como: desórdenes hematológicos, cáncer, diabetes, y enfermedad de Fabry, entre otras (Curinha *et al*., 2014).

Por ejemplo, la existencia de mutaciones puntuales en la señal de poliadenilación se ha relacionado con desórdenes hematológicos como las talasemias, esto debido a que la RNAP pol II depende esencialmente de la señal de poliadenilación para la terminación de la transcripción. Al encontrarse mutada esta señal, no es reconocida adecuadamente y se producen isoformas alargadas del ARNm que codifica para globina, causando una producción anormal de hemoglobina (Orkin *et al.,* 1985; Rund *et al.,* 1992).

Con respecto al cáncer, recientes estudios han relacionado los hallazgos de tumores proliferativos con la presencia de ARNm que tienen 3'UTRs cortos. El acortamiento de los 3'UTRs conlleva a la pérdida de secuencias esenciales para la estabilidad de los transcritos y su expresión, lo que les permite evadir la regulación de la expresión por microARNs y por proteínas reguladoras (RBPs). Tanto 3'UTRs cortos o largos, dependiendo del tipo de cáncer y del estado de desarrollo del mismo, han sido reportados. Curiosamente dos diferentes líneas celulares de cáncer de seno (MCF7 y MB231) presentan un patrón de tamaño de 3'UTR opuesto. La línea MCF7 demuestra una alta producción de isoformas de transcritos con 3'UTRs cortos, mientras que la línea MB231 presenta isoformas con 3'UTRs largas (Stacey *et al.,* 2011). En el caso de genes anticancerígenos con 3'UTRs largos, la presencia de 3'UTRs largas y por lo tanto de más elementos de respuesta a microARNs (MREs) puede promover la inhibición de la expresión

génica, favoreciendo de esta manera los procesos cancerígenos (Li *et al.,* 2014).

Estos ejemplos muestran que el proceso de corte/poliadenilación tiene una gran relevancia para la correcta expresión de muchos genes y que su alteración puede provocar el desarrollo de enfermedades y una desregulación en la homeostasis celular.

1.2. Proceso de corte y poliadenilación en Eucariotas

Todos los pre-ARNm en eucariotas son procesados en el extremo 3' de la región no codificante (UTR del inglés "*Unstranslated region*") antes de ser exportados hacia el citoplasma, con la única excepción de los pre-ARNm que codifican para las histonas que solo precisan de la hidrólisis de un solo enlace fosfodiester en el transcrito naciente (Gilmartin *et al.,* 2005).

El procesamiento de los 3'UTR implica dos pasos que se llevan a cabo de forma coordinada: el corte y la poliadenilación del pre-ARNm. Estas reacciones requieren de la participación de varios complejos multi-proteicos, así como de secuencias en *cis* propias del 3´UTR de los pre-ARNm. Así, la señal de poliadenilación hexamérica altamente conservada AAUAAA que se encuentra 10-30 nucleótidos (nt) río arriba del sitio de corte es reconocida por el factor específico de corte y poliadenilación (CPSF) que se compone de seis subunidades (CPSF160, CPSF100, CPSF73, CPSF30, Fip1, y WDR33). CPSF30 y WDR33 interactúan directamente con la secuencia AAUAAA, mientras que CPSF73 es la endonucleasa que lleva a cabo el corte del 3´UTR en el sitio que generalmente se compone por un dinucleótido CA. Una región menos conservada rica en G/U, localizada 20-40 nt río abajo del sitio de corte es reconocida por el factor estimulante del corte (CstF) que consiste de tres subunidades (CstF77, CstF64, CstF50). CstF64 realiza contacto directo con esta secuencia e interactúa con CstF77 que a su vez interactúa con CstF50. Por otra parte, el motivo UGUA que se presenta en una o más copias a una distancia variable río arriba del sitio de corte es reconocido por

la subunidad de 25 kDa del factor de corte Im que está formado por cuatro subunidades (CFIm25, CFIm59, CFIm68 y CFIm72) (Brown *et al.,* 2003). EhCFIm también interactúa con el factor de corte IIm (CFIIm) que tiene dos subunidades (Pcf11 y Clp1). Otros factores individuales ayudan a la formación del complejo por medio de interacciones proteína-proteína, entre los que destacan: la simplequina que es una proteína *scaffold* importante para el reclutamiento de factores al complejo de corte/poliadenilación, la poliA polimerasa (PAP) que adiciona de 200 a 250 residuos de adenina en el extremo 3' cortado y la proteína de unión a poliA (PABP) que se une a los residuos de adenina adicionados para protegerlos de la degradación por exonucleasas. PABP participa también en diferentes procesos celulares como son el transporte de ARNm del núcleo al citoplasma, el inicio y la terminación de la traducción, y la degradación de los ARNm (Kûhn & Wahle 2004). En realidad, la maquinaria de corte-poliadenilación del extremo 3´ de los transcritos puede llegar a incluir por lo menos 20 complejos proteicos en levaduras y hasta 80 en células humanas (Shi 2015).

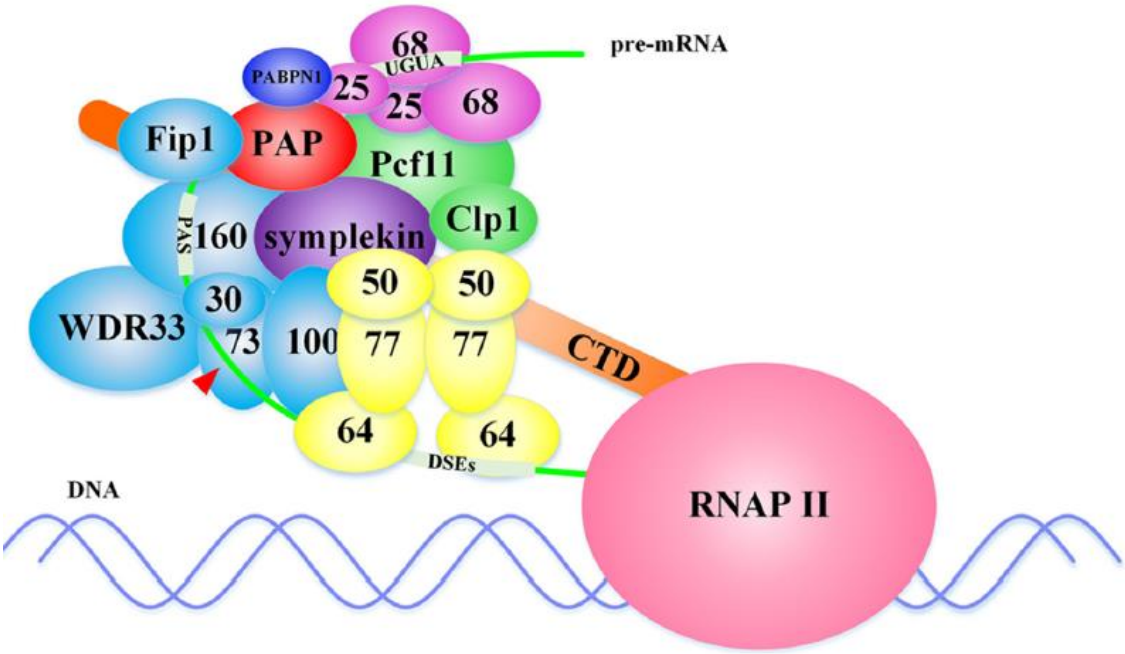


Figura 1. Modelo delos principales factores participantes en la maquinaria de corte/poliA del extremo 3'UTR de los ARNm en humano. (Xiang *et al.,* 2014)

1.3. Factor de corte CFIm

El Factor de corte CFIm se asocia tempranamente con el complejo de elongación de la transcripción junto con CPSF y CstF. Además, su unión al pre-ARNm ayuda a la estabilización de la unión de CPSF al ARN. CFIm es un complejo heterotetramérico formado por un homodímero de la subunidad pequeña de 25 kDa y un homodímero de alguna de las subunidades mayores: CFIm59 y CFIm68, codificados por dos genes parálogos, o bienCFIm72 que es una isoforma de CFIm68 generada por medio de *splicing* alternativo. Las subunidades mayores que tienen un motivo de reconocimiento de ARN (RRM) parecen tener un rol redundante ya que se ha demostrado en experimentos *in vitro* que CFIm25 y CFIm68 por sí solos tienen la capacidad de llevar a cabo la función del complejo CFIm (Dettwiler *et al.,* 2004). En levaduras, este complejo no tiene equivalente, por lo que anteriormente, se pensaba que CFIm solo estaba presente en metazoos; pero recientemente se ha encontrado secuencias génicas que codifican para proteínas homólogas en parásitos protozoarios como *T. brucei* y *E. histolytica* (Koch *et al.,* 2015, Pezet *et al.,* 2013).

1.4. CFIm25

La subunidad de 25 kDa del CFIm es codificada por el gen CPSF5 en humanos, tiene una longitud de 255 aminoácidos (aa) y un peso molecular de 25 kDa. La proteína contiene un dominio Nudix (Nucleótido difosfato unido a una molécula X) en la region 76-201 aa que depende de un motivo o caja nudix de 23 aa (GXXXXXEXXXXXXXREUXEEXGU), donde U puede ser isoleucina, leucina o valina y X, cualquier aminoácido) para su actividad catalítica (Coseno *et al.,* 2008; Yang *et al.,* 2010) Generalmente las proteínas que pertenecen a la superfamilia Nudix están presentes en todos los reinos de la naturaleza y tienen actividad de hidrolasa (McLennan 2005). Sin embargo, CFIm25 carece de dos glutamatos en el motivo Nudix, y esto podría explicar porque solo tiene la función de unirse al ARN, sin poder realizar su hidrólisis. El plegamiento típico $\alpha/\beta/\alpha$ de las proteínas Nudix le permite a CFIm25 unirse al ARN (Coseno *et al.,* 2008).

CFIm25 interactúa con proteínas no solo de la maquinaria de poliadenilación como son; PAP, PABP (Dettwiler *et al.,* 2004; Kim & Lee 2001) y CPSF (Venkataraman *et al.,* 2005), sino también con proteínas de otros procesos relacionados con el ARN, como son el *splicing* (CLP1, SRSF3, SNRNP70) (Hein *et al.,* 2015) y el transporte al citoplasma (HNRNPK) (Vinayagam *et al.,* 2011) por mencionar algunos ejemplos. Por lo que CFIm25 representa un enlace funcional entre estos eventos del metabolismo del ARN.

Para demostrar la importancia de la proteína CFIm25 en el proceso de corte/poliA se han llevado a cabo diferentes trabajos de investigación. Por ejemplo, Kubo *et al*. (2006) desarrollaron ARN de interferencia (ARNi) específicos para inhibir la expresión de CFIm25 en células HeLa y analizar las consecuencias en el ARNm de varios genes reporteros (TIMP-2, syndecan2, ERCC6 y DHFR) por ensayos 3'RACE y Northern blot. Sus resultados evidenciaron que al silenciar CFIm25, se produce preferentemente ARNm con 3'UTRs cortos, es decir, se seleccionan los sitios de poliA proximales y no los sitios de poliA distales. Por otro lado, Brown *et al*. (2003) desarrollaron aptámeros específicos que permiten bloquear la función de la proteína CFIm25 y recrearon *in vitro* las diferentes etapas del proceso de corte/poliA. Los investigadores mostraron que el bloqueo de CFIm25 por los aptámeros específicos altera cada uno de los pasos del proceso de corte/poliA, es decir el reclutamiento de la maquinaria de poliadenilación, así como las reacciones de corte y poliadenilación. En conjunto, estos trabajos muestran que CFIm25 regula la selección del sitio de corte/poliA y tiene un rol esencial para síntesis de la cola de poliA, por lo que representa una proteína clave en la regulación de la expresión génica.

A través de ensayos tipo SELEX (Systematic Evolution of Ligands by EXponential enrichment) se identificó que CFIm25 reconoce el motivo UGUAN en el 3'UTR de los pre-ARNm (Manley 2013). La cristalización de la proteína CFIm25 en complejo con el motivo UGUA, cuya estructura se encuentra depositada en el Protein Data Bank (PDB) con el código 3MDG, revela que la interacción entre CFIm25 y UGUAN se lleva a cabo

principalmente a través de puentes de hidrógeno con átomos de la cadena principal y de la cadena lateral, por apilamiento aromático y apilamiento de enlaces peptídicos. La mayoría de los residuos necesarios para la interacción con el ARN se encuentran en el dominio Nudix, pero otros que también participan en la interacción están ubicados en el extremo N-terminal de la proteína. El análisis del complejo CFIm25/UGUA reveló que 14 residuos están implicados en la interacción. U1 forma tres puentes de hidrógeno intramoleculares a través de su cara Watson-Crick. Esta base es reconocida por Phe104, cuya cadena principal interactúa con O2 y N3, por Glu81 cuya cadena principal estabiliza a O4, y por Leu106 a través de una molécula de glicerol. U1 también forma un puente de hidrógeno con Thr102 a través del hidroxilo de O2'. U1 también es estabilizado por apilamiento con Tyr208 y Gly209. G2 participa en interacciones por puentes de hidrógeno no solo con la proteína sino también con A4 por contacto intramolecular. El grupo amino N2 se une a la cadena lateral de Glu55 mientras que el grupo N1 interactúa con Glu55 por medio de una molécula de agua. También interactúa con Thr102 y Phe103 por medio de puentes de hidrógeno. Contactos de Van der Waals con Leu99 fortalecen el contacto molécula-molécula. U3 forma puentes de hidrógeno con Arg63, y una molécula de agua media la interacción entre U3 y A4, Glu55 también se conecta por una molécula de agua con U3. A4 tiene contacto con Phe103 y Glu55 y A5 interactúa con Phe104 y Tyr208 (Yang *et al.,* 2010).

1.5. Aptámeros como una novedosa estrategia diagnóstica y terapéutica

Los aptámeros son oligonucleótidos de ADN o ARN con una estructura 3D única que les permite interactuar con un blanco específico con alta afinidad y especificidad. El término aptámero proviene del latín *“aptus”* que significa “que encaja” y del griego *“mers”* que significa “molécula”. A pesar de que los aptámeros fueron descritos primero como moléculas sintéticas, luego se descubrieron de forma natural y fueron nombrados como “riboswitches” que pueden regular la transcripción o la traducción (Winkler *et al.,* 2002).

La estrategia para obtener los aptámeros de forma sintética es conocida como SELEX por sus siglas en inglés ***S****ystematic* ***E****volution of* ***L****igands by* ***EX****ponential Enrichment*. Esta estrategia fue publicada por primera vez en el año 1990 por dos grupos de investigación independiente pertenecientes a la Universidad de Harvard y a la Universidad de Colorado en Estados Unidos (Ellington & Szostak 1990; Tuerk & Gold 1990). Brevemente, la estrategia SELEX comienza con la generación de una biblioteca de ADN o ARN de entre 10^4 y 10^{15} secuencias al azar. Los oligonucleótidos que interactúan con su molécula blanco (generalmente proteínas inmovilizadas en una matriz inerte) son seleccionados y los que no, son descartados. Posteriormente los que han sido elegidos son amplificados por PCR o RT-PCR (dependiendo del tipo de ácido nucleico) para enriquecer estas secuencias y posteriormente continuar con una nueva ronda de selección hasta que el 90% de los oligonucleótidos se unan a la molécula blanco. Finalmente, los fragmentos seleccionados se clonan y se valida su interacción con la molécula blanco mediante ensayo de interacción ácidos nucleicos-proteína (EMSA o REMSA) por ejemplo. Si el blanco es una proteína de unión al ARN o ADN, la comparación de las secuencias de los aptámeros así seleccionados permite identificar un motivo consenso que corresponde al sitio de unión de la proteína. La afinidad y la especificidad de un aptámero por su blanco varían dependiendo de la molécula blanco, pero en general oscila en rangos pM to nM.

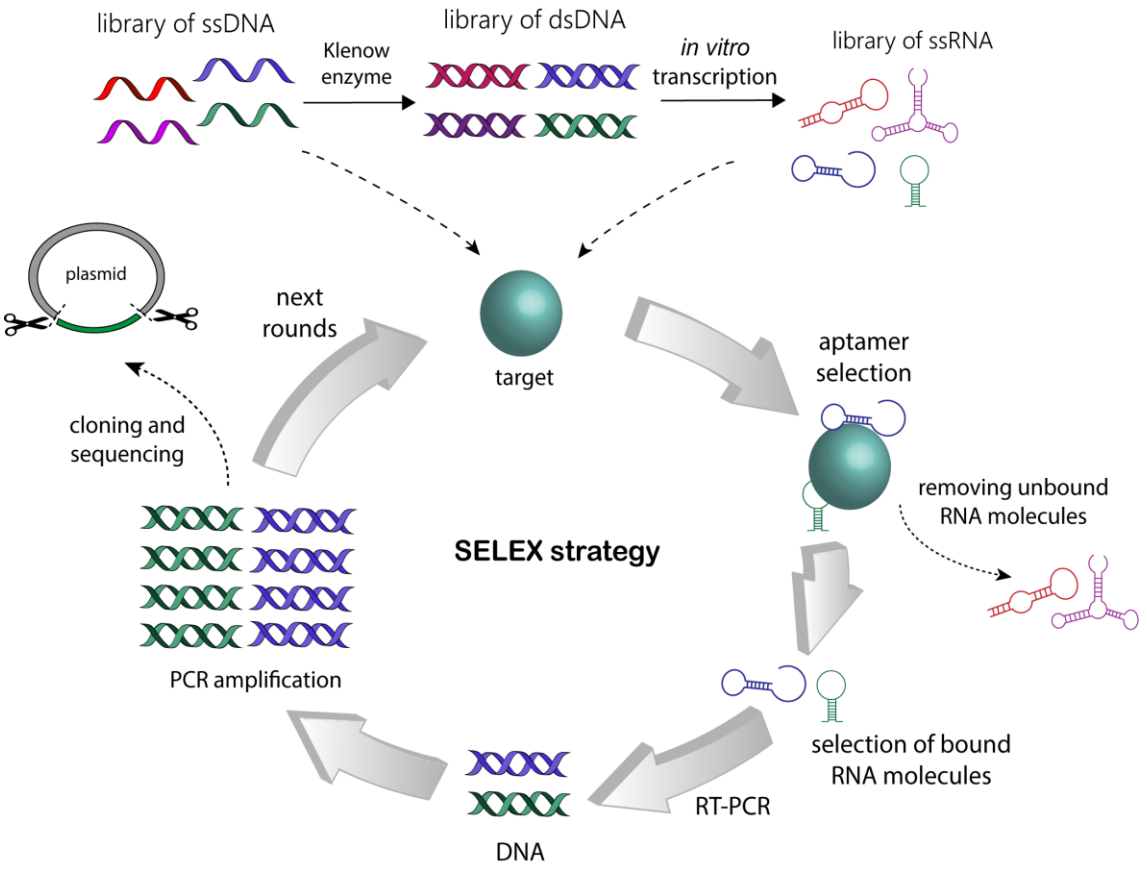


Figura 2. Representación esquemática de la estrategia SELEX (Ospina-Villa *et al*., 2016)

Los aptámeros también son conocidos como anticuerpos químicos ya que tienen afinidad por un blanco específico al igual que los anticuerpos. Sin embargo, presentan varias ventajas sobre los anticuerpos que comúnmente se usan para identificar o inhibir una proteína. Los aptámeros pueden ser seleccionados contra casi cualquier tipo de blancos ya sea pequeñas moléculas, toxinas, péptidos, proteínas, virus, bacterias e incluso células completas. De hecho, el blanco de los aptámeros no tiene que ser una molécula que presente algún grado de inmunogenicidad. Los aptámeros pueden obtenerse *in vitro* químicamente sin necesidad de usar modelos animales, lo que hace que su producción sea más rápida, sencilla, y económica. Su estabilidad térmica es mucho mayor que los anticuerpos, y los aptámeros pueden ser modificados químicamente para mejorar su resistencia a nucleasas y evitar su eliminación y degradación en el ambiente celular (Keefe *et al.,* 2010; Zhou & Rossi 2017).

Debido a todo eso, existe una gran variedad de potenciales usos de los aptámeros como son: desarrollo de nuevos medicamentos y herramientas terapéuticas, entrega de fármacos (del inglés "*drug delivery*"), diagnóstico de enfermedades, bioimágenes, análisis químicos, detección de sustancias peligrosas, inspección de alimentos, entre muchos otros. Es por eso que diversos grupos de investigación han tratado de emplear la tecnología de selección de aptámeros para el tratamiento de infecciones parasitarias a través de diferentes mecanismos, por ejemplo, el bloqueo de la interacción huésped-parásito en *Plasmodium falciparum* (Bradford *et al*., 2009), y el bloqueo de proteínas extracelulares o intranucleares en *Leishmania sp.* (Ramos *et al*., 2007).

Los estudios más relevantes acerca de los aptámeros se describen en el artículo de revisión publicado por nuestro grupo de investigación que se incluye en el Capítulo 1 del presente trabajo (Ospina-Villa *et al.,* 2016). Cabe mencionar que, a la fecha, no se han reportado aptámeros que reconozcan proteínas de la amiba patógena, *Entamoeba histolytica*.

2. ANTECEDENTES

2.1. Amibiasis

La amibiasis humana es la infección causada por el parásito protozoario *Entamoeba histolytica* que ocasiona una alta morbilidad y mortalidad en países en vía de desarrollo generando que el 10% de la población se encuentre infectada con este parásito y se reportan 100.000 muertes cada año (OMS 2012). Los factores que permiten la diseminación e incremento de la transmisión incluyen las malas condiciones de salubridad, la falta de acceso al agua potable, la ignorancia, la pobreza, la desnutrición, y la sobrepoblación (Bansal *et al.,* 2006). La amibiasis se considera la tercera infección parasitaria responsable de muertes en el mundo después de la malaria y la esquistosomiasis (Haque 2007).

E. histolytica tiene dos estadios principales de vida: el quiste que es la forma infectiva y resistente, y el trofozoíto que es la forma invasiva. El quiste mide de 10-15 µm y el trofozoíto 15-20 µm. El ciclo de vida del parásito comienza con la ingestión de quistes maduros que se encuentran en alimentos, manos o agua contaminada con heces fecales. En el intestino delgado ocurre la esquistación donde se liberan los trofozoítos que viajan al intestino grueso. Estos se dividen por fisión binaria y producen quistes, los cuales son liberados a través de las heces. Generalmente los quistes pueden ser encontrados en heces formadas, mientras que los trofozoítos se encuentran en heces diarreicas. Los quistes pueden sobrevivir días e incluso semanas en el medio ambiente externo debido a la protección que le confieren sus paredes y así pueden contaminar a otro hospedero. Los trofozoítos que salen en las heces son rápidamente destruidos y si fueran ingeridos son destruidos rápidamente por el medio ambiente gástrico (Espinosa-Cantellano, M., & Martínez-Palomo 2000).

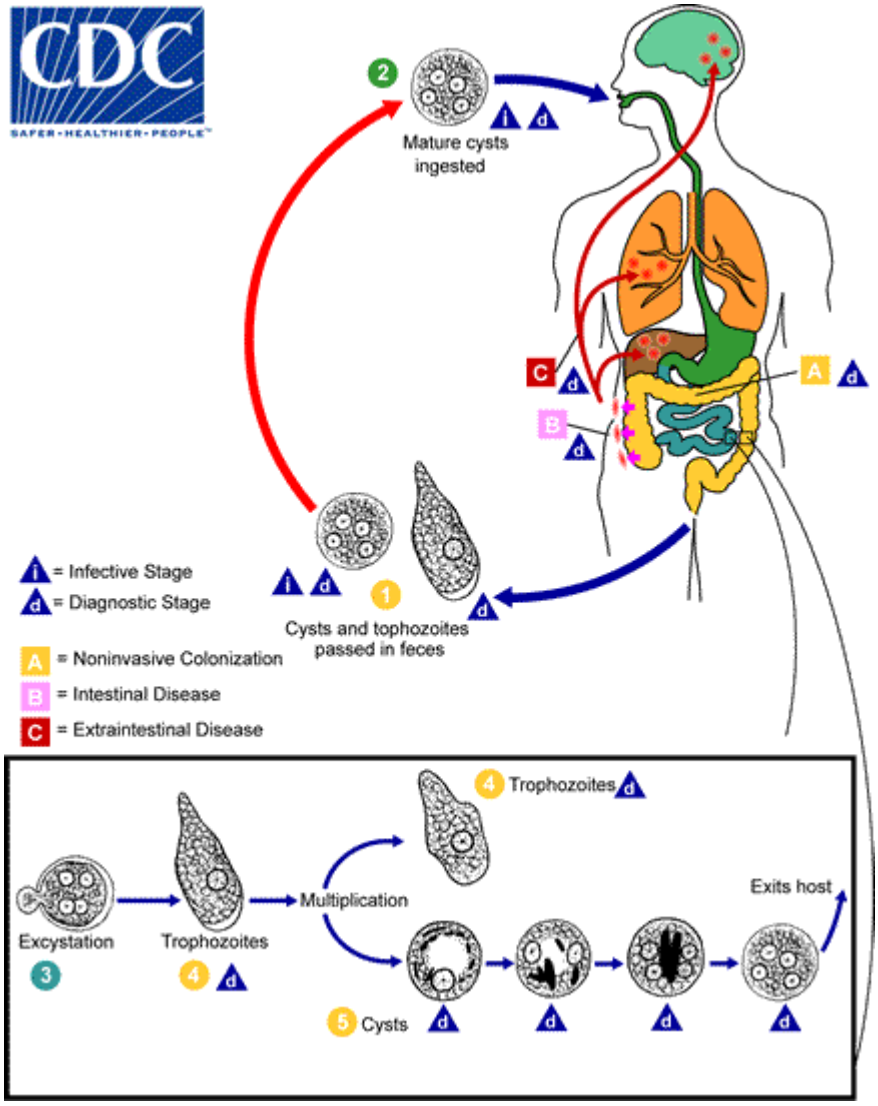


Figura 3. Ciclo de vida de *E. histolytica* (CDC 2016)

La gran mayoría de los trofozoítos permanece en el lumen intestinal produciendo una infección no invasiva. El hospedero entonces será portador con potencial para infectar a otros individuos. En la mayoría de los casos, los trofozoítos tienen la capacidad de infectar la mucosa intestinal produciendo una infección sintomática caracterizada por diarrea o disentería: En ocasiones, los trofozoítos pueden pasar la barrera intestinal y viajar a través del torrente sanguíneo para llegar hasta otros órganos como el hígado, el cerebro y los pulmones, formando abscesos que pueden ser fatales para el paciente. Aún no está claro por qué se presentan diferencias en la evolución clínica. Se tienen evidencias que estas diferencias podrían deberse al

genotipo del parásito, lo que se ha convertido ahora en el reto de la era postgenómica (Ali *et al*., 2008). También ha sido documentada la infección vía sexual a través de contacto con material fecal (CDC 2016).

La infección por *E. histolytica* ocurre a nivel mundial pero especialmente en países tropicales. Aproximadamente 480 millones de personas están infectadas (10%-12% de la población mundial) (OMS 2012). En México, la amibiasis es una de las 20 principales causas de morbilidad (González-Vásquez *et al.,* 2012). Un estudio seroepidemiológico demuestra que el 8.41% de la población mexicana tiene anticuerpos circulantes contra *E. histolytica*. Los niños mexicanos entre 5 y 9 años son los más afectados por este parásito y los casos de amibiasis intestinal se calculan entre 1 y 5 mil casos por cada 100 mil habitantes y de absceso hepático amibiano 10 casos por cada 100 mil habitantes (González-Vásquez *et al.,* 2012). Se cree que el subregistro es enorme ya que desde el año 2002 el absceso hepático no es una enfermedad de reporte obligatorio en México.

El tratamiento de elección para la amibiasis es el metronidazol descubierto en los años 60's, junto con otros derivados del nitromidazole como: tinidazol, secnidazol y ornidazol. El metronidazol se introduce entre las cadenas de ADN provocando la muerte de los trofozoítos. Pero no es efectivo para el tratamiento de los quistes. Por otro lado, el metronidazol produce una gran cantidad de efectos secundarios como son: taquicardia, dolor epigástrico, náuseas, vómitos, diarrea, mucositis oral, anorexia, angioderma, shock anafiláctico, rash, prúrito, sofocos, entre otros (https://www.drugs.com/sfx/metronidazole-side-effects.html). Además, el metronidazol es considerado como potencial carcinógeno por el *National Toxicology Program (NTP)* de los Estados Unidos (Bendesky *et al.,* 2002). Otros tratamientos alternativos de la amibiasis son las cloroquinas que actúan en las formas vegetativas del parásito inhibiendo la síntesis de ADN, y la emetina que inhibe la síntesis de proteínas. La emetina raramente es usada debido a su alta toxicidad, la cloroquina se usa junto con el metronidazol y/o emetina para el tratamiento del acceso hepático amebiano (Bansal *et al.,* 2006). Por otra parte, diferentes reportes han venido demostrando la

existencia de cepas de *E. histolytica* con resistencia a los fármacos tanto *in vitro* como *in vivo (*Wassmann *et al.,* 1999*).* Por otro lado, el mal uso de los medicamentos antiamibianos también puede generar parásitos resistentes en pacientes*.* Actualmente se sabe que los mecanismos implicados para la resistencia del parásito a los medicamentos tienen que ver con el flujo o expulsión de los mismos por medio de P-glicoproteínas (Pgp), o por ATPasas, la alteración del blanco, o por pérdida de la activación del medicamento.

Todo lo anterior indica que es necesario la búsqueda de nuevos fármacos o herramientas terapéuticas de baja toxicidad que permitan controlar la infección por *E. histolytica*. La caracterización de proteínas envueltas en la muerte celular, fagocitosis, adherencia y sobrevida del parásito continúan siendo un tema de investigación para la elaboración de nuevos medicamentos y potenciales vacunas contra este parásito (*Solis et al.*, 2009). Debido a su relevancia en la expresión génica, la poliadenilación de los transcritos y las proteínas que participan en este proceso, también pueden representar un proceso molecular interesante para la búsqueda de blancos moleculares para el desarrollo de nuevos tratamientos o vacunas contra *E. histolytica*.

2.2. Proceso de corte/poliadenilación en *E. histolytica*

Previamente en nuestro grupo de investigación, se analizó a través de herramientas bioinformáticas, el genoma del parásito *E. histolytica* para identificar las secuencias y los factores necesarios para el proceso de corte/poliadenilación de los transcritos. Las secuencias identificadas incluyen la señal de poliadenilación (UA(A/U)UU), el sitio de corte y poliadenilación que generalmente es un residuo U que es flanqueado por elementos ricos en U tanto río abajo como río arriba y un nuevo elemento rico en A que es único para *E. histolytica*. (Zamorano *et al.,* 2008). Estos hallazgos fueron posteriormente confirmados por medio de RNA-Seq por Hon *et al.* en el 2013.

En cuanto a las proteínas, se identificaron genes que codifican para los principales complejos conocidos en *Homo sapiens*, es decir CPSF, CstF, CFIm y CFIIm, así como los monómeros PAP, Ssu72, Mpe1, simplequina, Psf2 y PC4 (López-Camarillo *et al.,* 2005; López Camarillo *et al.,* 2014) (Figura 1). Estos resultados indican que el proceso de corte/poliadenilación es altamente conservado ya que no existe una gran diferencia entre las secuencias y las proteínas necesarias para llevar a cabo este proceso entre *Homo sapiens* y *E. histolytica*.

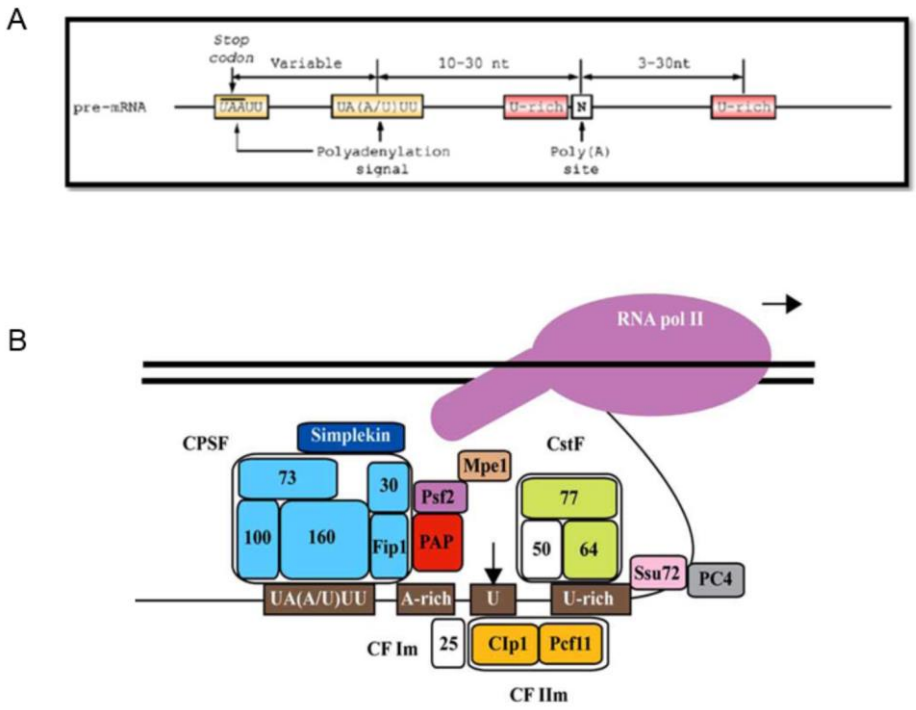


Figura 4. Maquinaria de corte/poliadenilación de *E. histolytica* (López Camarillo *et al.,* 2005) A. Secuencias en cis conservadas en los 3'UTR de los ARNm. B. Complejos proteicos y monómeros conservados en *E. histolytica.*

Hablando específicamente del complejo CFIm, en *H. sapiens* mencionamos anteriormente que existen cuatro subunidades de 25, 59, 68 y 72 kDa y que son necesarias por lo menos dos subunidades para formar un complejo heterotetramérico funcional, mientras que en *E. histolytica* solo existe la subunidad de 25 kDa. La ausencia de subunidades de alto peso molecular para el complejo CFIm en amiba sugiere que la poliadenilación de los ARNm

tiene un mecanismo particular en este parásito, en el cual EhCFIm25 podría tener un rol central.

2.3. Proteína EhCFIm25

El gen *EhCFIm25* reportado en el locus EHI_077110 en la base de datos Amoeba (http://amoebadb.org/amoeba/) cuenta con 768 nt sin intrones. Codifica para una proteína de 256 a.a identificada con el código C4M2T1 en el *Protein Data Bank* que tiene una identidad de 27-35% con proteínas homólogas en diversos organismos en la escala evolutiva, desde plantas hasta humano, incluyendo otros parásitos protozoarios. Como la proteína humana, EhCFIm25 es una proteína Nudix no convencional ya que la caja Nudix carece de tres de los cuatro glutamatos que son reemplazados por dos residuos de lisina y uno de serina. Estos cambios, al igual que en la proteína humana, podrían afectar su función como hidrolasa, pero no cambian el típico plegamiento α/β/α de las proteínas Nudix (Pezet-Valdez *et al.,* 2013). Por otra parte, al comparar las secuencias de amino ácidos de las proteínas CFIm25 de amiba y de humano, se observan otros residuos conservados, como la lisina en las posiciones 45 y 26 respectivamente, cuya acetilación modula la interacción entre CFIm25 y PAP en humano (Shimazu *et al.,* 2007), así como la tirosina ubicada en la posición 62 y 43, respectivamente, que es fosforilada en la proteína humana (Rush *et al.,* 2005).

Al igual que la proteína humana, EhCFIm25 tiene la capacidad de unirse al ARN *in vitro.* Específicamente, estudios de interacción RNA-proteína muestran que la proteína EhCFIm25 se une al 3'UTR del gen *EhPgp5* usado como modelo. Este fragmento de ARN contiene las secuencias en *–cis* previamente reportadas en este parásito, es decir dos sitios de corte, dos sitios de poliadenilación y dos sitios ricos en U. Notablemente, el sitio UGUA de reconocimiento para la proteína humana CFIm25 no se encuentra en este fragmento, lo que sugiere que el motivo reconocido por EhCFIm25 en el ARN es diferente al reconocido por la proteína humana CFIm25 (Pezet-Valdez *et al.,* 2013).

Por otro lado, la comparación de la secuencia de amino ácidos de proteínas CFIm25 de un gran número de organismos reveló que dos de los residuos que son importantes para la interacción de la proteína humana con el motivo UGUA están altamente conservados, específicamente Leu106 y Tyr208 en humano (Yang *et al.,* 2010). Estos residuos que corresponden a Leu135 y Tyr236 en EhCFIm25, también tienen un rol fundamental en la interacción de la proteína amebiana con el ARN ya que las mutaciones L135A, Y236A y Y236F inhiben totalmente la capacidad de la proteína EhCFIm25 de interactuar con el ARN, mientras que existe una reducción significativa (70%) de dicha interacción para la proteína mutante EhCFIm25*L135T. Leu135 está localizada dentro de una cavidad con carga negativa formada por aminoácidos polares que forman un ambiente que evita la entrada de agua y fortalece las interacciones débiles entre la proteína y el ARN. Tyr236 se encuentra en la parte externa de la proteína y podría contribuir a la formación de puentes de hidrógeno con el ARN para promover el posicionamiento y aumentar la unión proteína-ARN. El grupo OH del anillo fenólico de Tyr236 parece ser esencial para unirse al ARN ya que al sustituir Tyr236Phe suprime totalmente la unión proteína-ARN. El cambio Leu135Thr no suprime totalmente la unión al ARN lo que sugiere que Leu135 no es esencial para la interacción directa con el ARN sino para el mantenimiento de la estructura 3D. No podemos descartar que otros residuos no estudiados todavía pudieran estar participando en la interacción proteína-ARN, como por ejemplo los residuos Glu55, Arg63 y Phe103 de la proteína humana (Ospina-Villa *et al.,* 2015).

Finalmente, a través de ensayos de Far-Western y Pull-Down, se pudo determinar que EhCFIm25 tiene la capacidad de interactuar con EhPAP, la proteína responsable de la síntesis de la cola de A (Pezet-Valdéz, et al., 2013). Además, nuestros resultados preliminares indican que interactúa con el coactivador positivo 4 (EhPC4) que ha sido relacionado con la virulencia, replicación del ADN y multinucleación en *E. histolytica* (Hernández de la Cruz *et al.,* 2014; Hernández de la Cruz *et al.,* 2016). Estos datos muestran que EhCFIm25 no solo juega un papel importante en el proceso de corte/poliA sino que también podría estar participando en otros procesos por medio de

su interacción con otras proteínas, por lo que esta proteína podría representar un blanco molecular interesante para el control del parásito.

JUSTIFICACIÓN

La reacción de corte y poliadenilación de los pre-ARNm es esencial para que se lleve a cabo de forma adecuada la expresión génica, y su alteración ocasiona defectos letales en organismos eucariotas.

En humano, la subunidad de 25 kDa del Factor de Corte heterotetramérico CFIm, es clave para reacción de corte/poliadenilación de los transcritos; además, posee una amplia red de interacciones con proteínas de otras maquinarias de procesamiento del ARNm. Por lo que juega un papel esencial en la regulación de la expresión génica.

En Entamoeba histolytica, el protozoario responsable de la amibiasis humana, solo existe la subunidad de 25 kDa del CFIm. La ausencia de las subunidades de alto peso molecular sugiere que EhCFIm25 podría tener un rol particularmente importante en la poliadenilación de los ARNm en este parásito.

Por lo tanto, la caracterización molecular y funcional de la proteína CFIm25 de *E. histolytica* nos permitirá entender cómo se lleva a cabo el proceso de corte/poliA en este parásito y potencialmente identificar la trascendencia de esta proteína en este proceso.

OBJETIVOS

Objetivo General

Caracterizar molecular y funcionalmente la proteína EhCFIm25 para determinar su trascendencia en el proceso de corte /poliA y su pertinencia como posible blanco bioquímico en la amibiasis.

Objetivos Específicos

1. Evaluar la relevancia de la proteína EhCFIm25 en *E. histolytica*
2. Identificar el motivo de RNA al cual se une la proteína EhCFIm25
3. Caracterizar el efecto del secuestro de EhCFIm25 con aptámeros específicos sobre la viabilidad celular del parásito
4. Determinar la relación estructura-actividad en la proteína EhCFIm25

ESTRATEGIA EXPERIMENTAL

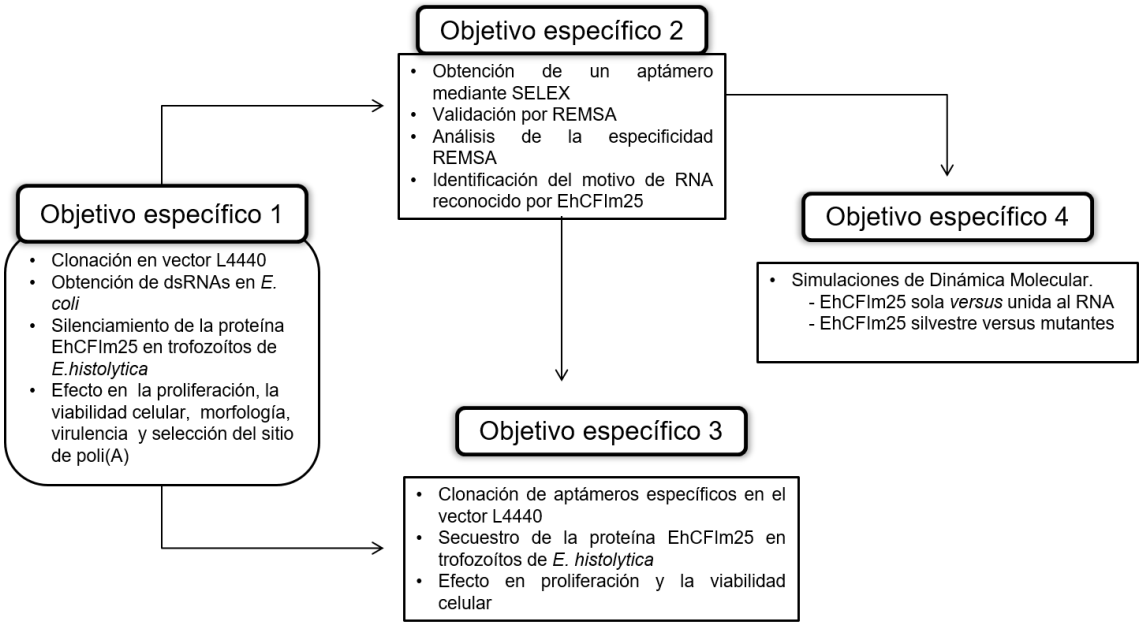


Figura 5. Estrategia experimental

RESULTADOS

CAPÍTULO 1:

Aptamers as a promising approach for the control of parasitic diseases.

Ospina-Villa, J.D., Zamorano-Carrillo, A., Castañón-Sánchez, C.A., Ramírez-Moreno, E., Marchat, L.A.

Braz J Infect Dis. 2016 ;20(6):610-618. DOI: 10.1016/j.bjid.2016.08.011.

BRAZ J INFECT DIS 2016;20(6):610–618

The Brazilian Journal of
INFECTIOUS DISEASES

www.elsevier.com/locate/bjid

Review article

Aptamers as a promising approach for the control of parasitic diseases

Juan David Ospina-Villa[a], *Absalom Zamorano-Carrillo*[a], *Carlos A. Castañón-Sánchez*[b], *Esther Ramírez-Moreno*[a], *Laurence A. Marchat*[a,*]

[a] *Instituto Politécnico Nacional, Escuela Nacional de Medicina y Homeopatía, Ciudad de Mexico, Mexico*
[b] *Hospital Regional de Alta Especialidad de Oaxaca, Subdirección de Enseñanza e Investigación, Oaxaca, Mexico*

ARTICLE INFO

Article history:
Received 14 June 2016
Accepted 16 August 2016
Available online 15 October 2016

Keywords:
Aptamer
Parasite control
Protozoan parasite
SELEX strategy

ABSTRACT

Aptamers are short single-stranded RNA or DNA oligonucleotides that are capable of binding various biological targets with high affinity and specificity. Their identification initially relies on a molecular process named SELEX (Systematic Evolution of Ligands by EXponential enrichment) that has been later modified in order to improve aptamer sensitivity, minimize duration and cost of the assay, as well as increase target types. Several biochemical modifications can help to enhance aptamer stability without affecting significantly target interaction. As a result, aptamers have generated a large interest as promising tools to compete with monoclonal antibodies for detection and inhibition of specific markers of human diseases. One aptamer-based drug is currently authorized and several others are being clinically evaluated. Despite advances in the knowledge of parasite biology and host–parasite interactions from "omics" data, protozoan parasites still affect millions of people around the world and there is an urgent need for drug target discovery and novel therapeutic concepts. In this context, aptamers represent promising tools for pathogen identification and control. Recent studies have reported the identification of "aptasensors" for parasite diagnosis, and "intramers" targeting intracellular proteins. Here we discuss various strategies that have been employed for intracellular expression of aptamers and expansion of their possible application, and propose that they may be suitable for the clinical use of aptamers in parasitic infections.

Aptamers and strategies for their identification

Aptamers are DNA or RNA oligonucleotides with a unique tridimensional structure that allows them for interacting with a specific target with high affinity and specificity. The term "aptamer" is derived from the Latin word "*aptus*" meaning "to fit" and a Greek word "*mers*" meaning "particle".[1] Although aptamers were first described as artificial molecules, they were later found as natural components of riboswitches that affect transcription or translation.[2,3] The term aptamer can also design a peptide with protein-binding properties;

* *Corresponding author.*
E-mail addresses: lmarchat@gmail.com, lmarchat@ipn.mx (L.A. Marchat).
http://dx.doi.org/10.1016/j.bjid.2016.08.011

CAPÍTULO 2:

Silencing the cleavage factor CFIm25 as a new strategy to control *Entamoeba histolytica* parasite.

Juan David Ospina-Villa, Nancy Guillén, Cesar López-Camarillo, Jacqueline Soto-Sánchez, Esther Ramírez-Moreno, Raul Garcia-Vazquez, Carlos A. Castañon-Sanchez, Abigail Betanzos, Laurence A.Marchat.

J Microbiol. **2017**, 55 (10). DOI 10.1007/s12275-017-7259-9

Journal of Microbiology (2017) Vol. 55, No. 10, pp.
DOI 10.1007/s12275-017-7259-9

eISSN 1976-3794
pISSN 1225-8873

Silencing the cleavage factor CFIm25 as a new strategy to control *Entamoeba histolytica* parasite[§]

Juan David Ospina-Villa[1], Nancy Guillén[2], Cesar Lopez-Camarillo[3], Jacqueline Soto-Sanchez[1], Esther Ramirez-Moreno[1], Raul Garcia-Vazquez[1], Carlos A. Castañon-Sanchez[4], Abigail Betanzos[5], Laurence A.Marchat[1]*

[1]*Instituto Politécnico Nacional – ENMH, Ciudad de México, Mexico*
[2]*Institut Pasteur, Unité d'Analyses d'Images Biologiques, Paris, France*
[3]*Universidad Autónoma de la Ciudad de México – Posgrado en Ciencias Genómicas, Ciudad de México, Mexico*
[4]*Hospital Regional de Alta Especialidad, Oaxaca, Mexico*
[5]*Cátedras, CONACYT, Departamento de Infectómica y Patogénesis Molecular, CINVESTAV-IPN, Ciudad de México, Mexico*

(Received Jun 27, 2017 / Revised Aug 16, 2017 / Accepted Aug 19, 2017)

The 25 kDa subunit of the Clevage Factor Im (CFIm25) is an essential factor for messenger RNA polyadenylation in human cells. Therefore, here we investigated whether the homologous protein of *Entamoeba histolytica*, the protozoan responsible for human amoebiasis, might be considered as a biochemical target for parasite control. Trophozoites were cultured with bacterial double-stranded RNA molecules targeting the *EhCFIm25* gene, and inhibition of mRNA and protein expression was confirmed by RT-PCR and Western blot assays, respectively. *EhCFIm25* silencing was associated with a significant acceleration of cell proliferation and cell death. Moreover, trophozoites appeared as larger and multinucleated cells. These morphological changes were accompanied by a reduced mobility, and erythrophagocytosis was significantly diminished. Lastly, the knockdown of *EhCFIm25* affected the poly(A) site selection in two reporter genes and revealed that *EhCFIm25* stimulates the utilization of downstream poly(A) sites in *E. histolytica* mRNA. Overall, our data confirm that targeting the polyadenylation process represents an interesting strategy for controlling parasites, including *E. histolytica*. To our best knowledge, the present study is the first to have revealed the relevance of the cleavage factor CFIm25 as a biochemical target in parasites.

Keywords: amoebiasis, gene knockdown, polyadenylation, protozoan parasite, virulence

*For correspondence. E-mail: lmarchat@gmail.com; lmarchat@ipn.mx; Tel.: +52-55 5729-6300 (ext. 55543)
[§]Supplemental material for this article may be found at http://www.springerlink.com/content/120956.

Introduction

The polyadenylation of pre-messenger mRNA (pre-mRNA) at the 3′-end is a fundamental process for gene expression regulation in eukaryotic cells; it establishes an important connection with transcription (Batt *et al.*, 1994), confers stability to mRNA (Barnhart *et al.*, 2013), plays a role in mRNA nuclear export, streamlines translation (Colgan and Manley 2016), and protects mRNA from degradation (Tourrière *et al.*, 2002). Alterations in polyadenylation have been related with several human illnesses like α and β thalassemia, neonatal diabetes, Fabry disease, and cancer (Curinha *et al.*, 2014), as well as with lethal defects in yeast (Wang *et al.*, 2005), demonstrating that this event is essential for accurate cell survival. The poly(A) tail formation requires the participation of protein complexes known as CPSF (Cleavage and Polyadenylation Specific Factor), CstF (Cleavage Stimulating Factor), CFIm and CFIIm (Cleavage Factor Im and IIm), as well as PAP (Poly(A)Polymerase) and PABPII (Poly(A) Binding Protein), that bind to specific motifs in 3′-UTR. In human cells, the CFIm complex is a heterotetrameric complex formed by a homodimer of 25 kDa subunits interacting with two larger subunits (72, 65, or 59 kDa). Each CFIm25 subunit (also known as CPSF5 or NUDT21) binds to the UGUA motif and affinity is increased by interaction with the RRM domain of larger subunits (Yang *et al.*, 2011). The knockdown of CFIm25 protein affects the poly(A) site selection (Kubo *et al.*, 2006), the recruitment of polyadenylation factors to pre-mRNA 3′-end, and the cleavage and polyadenylation reactions (Brown and Gilmartin, 2003), which highlights the relevance of this subunit in poly(A) tail synthesis. Our search in the Nnified Human Interactome database at http://www.unihi.org/ (Kalathur *et al.*, 2013) showed that CFIm25 interacts with splicing factors (U2AF1, SF3B1, SNRNP70, and others), and export (NXF1) and transcription factors (GTF2F1, HSF4, TCERG1). Some of these interactions have been previously reported (De Vries *et al.*, 2000; Awasthi and Alwine, 2003; Ingham *et al.*, 2005; Vinayagam *et al.*, 2011), confirming that CFIm25 establishes a functional link between the different molecular events of mRNA synthesis and processing.

We previously reported the polyadenylation machinery of *Entamoeba histolytica*, the protozoan parasite responsible for the human amoebiasis that affects 50 million individuals per year (Ralston and Petri, 2011). By extensive analyses of parasite genome sequences, we described conserved motifs in pre-mRNA 3′-ends, namely the poly(A) signal A(U/A)UU, and the U-rich and A-rich elements (Zamorano *et al.*, 2008). We also identified polyadenylation factors that contain the functional domains described in homologous proteins in

CAPÍTULO 3:

Targeting the polyadenylation factor EhCFIm25 with RNA aptamers controls survival in *Entamoeba histolytica.*

Juan David Ospina-Villa, Alexandre Dufour, Christian Weber, Esther Ramírez-Moreno, Absalom Zamorano, Nancy Guillen, César López-Camarillo, Laurence A Marchat.

Enviado a Scientific Reports

Targeting the polyadenylation factor EhCFIm25 with RNA aptamers controls survival in *Entamoeba histolytica*

Juan David Ospina-Villa[1], Alexandre Dufour[2,3], Christian Weber[3,4], Esther Ramirez-Moreno[1], Absalom Zamorano-Carrillo[1], Nancy Guillen[5], César Lopez-Camarillo[6], Laurence A Marchat[1,*]

[1]Instituto Politécnico Nacional, Escuela Nacional de Medicina y Homeopatía,Guillermo Massieu Helguera 239, Fracc. La Escalera Ticoman, CP 07320, Ciudad de México, México.
[2]Institut Pasteur, Unité d'Analyse d'Images Biologiques, 25 Rue du Dr Roux F-75015 Paris, France.
[3]Centre National de la Recherche Scientifique CNRS UMR 3691, 25 Rue du Dr Roux F-75015 Paris, France
[4]Institut Pasteur, Unité d'Imagerie et Modélisation, 28 rue du Docteur Roux, 75015 Paris, France
[5]Centre National de la Recherche Scientifique, CNRS-ERL9195, 25 Rue du Dr Roux, F-75015 Paris, France.
[6]Universidad Autónoma de la Ciudad de México, Posgrado en Ciencias Genómicas, San Lorenzo 290, Col. Del Valle, CP 03100, Ciudad de México, México.

Correspondence
L.A. Marchat. Sección de Estudios de Posgrado e Investigación, ENMH, Instituto Politécnico Nacional, Guillermo Massieu Helguera 239, Fracc. La Escalera, CP 07320, Mexico City, Mexico.
Telephone number: (52-55) 5729-6300 ext. 55543; Email: lmarchat@gmail.com

ABSTRACT

Messenger RNA 3´-end polyadenylation is an important regulator of gene expression in eukaryotic cells. In our search for new ways of treating parasitic infectious diseases, we looked at whether or not alterations in polyadenylation might control the survival of *Entamoeba histolytica* (the agent of amoebiasis in humans). We used molecular biology and computational tools to characterize the mRNA cleavage factor EhCFIm25, which is essential for polyadenylation in *E. histolytica*. By using a strategy based on the systematic evolution of ligands by exponential enrichment, we identified single-stranded RNA aptamers that target EhCFIm25. The results of RNA-protein binding assays showed that EhCFIm25 binds to the GUUG motif *in vitro*, which differs from the UGUA motif bound by the homologous human protein. Accordingly, docking experiments and molecular dynamic simulations confirmed that interaction with GUUG stabilizes EhCFIm25. Incubating *E. histolytica* trophozoites with selected aptamers inhibited parasite proliferation and rapidly led to cell death. Overall, our data indicate that targeting EhCFIm25 is an effective way of limiting the growth of *E. histolytica in vitro*. The present study is the first to have highlighted the potential value of RNA aptamers for controlling this human pathogen.

INTRODUCTION

The processing of pre-messenger RNA (pre-mRNA) at the 3´-untranslated region (3´-UTR) is an essential maturation step; it increases mRNA stability, facilitates export of pre-mRNA from the nucleus to the cytoplasm, and enhances mRNA translation efficiency[1-3]. The key steps in this essential event in the life of an RNA consist of pre-mRNA 3´-end cleavage and then polyadenylation. These reactions are determined by a multistep mechanism in which specific RNA sequence motifs are recognized by multiprotein complexes, including cleavage and polyadenylation specificity factor (CPSF), cleavage stimulating factor (CstF), and cleavage factors Im and IIm (CFIm and CFIIm)[4]. CPSF, CstF and CF subunits are conserved among all eukaryotic organisms, including several unicellular parasites[5,6]. We focused on *Entamoeba histolytica*. This parasite is the causative agent of human amoebiasis, an infectious disease that still represents a major health problem in many developing countries worldwide. The symptoms of human amoebiasis range from asymptomatic infection, diarrhoea, dysentery, fulminant colitis and peritonitis to the development of potentially lethal extraintestinal abscesses (mainly in the liver)[7,8]. Moreover, the severe adverse drug reactions associated with currently available pharmacological treatments (metronidazole and other nitroimidazole compounds) and the decrease in *E. histolytica* drug susceptibility have created an urgent need for alternative, novel, specific treatments [9,10].

The pathogenicity of *E. histolytica* has been linked to both host factors and genomic and transcriptomic factors in the parasite [11,12]. However, little is currently known about how gene expression is regulated in this pathogen. Nevertheless, DNA/RNA motifs and nuclear factors involved in transcription, splicing, and mRNA 3´ end processing have been described[5,13,14]. In particular, protein amino-acid (aa) sequence comparisons suggest that the amoeba's pre-mRNA 3´-end processing machinery is in an intermediate evolutionary position between mammals and yeast. Furthermore, the presence of non-canonical poly(A) polymerases adds complexity to the mRNA 3' end formation process in this single-celled eukaryote[5]. *E. histolytica* only has the 25 kDa subunit of CFIm (EhCFIm25)[5], whereas active CFIm in humans is a heterotetramer complex comprising two 25 kDa subunits that interact with a dimer of 59 or 68 kDa subunits[15-17]. CFIm25 belongs to the Nudix hydrolase superfamily, and is an

essential regulator of poly(A) site selection, and polyadenylation/cleavage reactions in eukaryotic cells[18,19,20]. The absence of higher molecular mass subunits in *E. histolytica*'s CFIm suggests that (i) mRNA polyadenylation has a particular mechanism in the parasite, and (ii) EhCFIm25 has central role[5]. Although EhCFIm25 conserves the characteristic features of its human orthologue, it possesses several important differences. Notably, three of the four glutamate residues at positions 154, 157 and 158 in the conserved Nudix box are replaced by lysine, and the last glycine residue of the motif is replaced by the hydrophilic residue serine. EhCFIm25 interacts with the poly(A) polymerase EhPAP[21] and the transcriptional coactivator EhPC4 (our unpublished data), which is related to virulence, DNA replication and multinucleation in *E. histolytica*[22,23]. EhCFIm25 also interacts with mRNA 3'UTR through the conserved Leu135 and Tyr236 residues[24], although the RNA binding motif has yet to be identified.

It has been demonstrated that several components of the polyadenylation machinery may be valuable therapeutic targets in protozoan parasite (namely CPSF-30 (CPSF4) in *Trypanosoma brucei* and CPSF-73 (CPSF3) in *Toxoplama gondii* and *Plasmodium* falciparum)[25-28]. Furthermore, we recently observed that EhCFIm25 silencing induces cell death and decreases virulence capacity (our unpublished data) – prompting us to hypothesize that EhCFIm25 may be a relevant target for *E. histolytica* control. To test this hypothesis, we applied two powerful molecular biology approaches: (i) the development of aptamers that bind to EhCFIm25 and (ii) delivery aptamers to trophozoites via soaking. Aptamers are single-stranded (ss) RNA or DNA oligonucleotides whose unique three-dimensional structure enables them to interact with a specific target molecule (*i.e.* EhCFIm25 in the present case). A large body of biomedical research has shown that aptamers are better than other tools (e.g. antibodies) at detecting and inhibiting target molecules in diagnostics, therapeutics, and drug development[29]. In the present work, we used a systematic evolution of ligands by exponential enrichment (SELEX) protocol to identify RNA aptamers that target the EhCFIm25 protein. Thanks to RNA-protein binding assays and molecular modelling, we discovered that EhCFIm25 bound to the aptamers' GUUG motif. Moreover, we demonstrated *in vitro* that ingestion of these aptamers dramatically inhibited *E. histolytica*'s growth.

METHODS

Cell cultures

E. histolytica trophozoites (HMI:IMSS) were grown at 37 °C in TYI-S-33 medium with 20% bovine serum, 100 U/ml penicillin and 100 µg/ml streptomycin[30]. RNAse III-deficient *Escherichia coli* strain HT115 (rnc14::DTn10) was grown at 37°C in LB broth for plasmid construction or 2YT broth for double-stranded (ds) RNA expression, in the presence of ampicillin (100 mg/ml) and tetracycline (10 mg/ml)[31].

Expression and purification of the recombinant EhCFIm25 protein

Competent *E. coli* BL21 (DE3) pLysS bacteria were transformed with the pRSET-*EhCFIm25* plasmid[21]. The EhCFIm25 was expressed with 1 mM isopropyl beta-D-thiogalactopyranoside (IPTG) and purified by Ni^{2+}-NTA affinity chromatography (Qiagen). The identity and integrity of the histidine-tagged EhCFIm25 protein was confirmed by 10% SDS-PAGE and Western blot assays using anti-6x-His tag antibodies (Roche) at 1:10000 dilution and the ECL Plus Western blotting detection system (Amersham) .

SELEX strategy

RNA aptamers targeting EhCFIm25 were obtained using the SELEX protocol with some modifications[32]. First, we designed library primers with a region of 20 random nucleotides flanked by two conserved sequences (5′-TTACAGCAACCACCGGGGATCCATGGGCACTATTTATATCAAC$(N)_{20}$AATGTCGTTGGTGGCCC-3′), a forward primer with a *Eco*RI site, the T7 promotor region and a complementary region to the conserved 3´ end of library primers, and a reverse primer with a *Bam*HI site and a complementary region to the conserved 5´ end of library primers (Fig. 1A). Then, library and forward primers (100 µM) were mixed with 10 mM dNTPs and 5 U Klenow enzyme (NEB) in Klenow buffer, for 1 h at 37 °C to obtain dsDNA. The ssRNA fragments potentially corresponding to about 4^{20} sequences were *in vitro* transcribed using 1U T7 RNA polymerase (NEB), purified with PCA (Phenol:Chloroform:Isoamyl Alcohol 25:24:1) and quantified. Finally, they were passed throughout a Ni^{2+}-NTA column previously coated with the histidine-tagged EhCFIm25 protein.

After washing, bound aptamers were eluted with NT2 buffer, purified with PCA and incubated with the reverse primer and the M-MLV Reverse Transcriptase enzyme (Thermo Fisher Scientific). The resulting cDNA fragments were PCR-amplified using forward and reverse primers with the Platinum® Taq DNA Polymerase High Fidelity (Thermo Fisher Scientific) as follows: 94 °C for 3 min; 30 cycles at 94 °C for 1 min, 40 °C for 1 min, 72 °C for 1 min; plus a final extension step at 72 °C for 20 min. Then, ssRNA were *in vitro* transcribed as described above to perform a new round of selection (R) (Supplementary Figure S1). Seven rounds of selection (R7) were performed to select aptamers with affinity for the EhCFIm25 protein. Finally, PCR products were cloned into the *Eco*RI and *Bam*HI sites of the pRSET A plasmid and sequenced using an Applied Biosystems 3500 Genetic Analyzer in the Unit of Molecular Biology at UNAM-Mexico.

RNA-Electrophoretic Mobility Shift Assays (REMSA)

Selected aptamers (C4 and C5), and the 3´UTR of *thioredoxin* (EHI_021560), and *60SRibL7* (EHI_025830) genes of *E. histolytica*, were *in vitro* transcribed from pGEMT-*Ehthio,* pGEM-*Ehrib* plasmids (GenScript), pRSET-*C4* and pRSET-*C5* plasmids, respectively, using the MEGAshortscript T7 Kit (Ambion). These molecules and RNA molecules selected from the seventh round of the SELEX protocol (R7), were labeled with the Biotin RNA Labeling Mix (ROCHE). Their size and integrity were verified by agarose gel electrophoresis and chemiluminescence (Chemiluminescent Nucleic Acid Detection Module, Pierce). Then, RNA probes (100 ng/µl) were mixed with EhCFIm25 (20 µg) or *E. histolytica* extracts (50 µg) at room temperature for 20 min; RNA-protein complexes were resolved at 100 V for 1 h on pre-electrophoresed 10% non-denaturing PAGE and detected using LightShift™ Chemiluminescent RNA EMSA Kit (Thermo Scientific). In some assays, proteinase K (20 U), tRNA (0.1 mg/mL), RNA fragments of the first round of SELEX (R0), the mutant protein (EhCFIm25*L135T)[24] and protein extracts from HeLa cells or *Trypanosoma cruzi* parasites were used as controls.

Modeling and docking experiments

The secondary structure of C4 and C5 aptamers was predicted using the Unified Nucleic Acid Folding and hybridization package (http://unafold.rna.albany.edu/)[33,34]. The three-dimensional structure of EhCFIm25 (C4M2T1, 255 residues) was predicted by homology modeling with the MODELLER package[35] (http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/) using as template the crystal structure of the human CFIm25 protein (chain A) in complex with the UUGUAU RNA molecule (chain C) (PDB 3MDG), and validated with the Verify_3D software[36] (http://services.mbi.ucla.edu/Verify_3D/). On the other hand, we used the RNAComposer online software (http://rnacomposer.cs.put.poznan.pl/) to obtain a PDB structure of the GUUG RNA motif. Then, both molecules were individually submitted to molecular dynamics (MD) simulations as described below to obtain the most relaxed structures. Finally, we used the NPDock (Nucleic acid-Protein Dock) web server (http://genesilico.pl/NPDock/) for modeling the GUUG-EhCFIm25 complex structure using default parameters[37]. The best scoring model was chosen to analyze contacts between EhCFIm25 residues and the GUUG molecule by the CMA (Contact Map Analysis) software (http://ligin.weizmann.ac.il/cma/). For some analyses, we used the Swiss-PdbViewer 4.10 software (http://www.expasy.org/spdbv/)[38] to obtain the mutant EhCFIm25*L135T protein.

Molecular dynamics simulation

MD simulations of EhCFIm25 protein alone or interacting with the GUUG motif were performed using the GROMACS 5.1 package and the CHARMM27 v2.0 force field for proteins[39,40]. Molecules were centered in a cubic box at 1.0 nm from edges, using periodic boundary conditions and three points equilibrated solvent model. Cl^- or Na^+ ions were added to neutralize the system that was equilibrated under a canonical or NVT ensemble to stabilize the temperature, and then under a NPT ensemble to equilibrate the pressure using the Parrinello-Rahman barostat[41]. Finally, we performed the position restraints and the MD production at 300 °K running a first step of 1 ns simulation followed by an extension step of 50 ns.

Analysis of MD trajectories

First, we used the Trjconv tool of GROMACS to correct any periodicity in the system. After that, the atomic characteristics of EhCFIm25 protein alone or with the GUUG motif were compared using the analysis tools included in GROMACS software. RMSD (root mean square deviation) and RMSF (root mean square fluctuations) values of Cα backbone were calculated. The electrostatic potential was determined using the PDB2PQR server and the APBS software package[42] (http://nbcr-222.ucsd.edu/pdb2pqr_2.1.1/). The simulation trajectory was visualized in the VMD 1.9.3 *beta1* software[43] (http://www.ks.uiuc.edu/Research/vmd/).

Blocking of EhCFIm25 by C4 and C5 aptamers

To deliver aptamers into *E. histolytica* trophozoites, we used bacterially expressed dsRNA and parasite-soaking experiments as described[44] Briefly, DNA sequences corresponding to C4 and C5 aptamers were PCR amplified from pRSET-*C4* and pRSET-*C5* plasmids, respectively, and cloned into the *Sma*I and *Xho*I sites of the pL4440 vector (Fig. 5A). DNA sequencing was performed to verify the resulting pL4440-*C4* and pL4440-*C5* plasmids. Then, competent *E. coli* HT115 cells were independently transformed with each plasmid and dsRNA synthesis was induced with 2 mM IPTG for 4 h at 37 °C. Bacterial pellet was mixed with 1 M ammonium acetate and 10 mM EDTA, incubated with phenol:chloroform:isoamyl alcohol (25:24:1) and centrifuged. Nucleic acids were washed with isopropanol and 70% ethanol. DNase I (Invitrogen) and RNase A (Ambion) were added to eliminate ssRNA and dsDNA molecules, *C4*-dsRNA and *C5*-dsRNA were washed with isopropanol and 70% ethanol, analyzed by agarose gel electrophoresis and quantified. Finally, purified dsRNA molecules (100 µg/ml) were added to trophozoites cultures (5.0×10^4) at 37 °C. Parasites grown in standard conditions or incubated with the *gfp*-dsRNA (green-fluorescent protein) were used as controls.

Cell proliferation and viability assays

Each day, *E. histolytica* trophozoites were counted in a Neubauer chamber to evaluate cell proliferation. Simultaneously, living trophozoites were identified

from the Trypan blue dye exclusion test. Experiments were performed twice in duplicate.

dsRNAs stability

Purified *C4*-dsRNA or *C5*-dsRNA were mixed with complete TYI-S medium (100 µg/ml) at 37 °C. Each day, dsRNA molecules were resolved through 1% agarose gel electrophoresis and GelRed staining to evaluate their integrity.

RESULTS

Identification and purification of RNA aptamers targeting EhCFIm25

Aptamers have been used to identify the RNA binding site in human CFIm25, and they inhibited the protein's functions[19]. We identified RNA aptamers against the EhCFIm25 protein by using a SELEX affinity strategy with the recombinant His-tagged protein and a random ssRNA library (Fig. 1A). The canonical steps of the SELEX process were incubation, separation, elution, amplification, and purification of the single-stranded oligonucleotide; this process was repeated for several rounds (Supplementary Figure S1). At each round of selection in the SELEX process, PCRs confirmed the recovery of aptamers (data not shown). The pool of aptamers obtained from round seven (R7) was biotin-labelled, and its ability to bind to EhCFIm25 was confirmed in an RNA-electrophoretic mobility shift assay (REMSA). The resulting RNA-protein complex disappeared when proteinase K was added but was maintained in the presence of tRNA (used as a nonspecific competitor). In contrast, an RNA-protein complex was not observed when we used the mutant EhCFIm25*L135T protein, in which the Leu135 residue had been replaced by Thr[24]. Similarly, EhCFIm25 did not form a complex with RNA fragments from SELEX R0 (Fig. 1C).

In vitro verification of isolated RNA aptamers.

Two aptamers (referred to as C4 and C5) from the last round of selection (R7) were chosen for further characterization because of their high expected specificity and affinity for the EhCFIm25 protein (Fig. 1B). The aptamers were reverse-transcribed and biotin-labelled for the REMSA. Both C4 and C5 interacted with the recombinant EhCFIm25 protein but not with the mutant EhCFIm25*L135T (Fig. 1D). Moreover, C4 and C5 were able to form a complex

with total protein extract from *E. histolytica*. The addition of anti-EhCFIm25 antibodies produced a supershift in the RNA-protein complex, indicating that C4 and C5 also bind to the endogenous EhCFIm25 protein (Fig. 1E). However, neither C4 nor C5 were able to bind to CFIm25 proteins from HeLa cells (UniProtKB/Swiss-Prot: O43809) or *T. cruzi* parasites (GenBank: EKG06513.1) used as controls, which share homology (32% and 30%, respectively) with EhCFIm25 (Fig. 1F and G).

EhCFIm25 protein binds to the GUUG motif *in vitro*

A set of 12 aptamers from R5, R6 and R7 (including C4 and C5) were cloned and sequenced. Nucleotide sequence comparison showed that all 12 aptamers comprised the GUUG motif, which may represent the RNA binding site for EhCFIm25. The fact that this motif is also present in the 3´UTR of the *EhPgp5* gene that interacts with EhCFIm25[21,24] prompted us to evaluate the relevance of this sequence in the aptamer-EhCFIm25 interaction. To this end, we performed a REMSA using the 3´UTR of two genes that carry the GUUG sequence (Fig. 2A). The *E. histolytica* genes coding for thioredoxin and 60S ribosomal protein L7 (accession number EHI_026340 and EHI_192110, respectively) were selected because RNA sequencing experiments had shown that they contain two functional alternative poly(A) sites[45]. The REMSA showed that the 3´UTR of both genes formed a complex with the endogenous protein contained in total protein extract from *E. histolytica,* as well as with the purified recombinant EhCFIm25 (Fig. 2B). To assess the relevance of the GUUG motif, we designed an oligonucleotide with a random sequence that contains the GUUG motif. As expected, EhCFIm25 formed a complex with this GUUG probe, whereas replacement of the GUUG sequence with the CAAC motif (in a CAAC probe) totally abolished EhCFIm25 binding. Moreover, the GUUG-probe acted as a specific competitor in REMSAs using the 3´UTR of the EHI_026340 and EHI_192110 genes as probes (Fig. 2C-E).

Protein and RNA modelling, molecular docking, and molecular dynamics simulations.

To gain insight into how the EhCFIm25 protein binds to the GUUG motif, we compared data generated in MD simulations of EhCFIm25 alone or bound to

the GUUG motif. We used the three-dimensional structure of the predicted RNA-protein complex with the best score in a docking assay. Results of MD simulations showed that the EhCFIm25-GUUG interaction mainly occurs through the nitrogenous base guanine 4 in the RNA, and some positively charged aa (Lys, Arg and His), some aromatic aa (Phe and Tyr), and the Leu135 residue previously known to be important for the RNA binding activity of EhCFIm25 (Fig. 3)[24]. We also compared the changes over time (during 50 ns) in the structure of free *vs.* RNA-bound EhCFIm25 at 300 K. Under both experimental conditions, the root mean square deviation (RMSD) reached equilibrium after 15 ns of simulation (Fig. 4A). The free protein presented fluctuations in the 0.3-0.4 nm range, while the GUUG-interacting protein fluctuated less (0.2-0.3 nm) - suggesting that the interaction with RNA stabilized the protein's structure as a whole. In contrast, the GUUG motif had a destabilizing effect on the mutant EhCFIm25*L135T protein, which confirms the poor RNA binding capacity observed *in vitro*[24]. Consistently, analysis of the trajectories in the MD simulations of the free and GUUG-bound proteins confirmed that EhCFIm25 is stabilized in the presence of the RNA molecule (Supplementary videos S1 and S2). Indeed, the average structure in the last 30 ns of the simulation showed that the GUUG-bound protein is more compact than the free EhCFIm25; the latter has a pair of unstable loops and disordered alpha helices (Fig. 4B).

To establish how the interaction with RNA affects the behaviour of the protein's aa, we compared the root mean square fluctuation (RMSF) of EhCFIm25 in the presence or absence of the GUUG fragment during the last 30 ns of the MD simulation (Fig. 4C). Under both conditions, the structure contained the same flexible regions (indicated by black filled arrows in the Figure). However, the GUUG-interacting protein displayed lower RMSF values in the 70-90 aa region (box). Examination of EhCFIm25's three-dimensional structure using the VMD software confirmed that the amino-terminal region and the 70-90 aa region (corresponding to an alpha helix and a loop) fluctuated the most, whereas the rest of the structure was more stable in both the bound and free proteins (Fig. 4D). In line with the RMSF data, the 70-90 aa region appeared to be stabilized by structure gain (transition from a loop to an alpha helix) in the GUUG-interacting EhCFIm25; this suggests that the 70-90 aa region is closely involved

in RNA binding (Supplementary video S1 and S2). Lastly, we calculated the electrostatic potential of EhCFIm25's surface. As shown in Fig. 4E, the RNA binding site (circle) is more electropositive in the GUUG-interacting protein than in the free protein - suggesting that electrostatic interactions are involved in the formation of the GUUG-EhCFIm25 complex.

C4 and C5 aptamers inhibit proliferation and induce the death of *E. histolytica* trophozoites

It has been previously reported that some vectors can allow the expression or delivery of aptamers inside cells, where they can reach their nuclear or cytoplasmic targets[46-48]. To deliver C4 and C5 aptamers into *E. histolytica*, we took advantage of our previous observation whereby soaking trophozoites in medium with dsRNA produced in bacteria results in uptake[44]. We cloned the cDNA sequences corresponding to the C4 or C5 aptamers into a pL4440 plasmid (Fig. 5A), expressed the aptamers in bacteria, and performed parasite soaking experiments with the resulting *C4*- and *C5*-dsRNA. Both *C4*- and *C5*-dsRNA were identified by electrophoresis in the culture medium on day four of the experiment - indicating that dsRNA was available for ingestion by trophozoites (Fig. 5B). Growth curves showed that both C4 and C5 aptamers significantly reduced cell proliferation, relative to control experiments. For example, proliferation at 96 h was reduced by ~80% in the presence of aptamers (Fig. 5C). Consistently, cultures exposed to *C4*- or *C5*-dsRNA contained a higher number of dead trophozoites (~80%) than control groups did (~20%) (Fig. 5D). We did not observe any significant differences between trophozoites grown in standard conditions or with *gfp*-dsRNA; hence, ingestion of unrelated dsRNA had no effect on the parasite's growth. This observation strengthens the relationship between the C4 and C5 aptamers and the observed phenotype. The hypothesis that C4 and C5 aptamers could act via RNA inhibition was rejected, since BLAST sequence analyses showed that neither the selected strand nor the complementary strand displayed homology with any of the parasite's gene.

DISCUSSION

In the present study, we developed RNA aptamers that targeted the EhCFIm25 protein in *E. histolytica* and reduced trophozoite survival. To the best of our knowledge, this is the first study to have applied aptamer-based approaches in *E. histolytica* and the second to have isolated aptamers that can distinguish between two proteins that differ only with regard to a single point mutation. By using a new contrast screening strategy with SELEX (five cycles), Chen et al. (2005) recently isolated an RNA aptamer bound more strongly to a mutant p53 protein (R175H) than to the wild-type protein[49]. Our use of seven rounds of selection in a conventional SELEX protocol enabled us to generate two aptamers (C4 and C5) with a high degree of specificity for the wild-type EhCFIm25, as confirmed by REMSAs with the mutant protein and extracts from HeLa cells or *T. cruzi*.

Since EhCFIm25 is an RNA-binding protein, its interaction with C4 and C5 depends on the RNAs' sequence and the three-dimensional structure. Given that all the isolated aptamers and all the 3'UTRs linked by EhCFIm25 contain the GUUG sequence, we hypothesize that the latter is the canonical RNA motif recognized *in vitro* by EhCFIm25. *In silico* analysis of the EhCFIm25-GUUG complex revealed that the interaction takes place in an RNA binding pocket that is in the same position as in the crystal structure of the human CFIm25 protein complexed with the UGUAU motif[50] - despite differences in the aa and nucleotide sequences (Supplementary Figure S2). On the basis of data obtained from the computational analysis of 32 crystallized RNA-protein complexes Jones et al. suggested that aromatic and positive charged aa have a fundamental role in RNA binding[51]. Moreover, the interaction with the GUUG sequence stabilizes EhCFIm25 - making it more compact, and structuring some regions of the protein. Notably, the 70-90 aa region that folds into an alpha helix and the loop seems to be involved in the positioning of the GUUG molecule in EhCFIm25's RNA binding pocket. These structural transitions from loops to alpha helices are considered to be "molecular switches" for interaction with nucleic acids[52]. Furthermore, the formation of a cluster of positively charged aa may promote electrostatic interaction between EhCFIm25 and the GUUG fragment. With a view to the future clinical applications for C4 and C5 aptamers, it must be kept in mind that the GUUG motif bound by EhCFIm25 differs from the UGUA sequence recognized by the human homolog (which displays

sequence identity of only 32% with the parasite protein)[19]. Although the human and parasite CFIm25 proteins have similar three-dimensional structures, there are marked differences between the respective aa sequences and proteins display sequence identity of only 32%[21]. Given that the C4 and C5 aptamers did not bind to human proteins (including CFIm25), they would be unlikely to affect the polyadenylation process in the hosts' cells. Additional ongoing experiments should confirm the specificity, selectivity and affinity of the interaction between C4/C5 and EhCFIm25, as a first step towards the aptamers' use as therapeutic tools.

Several researchers have reported that disrupting 3´UTR processing provides opportunities for the development of new anti-parasitic compounds[25-28]. In *E. histolytica*, we have observed that EhCFIm25 silencing alters poly(A) site selection and parasite survival (our unpublished data). Since EhCFIm25 controls the selection of poly(A) sites, EhCFIm25 sequestration is expected to modify the stability of the mRNAs produced. This may result in an overall alteration of gene expression and then rapid death of the parasite. Moreover, the presence of aptamers might also affect the availability and/or function of proteins that interact with EhCFIm25 (such as EhPAP[21] and EhPC4) and proteins involved in other mRNA processing events (as has been described in humans)[53,54]. Consistently, our molecular protein-protein docking experiments indicated that EhCFIm25's interaction with GUUG, EhPAP and EhPC4 involves distinct binding sites (Supplementary Figure S3). Hence, binding of EhCFIm25 by C4 or C5 aptamers (leading to an increase in protein stability) may also have additional effects on EhCFIm25's interaction with EhPAP and EhPC4 - both of which have fundamental roles in gene expression. We conclude that aptamers represent a powerful new biotechnological tool for blocking polyadenylation in *E. histolytica*.

Most conventional means of introducing nucleic acids to cells are based on lipofectamine transfection systems. In the present work, we observed that the dsRNA soaking method (originally used to silence gene expression in *E.* histolytica)[44] represents an efficient, reproducible, fast, easy-to-implement strategy for blocking a nuclear target protein like EhCFIm25. A single inoculation of C4 or C5 aptamer had a huge inhibitory effect on the proliferation and viability of *E. histolytica* trophozoites.

In conclusion, we developed RNA aptamers that contain the GUUG motif (recognized by the EhCFIm25 factor) and that effectively control *E. histolytica*'s survival *in vitro*. These valuable findings suggest that aptamers could be used to control *E. histolytica* during the development of amoebiasis or to eradicate residual trophozoites during antibiotic treatment. In the context of parasitic infections, several other groups have developed aptamers that specifically identify *T. cruzi* and *Plasmodium* in blood, and *Leishmania* in the sand fly vector[55-60]. Other researchers have used aptamers to assess drug efficacy in Chagas disease[61]. In *Leishmania*, the SELEX process has been used to generate aptamers that bind to nuclear proteins such as histone H2A [62-64] and the poly(A)-binding protein[65]; however, the researchers did not evaluate the aptamers' effect on parasite survival. Future investigations will focus on clinical strategies for aptamer delivery against human pathogens *in vivo*. The results of the present proof-of-concept study in *E. histolytica* may promote the use of aptamers to treat globally challenging parasitic diseases.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

J.D.O.V., C.L.C., A.Z.C, N.G., A.D. and L.A.M conceived the study. J.D.O.V., A.D., C.W., A.Z.C, E.R.M. and L.A.M. designed the experiments. J.D.O.V., C.L.C., A.Z.C, N.G. and L.A.M. analyzed the data. C.W., E.R.M. and A.D. contributed with reagents, materials and analysis tools. J.D.O.V., C.L.C., N.G. and L.A.M. wrote the manuscript.

FIGURE LEGENDS

Figure 1. Identification of aptamers against EhCFIm25. A) Schematic representation of ssDNA oligonucleotides used to generate the ssRNA library for the SELEX protocol. B) The predicted secondary structures of the C4 and C5 aptamers. C-G) The RNA-electrophoretic mobility shift assay (REMSA). Proteins were incubated with a biotin-labelled RNA probe, and the RNA-protein complexes were resolved via PAGE and chemiluminescence assays. C) A REMSA of the R7 aptamer population with wild-type EhCFIm25 and mutant

EhCFIm25*L135T proteins. Proteinase K (1 µg), unspecific competitor (tRNA) and RNA molecules from the first round of SELEX (R0) were used as controls. D) A REMSA of C4 and C5 aptamers with wild-type EhCFIm25 and mutant EhCFIm25*L135T proteins. E) A REMSA of C4 and C5 aptamers with *E. histolytica* protein extracts. Anti-EhCFIm25 antibodies were used as controls. F and G) A REMSA of C4 and C5 aptamers with protein extracts from HeLa cells (F) and *Trypanosoma cruzi* parasites (G). *Eh: E. histolytica*; *Tc: T. cruzi*; single arrowhead: the RNA-protein complex; double arrowhead: the RNA-protein-antibody complex; asterisk: the free probe.

Figure 2. EhCFIm25 binds to the GUUG motif. A) Sequence of the *Ehthio*- and *Ehrib*-3'UTR probes used in REMSAs. Box: the GUUG motif; circle: the stop codon: bold and underlined nucleotides: the cleavage and polyadenylation site. B-D) The REMSA results. Proteins were incubated with a biotin-labelled RNA probe, and RNA-protein complexes were resolved via PAGE and chemiluminescence assays. B) A REMSA of *Ehthio*- and *Ehrib*-3´UTR with *E. histolytica* extracts and the recombinant EhCFIm25 protein. Anti-EhCFIm25 antibodies were used as a control. C) A REMSA of *Ehthio*- and *Ehrib*-3´UTR with the recombinant EhCFIm25 protein. The GUUG-containing oligonucleotide (GUUG-probe) was used as a competitor. D) A REMSA of the GUUG-probe with recombinant EhCFIm25. The CAAC-containing oligonucleotide (CAAC-probe) was used as control. *Eh: E. histolytica*; single arrowhead: the RNA-protein complex; double arrowhead: the RNA-protein-antibody complex; asterisk: the free probe. E) Sequence of the GUUG- and CAAC-probes. Box: the GUUG motif; the T7 promoter sequence is underlined.

Figure 3. Interactions between EhCFIm25 and the GUUG motif. A) Graphic representation of the average structure of EhCFIm25 complexed to the GUUG motif. B) Contact map between the aa in EhCFIm25 and the nucleotides (U2, U3 and G4) in the GUUG motif. C) A close-up view of the interaction zone between selected aa in EhCFIm25 (in red) and U2, U3 and G4 in the GUUG fragment (in black).

Figure 4. Molecular dynamics simulations of free and GUUG-binding EhCFIm25 proteins. A) Change over time in root mean square deviation of Cα (RMSD). Data on the mutant EhCFIm25*L135T protein were included as control. B) Average structure of the free (left) and GUUG-binding (right) proteins. Black arrow: unstable loop; white arrow: disordered alpha helix. C) Root mean square fluctuations of Cα coordinates (RMSF). Arrows show the most flexible regions in both systems. The box indicates the region (70-90 aa) that differs in flexibility when comparing the two systems. D) Fluctuations of aa in the average structure of free (left) and GUUG-binding (right) protein. Black arrows indicate the 70-90 aa region with the greatest fluctuation. E) The surface electrostatic potential of the free (left) and GUUG-binding (right) proteins. Upper panel: front view; lower panel: rear view. The circle shows the RNA-interacting area.

Figure 5. Effect of C4 and C5 aptamers on the proliferation and viability of *E. histolytica* trophozoites. A) pL4440-C4/C5 plasmid constructs. B) Detection of dsRNA in TYI-S-33 medium. C- D) Trophozoites were treated with *C4*- or *C5*-dsRNA (100 µl/ml) and incubated at 37°C. Each day, the cell count was determined (C), and cell viability was assessed in a Trypan blue assay (D). Trophozoites grown in the absence of dsRNA (control) or in the presence of *gfp*-dsRNA were also used. Data were analysed in a two-way analysis of variance or a T-test, as appropriate. **p<0.01, ***p<0.001 and ****p<0.0001.

ADDITIONAL INFORMATION

Supplementary Figure S1. Schematic representation of the SELEX strategy. The first step involves the conversion of a library of single-stranded (ss) oligonucleotides into double-stranded (ds) DNA, using the Klenow fragment of DNA Polymerase I. Next, the dsDNA library is reverse-transcribed *in vitro* into RNA that interacts with the recombinant EhCFIm25 protein immobilized on a nickel nitrilotriacetic acid column. ssRNAs that do not interact with the protein are discarded, whereas interacting ssRNAs are amplified by RT-PCR and transcribed *in vitro* to generate a new ssRNA library for subsequent rounds of selection.

Supplementary Figure S2. Comparison of amoeba and human RNA-CFIm25 complexes, showing the three-dimensional structures of EhCFIm25+GUUG complex obtained by docking (A) and the crystal structure of the CFIm25 protein interacting with the UGUAUU Motif (B). Contact maps are shown on the right of each panel. C) Amino acid sequence alignment of EhCFIm25 and human CFIm25 proteins. Black shadows: EhCFIm25 aa that are in contact with the GUUG motif; black arrows: aa in each protein that interact with RNA.

Supplementary Figure S3. Interaction domains in EhCFIm25. A-C) Molecular docking was used to predict sites involved in the interaction between the EhCFIm25 protein (tan shadow) and the GUUG fragment (black bead) (A), EhPAP (blue shadow) (B) and EhPC4 (orange shadow) (C). D) Graphical representation of EhCFIm25 (tan) interacting with the GUUG sequence (black surface), EhPAP (blue surface), and EhPC4 (orange surface).

Supplementary video S1. A molecular dynamics simulation of free EhCFIm25 protein.

Supplementary video S2. A molecular dynamics simulation of EhCFIm25 interacting with the GUUG fragment.

REFERENCES

1. Colgan, D. & Manley, J. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11**, 2755-2766 (1997).
2. Zhao, J., Hyman, L., & Moore, C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev*. **63**, 405-45. (1999).
3. Xiang, K., Tong, L. &Manley, J. L. Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. *Mol. Cell. Biol. J.* **34**, 1894-1910. (2014).
4. Mandel, C., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3’-end processing. *Cell. Mol. Life Sci*. **65**, 1099-1122 (2007).
5. López-Camarillo, C., López-Rosas, I., Ospina-Villa, J. & Marchat, L. Deciphering molecular mechanisms of mRNA metabolism in the deep-

branching eukaryote Entamoeba histolytica. *Wiley Interdiscip. Rev. RNA.* **5**, 247-262 (2013).

6. del-Moral-Stevenel, M. *et al.* Transcriptional profile of processing machinery of 3′ end of mRNA in Trichomonas vaginalis. *Genes Genom.* **37**, 399-408 (2015).
7. Stanley, S. Amoebiasis. *The Lancet.* **361**, 1025-1034 (2003).
8. Ralston, K. & Petri, W. Tissue destruction and invasion by Entamoeba histolytica. *Trends Parasitol*. **27**, 254-263 (2011).
9. Kapoor, K., *et al.* Evaluation of metronidazole toxicity: a prospective study. *Int. J. Clin. Pharmacol. Res.* **19**, 83-8 (1999).
10. Hanna, R., Dahniya, M., Badr, S. & El-Betagy, A. Percutaneous catheter drainage in drug-resistant amoebic liver abscess. *Trop. Med. Int. Health.* **5**, 578-581 (2000).
11. Ali, I. *et al.* Evidence for a Link between Parasite Genotype and Outcome of Infection with Entamoeba histolytica. *J. Clin. Microbiol.* **45**, 285-289 (2006).
12. Duggal, P. *et al.* Influence of Human Leukocyte Antigen Class II Alleles on Susceptibility toEntamoeba histolytica Infection in Bangladeshi Children. *J. Infect. Dis.* **189**, 520-526 (2004).
13. Valdés, J. *et al.* Proteomic analysis of Entamoeba histolytica in vivo assembled pre-mRNA splicing complexes. *J. Proteomics*. **111**, 30-45 (2014).
14. Pearson, R. J. & Singh, U. Approaches to characterizing Entamoeba histolytica transcriptional regulation. *Cell. Microbiol.* **12**, 1681-1690. (2010).
15. Kim, S. *et al.* Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells.* **15**, 1003-1013 (2010).
16. Li, H. *et al.* Structural basis of pre-mRNA recognition by the human cleavage factor Im complex. *Cell Res.* **21**, 1039-1051 (2011).
17. Yang, Q., Coseno, M., Gilmartin, G. & Doublié, S. Crystal Structure of a Human Cleavage Factor CFIm25/CFIm68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping. *Structure*. **19**, 368-377 (2011).
18. Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A. & Handa, H. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs. *Nucleic Acids Res.* **34**, 6264-6271 (2006).

19. Brown, K. & Gilmartin, G. A Mechanism for the Regulation of Pre-mRNA 3'Processing by Human Cleavage Factor Im. *Mol. Cell.* **12**, 1467-1476 (2003).
20. Hardy, J. G.& Norbury, C. J. Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation. *Biochem. Soc. Trans*. **44**, 1051-1057 (2016).
21. Pezet-Valdez, M. *et al.* The 25 kDa Subunit of Cleavage Factor Im Is a RNA-Binding Protein That Interacts with the Poly(A) Polymerase in Entamoeba histolytica. *PLoS ONE.* **8**, e67977 (2013).
22. Hernandez de la Cruz, O. *et al.* Proteomic profiling reveals that EhPC4 transcription factor induces cell migration through up-regulation of the 16-kDa actin-binding protein EhABP16 in Entamoeba histolytica. *J. Proteomics.* **111**, 46-58 (2014).
23. Hernandez de la Cruz, O. *et al.* Multinucleation and Polykaryon Formation is Promoted by the EhPC4 Transcription Factor in Entamoeba histolytica. *Sci. Rep.* **6**, 19611. (2016).
24. Ospina-Villa, J. *et al.* Amino acid residues Leu135 and Tyr236 are required for RNA binding activity of CFIm25 in Entamoeba histolytica. *Biochimie.* **115**, 44-51 (2015).
25. Hendriks, E., Abdul-Razak, A. & Matthews, K. tbCPSF30 Depletion by RNA Interference Disrupts Polycistronic RNA Processing in Trypanosoma brucei. *J. Biol. Chem.* **278**, 26870-26878 (2003).
26. Sidik, S. *et al.* A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell.* **166**, 1423-1435.e12 (2016).
27. Palencia, A. *et al.* Targeting Toxoplasma gondii CPSF3 as a new approach to control toxoplasmosis. *EMBO Mol. Med.* **9**, 385-394 (2017).
28. Sonoiki, E. *et al.* A potent antimalarial benzoxaborole targets a Plasmodium falciparum cleavage and polyadenylation specificity factor homologue. *Nat. Commun.* **8**, 14574 (2017).
29. Ospina-Villa, J., Zamorano-Carrillo, A., Castañón-Sánchez, C., Ramírez-Moreno, E. & Marchat, L. Aptamers as a promising approach for the control of parasitic diseases. *Braz. J. Infect. Dis.* **20**, 610-618 (2016).
30. Diamond, L., Harlow, D. & Cunnick, C. A new medium for the axenic cultivation of Entamoeba histolytica and other Entamoeba.*Trans*. *R. Soc. Trop. Med. Hyg.* **72**, 431-432 (1978).

31. Takiff, H., Chen, S. & Court, D. Genetic analysis of the rnc operon of Escherichia coli. *J. Bacteriol.* **171**, 2581-2590 (1989).
32. Manley, J. SELEX to Identify Protein-Binding Sites on RNA. *Cold Spring Harb. Protoc.* **2013**(2):156-163 (2013).
33. Markham, N. R. & Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res*. **33**, W577-W581 (2005).
34. Markham, N. R. & Zuker, M. UNAFold: software for nucleic acid folding and hybridization. In Keith, J. M., editor, Bioinformatics, Volume II. Structure, Function and Applications, number 453 in Methods in Molecular Biology, chapter 1, pages 3–31. Humana Press, Totowa, NJ. ISBN 978-1-60327-428-9 (2008).
35. Lambert, C., Leonard, N., De Bolle, X. & Depiereux, E. ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics.* **18**, 1250-1256 (2002).
36. Bowie, J.U., Lüthy, R. & Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* **253**(5016):164-70 (1991).
37. Tuszynska, I., Magnus, M., Jonak, K., Dawson, W. & Bujnicki, J. NPDock: a web server for protein–nucleic acid docking. *Nucleic Acids Res.* **43**, W425-W430 (2015).
38. Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723 (1997).
39. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435-447 (2008).
40. MacKerell, A., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56**, 257-265 (2000).
41. Parrinello, M. & Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **45**, 1196-1199 (1980).
42. Dolinsky, T., Nielsen, J., McCammon, J. & Baker, N. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665-W667 (2004).
43. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38 (1996).

44. Solis, C., Santi-Rocca, J., Perdomo, D., Weber, C. & Guillén, N. Use of Bacterially Expressed dsRNA to Downregulate Entamoeba histolytica Gene Expression. *PLoS ONE.* **4**, e8424. (2009).
45. Hon, C. *et al.* Quantification of stochastic noise of splicing and polyadenylation in Entamoeba histolytica. *Nucleic Acids Res.* **41**, 1936-1952 (2012).
46. Good, P. *et al.* Expression of small, therapeutic RNAs in human cell nuclei. *Gene Therapy.* **4**, 45-54 (1997).
47. Thomas, M. *et al.* Selective Targeting and Inhibition of Yeast RNA Polymerase II by RNA Aptamers. *J. Biol. Chem.* **272**, 27980-27986 (1997).
48. Mayer, G. *et al.* Controlling small guanine-nucleotide-exchange factor function through cytoplasmic RNA intramers. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4961-4965 (2001).
49. Chen, L. *et al.* The isolation of an RNA aptamer targeting to p53 protein with single amino acid mutation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10002-10007 (2015).
50. Yang, Q., Gilmartin, G. & Doublie, S. Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3' processing. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10062-10067 (2010).
51. Jones, S. Protein-RNA interactions: a structural analysis. *Nucleic Acids Res.* **29**, 943-954 (2001).
52. Sandhu, K. & Dash, D. Dynamic α-helices: Conformations that do not conform. *Proteins: Struct. Funct. Bioinf.* **68**, 109-122 (2007).
53. de Vries, H. Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.* **19**, 5895-5904 (2000).
54. Millevoi, S. *et al.* An interaction between U2AF 65 and CF Im links the splicing and 3′ end processing machineries. *EMBO J.* **25**, 4854-4864 (2006).
55. Nagarkatti, R. *et al.* Development of an Aptamer-Based Concentration Method for the Detection of Trypanosoma cruzi in Blood. *PLoS ONE.* **7**, e43533 (2012).
56. Cheung, Y. *et al.* Structural basis for discriminatory recognition of Plasmodium lactate dehydrogenase by a DNA aptamer. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15967-15972 (2013).

57. Lee, S., Manjunatha, D., Jeon, W. & Ban, C. Cationic Surfactant-Based Colorimetric Detection of Plasmodium Lactate Dehydrogenase, a Biomarker for Malaria, Using the Specific DNA Aptamer. *PLoS ONE.* **9,** e100847 (2014).
58. Nagarkatti, R., de Araujo, F., Gupta, C. & Debrabant, A. Aptamer Based, Non-PCR, Non-Serological Detection of Chagas Disease Biomarkers in Trypanosoma cruzi Infected Mice. *PLoS Negl. Trop. Dis.* **8**, e2650 (2014).
59. Birch, C., Hou, H., Han, J. & Niles, J. Identification of malaria parasite-infected red blood cell surface aptamers by inertial microfluidic SELEX (I-SELEX). *Sci. Rep.* **5**, 11347 (2015).
60. Bruno, J. A Review of Therapeutic Aptamer Conjugates with Emphasis on New Approaches. *Pharmaceuticals.* **6**, 340-357 (2013).
61. de Araujo, F., Nagarkatti, R., Gupta, C., Marino, A. & Debrabant, A. Aptamer-Based Detection of Disease Biomarkers in Mouse Models for Chagas Drug Discovery. *PLoS Negl. Trop. Dis.* **9**, e3451 (2015).
62. Ramos, E. *et al.* A DNA aptamer population specifically detects Leishmania infantum H2A antigen. *Lab. Invest.* **87**, 409-416 (2007).
63. Ramos, E., Moreno, M., Martín, M., Soto, M. & Gonzalez, V. In Vitro Selection o fLeishmania infantum H3-Binding ssDNA Aptamers. *Oligonucleotides* **20**, 207-213 (2010).
64. Martín, M. *et al.* DNA Aptamers Selectively Target Leishmania infantum H2A Protein. *PLoS ONE.* **8**, e78886. 10.1371/journal.pone.0078886 (2013).
65. Guerra-Pérez, N. *et al.* Molecular and Functional Characterization of ssDNA Aptamers that Specifically Bind Leishmania infantum PABP. *PLOS ONE.* **10**, e0140048. 10.1371/journal.pone.0140048 (2015).

Fig1

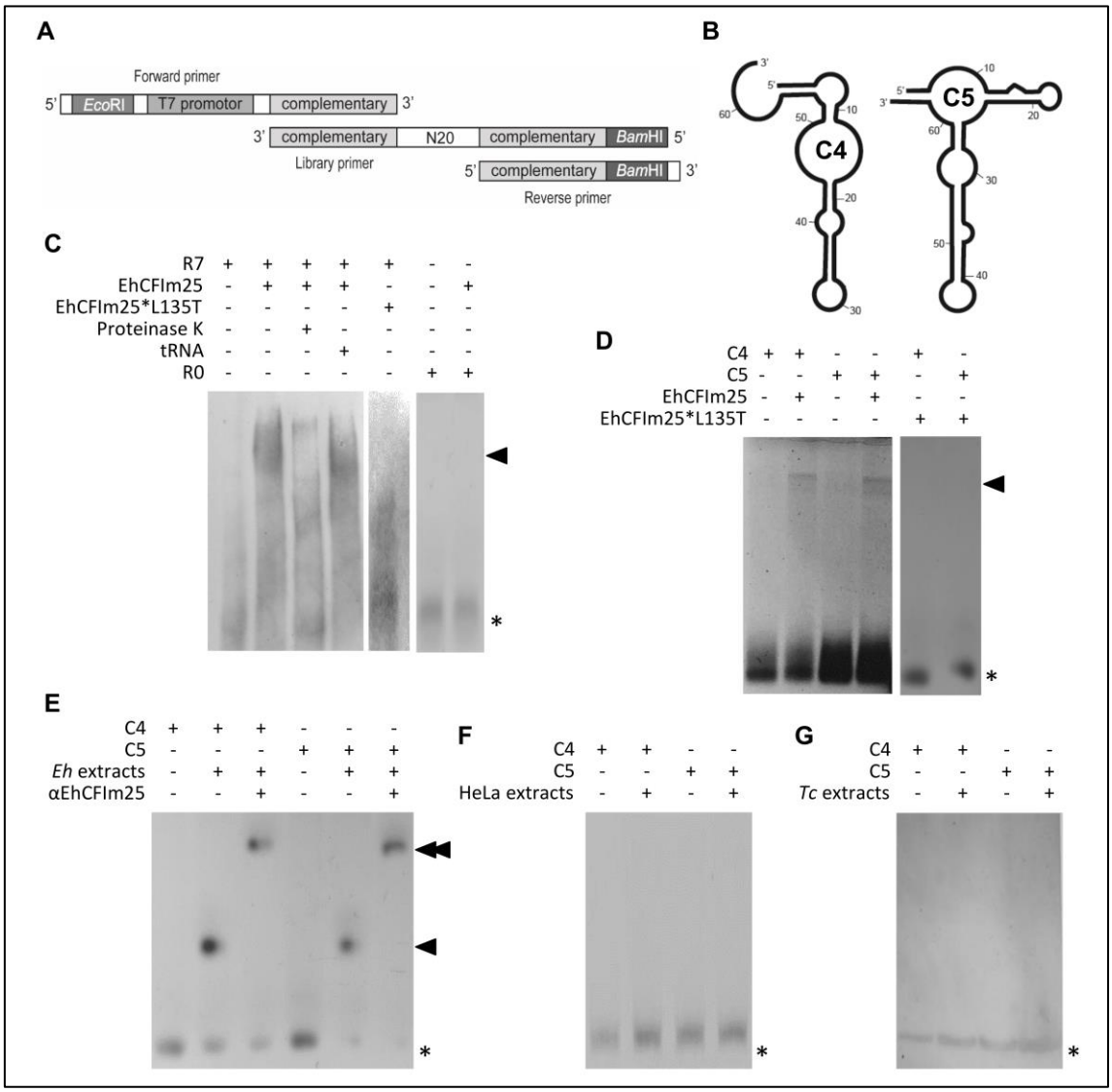

A
Forward primer
5'
EcoRI
T7 promotor
complementary
3'
3'
complementary
N20
complementary
BamHI
5'
Library primer
5'
complementary
BamHI
3'
Reverse primer
B
C4
C5
C
R7
EhCFIm25
EhCFIm25*L135T
Proteinase K
tRNA
R0
D
C4
C5
EhCFIm25
EhCFIm25*L135T
E
C4
C5
Eh extracts
αEhCFIm25
F
C4
C5
HeLa extracts
G
C4
C5
Tc extracts

Fig2

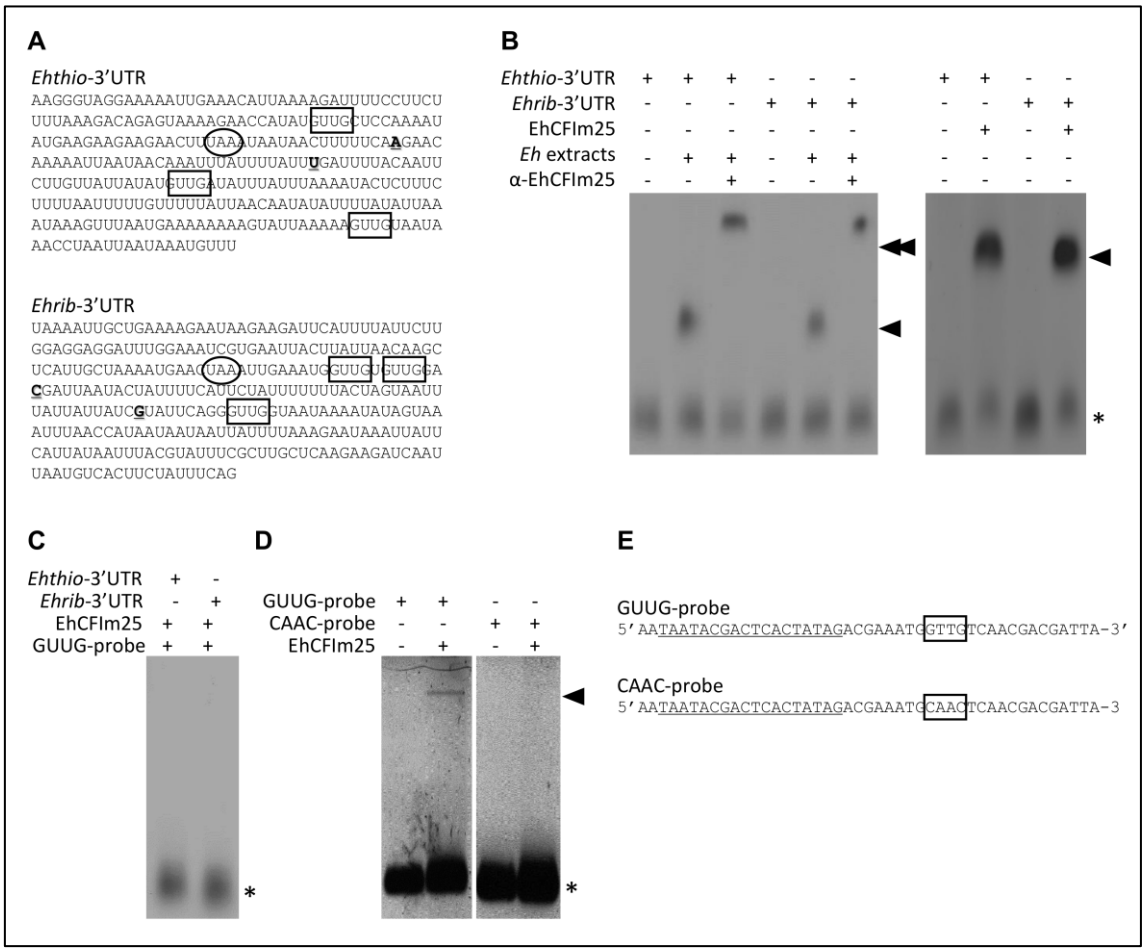

A
Ehthio-3'UTR
AAGGGUAGGAAAAAUUGAAACAUUAAAAGAUUUUCCUUCU
UUUAAAGACAGAGUAAAAGAACCAUAUGUUGCUCCAAAAU
AUGAAGAAGAAGAACUUUAAAUAAUAACUUUUUCAAGAAC
AAAAAUUAAUAACAAAUUUAUUUUAUUUGAUUUUACAAUU
CUUGUUAUUAUAUGUUGAUAUUUAUUUAAAAUACUCUUUC
UUUUAAUUUUUGUUUUUAUUAACAAUAUAUUUUAUAUUAA
AUAAAGUUUAAUGAAAAAAAAGUAUUAAAAAGUUGUAAUA
AACCUAAUUAAUAAAUGUUU
Ehrib-3'UTR
UAAAAUUGCUGAAAAGAAUAAGAAGAUUCAUUUUAUUCUU
GGAGGAGGAUUUGGAAAUCGUGAAUUACUUAUUAACAAGC
UCAUUGCUAAAAUGAACUAAAUUUGAAAUGGUUGUGUUGGA
CGAUUAAUACUAUUUUCAUUCUAUUUUUUUACUAGUAAUU
UAUUAUUAUCGUAUUCAGGGUUGGUAAUAAAAUAUAGUAA
AUUUAACCAUAAUAAUAAUUAUUUUAAAGAAUAAAUUAUU
CAUUAUAAUUUACGUAUUUCGCUUGCUCAAGAAGAUCAAU
UAAUGUCACUUCUAUUUCAG
B
Ehthio-3'UTR + + + - - - + + - -
Ehrib-3'UTR - - - + + + - - + +
EhCFIm25 - - - - - - - + - +
Eh extracts - + + - + + - - - -
α-EhCFIm25 - - + - - + - - - -
C
Ehthio-3'UTR + -
Ehrib-3'UTR - +
EhCFIm25 + +
GUUG-probe + +
D
GUUG-probe + + - -
CAAC-probe - - + +
EhCFIm25 - + - +
E
GUUG-probe
5'AATAATACGACTCACTATAGACGAAATGGTTGTCAACGACGATTA-3'
CAAC-probe
5'AATAATACGACTCACTATAGACGAAATGCAACTCAACGACGATTA-3

Fig3

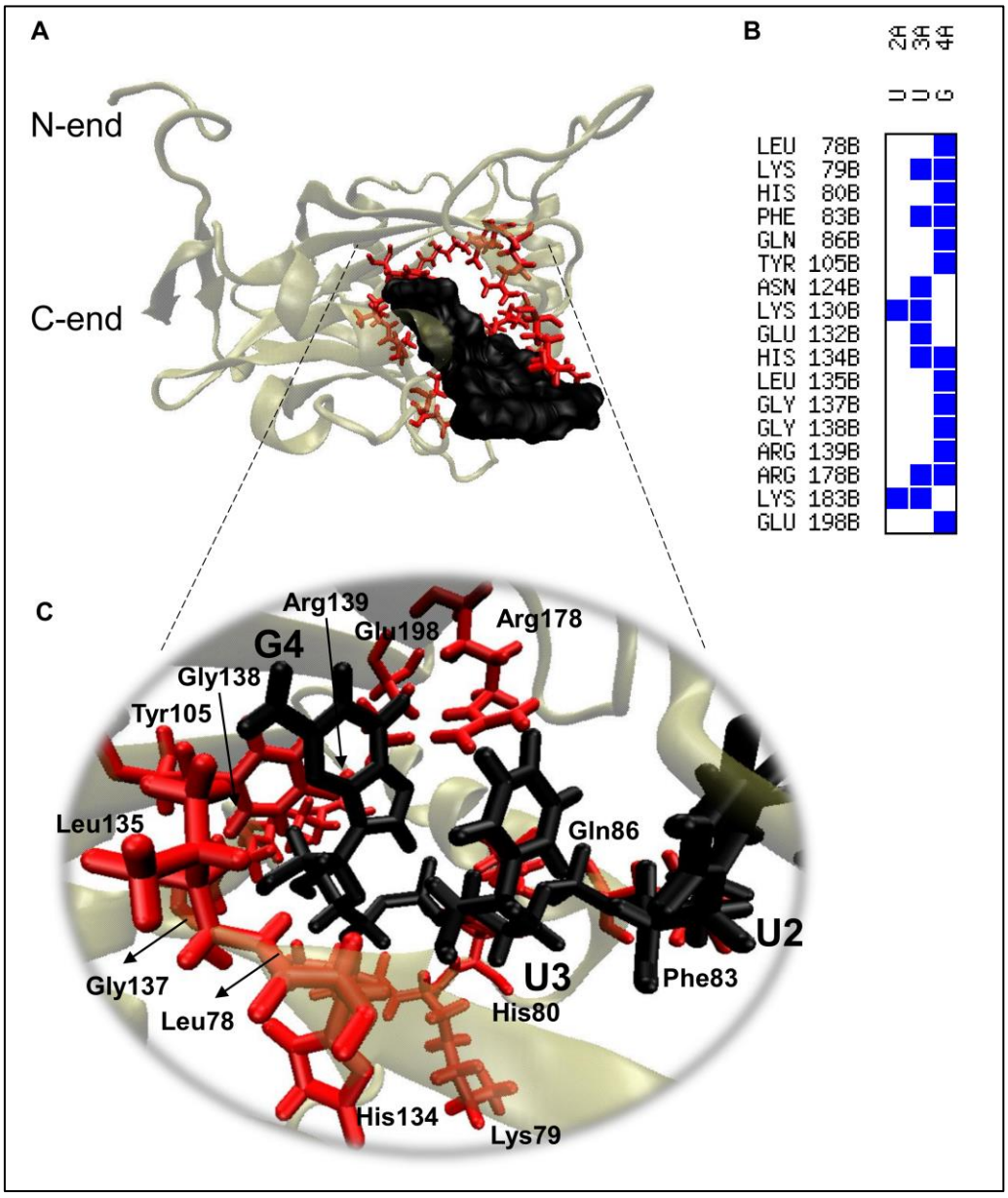

A
N-end
C-end
B
2A
3A
4A
U
U
G
LEU 78B
LYS 79B
HIS 80B
PHE 83B
GLN 86B
TYR 105B
ASN 124B
LYS 130B
GLU 132B
HIS 134B
LEU 135B
GLY 137B
GLY 138B
ARG 139B
ARG 178B
LYS 183B
GLU 198B
C
Arg139
Glu198
Arg178
G4
Gly138
Tyr105
Leu135
Gln86
U2
U3
Phe83
Gly137
Leu78
His80
His134
Lys79

Fig4

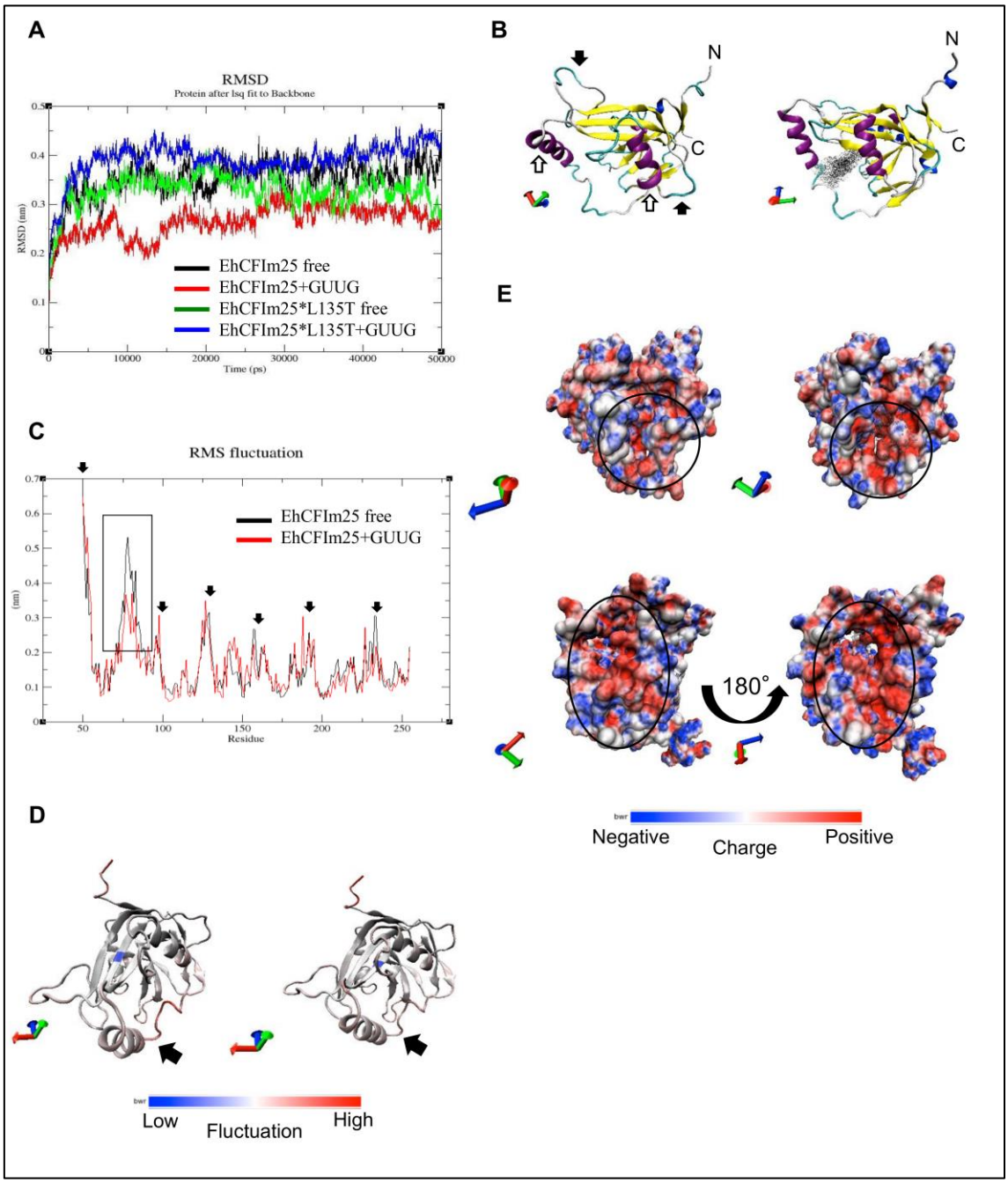
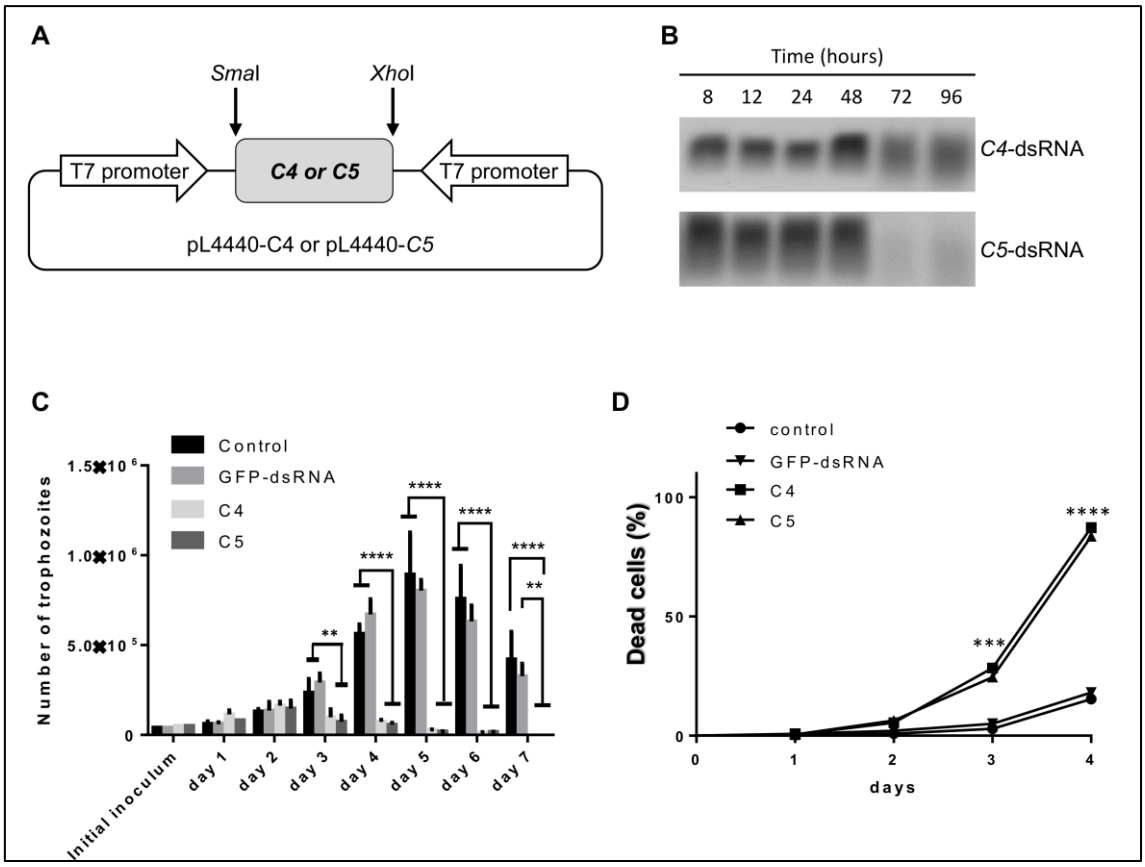
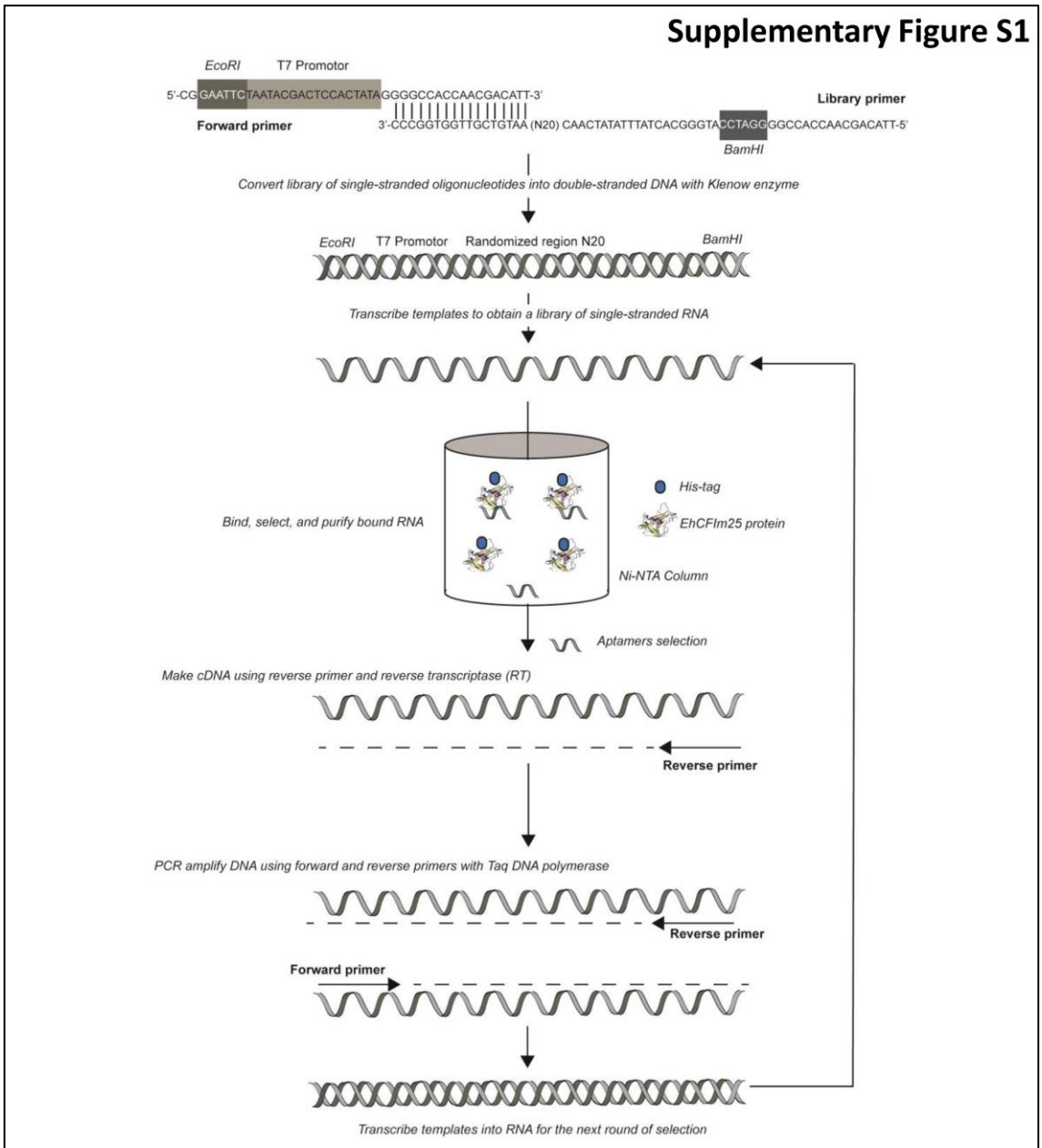

A
RMSD
Protein after lsq fit to Backbone
RMSD (nm)
Time (ps)
EhCFIm25 free
EhCFIm25+GUUG
EhCFIm25*L135T free
EhCFIm25*L135T+GUUG
B
N
C
N
C
C
RMS fluctuation
(nm)
Residue
EhCFIm25 free
EhCFIm25+GUUG
D
Low
Fluctuation
High
E
180°
bwr
Negative
Charge
Positive

Fig5


A
SmaI
XhoI
T7 promoter
C4 or C5
T7 promoter
pL4440-C4 or pL4440-C5
B
Time (hours)
8 12 24 48 72 96
C4-dsRNA
C5-dsRNA
C
Control
GFP-dsRNA
C4
C5
Number of trophozoites
1.5✖10 6
1.0✖10 6
5.0✖10 5
0
Initial inoculum
day 1
day 2
day 3
day 4
day 5
day 6
day 7
D
control
GFP-dsRNA
C4
C5
Dead cells (%)
100
50
0
days
0 1 2 3 4

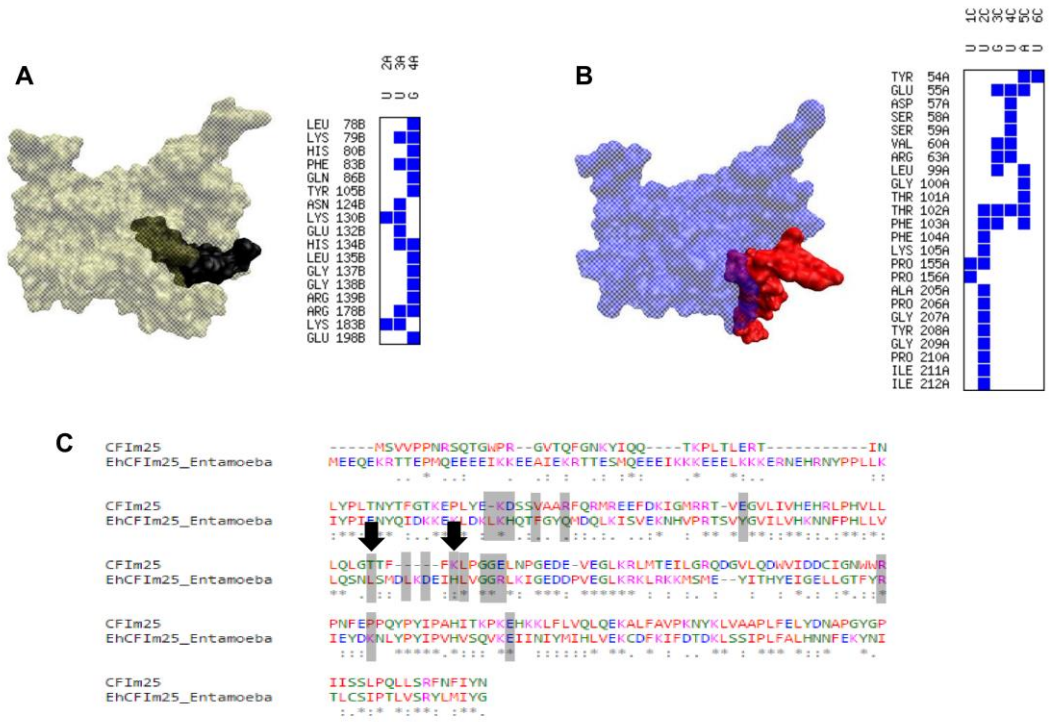
Suppl FigS1



Supplementary Figure S1. Schematic representation of the SELEX strategy. The first step involves the conversion of a single-stranded (ss) oligonucleotides library into double-stranded (ds) DNA with the Klenow enzyme. Then, the dsDNA library is *in vitro* transcribed into RNA that interacts with the recombinant EhCFIm25 protein immobilized on a Ni-NTA column. ssRNA that do not interact with the protein are discarded, while interacting ssRNA are amplified by RT-PCR and *in vitro* transcribed to generate a new ssRNA library for subsequent rounds of selection.

Suppl FigS2

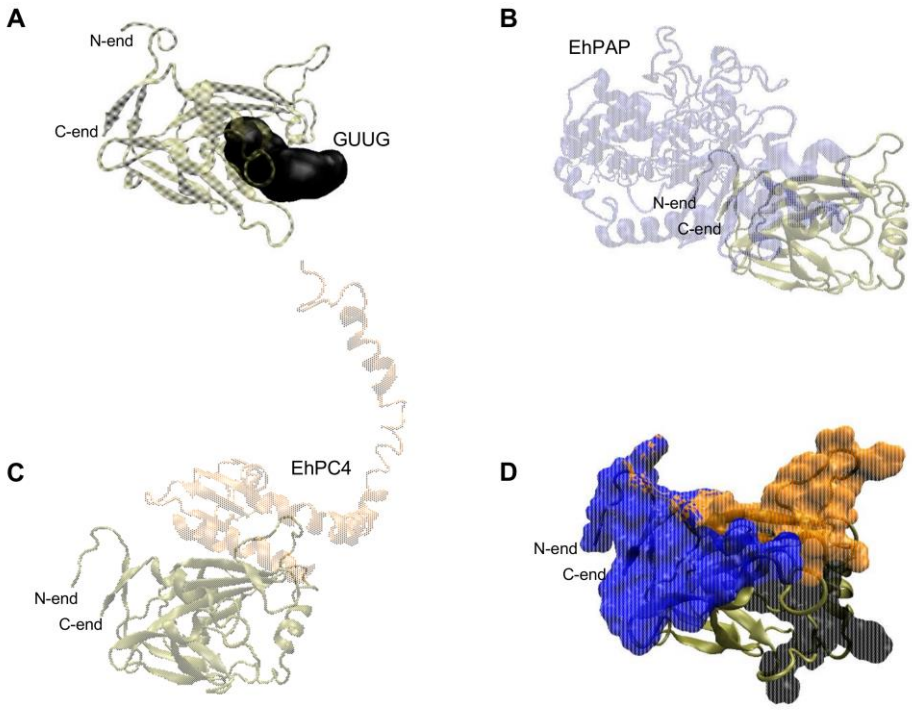
Supplementary Figure S2



Supplementary Figure S2. Comparison between amoeba and human RNA-CFIm25 complexes. Comparison between the three-dimensional structures of EhCFIm25+GUUG complex obtained by Docking (A) and the crystal structure of the CFIm25 protein interacting with the UGUAUU Motif (B). Contact maps are at the right in each panel. C) Alignment of amino acid sequences of EhCFIm25 and human CFIm25 proteins. Black shadows, position of amino acids of EhCFIm25 that are in contact with the GUUG motif; black arrows, amino acids of both proteins that interact with RNA

Suppl FigS3

Supplementary Figure S3



Supplementary Figure S3. Interactions domains in EhCFIm25. A-C) Molecular Docking to predict interaction sites of the EhCFIm25 protein (tan shadow) with the GUUG fragment (black bead) (A), EhPAP (blue shadow) (B) and EhPC4 (orange shadow) (C). D) Graphical representation of EhCFIm25 (tan) interacting withthe GUUG sequence (black surface), EhPAP (blue surface), and EhPC4 (orange surface).

CAPÍTULO 4:

Importance of amino acids Leu135 and Tyr236 for the interaction between EhCFIm25 and RNA: a molecular dynamics simulation study.

Juan David Ospina-Villa, Juan García-Contreras, Jorge Luis Rosas-Trigueros, Esther Ramírez-Moreno, César López-Camarillo, Beatriz Zamora-López, Laurence A, Marchat, Absalom Zamorano-Carrillo.

Enviado a Journal of Molecular Modeling

Title:

Importance of amino acids Leu135 and Tyr236 for the interaction between EhCFIm25 and RNA: a molecular dynamics simulation study.

Authors:

Juan David Ospina-Villa[1], Juan García-Contreras[1], Jorge Luis Rosas-Trigueros[2], Esther Ramírez-Moreno[1], César López-Camarillo[3], Beatriz Zamora-López[4], Laurence A, Marchat[1], Absalom Zamorano-Carrillo[1*]

[1]Programa Institucional de Biomedicina Molecular, Programa de Doctorado en Ciencias en Biotecnología, ENMH, Instituto Politécnico Nacional, Guillermo Massieu Helguera 239, Fracc. La Escalera, Ticomán, Del. Gustavo A. Madero, CP 07320, Ciudad de México, México.

[2]Laboratorio Transdisciplinario de Investigación en Sistemas Evolutivos, ESCOM, Instituto Politécnico Nacional, Av. Juan de Dios Bátiz esq. Miguel Othón de Mendizábal, Col. Lindavista, Del. Gustavo A. Madero, CP 07738, Ciudad de México, México.

[3]Posgrado en Ciencias Genómicas, Universidad Autónoma de la Ciudad de México, San Lorenzo 290, Colonia del Valle, CP 03100, Ciudad de México, México.

[4]Departamento de Psiquiatría y Salud Mental, Facultad de Medicina, UNAM, Circuito Interior y Cerro del Agua, CP 04510, Ciudad de México, México.

*Corresponding author. Tel. (52 55) 5729 6300

ORCID numbers:

Juan David Ospina-Villa: 0000-0003-1679-3036; Juan García-Contreras: 0000-0001-9206-0446; Jorge Luis Rosas-Trigueros: 0000-0003-2030-9817; Esther Ramírez-Moreno: 0000-0001-6924-6541; César López-Camarillo: 0000-0002-9417-2609; Beatriz Zamora-López: 0000-0003-4052-8849; Laurence A. Marchat: 0000-0003-1615-8614; Absalom Zamorano-Carrillo: 0000-0001-6146-195X

Acknowledgments

This work was supported by ECOS-France under Grant M14S02, CONACyT-Mexico under Grants 178550 and 249554, and SIP-IPN-Mexico under Grants 20150720, 20160801, 20170555 and 20170520. AZC, JLTR, ERM, and LAM were supported by COFAA-IPN. JDOV was a scholarship recipient from Mexican BEIFI-IPN and CONACyT programs.

Abstract

The CFIm25 subunit of the heterotetrameric Cleavage Factor Im (CFIm) is a key factor in the formation of poly(A) tail at mRNA 3´ end, regulating the recruitment of polyadenylation factors, poly(A) site selection, and cleavage/polyadenylation reactions. We previously reported the homologous protein (EhCFIm25) in *Entamoeba histolytica*, the protozoan causing human amoebiasis, and showed the relevance of conserved Leu135 and Tyr236 residues for RNA binding. We also identified the GUUG sequence as the recognition site of EhCFIm25. To understand the interactions network that allows the EhCFIm25 to maintain its three-dimensional structure and function, here we performed molecular dynamics simulations of wild type (WT) and mutant proteins, alone or interacting with the GUUG molecule. Our results indicated that in the presence of the GUUG sequence, WT converged more quickly to lower RMSD values in comparison with mutant proteins. However, RMSF values showed that movements of amino acids of WT and EhCFIm25*L135T were almost identical, interacting or not with the GUUG molecule. Interestingly, EhCFIm25*L135T, which is the only mutant with a slight RNA binding activity, presents the same stabilization of bend structures and alpha helices as WT, notably in the C-terminus. Moreover, WT and EhCFIm25*L135T presented almost the same number of contacts that mainly involve Lysine residues interacting with the G4 nucleotide. Overall, our data proposed a clear description of the structural and mechanistic data that govern the RNA binding capacity of EhCFIm25.

Keywords: Molecular dynamics, Cleavage Factor, *Entamoeba histolytica*, polyadenylation, protein-RNA interaction.

Introduction

Cleavage and polyadenylation of pre-messenger RNA (pre-mRNA) 3´ end are fundamental processes for transcript maturation and gene expression regulation in eukaryotes. Poly (A) tail is required for nuclear-cytoplasmic export, translation efficiency, transcript stability and transcript degradation prevention [1-3]. Also, most human genes have multiple cleavages and polyadenylation sites whose alternative utilization produces mRNA isoforms with 3´UTRs of different sizes, potentially affecting target sequences for microRNAs and RNA-binding proteins [4]. Consequently, alterations in 3´UTR length also have implications in many human diseases [5]. Poly(A) tail formation involves cis-acting sequences that are recognized by protein complexes, mainly the Cleavage Factors I and IIm (CFI and CFIIm), the Cleavage and Polyadenylation Specificity Factor (CPSF) and the Cleavage Stimulating Factor (CstF) [6-8]. CFIm25, the smallest subunit of the heterotetrameric CFIm complex, is a highly conserved polypeptide that specifically recognizes the UGUAN sequence in 3´UTR [9,10]. Notably, CFIm25 is a key factor for polyadenylation machinery recruitment, poly(A) site selection, as well as cleavage and polyadenylation reactions [11,12]. Analysis of the human CFIm25-UGUA complex revealed that the central Nudix domain of the protein (residues 76-201) with the conserved Nudix box -GHVEPGEDDLETALRETQEEAGI- (residues 43-65) includes critical amino acids for regulating protein-RNA interaction. An analysis of the crystallographic structure of CFIm25 bound to the UGUA RNA sequence showed that 14 amino acids interact directly or indirectly with the different ribonucleotides, namely, Glu55, Asp57, Ser58, Arg63, Glu81, Leu99, Gly100, Thr101, Thr102, Phe103, Phe104, Leu106, Tyr208, and Gly209. In particular, U1 formed three hydrogen bonds, *id est*, O2 and N3 interacted with the main -chain and carbonyl groups of Phe104, respectively; the O4 was stabilized by the side chain of Glu81; and the main chain amide, via a glycerol, interacted with Leu106. Another hydrogen bond was reported between O2 and the hydroxyl of the ribose and main-chain carbonyl of the Thr102. Also, the U1 was stabilized by stacking with the plane constituted by the peptide bond between Tyr208 and Gly209. Intramolecular interactions also play a fundamental role in this process [9,10,13].

Entamoeba histolytica, the protozoan parasite causative of human amoebiasis that affects about 50 million people worldwide [14], only has the small 25 kDa subunit of CFIm that interacts with the *EhPgp5* mRNA 3′ UTR [15,16]. In an attempt to describe this interaction, we identified that two of the 14 amino acid residues that are critical for RNA binding in the human CFIm25, Leu106 and Tyr208 residues, correspond to conserved Leu135 and Tyr236 in EhCFIm25. By site-directed mutagenesis and RNA-

binding assays, we showed that a single change of either Leu135 or Tyr236 amino acids to Ala, as well as replacement of Tyr236 to Phe, entirely abolish the RNA binding ability of EhCFIm25 *in vitro*; in contrast, a slight RNA affinity remained when Thr replaced Leu135. These data suggested that chemical groups of Leu135 and Tyr236 side chains, as well as the OH group of the phenol ring at position 236, are essential for RNA interaction [17]. Recently, we identified the GUUG sequence as the recognition site of EhCFIm25 (our unpublished data). However, the impact of each single amino acid mutations in the three-dimensional structure, intramolecular interactions, and protein stability of EhCFIm25 remains to be explored including if they affect the GUUG binding.

Molecular dynamics (MD) simulation represents a useful tool to elucidate the impact of amino acid substitutions on polypeptide folding and intramolecular interactions that may affect protein functions as a consequence of significant changes regarding flexibility, number of hydrogen bonds, and stability [18]. Although the MD studies for a single molecule, in particular for proteins, are published more frequently, the understanding of the RNA-protein interaction represents a challenge for this computational technique. However, some *in silico* studies have proposed that the electrostatic interactions or the number of hydrogen bonds could regulate this binding [19-21]. Here, we performed MD simulations of wild type and mutant EhCFIm25 proteins to understand the intramolecular interactions network that allows binding to the GUUG motif. The description of structural and mechanistic information regarding EhCFIm25 and its interactions with RNA should contribute to the understanding of mRNA 3´ end formation and stabilization in *E. histolytica.*

Methods

Molecular modeling and docking

The three-dimensional structure of EhCFIm25 (UniProt: C4M2T1, 255 residues) was predicted by homology modeling with the MODELLER package (http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/) [22], using as template the crystal structure of the human CFIm25 protein (chain A) in complex with the UUGUAU RNA molecule (chain C) (PDB 3MDG). The quality of the modeled structure was validated with the Verify_3D software and submitted to a process of energy minimization (http://services.mbi.ucla.edu/Verify_3D/) [23]. On the other hand, we used the RNAComposer online software (http://rnacomposer.cs.put.poznan.pl/) to construct the PDB structure of the GUUG RNA motif. Then, both molecules were individually submitted to MD simulations for one ns as described below to obtain the most relaxed structures. Later, they were subjected to NPDock (Nucleic Acid-Protein

Dock) web server (http://genesilico.pl/NPDock/) for modeling the plausible GUUG-EhCFIm25 complex using the default parameters (5Å of RMSD cutoff for clustering, 1000 steps of simulation, initial temperature of 15,000 K and 295 K for the last step of simulation) and a blind docking approach. This docking uses specific protein–nucleic acid statistical potentials for scoring and selecting of modeled complexes [24,25]. Finally, we used the Swiss-PdbViewer 4.10 software [26] (http://www.expasy.org/spdbv/) to introduce individual mutations in the EhCFIm25 protein (with or without the GUUG motif) by choosing the alpha carbon of Leu135 and Tyr236 residues to substitute them for Ala or Thr, and Ala or Phe, respectively.

Molecular Dynamics (MD) Simulation

MD simulations of wild type (WT) and mutant EhCFIm25 proteins, with or without the GUUG motif, were performed using the GROMACS 5.0 package [27] and the CHARMM27 v2.0 force field [28]. We used a cubic box with 10 Å as margin for free protein systems and 15 Å for RNA-protein complexes. The systems/boxes were solvated with 12,825 and 28,390 water molecules respectively, and three Cl- ions were added to equilibrate the electric charges. We used a 3-point solvent model called spc216 as a model of water [29]. Minimization of both systems, with or without the GUUG RNA sequence, was achieved using the conjugate gradient method. Later, the system equilibration was completed in two phases: the first phase was conducted under the NVT (isothermal-isochoric) ensemble for 100-ps to stabilize the temperature at 300 K; then, the equilibrium of pressure was achieved under the NPT (isothermal-isobaric) ensemble for 100-ps using the PME algorithm. Finally, MD simulations were continued for 50 ns.

Analysis of MD trajectories

The atomic characteristics of WT and mutant EhCFIm25 proteins, with or without the GUUG motif, were compared using the analysis tools included in the GROMACS software. Values of root mean square deviation (RMSD) of α-carbons, root mean square fluctuations (RMSF) of α-carbons, radius of gyration (RG) and solvent-accessible surface area (SASA) were calculated. For these analyses, the time when the RMSD converged was considered as the initial point (10 ns) of the production simulations. The evolution of the secondary structures of both free and RNA-bound proteins (WT and mutants) was followed using the do_dssp tool of the GROMACS software.

Analysis of the interaction EhCFIm25-GUUG

The contact analysis between protein and RNA were performed with an online software of the Weizmann Institute of Science in Israel (http://bip.weizmann.ac.il/oca-bin/lpccsu/) [30] using the average structure of the protein with or without RNA.

Results

Three-dimensional structure of EhCFIm25 and its variants, in presence and absence of GUUG.

We previously found that the EhCFIm25 interaction with RNA involves Leu135 and Tyr236 residues [17] and the GUUG motif (our unpublished data). To explain the structural effect of L135A, L135T, Y236A and Y236F single point mutations, and their consequences in the stability and functionality, specifically the RNA binding activity of the protein, we compared MD simulations of WT and mutant proteins, alone and in the presence of the GUUG sequence. Because the experimental structure of the EhCFIm25 protein has not been determined, a three-dimensional model was predicted using the crystal structure of the human CFIm25 protein as a template (PDB 3MDG). The predicted model of EhCFIm25, comprising residues 50 to 255, and the verification of its 3D structure with the Verify3D software showed an 83.50% of the residues with a score > 0.2 in the 3D/1D profile, suggesting a suitable structure for the minimization energy process. Similarly, we obtained the RNA sequence GUUG in a PDB format. Later, relaxed three-dimensional models of both molecules were employed to get the EhCFIm25-GUUG complex structure by blind docking studies with the NP Dock program using default parameters. Five protein-RNA complexes were obtained, and the structure with the lowest interaction energy served for further simulations. Substitutions in the residues Leu135 and Tyr236 were produced by the Swiss-PdbViewer 4.10 software as follows: 1) Leu135 was replaced by Ala (EhCFIm25*L135A) or Thr (EhCFIm25*L135T); 2) Tyr236 by Ala (EhCFIm25*Y236A) or Phe (EhCFIm25*Y236F) (Figure 1A-F). The change to alanine shortened the side chain of both residues, whereas replacement by threonine and phenylalanine did not significantly affect the side chain length but changed the polarity of the residue.

Overall stability and flexibility

To evaluate the effect of each mutation on EhCFIm25 protein conformation, in the presence or absence of the GUUG RNA fragment, we investigated the differences among structural parameters obtained from the trajectories. To get a characterization of the global movement in comparison with their initial structures, the RMSD values were calculated. All free proteins showed an increase in RMSD during the first ten ns (stabilization step); then, they oscillated in an interval of 0.35-0.45 nm (production step)

without any significant differences between WT and mutant proteins. However, EhCFIm25*L135T bore the closest resemblance to WT and presented the lowest values (Figure 1G). In contrast, in the presence of the GUUG sequence, the mutants converged to RMSD values around 0.4 nm, but the WT protein-RNA system (EhCFIm25+GUUG) reached convergence more quickly and at a lower value (0.2-0.3 nm), suggesting that the atomic positions of the WT were more stable due to the interaction with RNA (Figure 1H).

Another dynamical parameter to study related to local movements of each residue was the RMSF. Figure 2 shows a comparison of the EhCFIm25 WT with the considered mutants, as well as the secondary structure. Most RMSF values for the WT protein were between 0.1 and 0.2 nm; however, we observed two mobile zones that span residues 60-100 (denoted as zone a) and residues 120-135 (zone b) with RMSF values around 0.45 nm and 0.25 nm, respectively. Globally, RMSF profiles of mutant proteins were quite similar to that of WT EhCFIm25. However, there were some specific changes for each variant. The secondary structure of the flexible zone an in EhCFIm25 includes a loop (61-63), a small alpha helix (64-66), a beta sheet (67-72), a loop (73-82), an alpha helix (83-95), and another loop (96-100). This region presented higher fluctuations (up to 0.7 nm) in EhCFIm25*L135A, while it presented a lower mobility (higher rigidity) in EhCFIm25*Y236F and EhCFIm25*Y236A (Figure 2A, 2C and 2D). Interestingly, EhCFIm25*L135T is the only mutant protein that presented a similar behavior concerning the WT in the zone a. On the other hand, the mobility of the beta-loop-beta conformation of zone b was also increased in EhCFIm25*L135A, while it was reduced in EhCFIm25*Y236F and EhCFIm25*L135T, in comparison with the WT protein (Figure 2B and 2D). Also, the mutant EhCFIm25*Y236F protein exhibited a reduced mobility in a region covering residues 150-170 (zone c) that corresponds to a loop-alpha helix (Figure 2D). This analysis showed that movements of amino acids of WT EhCFIm25 and EhCFIm25*L135T are almost identical, indicating that the three-dimensional structures of both proteins are not significantly different during the entire simulation (Figure 2B). Later, we evaluated the effect of the EhCFIm25-GUUG interaction on the RMSF values. The results showed that the corresponding values of the mutants in the zone a were similar among them (approximately 0.3 nm), except for EhCFIm25*L135A that presented a decrease (0.15 nm) (Figure 3A). In zone b, the RMSF value decreased in all mutants, but in EhCFIm25*Y236F the result was similar to WT (Figure 3D). Regarding zone c, only the variant EhCFIm25*Y236A exhibited significant variation compared with the WT, surrounding residues that involve an alpha helix spanning from Val148 to Lys154 (Figure 3C). Finally, we observed the appearance of a highly flexible zone (up to 0.3 nm) denoted as zone d (residues 225-

240) in the EhCFIm25*L135A protein (Figure 3A). Considering the EhCFIm25*L135T, except for zone b, its RMSF values are similar to WT; where the two peaks in the zone a are also reached, although a slight increase in the fluctuation in the alpha-helix is detected (Figure 3B). Overall, RMSD and RMSF values of the WT and EhCFIm25*L135T proteins suggest a similar behavior between them, interacting or not with the GUUG.

Secondary structure

To gain insights into the effect of individual mutations on protein folding, we evaluated the evolution of the secondary structure of the protein and its variants (Figure 4). Although beta-sheets remained almost constant in all conditions, some significant differences could be observed when comparing protein variants. Regarding the effect of the mutations, all the proteins mostly retained the secondary structure of the WT protein. However, some differences could be noted such as a tendency to increase the alpha helix conformation at the C-end for EhCFIm25*L135A and EhCFIm25*Y236A. The opposite is observed for this last protein around residues 50-75, where the helical content seemed to be replaced by turn and bend structures (Figure 4, left panel).
In the presence of the RNA sequence GUUG, the wild type protein seemed to be more structured in the regions with an alpha-helix conformation. The region spanning residues 50-75 showed a tendency to change from a turn to a bend structure, and importantly, three beta-sheets and a series of alpha-helices at the C-terminus of the protein were consolidated. Interestingly, EhCFIm25*L135T presents the same stabilization of bend structures and alpha helices as described for the WT protein. A different behavior is observed for the Y236A variant, where the helices are smaller at the C-terminal region but larger in the region of residues 50-75.
Altogether, these data suggested that the C-terminus of WT and that of the EhCFIm25*L135T protein were more structured in the presence of the GUUG molecule (Figure 4, right panel).

EhCFIm25-GUUG interaction

To obtain a more detailed description of the EhCFIm25-GUUG interaction, we quantified the frequencies of the contacts between amino acids residues and the ribonucleotides involved. Table 1 shows the distribution of the interactions of G1, U2, U3, G4 with each protein variant. Particularly, EhCFI25*L135T presented approximately the same number of contacts that WT, 25 and 23 respectively, whereas the other variants obtained greater contacts (37, 33 and 41). Moreover, the 56% of the interactions of WT or EhCFIm25*L135T with the GUUG were by G4. Regarding the

amino acid residues, the 26.8% of the interactions of WT with GUUG involved the lysines at position 79, 130, and 183. Similarly, for EhCFIm25*L135T the lysine residues at 130 and 183 participated in 20% of the contacts with RNA sequence, although isoleucines 133, 200, 202, and 243 also contributed with other 20% of the interactions. In the variants EhCFIm25*L135A, EhCFIm25*Y236A and EhCFIm25*Y236F is also remarkable the significant role of the G4 and lysines but to a lesser extent or with a different distribution (Table 1). In Figure 2, we delimited four zones of interest for their variations in the RMSF values, these regions contain residues of interest revealed for Table 1 as follows: 1) Zone a (70-90): 78, 79, 80, 81, 82, 83, 84, 86, 87; Zone b (120-135): 123, 124, 129, 130, 131, 132, 133, 134, 135; Zone c (140-200): 158, 178, 181, 182, 183, 184, 185, 198, 200; and Zone d (225-240): 233, 234, 235, 236, 237, 239, 240. A three-dimensional representation of the area of GUUG interaction in EhCFIm25 is represented by black dots, which correspond to the region defined by the residues listed in Table 1. Interestingly, this cleft of interaction with GUUG is surrounded by the zones a, b, c, and d (Figure 5A), which contains the lysine 79, 130 and 183, respectively (Figure 5B).

Discussion

Complex secondary structure elements of RNA such as stem–loops and bulges tend to be a target of interaction for proteins [31]. Additionally, non-Watson– Crick base pairing occurs in loop regions of RNA structures, and proteins preferentially identify these regions [32]. However, the focus of the RNA-protein interactions has been the identification of recurring RNA recognition motifs such as the ribonucleoprotein (RNP) and arginine-rich motifs and the particular interactions within individual complexes [31,33,34].

In higher eukaryotes, cleavage, and polyadenylation of pre-messenger RNA (pre-mRNA) 3´ end, a fundamental process for transcript maturation and gene expression regulation, is regulated by several RNA binding proteins that interact with RNA 3’ UTR [1-3]. Although protein-RNA complexes present some diversity in binding sites, van der Waals and hydrogen bond contacts and interactions with guanine and uracil are the most representative. Also, amino acids such as arginine (positive), phenylalanine (aromatic), and tyrosine (aromatic) have been reported as key participants in the binding sites [35]. In the case of the Cleavage factor CFIm25 that binds to the UGUA sequence, the Nudix domain contains the majority of the residues involved in RNA binding activity through hydrogen-bonding via main-chain and side-chain atoms, aromatic stacking, and peptide bond stacking. Some residues located outside the Nudix domain are also essential for RNA binding activity [10]. We previously indicated

that only substitution of two conserved Leu and Tyr residues at position 135 and 236, respectively, affects the RNA binding capacity of the EhCFIm25 protein of *E histolytica* [17]. To better understand how these mutations affect the structural behavior of EhCFIm25 and its function, we performed molecular dynamics (MD) simulations of WT and mutant proteins, alone or in the presence of the GUUG RNA fragment. Although punctual mutations did not affect the predicted three-dimensional structure of EhCFIm25, differences in the RMSD trajectories of each variant suggest the existence of small modifications in the path of transition of structures from the starting conformation to the final state. RMSF data evidenced that L135A and Y236F changes produce a significant increase in the flexibility of the molecule, mainly in the zone a, but also in other regions of the polypeptide. These alterations could explain the loss of RNA binding activity of EhCFIm25*L135A, and EhCFIm25*Y236F, respectively. In contrast, EhCFIm25*L135T and EhCFIm25* Y236A conserved a three-dimensional structure more similar to the WT protein. Interestingly, EhCFIm25*L135T was the only mutant that retained a slight RNA binding activity. Similar to other Nudix proteins, EhCFIm25 adopts a typical α/β/α fold [36]. In human homologous protein CFIm25, it has been previously reported that the loop connecting β2 and α1 (residues 51–60) acts as a strap that occludes the canonical Nudix substrate-binding pocket, which makes it an essential part of the RNA recognition pocket [10]. Regarding *E. histolytica*, our results suggested that the “a” zone stability is important for EhCFIm25 to interact with the RNA, also for EhCFIm25*L135T although to a lesser extent. On the other hand, Sandhu and Dash [37] suggested that changes in the conformation from a coil to helix regulate the recognition of nucleic acid acting as ‘‘molecular switches’’. This mechanism was mainly observed in the WT protein and EhCFIm25*L135T, but with minor intensity in the other mutant proteins which have lost their ability to interact with the RNA.

We hypothesized that the loss of protein function relies on alterations in intramolecular interactions in EhCFIm25*L135A, EhCFIm25*Y236A, and EhCFIm25*Y236F proteins. In agreement with this assumption, the presence of Ala at position 135 caused that neighboring residues occupy the intramolecular space left vacant by the absence of the side chain of the original Leu residue, which led to a molecular rearrangement in the EhCFIm25*L135A mutant protein. In contrast, both Leu and Thr have a side chain of similar size; as a result, the intramolecular space around this position was not significantly affected, and the amino acid structure did not vary in the mutant EhCFIm25*L135T protein, explaining why this variant keeping a slight RNA binding capacity. However, the difference between Leu and Thr polarity could somewhat produce a rearrangement of the internal protein structure and modify the relative

position of several amino acids. Typically, leucine's side chain prefers to be buried in protein hydrophobic cores, while threonine is more reactive due to its hydroxyl group [38]. Our results suggested that Leu135, namely its side chain, is essential for the structural equilibrium of the EhCFIm25 protein rather than for direct interaction with RNA.

On the other hand, a minimum change either in the side chain or the polarity of the residue at position 236 disrupted the RNA-protein interaction, which indicated that a direct interaction might occur between the Y236 residue and the RNA molecule. It has been described that the hydroxyl group of the Tyrosine side chain contributes favorably to protein stability through hydrogen bond formation [39]. Significant structural changes occurred when Y236 was replaced by either Ala or Phe, suggesting that these mutations produce substantial variations in the RNA pocket conformation. Some of the interactions that were evidenced by the MD simulations were those of the lysines 79, 130 and 183 interacting with the GUUG, particularly the electrical charge might be determinant. These data showed that intramolecular interactions and therefore protein stability is altered in the mutant proteins, and pointed out the relevant role of the hydroxyl group of Tyr236 for RNA-protein interaction.

In conclusion, MD simulations of EhCFIm25 variants and analysis of their intramolecular interactions network contributed to explaining the loss of RNA binding capacity of the mutant proteins. Notably, they confirmed that the side-chain of Leu135 residue is essential for the structural equilibrium of the EhCFIm25 protein, while the hydroxyl group of Tyr236 is directly relevant for RNA binding. Experiments currently in progress will determine the kinetics and affinity of protein-RNA interaction, which would contribute to gain insight into the role of EhCFIm25 in mRNA 3´end formation.

REFERENCES

1. Jacobson A, Peltz SW (1996). Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. Annu Rev Biochem. 65:693-739. doi: 10.1146/annurev.bi.65.070196.003401.
2. Wickens M, Anderson P, Jackson RJ (1997). Life and death in the cytoplasm: messages from the 3´ end. Curr Opin Genet Dev. 2:220-32. https://doi.org/10.1016/S0959-437X(97)80132-3.
3. Garneau NL, Wilusz J, Wilusz CJ (2007). The highways and byways of mRNA decay, Nat Rev Mol Cell Biol. 2:113-26. doi:10.1038/nrm2104.
4. Elkon R, Ugalde AP, Agami R (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet. 7:496-506. doi: 10.1038/nrg3482.

5. Rehfeld A, Plass M, Krogh A, Friis-Hansen, L (2013). Alterations in polyadenylation and its implications for endocrine disease. Front Endocrinol (Lausanne). 4;53. doi: 10.3389/fendo.2013.00053.
6. McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL. (1997). 5´-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. 24:3306-18. doi:10.1101/gad.11.24.3306.
7. Hirose Y, Manley JL (1998). RNA polymerase II is an essential mRNA polyadenylation factor. *Nature,* 6697:93-6. doi:10.1038/25786.
8. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, Frank J, Manley JL (2009). Molecular architecture of the human pre-mRNA 30 processing complex. Mol Cell. 3:365–376. doi: 10.1016/j.molcel.2008.12.028.
9. Yang Q, Coseno M, Gilmartin GM, Doublié S (2011). Crystal structure of a human cleavage factor CFIm25/CFIm68/RNA complex provides an insight into poly(A) site recognition and RNA looping. Structure, 19: 368-377. doi:10.1016/j.str.2010.12.021.
10. Yang Q, Gilmartin GM, Doublié S. (2010). Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3′ processing. Proc Natl Acad Sci U S A. 22:10062-7. doi: 10.1073/pnas.1000848107.
11. Kubo T, Wada T, Yamaguchi Y, Shimizu A, Handa H (2006). Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 30-UTRs. Nucleic Acids Res. 21: 6264–6271. doi: 10.1093/nar/gkl794.
12. Gilmartin GM (2005). Eukaryotic mRNA 3' processing: a common means to different ends. Genes Dev. 21:2517-2521. doi: 10.1101/gad.1378105.
13. Li H, Tong S, Li X, Shi H, Ying Z, Gao Y, Ge H, Niu L, Teng M (2011). Structural basis of pre-mRNA recognition by the human cleavage factor Im complex. Cell Res. 7:1039-51. doi: 10.1038/cr.2011.67.
14. Ali IK, Clark CG, Petri WA (2008). Molecular epidemiology of amebiasis. Infect Genet Evol. 5:698-707. doi: 10.1016/j.meegid.2008.05.004.
15. López-Camarillo C, Luna-Arias JP, Marchat LA, Orozco E (2003). EhPgp5 mRNA stability is a regulatory event in the Entamoeba histolytica multidrug resistance phenotype. J Biol Chem. 13:11273-80. doi: 10.1074/jbc.M211757200.
16. Pezet-Valdez M, Fernández-Retana F, Ospina-Villa JD, Ramírez-Moreno ME, Orozco E, Charcas-López S, Soto-Sánchez J, Mendoza-Hernández G, López-Casamicha M, López-Camarillo C, Marchat LA (2013). The 25 kDa Subunit of Cleavage Factor Im Is a RNA-Binding Protein That Interacts with the Poly(A)

Polymerase in Entamoeba histolytica. PLoS One. 6:e67977. doi: 10.1371/journal.pone.0067977.

17. Ospina-Villa JD, Zamorano-Carrillo A, López-Camarillo C, Castañon-Sanchez CA, Soto-Sánchez J, Ramírez-Moreno E, Marchat LA (2015). Amino acid residues Leu135 and Tyr236 are required for RNA binding activity of CFIm25 in Entamoeba histolytica. *Biochimie,* 115; 44-51. doi: 10.1016/j.biochi.2015.04.017.
18. Priya Doss CG, Rajith B, Garwasis N. Mathew PR, Raju AS, et al. (2012). Screening of mutations affecting protein stability and dynamics of FGFR1-A simulation analysis. Appl Transl Genom. 1:37-43. doi:10.1016/j.atg.2012.06.002
19. Zhang YJ, Ding JN, Zhong H, Han JG (2017). Exploration micromechanism of VP35 IID interaction and recognition dsRNA: A molecular dynamics simulation. Proteins. 6; 1008-1023, doi: 10.1002/prot.25269.
20. Chang S. Zhang DW, Xu L, Wan H, Hou TJ, Kong R (2016). Exploring the molecular basis of RNA recognition by the dimeric RNA-binding protein via molecular simulation methods. RNA Biology, 11; 1133-1143. doi:10.1080/15476286.2016.1223007.
21. Krepl M, Cléry A, Blatter M, Allain FH, Sponer J (2016). Synergy between NMR measurements and MD simulations of protein/RNA complexes: application to the RRMs, the most common RNA recognition motifs. Nucleic Acids Res. 13:6452-70. doi: 10.1093/nar/gkw438.
22. Lambert C, Leonard N, De Bolle X, Depiereux E (2002). ESyPred3D: Prediction of proteins 3D structures. Bioinformatics*,* 18; 1250-1256. doi: 10.1093/bioinformatics/18.9.1250.
23. Bowie JU, Lüthy R, Eisenberg D (1991). A method to identify protein sequences that fold into a known three-dimensional structure. Science, 253; 164-170. doi: 10.1126/science.1853201.
24. Tuszynska I, Bujnicki JM (2011). DARS-RNP and QUASI-RNP: New statistical potentials for protein-RNA docking. BMC Bioinformatics 12:348, https://doi.org/10.1186/1471-2105-12-348.
25. Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM (2015). NPDock: a web server for protein-nucleic acid docking. Nucleic Acids Res. W1;W425-30. doi: 10.1093/nar/gkv493.
26. Guex N, Peitsch MC (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 15; 2714-23. doi: 10.1002/elps.1150181505.

27. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput. 3; 435-47. doi: 10.1021/ct700301q.
28. Foloppe N, MacKerell AD (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J. Comput. Chem. 21; 2. doi: 10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G.
29. Smith PE, van Gunsteren WF (1993). The viscosity of SPC and SPC/E water at 277 and 300 K. Chem. Phys. Lett. 4; 315-318. doi: 10.1016/0009-2614(93)85720-9.
30. Sobolev V Eyal E, Gerzon S, Potapov V, Babor M, Prilusky J, Edelman M (2005). SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment. Nucleic Acids Res 33; 39-43. doi: 10.1093/nar/gki398.
31. Nagai K (1996). RNA-protein complexes. Curr Opin Struct Biol. 1; 53-61. https://doi.org/10.1016/S0959-440X(96)80095-9.
32. Steitz TA (1999). RNA Recognition by Proteins In: Gesteland RF (ed) The RNA World, 2nd edition, Cold Spring Harbor Monograph, pp 427-450.
33. Fornerod M (2012). RS and RGG repeats as primitive proteins at the transition between the RNA and RNP worlds. Nucleus. 3; 4-5. http://dx.doi.org/10.4161/nucl.18631.
34. De Guzman RN, Turner RB, Summers MF (1998). Protein-RNA recognition. Biopolymers. 2-3;181-95. doi: 10.1002/(SICI)1097-0282(1998)48:2<181::AID-BIP7>3.0.CO;2-L.
35. Jones S, Daley DT, Luscombe NM, Berman HM, Thornton JM. (2001). Protein-RNA interactions: a structural analysis. Nucleic Acids Res. 4;943-54. https://doi.org/10.1093/nar/29.4.943.
36. Shen V, Kiledjian M (2006). Decapper Comes into Focus. Structure. 2; 171–172. doi:10.1016/j.str.2006.01.002.
37. Sandhu KS, Dash D (2007). Dynamic alpha-helices: conformations that do not conform. Proteins. 1; 109-22. doi:10.1002/prot.21328.
38. Betts MJ, Russell RB (2003) Amino Acid Properties and Consequences of Substitutions, in Bioinformatics for Geneticists (eds M. R. Barnes and I. C. Gray), John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/0470867302.ch14.
39. Pace CN, Horn G, Hebert EJ, Bechert J, Shaw K, Urbanikova L, Scholtz JM, Sevcik J (2001). Tyrosine Hydrogen Bonds Make a Large Contribution to Protein Stability. J Mol Biol. 2:393-404. doi:10.1006/jmbi.2001.4956.

Conflict of interest

All authors agree with the version that is submitted here and declare no conflict of interest.

Disclosure

The work described has not been published previously, and it is not under consideration for publication elsewhere. The funding sources had no involvement study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

Figure legend

Figure 1. Three-dimensional models of EhCFIm25 variants interacting with the GUUG molecule and time evolution of α-carbon RMSD. A) Three-dimensional structure of the EhCFIm25-GUUG complex obtained from the molecular docking test and close up view showing the protein-RNA interaction pocket with the position of Leu135 and Tyr236 residues. B-F) Close up views of the RNA binding pocket in the EhCFIm25variants. B) EhCFIm25-WT, C) EhCFIm25*L135A, D) EhCFIm25*L135T, E) EhCFIm25*Y236A, and F) EhCFIm25*Y236F. Residues selected for mutation are indicated; the RNA molecule is shown in black. G) and H) RMSD values of the structures without and with GUUG, respectively.

Figure 2. RMSF values of free EhCFIm25 variants. Comparison of RMSF values of wild-type EhCFIm25 protein *versus* mutant EhCFIm25*L135A (A), EhCFIm25*L135T (B), EhCFIm25*Y236A (C) and EhCFIm25*Y236F (D) proteins (upper panels). Representation of the secondary structure of the wild-type protein is shown on each plot (lower panels). Flexible regions are indicated in yellow (zone a: 70-90 amino acids), blue (zone b: 120-135 amino acids) and red (zone c: 150-170 amino acids).

Figure 3. RMSF values of EhCFIm25 variants interacting with the GUUG molecule. Comparison of RMSF values of wild-type EhCFIm25 protein *versus* mutant EhCFIm25*L135A (A), EhCFIm25*L135T (B), EhCFIm25*Y236A (C) and EhCFIm25*Y236F (D) proteins (upper panels). Representation of the secondary structure of the wild-type is shown on each plot (lower panels). Flexible regions are indicated in yellow (zone a: 70-90 amino acids), blue (zone b: 120-135 amino acids), red (zone c: 140-200 amino acids), and green (zone d: 225-240 amino acids).

Figure 4. Variation in the secondary structure of EhCFIm25. Data of EhCFIm25 variants, free (left panels) and bound to the GUUG RNA sequence (right panels) during MD simulations. The name of each protein is indicated at the left. Each secondary structure is represented according to the color code shown at the bottom. X axis, time of simulation; Y axis, residues in each protein.

Figure 5. RNA-Interacting region in the EhCFIm25. Schematic representation of the 3D structure of EhCFIm25 without (A) and with (B) the GUUG RNA motif. The four regions detected in Figures 2 and 3 are illustrated as follow: a (yellow), b (blue), c (red), and d (green).

Table 1. Frequency distribution of contacts between the amino acid residues of EhCFIm25 variants and the GUUG RNA molecule.

Fig 1

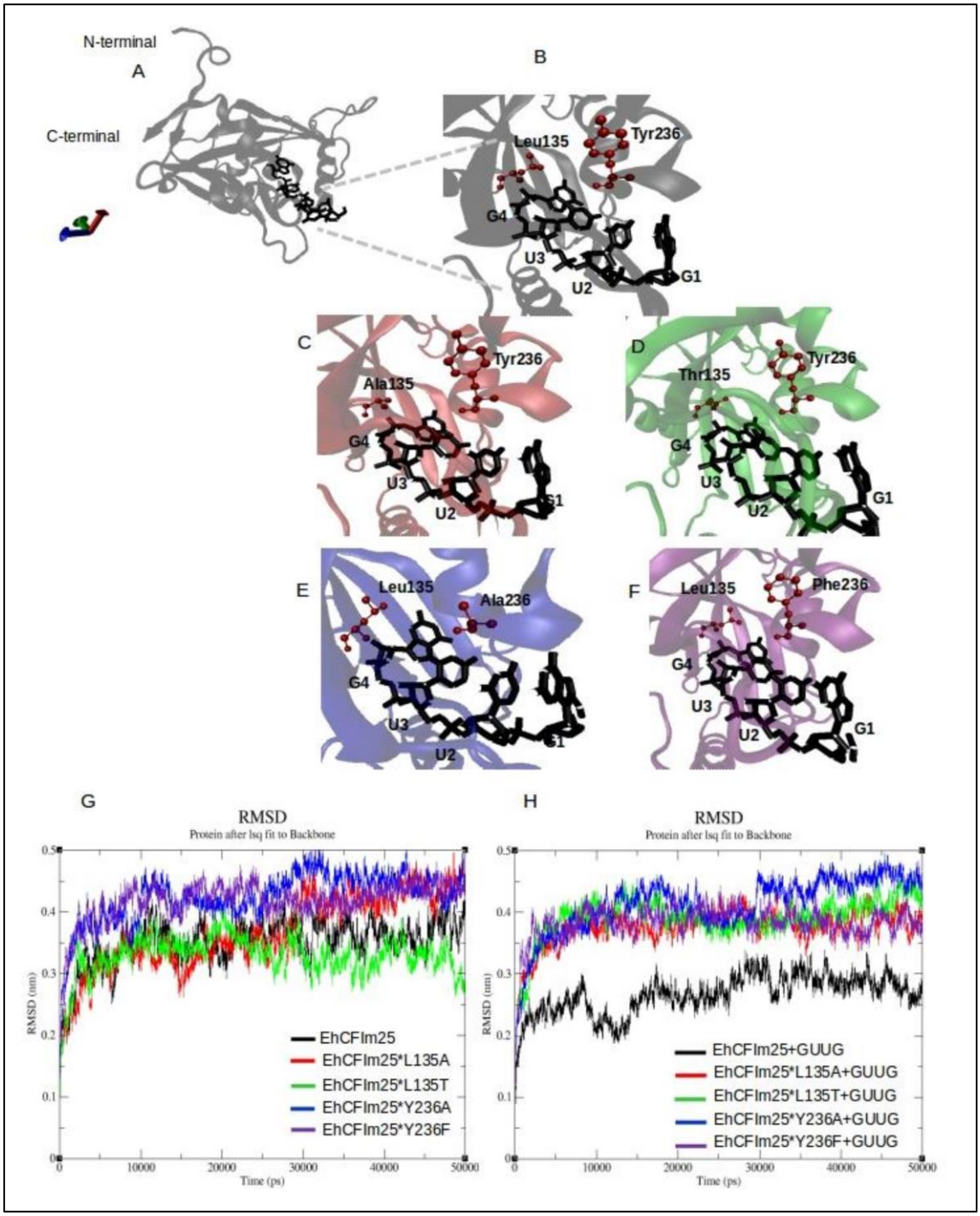

N-terminal
A
C-terminal
B
Leu135
Tyr236
G4
U3
U2
G1
C
Tyr236
Ala135
G4
U3
U2
G1
D
Tyr236
Thr135
G4
U3
U2
G1
E
Leu135
Ala236
G4
U3
U2
F
Leu135
Phe236
G4
U3
U2
G1
G
RMSD
Protein after lsq fit to Backbone
RMSD (nm)
Time (ps)
EhCFIm25
EhCFIm25*L135A
EhCFIm25*L135T
EhCFIm25*Y236A
EhCFIm25*Y236F
H
RMSD
Protein after lsq fit to Backbone
RMSD (nm)
Time (ps)
EhCFIm25+GUUG
EhCFIm25*L135A+GUUG
EhCFIm25*L135T+GUUG
EhCFIm25*Y236A+GUUG
EhCFIm25*Y236F+GUUG

Fig 2

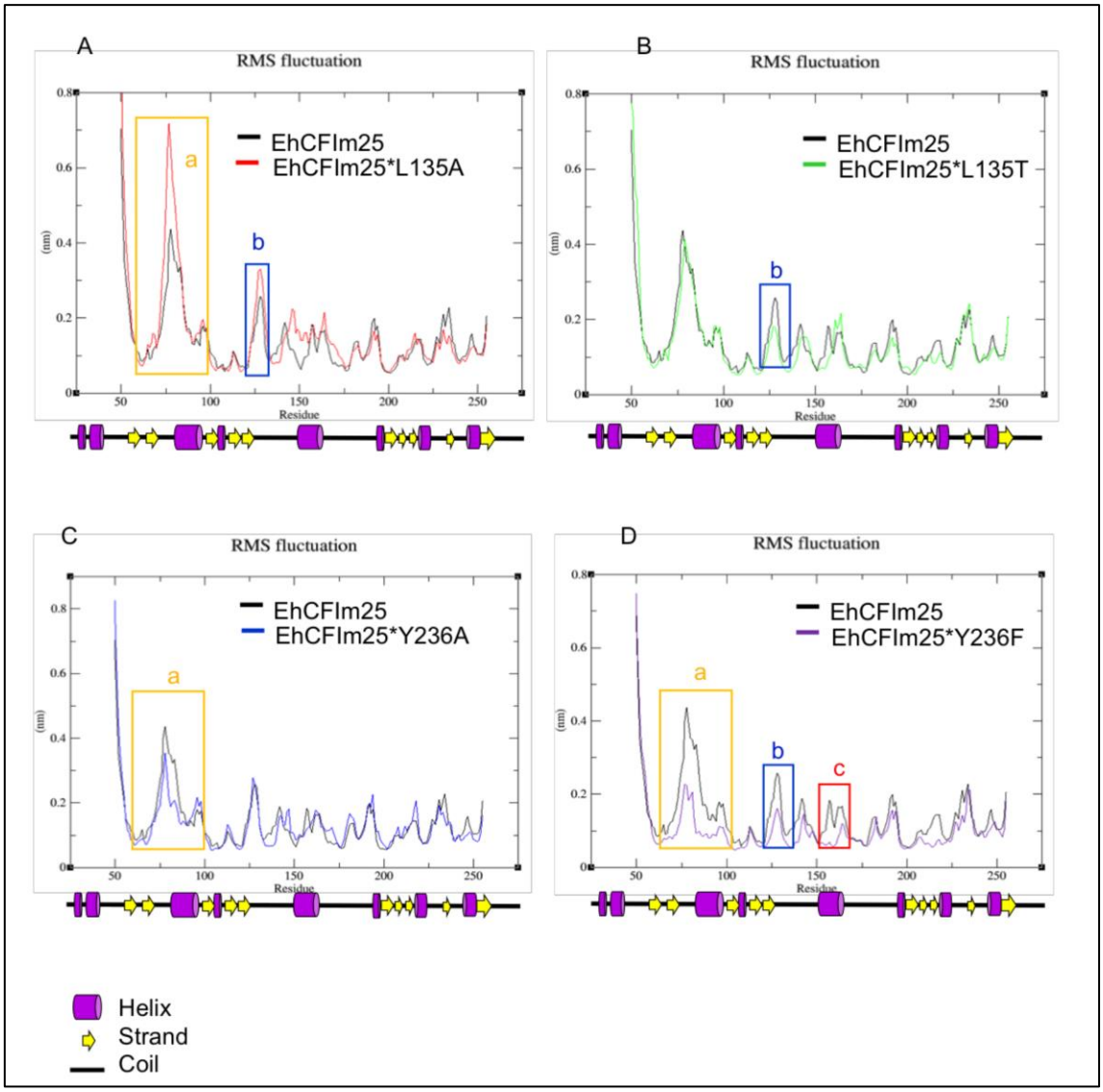

A
RMS fluctuation
EhCFIm25
EhCFIm25*L135A
a
b
(nm)
Residue
B
RMS fluctuation
EhCFIm25
EhCFIm25*L135T
b
(nm)
Residue
C
RMS fluctuation
EhCFIm25
EhCFIm25*Y236A
a
(nm)
Residue
D
RMS fluctuation
EhCFIm25
EhCFIm25*Y236F
a
b
c
(nm)
Residue
Helix
Strand
Coil

Fig 3

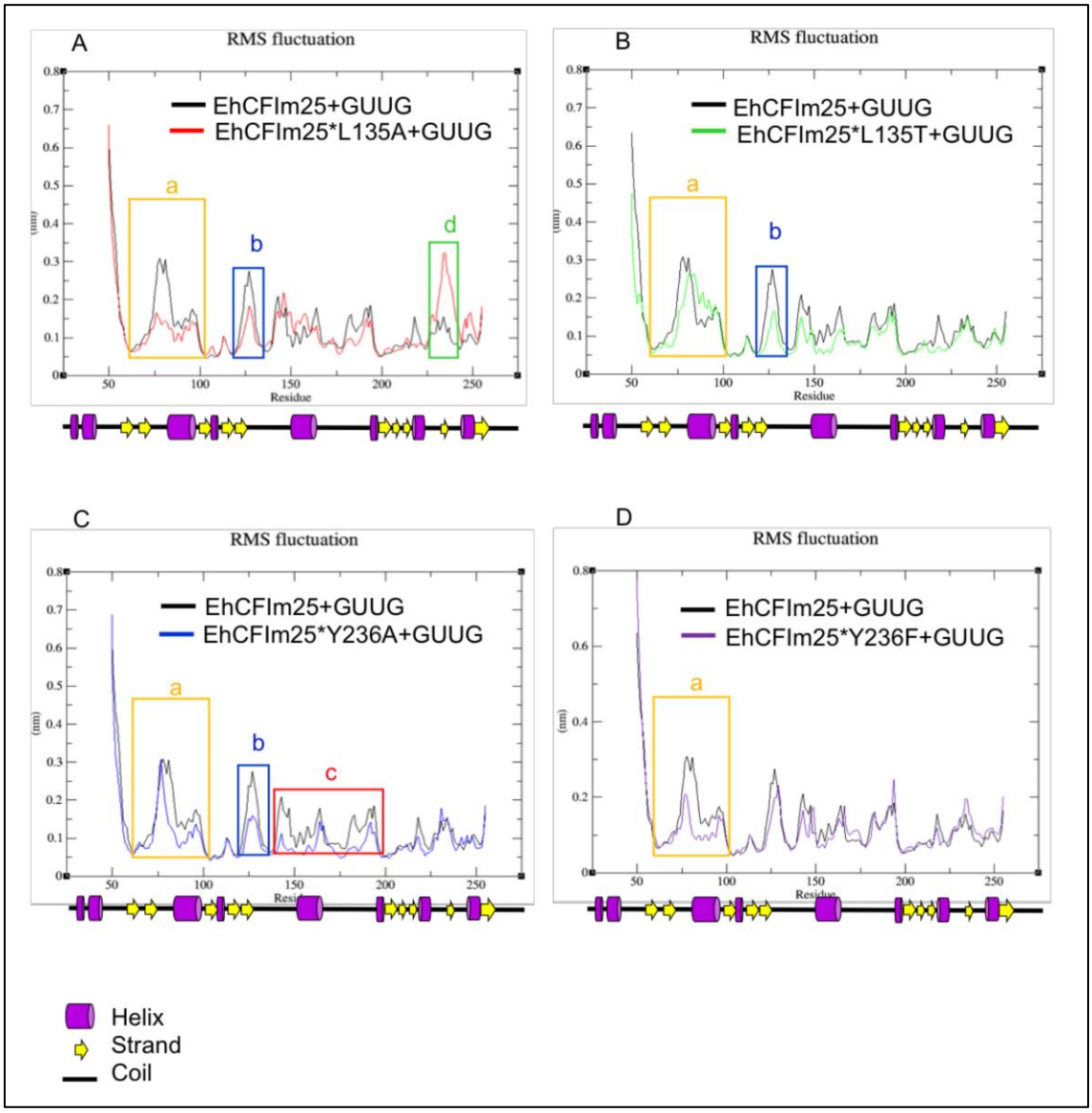

A
RMS fluctuation
EhCFIm25+GUUG
EhCFIm25*L135A+GUUG
a
b
d
Residue
B
RMS fluctuation
EhCFIm25+GUUG
EhCFIm25*L135T+GUUG
a
b
Residue
C
RMS fluctuation
EhCFIm25+GUUG
EhCFIm25*Y236A+GUUG
a
b
c
D
RMS fluctuation
EhCFIm25+GUUG
EhCFIm25*Y236F+GUUG
a
Residue
Helix
Strand
Coil

Fig4

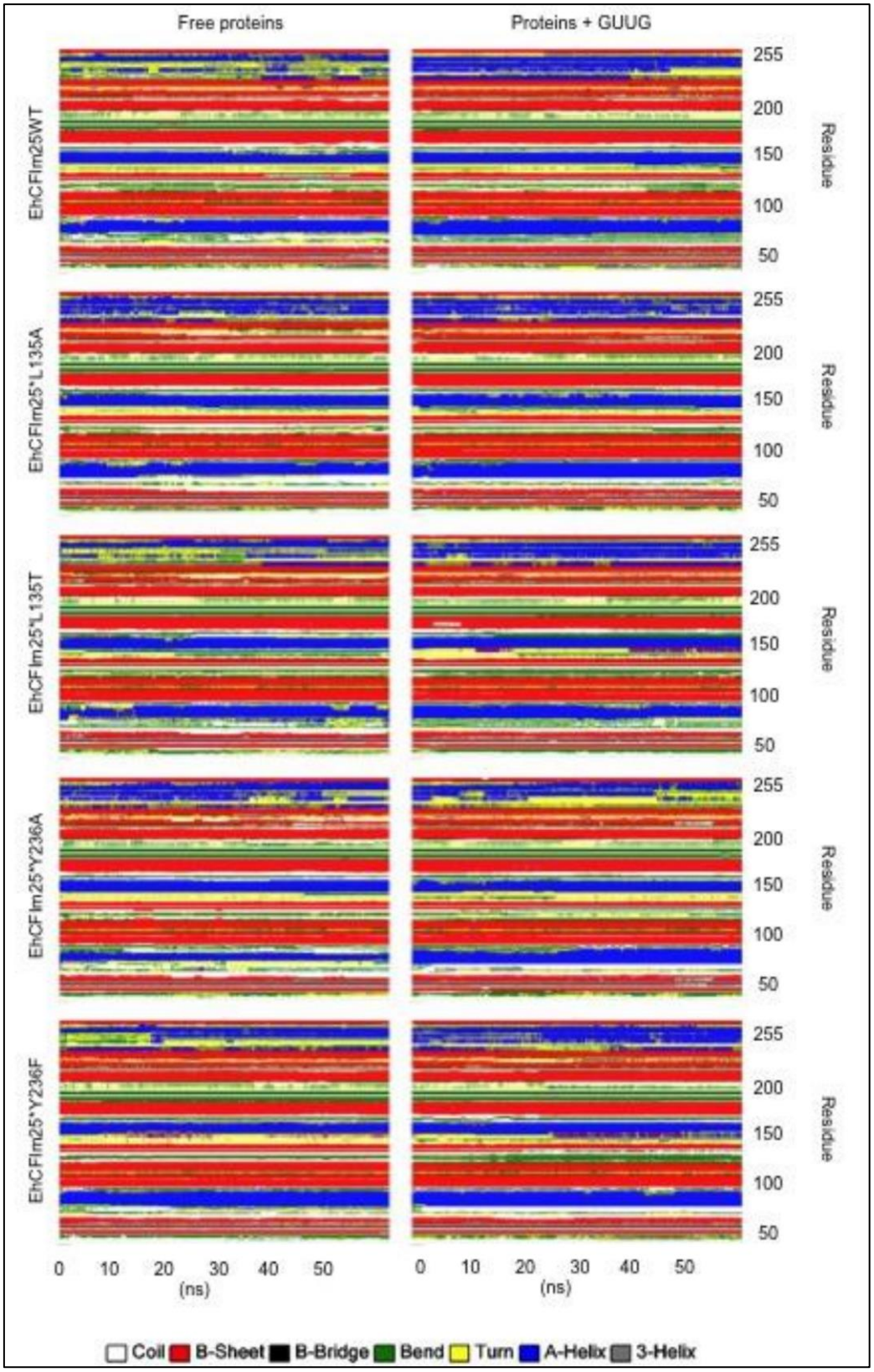

Free proteins
Proteins + GUUG
EhCFIm25WT
EhCFIm25*L135A
EhCFIm25*L135T
EhCFIm25*Y236A
EhCFIm25*Y236F
Residue
255
200
150
100
50
0
10
20
30
40
50
(ns)
Coil
B-Sheet
B-Bridge
Bend
Turn
A-Helix
3-Helix

Fig5

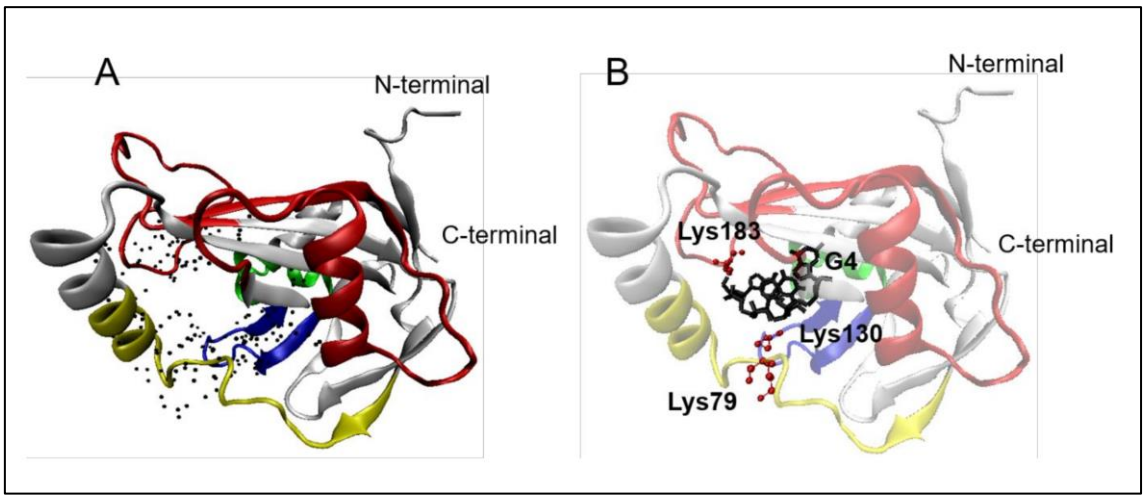

A
N-terminal
C-terminal
B
N-terminal
Lys183
G4
C-terminal
Lys130
Lys79

Table 1

Table 1. Frequency distribution of contacts between the amino acid residues of EhCFIm25 variants and the GUUG RNA molecule.

Residues	G1	U2	U3	G4	Total	Frequency (%)
EhCFIm25WT+GUUG						
Gly (137,138)	0	0	0	2	2	8,69
Leu (78,135)	0	0	0	2	2	8,69
Phe (83)	0	0	1	1	2	8,69
Asn (124)	0	0	1	0	1	4,34
Gln (86)	0	0	0	1	1	4,34
Tyr (105)	0	0	0	1	1	4,34
Lys (79,130,183)	0	2	**3**	1	**6**	**26,08**
Arg (139,178)	0	0	1	2	**3**	13,04
His (80,134)	0	0	1	2	**3**	13,04
Glu (132, 198)	0	0	1	1	2	8,69
Total	0	2	8	13	23	100
Frequency (%)	0	8,69	34,78	**56,52**	100	-
EhCFIm25*l135A+GUUG						
Gly (138)	0	0	0	1	1	2,70
Leu (129,240)	2	0	0	1	3	8,70
Phe (83)	0	1	1	0	2	5,40
Asn (237)	1	0	0	0	1	2,70
Gln (81)	0	1	2	1	4	10,81
Tyr (105,236)	1	0	0	1	2	5,40
Lys (79,130,158,183,235)	**3**	2	1	1	**7**	**18,91**
Arg (101)	0	0	0	1	1	2,70
His (80, 134)	0	1	2	1	4	10,81
Glu (132)	0	1	1	0	2	5,40
Thr (82,239)	1	0	1	1	3	8,70
Asp (131)	1	1	0	0	2	5,40
Ile (133)	1	1	1	1	4	10,81
Ala (135)0	0	0	0	1	1	2,70
Total	10	8	9	10	37	100
Frequency (%)	**27,02**	21,62	24,32	**27,02**	100	-
EhCFIm25*L135T+GUUG						
Leu (129,185,240)	1	1	0	2	4	16
Asn (237)	0	1	1	0	2	8
Tyr (105)	0	0	0	1	1	4
Lys (130, 183)	0	2	2	1	**5**	**20**
Arg (178)	0	0	0	1	1	4
His (134)	0	0	0	1	1	4
Glu (132)	0	0	1	0	1	4
Thr (135,239)	0	0	0	2	2	8
Asp (131)	0	0	1	0	1	4
Ile (133,200,202,243)	0	0	1	4	**5**	**20**
Val (107)	0	0	0	1	1	4
Ser (242)	0	0	0	1	1	4
Total	1	4	6	14	25	100
Frequency (%)	4	16	24	**56**	100	-
EhCFIm25*Y236A+GUUG						
Gly (137,138)	0	0	0	2	2	6,06
Leu (135,185)	0	0	2	2	**4**	**12,12**
Phe (83)	0	0	1	1	2	6,06
Asn (184,237)	1	2	0	0	3	9,09
Tyr (105)	0	0	0	1	1	3,03
Lys (130,183,235)	1	2	1	0	**4**	**12,12**

Arg (178)	0	0	1	1	2	6,06
His (134)	0	0	1	1	2	6,06
Glu (132,198,234)	1	0	1	2	**4**	**12,12**
Thr (239)	1	0	0	1	2	6,06
Asp (131)	0	1	0	0	1	3,03
Ile (133,200,202)	0	0	1	2	3	9,09
Ala (236)	1	0	0	0	1	3,03
Gln (81,86)	0	0	0	2	2	6,06
Total	5	5	8	15	33	100
Frecquency (%)	15,15	15,15	24,24	**45,45**	100	-
EhCFIm25*Y236F+GUUG						
Gly (84,137,138)	1	0	0	2	3	7,31
Leu (135,185)	0	0	0	2	2	4,87
Asn (237)	0	1	1	1	3	7,31
Tyr (105,181)	1	0	0	1	2	4,87
Lys (130,183,235)	1	3	3	0	**7**	**17,07**
Arg (178)	0	0	0	1	1	2,43
His (80,134)	0	0	2	2	4	9,75
Glu (132,234)	0	1	1	0	2	4,87
Thr (82,239)	0	0	1	2	3	7,31
Asp (131,182)	1	0	1	0	2	4,87
Ile (133,200,202)	0	0	1	2	3	7,31
Gln (81,86)	0	0	0	2	2	4,87
Phe (83,233,236)	1	1	2	1	**5**	12,19
Met (87)	1	0	0	0	1	2,43
Ser (123)	0	0	1	0	1	2,43
Total	6	6	13	**16**	41	100
Frequency (%)	14,63	14,63	31,70	**39,02**	100	-

DISCUSIÓN GENERAL

La amibiasis intestinal ocasionada por el parásito protozoario *Entamoeba histolytica* continúa siendo un problema grave de salud pública en numerosos países en vías de desarrollo. Las reacciones adversas asociadas a los medicamentos actuales y la existencia de casos de resistencia a los mismos (Bansal, *et al.,* 2006), hacen necesaria la identificación de blancos bioquímicos para la búsqueda de nuevos tratamientos o vacunas que permitan el control de la amibiasis. Debido a la relevancia de la poliadenilación de los transcritos en la expresión génica, en este trabajo nos propusimos analizar molecular y funcionalmente al factor de poliadenilación EhCFIm25 para evaluar su interés como blanco terapéutico.

El silenciamiento o inhibición de la expresión génica es una herramienta que permite entender la función de los genes e identificar potenciales blancos para la creación de nuevos fármacos. El genoma de *E. histolytica* fue publicado en 2005 (Loftus et al., 2005), y de este se predijeron 9.938 genes que tienen un tamaño promedio de 1.17 kb y corresponden al 49% del genoma. Particularmente, el parásito posee algunos genes que participan en el procesamiento de ARN de interferencia (ARNi) como son: tres Ago (de la familia Argonauta) (EHI_125650, EHI_186850 y EHI_177170) que contienen los dominios PAZ y Piwi, un gen RdRP (EHI_139420) y otro gen con un dominio RdRP parcial (EHI_179800). *E. histolytica* carece de genes que codifican para la proteína Dicer (Zhang *et al.,* 2011). Morgado *et al.,* en una revisión del 2016 reporta que *E. histolytica* tiene una vía de ARNi endógena robusta y no canónica que regula la expresión génica. Recientemente se caracterizó la EhARNasaIII, una proteína Dicer no canónica que tiene la capacidad de procesar los dsARN en pequeños fragmentos de ARN que contribuyen de forma productiva al silenciamiento génico (Pompey *et al.,* 2015). En *E. histolytica* ya se ha demostrado que los genes blancos de ARN pequeños son efectivamente silenciados (Bracha *et al.,* 2003). Así, los ARN pequeños que silencian el gen *Ehap*-a (ameboforo) en la clona G3 de *E. histolytica* se

observaron en una localización nuclear y el gen silenciado se encontró en contacto con EhAgo2-2 (Zhang et al., 2011). Por otro lado, la síntesis de dsARN por medio de bacterias *E. coli* HT115 transformadas con el vector plasmídico L4440 es una técnica económica, efectiva, rápida y fácil de realizar que permite obtener cantidades ilimitadas de moléculas de dsARN dirigidas contra el blanco seleccionado, a diferencia de los ARNi comerciales que son costosos y distribuidos en una cantidad limitada. (Solis et al., 2009). En conjunto, estos datos muestran que los trofozoítos tienen la capacidad de procesar las moléculas de dsRNA para que puedan llevar a cabo su función para silenciar o bloquear EhCFIm25.

En el presente estudio, mostramos que las moléculas de *EhCFIm25*-dsARN producidas en bacterias permiten obtener una reducción significativa en la expresión génica de la proteína blanco EhCFIm25 (20%, 80%, y 100% a los días 2, 4 y 6 respectivamente), al adicionar una sola dosis de 100 mg/ml en el medio de cultivo para que los trofozoítos los interiorizan por medio de fagocitosis (método conocido como *soaking* en inglés o remojo en español) y los procesen mediante el uso de las proteínas antes mencionadas. De manera interesante, el silenciamiento de la proteína EhCFIm25 disminuye la viabilidad celular y acelera la muerte de los trofozoítos, lo que indica que es una proteína esencial para la sobrevivencia del parásito.

La ausencia de EhCFIm25 también afecta sus propiedades de virulencia, alterando la movilidad y la eritrofagocitosis. Estos cambios correlacionan con el hecho de que los trofozoítos aumentan de tamaño, lo que probablemente conlleva a una disminución de la flexibilidad de la membrana plasmática, ya sea para el movimiento del parásito, la fagocitosis o la división celular. De acuerdo a eso, la multinucleación de los trofozoítos gigantes podría resultar de fallas en el proceso de la citocinesis, como se ha demostrado por la sobreexpresión de EhPC4 (Hernández de la Cruz *et al.,* 2014; Hernández de la Cruz *et al.,* 2016).

Por otra parte, al inhibir la expresión de EhCFIm25 en los trofozoítos de *E. histolytica,* se observó que el corte del 3´UTR de los ARNm modelos utilizados

se lleva a cabo preferentemente en el sitio más proximal al codón de paro, demostrando de esta manera que EhCFIm25 juega un rol importante en el mecanismo de poliadenilación alternativa de los transcritos. El uso del sitio de corte/poliadenilación proximal ocasiona la pérdida de secuencia en el 3'UTR que puede ser potencialmente importantes para la estabilidad y traducción de los ARNm. Aunque solo una pequeña porción de genes en *E. histolytica* utilizan la poliadenilación alternativa (1.9% - 2.4%) (Hon *et al.,* 2013), los genes de amiba con múltiples sitios de corte/poliadenilación participan en procesos fundamentales para la célula como son: unión al ADN, traducción, *splicing*, unión al ARNm, plegamiento de proteínas y transporte de proteínas. Entonces, la ausencia de la proteína EhCFIm25 produce cambios en la expresión de estos genes, provocando defectos críticos que explican el fenotipo observado.

Es importante mencionar también que EhCFIm25 interactúa con otras proteínas de otras maquinarias de procesamiento del ARN como EhPAP y de proteínas coactivadoras de la transcripción como EhPC4, lo que indica que EhCFIm25 no solo participa en la poliadenilación de los transcritos, sino también en otros procesos de maduración del ARN, así como en la transcripción del ADN. El silenciamiento de EhCFIm25 podría entonces tener un impacto más generalizado que resulte en la muerte de los trofozoítos.

Los RNAi generalmente no son usados como herramienta terapéutica debido a su corta vida media, unión a blancos inespecíficos, toxicidad, y falta de métodos seguros para conducirlos al sitio de acción (Ambesajir et al., 2012). Por eso nuestra atención se fijó en otras moléculas de ácidos nucleicos conocidas como aptámeros, como una alternativa para alterar la expresión y función de EhCFIm25. El estudio de los aptámeros ha tomado fuerza en las últimas décadas por su gran potencial de aplicación en las ciencias biomédicas, como lo demuestra que el aumento en el número de publicaciones científicas y en el número de moléculas en desarrollo o en estudios clínicos en diferentes etapas (Rozenblum, *et al.,* 2015; Sundaram *et al.,* 2013; Yu *et al.,* 2016). Estas moléculas de ADN o ARN, también llamadas "anticuerpos químicos", tienen grandes ventajas a nivel de productividad, estabilidad, inmunogenicidad del blanco, amplia variedad blancos y posibles modificaciones químicas. Por lo que

muchas empresas farmacéuticas y biotecnológicas han fijado su atención en estas moléculas para reemplazar o mejorar las propiedades de los anticuerpos comerciales. En parasitología existen muy pocos antecedentes acerca de la identificación de aptámeros. Algunos aptámeros ya han sido seleccionados y usados en diferentes parásitos protozoarios como *T. cruzi*, *Leishmania sp*., y *Plasmodium sp*.; sin embargo, todos estos trabajos están enfocados a la búsqueda de marcadores diagnósticos, y la sobrevivencia parasitaria por inhibición de blancos por medio de aptámeros no ha sido todavía evaluada (Moreno y González 2011; Nagarkatti, *et al.,* 2014; Guerra-Pérez, *et al.,* 2015).

En nuestro trabajo, usamos la estrategia SELEX para identificar un par de aptámeros de RNA (C4 y C5) que tienen como blanco el factor de poliadenilación EhCFIm25. Ni C4 ni C5 son capaces de reconocer a la proteína mutante EhCFIm25*L135T que tiene una única mutación puntual. Hasta donde sabemos, solo existe un estudio previo al nuestro en el que se aísla un aptámero que puede seleccionar entre dos proteínas que solo se diferencian por un residuo de amino ácido (Chen *et al.,* 2015). Los aptámeros C4 y C5 tampoco reconocen la proteína homóloga de humano de o de otro protozoario, lo que confirma la especificidad de ambos aptámeros y sugiere su posible uso en el diagnóstico o tratamiento de la amibiasis.

Debido a que EhCFIm25 es una proteína que se une al RNA, el análisis de las secuencias de los aptámeros y la posterior validación por ensayos tipo REMSA permitió determinar que el motivo GUUG es el sitio de interacción de EhCFIm25 al RNA. De manera interesante, los estudios de dinámica molecular evidenciaron que la interacción de la proteína EhCFIm25 al motivo GUUG aumenta su estabilidad. Esto se da particularmente en una región de 20 a.a (posición 70-90) que cambia su conformación de loop a alpha hélice cuando la proteína pasa de estar libre a interactuar con el ARN. Este comportamiento ya ha sido observado en otros estudios (Sandhu y Dash 2007) donde el cambio de conformación de loop a hélice alfa actúa como “interruptor molecular” para el reconocimiento de ácidos nucleicos por parte de las proteínas. También se observa que la proteína unida al ARN adquiere una carga positiva mayor en el pocket de interacción, lo que favorece los contactos con la carga negativa de

los ácidos nucleicos. La proteína mutante EhCFIm25*L135T que conserva un 30% de su capacidad de interacción con el ARN (Ospina-Villa *et al.,* 2015) tiene un comportamiento más similar a la proteína EhCFIm25, en comparación con otras mutantes que no son capaces de unirse al RNA. La mayoría (56%) de los contactos entre EhCFIm25 y la secuencia UUG se establecen entre el nucleótido G4 y residuos hidrofóbicos no polares (Gly 137 y 138, Leu 78 y 135) o polares cargados positivamente (Arg 139 y 178, His 80 y 134). Por otra parte, el 26.8% de los contactos involucra residuos de Lys (79,130,183) interactuando con las bases U2 y U3. Lys es un aminoácido cargado con capacidad de formar puentes de hidrógeno con el ARN. La proteína mutante EhCFIm25*L135T tiene un comportamiento similar a la proteína silvestre con respecto a los contactos con la base G4 y los residuos involucrados en dichos contactos. Además, la ausencia de la cadena lateral del residuo Leu original permite que los residuos vecinos ocupen el espacio vacante, lo que conlleva a cambios en la estructura de EhCFIm25*L135A, mientras que los cambios son menos importantes en la proteína EhCFIm25*L135T mutante que conserva cierta capacidad de unión al RNA. La cadena lateral de los residuos de Leu se encuentran generalmente escondidos en el núcleo hidrofóbico de las proteínas, mientras que Thr es más reactivo debido a su grupo hidroxilo (Betts, & Russell, 2003). Se ha reportado que el grupo hidroxilo de Tyr contribuye favorablemente a la estabilidad de las proteínas vía la formación de enlaces de hidrogeno (Pace, Horn, Hebert, Bechert, & Shaw, et al. 2001). En conjunto estos datos sugieren que la cadena lateral de Leu es esencial para el equilibrio estructural, mientras que el grupo hidroxilo del Tyr participa directamente en la interacción con el RNA.

La misma estrategia SELEX permitió también identificar que el motivo UGUA es la secuencia reconocida por la proteína humana CFIm25 (Venkataraman, et al. 2005, Brown, et al. 2003). El hecho de que ambas proteínas homólogas reconozcan motivos de RNA diferente es muy importante pensando en las potenciales aplicaciones terapéuticas de los aptámeros ya que un aptámero que contenga el motivo GUUG podría inhibir o secuestrar específicamente a la proteína EhCFIm25 del parásito, sin afectar la proteína CFIm25 de las células del huésped humano. En ese contexto, los aptámeros C4 y C5 tienen la

capacidad de inhibir no solo la función de EhCFIm25 sino también la función de las proteínas que interactúan con ella, a diferencia de los dsRNAs específicos que silencian la expresión de la proteína EhCFIm25. Los experimentos de docking molecular muestran que las proteínas EhPAP y EhPC4 interactúan con EhCFIm25 en sitios diferentes al “bolsillo” de interacción con el motivo GUUG. Esto permite inferir que los aptámeros C4 y C5 no solo bloquearían la función de EhCFIm25, sino también la de las proteínas EhPAP y/o EhPC4 interactuando con ella, generando de tal manera alteraciones mayores en la regulación de la expresión génica.

Efectivamente, tanto C4 y C5 producen una disminución en la viabilidad celular y una acelerada muerte de los trofozoítos de *E. histolytica*, mucho mayor a la observada con los dsRNA producidos en bacterias, lo que, aunado a su falta de efecto en las células del huésped, los convierte en herramientas terapéuticas particularmente atractivas para el control del parásito. Sin embargo, es necesario realizar estudios complementarios que avalen este hallazgo y permitan el uso clínico de os aptámeros comenzando por probar su toxicidad y evaluar su efectividad en modelos animales. Los aptámeros también tienen un gran potencial como herramientas diagnósticas y terapéuticas que debería ser probado para el mejoramiento del diagnóstico e investigación en *E. histolytica*. Esperamos que con este estudio se dé un primer paso para el uso de aptámeros como herramienta diagnóstica, de investigación y terapéutica para el tratamiento de enfermedades parasitarias.

En conjunto nuestros resultados muestran que alterar la poliadenilación y particularmente el factor de poliadenilación EhCFIm25 es una estrategia interesante para el desarrollo de tratamientos contra *E. histolytica.* Esta observación concuerda con los reportes de Guerra-Pérez *et al. (*2015), Hendriks *et al.* (2003), Palencia *et al.* (2017), Sidik *et al.* (2016) y Sonoiki *et al*. (2017) que han demostrado el interés de otros factores de poliadenilación como son PABP y CPSF, para la viabilidad de los protozoarios *Leishmania , Toxoplasma, Trypanosoma* and *Plasmodium*. Estos datos abren nuevas oportunidades para la identificación de nuevos blancos moleculares para el

desarrollo de nuevos tratamientos contra los protozoarios que afectan la salud humana.

CONCLUSIONES GENERALES

- El factor de poliadenilación EhCFIm25 participa en la regulación de la selección del corte/poliadenilación en los transcritos de los trofozoítos de *E. histolytica,* y es esencial para la sobrevida y virulencia del patógeno *in vitro,* por lo que representa un blanco terapéutico interesante para el desarrollo de nuevo métodos de control de este parásito.
- Mediante el ensayo SELEX, identificamos dos aptámeros de RNA, llamados C4 y C5, que reconocen de manera específica la proteína EhCFIm25.
- La comparación de las secuencias de los aptámeros y los ensayos de interacción RNA-proteína, revelaron que el sitio reconocido por la proteína EhCFIm25 en el ARN es el motivo GUUG.
- La interacción EhCFIm25 con el motivo GUUG está determinada por interacciones electrostáticas y estabilización de la estructura y del *pocket* de interacción a través del loop que abarca los residuos de amino ácidos r70-r90.
- La especificidad de la interacción entre los aptámeros C4 y C5 y la proteína EhCFIm25 sugiere que podrían ser usados como una novedosa herramienta terapéutica para el control de la amibiasis.

PERSPECTIVAS

- Evaluar el efecto del silenciamiento o bloqueo del factor de corte EhCFIm25 en los otros eventos de la poliadenilación, es decir el reclutamiento de la maquinaria de poliadenilación y la síntesis de la cola de poliA.
- Analizar el efecto del silenciamiento o bloqueo del factor de corte EhCFIm25 sobre el transcriptoma y/o proteoma del parásito, para identificar los genes afectados por las alteraciones en la poliadenilación y conocer las rutas y procesos moleculares que pudieran explicar el fenotipo de los trofozoítos.
- Caracterizar la afinidad, especificidad y sensibilidad de los aptámeros C4 y C5 mediante ensayos de ELONA y Slot-Blot
- Nanoencapsular los aptámeros C4 y C5 para mejorar su actividad y vida media.
- Evaluar el efecto de estos aptámeros nanoencapsulados en modelos animales de amibiasis.

BIBLIOGRAFIA

Adamson, T., Shutt, D. & Price, D. Functional Coupling of Cleavage and Polyadenylation with Transcription of mRNA. *J. Biol. Chem.* **280**, 32262-32271, (2005).

Ali, I. *et al.* Tissue Invasion by Entamoeba histolytica: Evidence of Genetic Selection and/or DNA Reorganization Events in Organ Tropism. *PLOS Negl. Trop. Dis*. **2**, (2008).

Ambesajir, A., Kaushik, A., Kaushik, J. & Petros, S. RNA interference: A futuristic tool and its therapeutic applications. *Saudi J Biol Sci* **19**, 395-403, (2012).

Bansal, D. *et al.* Multidrug resistance in amoebiasis patients. *Indian J. Med. Res* , **124**, 189–194, (2006).

Barfod, A., Persson, T. & Lindh, J. In vitro selection of RNA aptamers against a conserved region of the Plasmodium falciparum erythrocyte membrane protein 1. *J Parasitol Res*. **105**, 1557-1566, (2009).

Bendesky, A., Menéndez, D., & Ostrosky-Wegman, P. Is metronidazole carcinogenic?. *Mutat Res*. **2**,133-44, (2002).

Brown, K. & Gilmartin, G. A Mechanism for the Regulation of Pre-mRNA 3′ Processing by Human Cleavage Factor Im Mol. *Cell,* **6**, 1467-76, (2003).

Chan, S. *et al.* CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. *Genes Dev.* **28**, 2370-2380, (2014).

Chen, H. *et al.* R331W Missense Mutation of Oncogene YAP1 Is a Germline Risk Allele for Lung Adenocarcinoma With Medical Actionability. *J. Clin. Oncol.* **33**, 2303-2310 (2015).

Coseno, M. *et al.,* Crystal structure of the 25 kDa subunit of human cleavage factor Im. *Nucleic Acids Res.* **10**, 3474–3483, (2008).

Curinha, A. *et al.* Implications of polyadenylation in health and disease. *Nucleus.* **6**, 508-19 (2014).

Dettwiler, S. *et al.,* Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *J Biol Chem.* **34**, 35788-97, (2004).

Ellington, A., Szostak, J. In vitro selection of RNA molecules that bind specific ligands

Nature, **346**, 818-822 (1990).

Espinosa-Cantellano, M., & Martínez-Palomo, A. Pathogenesis of Intestinal Amebiasis: From Molecules to Disease. *Clin. Microbiol. Rev.***13**. 318-31. (2000).

Gilmartin, G. Eukaryotic mRNA 3' processing: a common means to different ends. *Genes Dev.* **19**, 2517-2521 (2005).

González-Vásquez, M. C. *et al.* De amibas y amebiasis: *Entamoeba histolytica. Elementos.* **87,** 13-18 (2012).

Guerra-Pérez, N. *et al.* Molecular and Functional Characterization of ssDNA Aptamers that Specifically Bind Leishmania infantum PABP. *PLoS One.* **10**, (2015).

Haque, R. Human Intestinal Parasites. *J Health Popul Nutr.* **4**, 387–391 (2007).

Hein, M. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163**, 712–723 (2015).

Hendriks, E., Abdul-Razak, A. & Matthews, K. tbCPSF30 Depletion by RNA Interference Disrupts Polycistronic RNA Processing in *Trypanosoma brucei*. *J. Biol. Chem.* **278**, 26870-26878 (2003).

Hernández de la Cruz, O. *et al.* Proteomic profiling reveals that EhPC4 transcription factor induces cell migration through up-regulation of the 16-kDa actin-binding protein EhABP16 in *Entamoeba histolytica*. *J. Proteomics* **111,** 46–58 (2014).

Hernández de la Cruz, O. *et al.* Multinucleation and Polykaryon Formation is Promoted by the EhPC4 Transcription Factor in Entamoeba histolytica. *Sci Rep.* **21**, 6:19611 (2016).

Hon, C. *et al.* Quantification of stochastic noise of splicing and polyadenylation in Entamoeba histolytica. *Nucleic Acids Res* **41**, 1936-1952 (2013).

Hu, J. *et al.* Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA* **11**, 1485-1493 (2005).

Keefe, A.D., Pai, S., & Ellington, A. Aptamers as therapeutics. *Nat Rev Drug Discov.* **7**, 537-50, (2010).

Kim, H., & Lee, Y. Interaction of poly(A) polymerase with the 25-kDa subunit of cleavage factor I. *Biochem Biophys Res Commun*. **2**, 513-8, (2001).

Koch, H., Raabe, M., Urlaub, H., Bindereif, A. & Preußer, C. The polyadenylation complex ofTrypanosoma brucei:Characterization of the functional poly(A) polymerase. *RNA Biology* **13**, 221-231 (2016).

Ku, T. *et al.* Nucleic Acid Aptamers: An Emerging Tool for Biotechnology and Biomedical Sensing. *Sensors* **15**, 16281-16313 (2015).

Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A. & Handa, H. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3′-UTRs. *Nucleic Acids Res* **34**, 6264-6271 (2006).

Kühn, U., & Wahle, E. Structure and function of poly(A) binding proteins. *Biochim Biophys Acta.* **2-3**, 67-84, (2004).

Li, L. *et al.* 3′UTR shortening identifies high-risk cancers with targeted dysregulation of the ceRNA network. *Sci Rep.* **4**, 5406 (2014).

Liu, Q. *et al.* Enhanced Photodynamic Efficiency of an Aptamer-Guided Fullerene Photosensitizer toward Tumor Cells. *Chem. Asian J.* **8**, 2370-2376 (2013).

Loftus, B. *et al.* The genome of the protist parasite Entamoeba histolytica. *Nature* **433**, 865-868 (2005).

López-Camarillo, C., Orozco, E. & Marchat, L. Entamoeba histolytica: Comparative genomics of the pre-mRNA 3′ end processing machinery. *Exp Parasitol.* **110**, 184-190 (2005).

López-Camarillo, C., López-Rosas, I., Ospina-Villa, J. & Marchat, L. Deciphering molecular mechanisms of mRNA metabolism in the deep-branching eukaryoteEntamoeba histolytica. Wiley Interdisciplinary Reviews: *RNA* **5**, 247-262 (2013).

Mandel, C. *et al.* Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953-956 (2006).

Manley, J. SELEX to Identify Protein-Binding Sites on RNA. Cold Spring Harb. Protoc. **2**, 156-163 (2013).

McLennan, A. The Nudix hydrolase superfamily. *Cell Mol Life Sci* **63**, 123-143 (2005).

Moreno, M. & González, V. Advances on Aptamers Targeting Plasmodium and Trypanosomatids. *Curr. Med. Chem.* **18**, 5003-5010 (2011).

Nagarkatti, R., de Araujo, F., Gupta, C. & Debrabant, A. Aptamer Based, Non-PCR, Non-Serological Detection of Chagas Disease Biomarkers in Trypanosoma cruzi Infected Mice. PLOS Negl. Trop. Dis. **1**, (2014).

Orkin, S.H., Cheng, T.C., Antonarakis, S.E., & Kazazian, H.H. Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *EMBO J.* **2**,453-6, (1985).

Ospina-Villa, J. *et al.* Amino acid residues Leu135 and Tyr236 are required for RNA binding activity of CFIm25 in Entamoeba histolytica. *Biochimie* **115,** 44-51 (2015).

Ospina-Villa, J., *et al.* Aptamers as a promising approach for the control of parasitic diseases. *Braz J Infect Dis*. **20**, 610-618 (2016).

Palencia, A. *et al.,* Targeting Toxoplasma gondii CPSF3 as a new approach to control toxoplasmosis. *EMBO Mol Med.* **3**, 385–394, (2017).

Pezet-Valdez, M. *et al.* The 25 kDa Subunit of Cleavage Factor Im Is a RNA-Binding Protein That Interacts with the Poly(A) Polymerase in Entamoeba histolytica. *PLoS ONE* **8**, (2013).

Ramos, E. *et al.* A DNA aptamer population specifically detects Leishmania infantum H2A antigen. *Lab. Invest.* **87**, 409–416 (2007).

Rozenblum, G.T., Lopez, V.G., Vitullo, A.D., & Radrizzani, M. Aptamers: current challenges and future prospects. *Expert Opin Drug Discov.* **2**, 127-35, (2016).

Rund, D. *et al.,* Two mutations in the beta-globin polyadenylylation signal reveal extended transcripts and new RNA polyadenylylation sites. *Proc Natl Acad Sci U S A*. **10,** 4324–4328, (1992).

Rush, M., Zhao, X. & Schwartz, S. A Splicing Enhancer in the E4 Coding Region of Human Papillomavirus Type 16 Is Required for Early mRNA Splicing and Polyadenylation as Well as Inhibition of Premature Late Gene Expression. *J. Virol.* **79**, 12002-12015 (2005).

Ryan, K. Calvo, O., & Manley, J. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* **10**, 565-573 (2004).

Sandhu, K. & Dash, D. Dynamic α-helices: Conformations that do not conform. *Proteins: Struct., Funct., Bioinf.* **68**, 109-122 (2007).

Shi, Y. Alternative polyadenylation: new insights from global analyses. *RNA.* **12**, 2105-17, (2012).

Sidik, S. *et al.* A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell.* **166**, 1423-1435, (2016).

Shimazu, T., Horinouchi, S. & Yoshida, M. Multiple Histone Deacetylases and the CREB-binding Protein Regulate Pre-mRNA 3′-End Processing. *J. Biol. Chem.* **282**, 4470-4478 (2006).

Soldevilla, M. *et al.* MRP1-CD28 bi-specific oligonucleotide aptamers: target costimulation to drug-resistant melanoma cancer stem cells. *Oncotarget* **7**, 23182-23196 (2016).

Solis, C., Santi-Rocca, J., Perdomo, D., Weber, C. & Guillén, N. Use of Bacterially Expressed dsRNA to Downregulate Entamoeba histolytica Gene Expression. *PLoS ONE* **4**, (2009).

Sonoiki, E. *et al.* A potent antimalarial benzoxaborole targets a cleavage and polyadenylation specificity factor homologue. *Nat. Commun.* **8**, 14574, (2017).

Stacey, S. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nature Genetics* **43**, 1098-1103 (2011).

Sundaram, P., Kurniawan, H., Byrne, M.E., & Wower, J. Therapeutic RNA aptamers in clinical trials. *Eur J Pharm Sci.* **1-2**, 259-71, (2013).

Tian, B. & Manley, J. Alternative cleavage and polyadenylation: the long and short of it. *TRENDS BIOCHEM SCI* **38**, 312-320 (2013).

Tuerk, C., Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science,* **249**, 505-510 (1990).

Venkataraman, K., Brown K., & Gilmartin, G. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.* **19**, 1315-1327 (2005).

Vinayagam, A. *et al.* A Directed Protein Interaction Network for Investigating Intracellular Signal Transduction. *Sci. Signal.* **4**, (2011).

Wang, S., Asakawa, K., Win, T., Toda, T. & Norbury, C. Inactivation of the Pre-mRNA Cleavage and Polyadenylation Factor Pfs2 in Fission Yeast Causes Lethal Cell Cycle Defects. *Mol. Cell. Biol.* **25**, 2288-2296 (2005).

Wassmann, C., Hellberg, A., Tannich, E. & Bruchhaus, I. Metronidazole Resistance in the Protozoan ParasiteEntamoeba histolyticaIs Associated with Increased Expression of Iron-containing Superoxide Dismutase and

Peroxiredoxin and Decreased Expression of Ferredoxin 1 and Flavin Reductase. *J. Biol. Chem.* **274**, 26051-26056 (1999).

Wiestner, A. *et al.* Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* **109**, 4599-4606 (2007).

Winkler, W., Cohen-Chalamish, S. & Breaker, R. An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15908-15913 (2002).

Xiang, K., Tong, L. & Manley, J. Delineating the Structural Blueprint of the Pre-mRNA 3'-End Processing Machinery. *Mol. Cell. Biol* **34**, 1894-1910 (2014).

Yang, Q., Gilmartin, G. & Doublie, S. Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3' processing. *Proc. Natl. Acad. Sci. U.S.A* **107**, 10062-10067 (2010).

Yang, Q., Gilmartin, G. & Doublié, S. The structure of human Cleavage Factor Imhints at functions beyond UGUA-specific RNA binding. *RNA Biology* **8**, 748-753 (2011).

Yang, X. *et al.* A Complex Containing the CPSF73 Endonuclease and Other Polyadenylation Factors Associates with U7 snRNP and Is Recruited to Histone Pre-mRNA for 3'-End Processing. *Mol. Cell. Biol* **33**, 28-37 (2013).

Yasuda, M., Shabbeer, J., Osawa, M., & Desnick, R. Fabry Disease: Novel a-Galactosidase A 3'-Terminal Mutations Result in Multiple Transcripts Due to Aberrant 3'-End Formation. *Am. J. Hum. Genet.* **73**,162–173 (2003).

Yu, Y. *et al.,* Molecular Selection, Modification and Development of Therapeutic Oligonucleotide Aptamers. *Int J Mol Sci.* **3,** 358, (2016).

Zamorano, A. *et al.* In silico analysis of EST and genomic sequences allowed the prediction of cis-regulatory elements for Entamoeba histolytica mRNA polyadenylation. *Comp. Biol. Chem* **32**, 256–263 (2008).

Zhang, H., Pompey, J. & Singh, U. RNA interference in Entamoeba histolytica: implications for parasite biology and gene silencing. *Future Microbiol.* **6**, 103-117 (2011).

Zhang, H., Alramini, H., Tran, V., Singh, U. Nucleus-localized antisense small RNAs with 5'-polyphosphate termini regulate long term transcriptional gene silencing in Entamoeba histolytica G3 strain. *J Biol Chem*. **52**, 44467-79, (2011).

Zhou, J., & Rossi, J. Aptamers as targeted therapeutics: current potential and challenges. *Nat Rev Drug Discov.* **3**, 181-202, (2017).

OTROS PRODUCTOS OBTENIDOS DURANTE EL DESARROLLO DEL DOCTORADO

Artículo publicado

Ospina-Villa JD, Zamorano-Carrillo A, Lopez-Camarillo C, Castañon-Sanchez CA, Soto-Sanchez J, Ramirez-Moreno E, Marchat LA. Amino acid residues Leu135 and Tyr236 are required for RNA binding activity of CFIm25 in Entamoeba histolytica. Biochimie. **2015** Aug;115:44-51. doi: 10.1016/j.biochi.2015.04.017.

Presentaciones en congresos

>2nd International ParaFrap Conference at Ile des Embiez Six-Fours-les-Plages, France on October 2-5, 2016.
Poster: “Evaluation of RNA aptamers against EhCFIm25 protein of *Entamoeba histolytica*”

>39° Congreso Nacional de Microbiología, Querétaro, México, 22-26 de marzo de 2015.
Poster: “Efecto de la inhibición de la proteína EhCFIm25 de *E.histolytica* sobre la viabilidad celular del parásito”

>International Congress of Parasitology. 10-15 de agosto, 2014. México.
Poster: “Molecular and Functional Characterization of the RNA Binding Capacity of the EhCFIm25”

>XXX Congreso Nacional de Bioquímica del 2 al 8 de noviembre de 2014 Guadalajara, México.
Poster: “Molecular dynamics simulation of EhCFIm25 protein and analysis of its ability to bind RNA”