



Guía del ARN peregrino

Señales del camino a la
mitocondria de *Trypanosoma cruzi*.

ÍNDICE

Resumen	2
Introducción	4
<i>Trypanosoma cruzi</i> y enfermedad de Chagas	4
Biología molecular del parásito	5
Mitocondria del parásito.....	10
Genoma mitocondrial.....	11
Antecedentes de productos nucleares con destino mitocondrial	13
Tránsito intracelular de mensajeros y sus actores.....	16
Objetivo general	20
Objetivos específicos	20
Materiales y métodos.....	21
Bases de datos	22
Métodos de análisis	23
Análisis estadístico	27
Resultados y discusión	28
Identificación y caracterización de genes nucleares que codifican para productos mitocondriales.	28
Selección y predicción de genes nucleares codificantes de productos mitocondriales.	28
Análisis composicional de transcritos de MiNT.....	31
Análisis de la base MiNT por categorías.....	36
MiNT-core	36
MiNT-MTS	42
MiNT-DiExTra-LC.....	52
MiNT-DiExTra-CC	65
MiNT-intra.....	83
Estudio de casos particulares	88
Conclusiones	102
Perspectivas	103
Bibliografía	104
Apéndice.....	116
Análisis de ARN de Transferencia.....	116
Agradecimientos	120

RESUMEN

La expresión génica en eucariotas depende marcadamente de la abundancia y funcionalidad de los mensajeros y se encuentra regulada tanto a nivel de tiempo como de espacio. La regulación temporal permite la expresión diferencial de la información genética estableciendo las características proteicas específicas de cada etapa. La regulación de la localización subcelular de ARN permite a la célula organizar y/o restringir la producción de proteínas a una región dada, jugando un rol fundamental tanto a nivel del desarrollo como de la fisiología celular. En estos mecanismos, las señales en *cis* de los ARNm constituyen actores principales operando ya sea como motivos que codifican péptido naciente o como señales reconocidas por proteínas que se encargan de transportar subgrupos de mensajeros a su destino intracelular. Si bien este mecanismo de regulación parece tener grandes efectos en la expresión génica, el mismo ha sido estudiado en forma discreta, focalizándose en transcritos específicos. Estudios más recientes han comenzado a hacer uso de estrategias del tipo *genome wide* con el objetivo de entender el alcance global de este mecanismo.

En el caso de *Trypanosoma cruzi*, un organismo con relevancia tanto a nivel básico como sanitario, la localización asimétrica de mensajeros podría constituir una herramienta de regulación fina que facilite la funcionalidad celular. Estudios previos han sugerido que mecanismos de localización podrían estar involucrados en la regulación de la expresión de proteínas estadio específica.

La mitocondria, que en el caso de estos parásitos es única, se presenta como un organelo dinámico a lo largo del ciclo de vida del parásito. Es de esencial importancia en el metabolismo energético del mismo así como en la respuesta al estrés oxidativo. En forma semejante a lo que sucede en eucariotas superiores, las proteínas codificadas a nivel de su genoma son escasas.

En este trabajo nos planteamos la búsqueda de señales, tanto a nivel de secuencia como de estructura, que sean compartidas por mensajeros que codifican proteínas con destino mitocondrial en *T. cruzi*.

En primer lugar, siguiendo diferentes criterios, se identificaron 1196 genes nucleares que codifican proteínas mitocondriales de *T. cruzi*. El número de genes de esta base de datos, que denominamos MiNT, es comparable con lo reportado para organismos relacionados. Encontramos que los transcritos de MiNT se caracterizan por tener CDSs de menor tamaño y menor grado de estructuración así como un mayor contenido de GC en la tercera posición del codón e índice de adaptación de codones, que las del resto de los transcritos de *T. cruzi*.

Para buscar las señales conservadas, se formaron subgrupos siguiendo varios criterios como ser: confianza en la asignación del destino mitocondrial, presencia o ausencia de señal de localización mitocondrial a nivel peptídico, perfil diferencial de expresión a lo largo del ciclo de vida y en las distintas fases del ciclo celular y distribución intramitocondrial. Asimismo se estudiaron grupos de genes particulares como el de los genes que codifican proteínas ribosomales y el de las aminoacil ARNt sintetasas.

Como resultado de estos análisis se definió una señal peptídica de localización mitocondrial consenso en *T. cruzi*.

Además, la búsqueda de señales nucleotídicas permitió identificar con alto nivel de confianza señales tanto a nivel de estructura primaria como secundaria. Entre ellas se destacan: un motivo lineal enriquecido en los genes que codifican proteínas que no cuentan con una señal peptídica consenso y que podría constituir una señal compensatoria de localización; un motivo lineal reconocido por la proteína TRRM3 de *T. cruzi* que podría ser en parte responsable por la localización diferencial de ciertos mensajeros, y una estructura en la región 3'UTR (*untranslated region*) que se encuentra enriquecida en los transcritos nucleares que codifican proteínas mitocondriales.

Por último, la búsqueda de las señales mencionadas nos permitió detectar errores en la anotación tanto a nivel de función como en la definición de la metionina inicial o de los marcos abiertos de lectura. De esta forma hemos contribuido a mejorar la anotación genómica del este parásito.

Los resultados obtenidos en este trabajo, junto con las perspectivas que se abren a partir de ellos, constituyen un aporte a la caracterización de la regulación de la expresión génica en *T. cruzi*.

Octubre 2016

INTRODUCCIÓN

Trypanosoma cruzi Y ENFERMEDAD DE CHAGAS

Trypanosoma cruzi es el protozoo parásito causante de la enfermedad de Chagas (Chagas 1909), también conocida como tripanosomiasis americana. Esta enfermedad constituye actualmente un problema para la población americana, donde se estima que hay de 6 a 7 millones de personas infectadas en 21 países de América Latina (WHO 2015). El 30% de las personas infectadas desarrollan alteraciones cardíacas y cerca de un 10% corren riesgo de presentar problemas digestivos, neurológicos o alteraciones mixtas que necesitan tratamientos específicos. A su vez se contabilizan cerca de 10.000 fallecimientos por año debido a esta afección. En las últimas décadas se han detectado casos en Estados Unidos, Canadá y algunos países europeos y del oeste del pacífico. Este fenómeno parece ser debido principalmente a la movilidad de la población de América latina hacia el resto del mundo (WHO 2015; Manne-Goehler et al. 2015).

El ciclo biológico de *T. cruzi* (Figura 1) comienza cuando el insecto hematófago se alimenta de sangre de un mamífero infectado, ingiriendo la forma tripomastigota sanguínea del parásito. Posteriormente, en el interior del insecto vector se pueden observar dos estadios con diferente morfología: los epimastigotas replicativos, en el tracto digestivo medio, y los tripomastigotas metacíclicos, acumulados en la ampolla rectal del insecto. Estos últimos constituyen la forma infectiva para el hospedero y son depositados con las heces del insecto en las cercanías de la picadura que se produce durante su alimentación. A través de la piel dañada, los tripomastigotas tienen una vía de entrada directa al torrente sanguíneo que les permite acceder a una gran variedad de tejidos. Dentro de las células, el parásito se diferencia a la forma amastigota, que luego de varias divisiones, se diferencia en tripomastigotas sanguíneos, ocasionando la lisis celular y siendo liberados al torrente sanguíneo (Tyler et al. 2002). Como mecanismo alternativo, los amastigotas pueden invadir directamente células de mamífero, con una infectividad comparable a la de los tripomastigotas (Mortara et al. 2005). Aunque, los tripomastigotas sanguíneos son casi indistinguibles morfológicamente de los metacíclicos, poseen diferencias notables a nivel de su biología.

Además de la ruta natural de transmisión del parásito, *T. cruzi* puede ser transmitido congénitamente, por trasplantes, por transfusiones o mediante el consumo de comida y fluidos contaminados (Nóbrega et al. 2009). De esta forma el aumento de las migraciones ha contribuido a diseminar la enfermedad aún en ausencia del insecto hematófago vector.

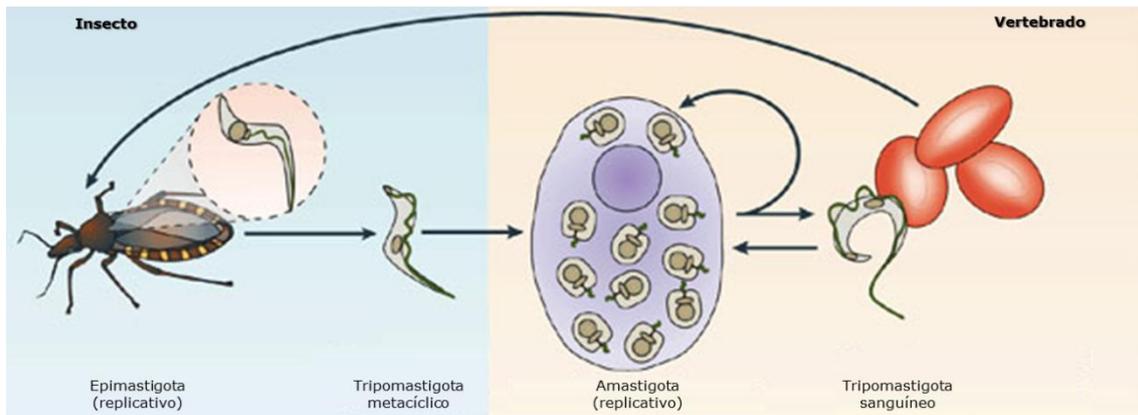


Figura 1. Ciclo de vida de *Trypanosoma cruzi*. Los tripomastigotas metacíclicos provenientes de los epimastigotas en el insecto vector son transmitidos al hospedero mamífero a través de las heces. Una vez en el mamífero, los tripomastigotas invaden células y son internalizados en una vacuola derivada del lisosoma (vacuola parasitófora). Dentro de la vacuola, los tripomastigotas se diferencian a amastigotas (2-8 horas), la vacuola parasitófora se rompe y los parásitos se localizan en el citoplasma de la célula (8-16 horas). Los amastigotas citosólicos comienzan a dividirse a las 24 h post-invasión y continúan dividiéndose cada 12 h por 5-6 días. Finalmente, los amastigotas se diferencian a tripomastigotas rompiendo la célula hospedera y, al entrar en la circulación sanguínea, diseminan la infección. Extraído y modificado de (Buscaglia et al. 2006).

La terapia contra esta parasitosis está basada únicamente en dos drogas: benznidazol y nifurtimox. Ambos compuestos tienen eficacia acotada y presentan graves efectos secundarios por lo que se están dedicando múltiples esfuerzos en el desarrollo de nuevas drogas dirigidas a las vías que sean esenciales para el parásito y/o permitan distinguirlo de sus hospederos (Álvarez et al. 2013; Papadopoulou et al. 2015; Buckner & Navabi 2010).

BIOLOGÍA MOLECULAR DEL PARÁSITO

T. cruzi, y los tripanosomátidos en general, presentan características biológicas excepcionales que los diferencian del resto de los eucariotas superiores (Smith & Parsons 1996) probablemente debido a la aparición temprana de los flagelados en la evolución (Woese et al. 1990).

Con el genoma de *T. cruzi* prácticamente secuenciado en su totalidad (El-Sayed et al. 2005), se tuvo de manera accesible una gran cantidad de información que ha aumentado la posibilidad de investigación sobre la biología de este parásito. Se encontró que aproximadamente 12.000 proteínas podrían ser codificadas por genoma haploide. En base a homología con genes correspondientes a otras proteínas previamente caracterizadas o por dominios funcionales conocidos, se pudo asignar una función probable al 50,8% de los genes que codificaban para estas proteínas (Choi & El-Sayed 2012). Esto significa que para más de 5.900 proteínas, su función es desconocida y se requiere mayor caracterización. También es de resaltar que más del 50% del genoma consiste en regiones de secuencias repetidas de retrotransposones como "hot spot" (RHS), genes repetidos en tándem, y genes de varias familias multigénicas como lo son

las transalidasas (TS/gp85), las mucinas, las proteínas de superficie asociadas a mucinas (MASP) y gp63. Cada una de estas familias incluye varios cientos de genes los cuales pueden ser expresados simultáneamente (Almeida & Nakayasu 2010).

El genoma de *T. cruzi* (60,3 Mb) se encuentra organizado en 41 pequeños cromosomas (Choi & El-Sayed 2012). Al igual que sucede con otros tripanosomátidos, los genes que codifican para proteínas, en principio no emparentadas funcionalmente, se encuentran dispuestos en grandes agrupamientos de policistrones (de 10 a cientos de genes) ubicados en la misma hebra. Estos agrupamientos direccionales, están separados por secuencias cortas de unas pocas kilobases, llamadas regiones de cambio de hebra (Figura 2).

Los arreglos en tándem de genes de ARN ribosomal (ARNr) se encuentran mayormente en las regiones de cambio de hebra. Los ARNr 5S en *T. brucei* y *T. cruzi* también están organizados en arreglos tipo tándem (Cordingley 1985).

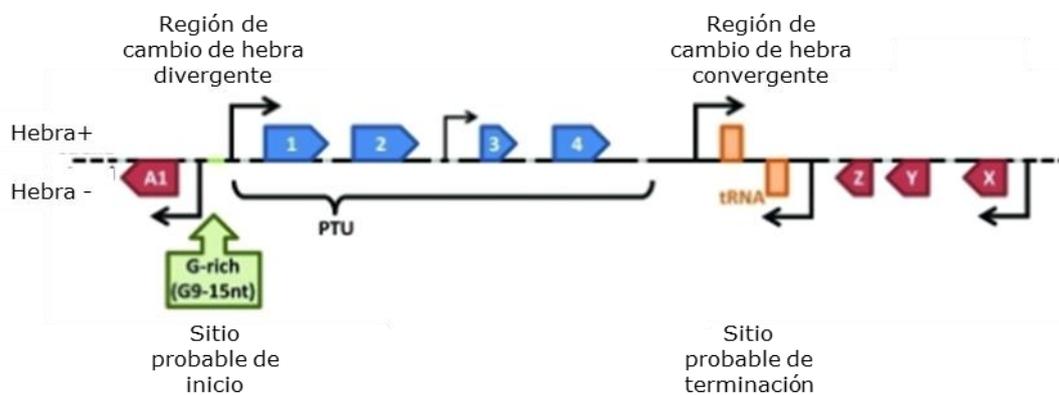


Figura 2. Organización de los genes en grupos con polaridad de hebra en tripanosomátidos. Con flechas negras se muestra el sentido de la transcripción. Los sitios de cambio de hebra divergente se postula que son sitios de inicio de la transcripción y los sitios de cambio de hebra convergentes se consideran sitios de terminación de la transcripción. Estos sitios discretos de inicio y terminación de la transcripción se encuentran enriquecidos en histonas modificadas. Se muestran en azul y rojo los diferentes genes (PTU= *polycistronic transcription units*). Extraído y modificado de (Campos et al. 2008).

Por otra parte, la mayoría de los genes de ARNt se encuentran organizados en agrupamientos de 2 a 10 genes posiblemente junto con otros genes que son transcritos por la ARN polimerasa III (Padilla-Mejía et al. 2009; Ivens et al. 2005). En contraste con otros eucariotas, la distribución de los genes de ARNt en el genoma de *L. major* y de *T. brucei* parece no ser totalmente al azar ya que los genes están confinados a un subgrupo de cromosomas (Padilla-Mejía et al. 2009).

A pesar del hecho de que estas especies divergieron hace más de 200 millones de años existe una alta conservación en el contexto genómico (sintenia) entre los cromosomas de los tripanosomátidos (Monnerat et al. 2004). En contraste con los genes codificantes, la mayoría de los

agrupamientos de ARNt no presentan sintenia alguna (Martinez-Calvillo et al. 2010).

La gran mayoría de los genes que codifican para proteínas carecen de intrones. Es de remarcar que el *cis-splicing* en estos parásitos solo se ha demostrado para el gen de la poli(A) polimerasa (Mair et al. 2000). Del mismo modo, solamente un isotipo de ARNt, el ARNt-Tyr, contiene un intrón en tripanosomátidos (Padilla-Mejía et al. 2009).

La ARN polimerasa II dirige la transcripción policistrónica de los genes que codifican para proteínas. La regulación en el inicio de la transcripción es incierta dada la ausencia de promotores canónicos para dicha enzima. Por lo que se considera que en estos parásitos, a diferencia de lo que sucede en eucariotas superiores, la transcripción de genes nucleares que codifican para proteínas es no sólo policistrónica sino que también constitutiva (Das & Bellofatto 2009). Se ha propuesto que la transcripción podría iniciar a partir de las regiones donde se produce el cambio en la hebra codificante y donde la estructura secundaria del ADN y la cromatina jugaría un rol importante (Martinez-Calvillo et al. 2010; Smircich et al. 2013; Reynolds et al. 2016; Respuela et al. 2008; Ekanayake & Sabatini 2011). A pesar de que la transcripción es policistrónica, los patrones de expresión de genes consecutivos es diferente, por lo que se considera que la regulación de la expresión génica en tripanosomas opera básicamente a nivel postranscripcional (Clayton & Shapira 2007).

De cualquier manera, los genes que codifican para ARNm que son regulados durante la respuesta de estrés térmico están selectivamente posicionados dentro de las unidades transcripcionales policistrónicas. Los genes que se encuentran disminuidos en su expresión están cerca de los sitios de inicio de la transcripción, mientras que aquellos que contrariamente presentan su expresión relativa aumentada están distantes (Kelly et al. 2012). El sesgo posicional que se describió inicialmente para genes de respuesta a estrés, se encontraría asociado también a otras funciones celulares. En efecto, mensajeros específicos de ciertas etapas del ciclo celular estarían regulados de la misma forma, por lo que su abundancia relativa podría ser dependiente de la localización de ciertos genes con respecto a los sitios de inicio de la transcripción (Kelly et al. 2012).

Los tripanosomátidos también están entre los pocos eucariotas inferiores con una estructura inusual de sus ARNm, debido a la estricta necesidad del procesamiento postranscripcional por *trans-splicing*. De esta manera, los transcritos policistrónicos dan lugar a cada ARNm maduro por adición de una secuencia de 39 nucleótidos idéntica en su extremo 5' terminal, denominada *splice leader* (SL). Además del universalmente conservado CAP (7 metil guanósina), los primeros 4 nucleótidos del SL están también metilados (Haile & Papadopoulou 2007). En *Leishmania* se demostró que

esta estructura es crítica para la asociación del mensajero con los polisomas (Ayub et al. 2009).

Los mecanismos postranscripcionales de regulación pueden actuar a nivel del procesamiento, transporte, estabilidad, localización subcelular y traducibilidad del mensajero, entre otros (Keene 2007). En eucariotas, se ha observado que existe una coordinación global de la expresión génica a nivel postranscripcional, en donde los ARNm que codifican para proteínas funcionalmente relacionadas son regulados coordinadamente formando operones y regulones postranscripcionales (Shepard et al. 2003; Keene 2007; Bolognani et al. 2010). Este tipo de regulación ocurre mediante la interacción de los ARNm con factores en *trans*, principalmente proteínas de unión al ARN, aunque también pueden ser ARNs no codificantes y metabolitos (Keene 2007). Probablemente, el ejemplo más citado de estos operones fue descrito para *Saccharomyces cerevisiae* (Shepard et al. 2003). Estos autores proponen que proteínas de unión al ARN de la familia pumilio estarían uniéndose distintivamente a transcritos que codifican para proteínas con función relacionada.

En *T. cruzi* los estudios sobre mecanismos postranscripcionales de regulación de la expresión génica se han focalizado principalmente en el análisis de procesos de estabilización/degradación de mensajeros (Araújo & Teixeira 2011; Li et al. 2012; Bayer-Santos et al. 2012; Guerra-Slampo et al. 2012) y su accesibilidad a la maquinaria traduccional activa (Cassola et al. 2007; Holetz et al. 2007; Cassola 2011; Kramer 2013). Más recientemente el control de la traducción y la eficiencia traduccional han sido reconocidos como un escalón adicional en el control de la expresión génica en *T. cruzi* (Smircich et al. 2015; da Silva Augusto et al. 2015).

Las proteínas de unión denominadas PUF (Pumilio) son ampliamente conocidas por sus roles en la división celular, en la diferenciación y en el desarrollo (Cho et al. 2006; Wreden et al. 1997). Si bien los ejemplares más caracterizados actúan como represores postranscripcionales (Wickens et al. 2002; Wharton & Aggarwal 2006), algunos estudios demuestran que pueden actuar como activadores y facilitando la localización subcelular de mensajeros (Saint-Georges et al. 2008; Jiang et al. 2010; Gerber et al. 2006). En particular, Puf3 se asocia con un subgrupo de 154 ARNm que casi exclusivamente codifican para proteínas mitocondriales, muchas de las cuales están involucradas en la síntesis de proteínas. Este antecedente, que también ha sido descrito en humanos (Raghavan et al. 2004), apoya la idea de que los ARNm que codifican para proteínas de organelos subcelulares podrían estar reguladas en espacio y tiempo por proteínas específicas de unión al ARN. De cualquier manera, esta regulación es compleja (Figura 3) y podría implicar la acción combinada de varios factores incluyendo aquellos que intervienen en la regulación de la exportación nuclear no convencional (Cenik et al. 2011). Datos derivados de diversos estudios sugieren que Puf3

podría regular un balance dinámico entre la traducción del ARNm, traslados del mensajero a la mitocondria, e importación cotraduccional de las proteínas (Quenault et al. 2011). En levaduras, también se encontró que otro miembro de la familia, Puf6, contribuye a la localización asimétrica de *ASH1* durante el transporte a la zona en gemación (Gu et al. 2004) y que Puf5 se encuentra involucrado en la localización del ARNm de PEX14 en el entorno de peroxisomas (Zipor et al. 2009). Al igual que Puf4 y Puf5, Puf 3 y Puf 6 se asocian con ARNm conteniendo un motivo consenso de unión (UGUANAUA) que es similar al definido para otras proteínas pumilio.

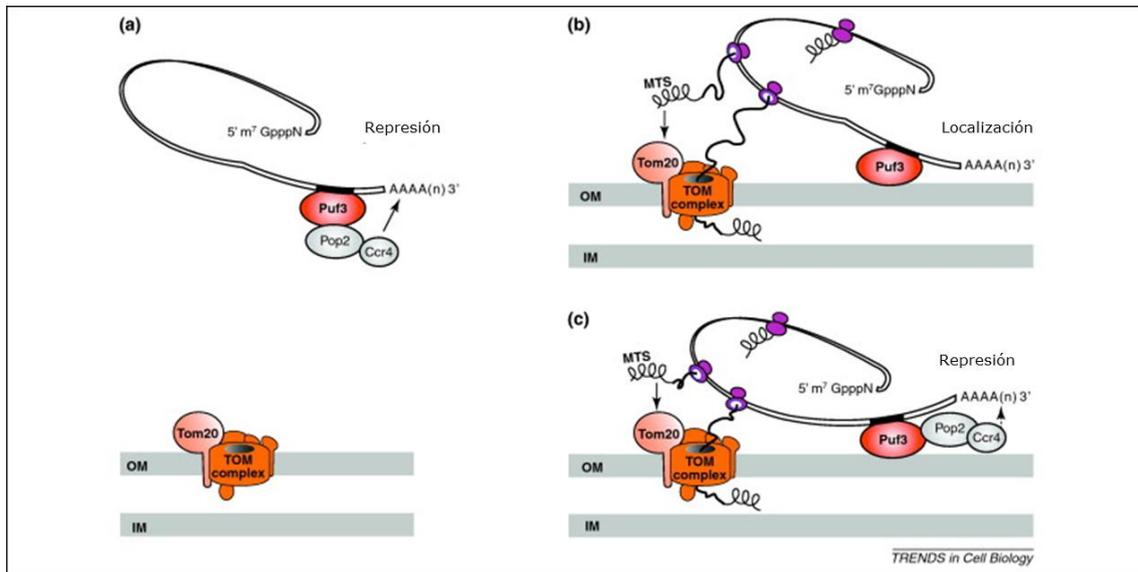


Figura 3. Modelos de la función de Puf3 en la represión del ARNm y la localización de los transcritos en la mitocondria. (a) Puf3 puede reprimir la expresión de los mensajeros con los que interactúa promoviendo por ejemplo la deadenilación de los ARNm blanco, vía el complejo Ccr4-Pop2-NTD y posiblemente mediante otros mecanismos que afecten la estructura del mensajero (b) Puf3 puede además reconocer sus mensajeros y servir para estabilizar la posición de los mismos en el entorno mitocondrial. (c) En un escenario alternativo la deadenilación puede ser modulada en el entorno mitocondrial reprimiendo la traducción del mensajero. Extraído y modificado de (Quenault et al. 2011).

Aunque estudios realizados tanto en *T. cruzi* como en *T. brucei*, determinaron que ciertos ARNm funcionalmente relacionados se regulan de forma coordinada apoyando la existencia de operones postranscripcionales en estos organismos (Dallagiovanna et al. 2008; Archer et al. 2009; Minning et al. 2009), aún no se han encontrado factores que cumplan una función semejante a la asociada a Puf3.

Por otra parte, se han caracterizado numerosos elementos en *cis*, tanto a nivel de secuencia como de estructura, que son blanco de proteínas reguladoras de unión específica. La mayor parte de estos elementos regulatorios se localizan en la región 3' no traducida de los mismos (3' UTR) (Hotz et al. 1997; Coughlin et al. 2000; Clayton & Shapira 2007; Haile & Papadopoulou 2007; Machado-Silva et al. 2010; Noé et al. 2008; Pérez-Díaz et al. 2013; De Gaudenzi et al. 2013; Pastro et al. 2013) aunque se han encontrado también en otras regiones (Clayton 2002; Keene 2007). Sin embargo, hasta la fecha no se han descrito señales de localización

subcelular para los mensajeros de origen nuclear que codifican para proteínas mitocondriales.

MITOCONDRIA DEL PARÁSITO

T. cruzi tiene varias particularidades a nivel de compartimientos/organelos subcelulares, algunos de los cuales no se encuentran presentes en eucariotas superiores y otros presentan características excepcionales. Entre ellos se destacan: los reservosomas, organelos similares a los lisosomas exclusivamente presente en la forma epimastigota del parásito (Vieira et al. 2005); los glicosomas, que se asemejan a los peroxisomas y contienen varias de las enzimas de la vía glicolítica (Michels et al. 2006); los acidocalcisomas, que se encuentran involucrados en el almacenamiento de calcio y fósforo (Docampo et al. 2005) y la mitocondria que característicamente posee su genoma confinado a una estructura definida, denominada kinetoplasto, (Shapiro & Englund 1995). Esta estructura fue estudiada en detalle en *T. cruzi* por Riou y colaboradores (Riou & Delain 1969; Riou & Yot 1977).

La mitocondria es un organelo interesante no sólo por lo variadas de las funciones celulares que cumple (síntesis aerobia del ATP por la vía oxidativa, mantenimiento del balance *redox* de la célula (Tomás & Castro 2013), etc.) sino porque además se encuentra bajo el control de dos genomas: el propio y el nuclear.

Los kinetoplástidos poseen una única, pero muy desarrollada mitocondria que se distribuye a lo largo del cuerpo celular del parásito. La mitocondria presenta los compartimientos clásicos: la membrana externa, el espacio intermembrana, la membrana interna que se invagina formando crestas y la matriz mitocondrial. Dependiendo de las fuentes nutricionales y del ambiente en el que se encuentre el parásito, la mitocondria puede ocupar una cantidad variable del volumen celular (de Souza et al. 2009).

En *T. cruzi* se observan variaciones de la mitocondria tanto a lo largo del ciclo de vida como durante los ciclos celulares de sus formas replicativas. Los tripomastigotas metacíclicos poseen la mitocondria en forma tubular ubicada del lado opuesto a la membrana ondulante, mientras que en tripomastigotas sanguíneos tiene un aspecto similar al de dos estructuras (Maria et al. 1972). En amastigotas y epimastigotas la mitocondria adopta forma de "pesuña" (Newberry & Paulin 1989).

Con respecto a la maquinaria de transporte de electrones, se demostró la presencia y participación del complejo II al V. Sin embargo, existe controversia con respecto a la contribución del complejo I (NADH: ubiquinona oxidoreductasa) al metabolismo energético del parásito (Oppendoes & Michels 2008). Por otra parte, se ha mostrado que mientras los componentes del ciclo de Krebs permanecen constitutivamente activos

durante el ciclo de vida de *T. cruzi*, algunos componentes de la cadena de transporte de electrones y chaperonas mitocondriales codificados en el núcleo, son regulados diferencialmente a lo largo del ciclo de vida, mostrando bajos niveles de expresión en la forma metacíclica de los tripomastigotas (Tyler et al. 2002).

Las proteínas mitocondriales codificadas a nivel nuclear son dirigidas a la mitocondria principalmente a través de señales en la secuencia peptídica de dichas proteínas, tanto en eucariotas superiores (Emanuelsson et al. 2001) como en tripanosomátidos (Herrmann & Neupert 2000a; Lee et al. 1999). En efecto, se mostró para *T. brucei* (Zhang et al. 2010), así como para *T. cruzi* que existe un alto número de proteínas mitocondriales codificadas a nivel nuclear que poseen péptido señal, típicamente hacia el extremo N-terminal (Herrmann & Neupert 2000b), sin embargo, otras carecen de estas señales.

Se han desarrollado múltiples métodos para predecir la localización mitocondrial de proteínas. Estos métodos utilizan diferentes características como composición aminoacídica (MitoPred (Guda et al. 2004), PLOC (Park & Kanehisa 2003), entre otros), propiedades fisicoquímicas, péptidos señal a nivel N-terminal y la ocurrencia de dominios (Horton et al. 2007; Emanuelsson et al. 2007). Sin embargo, estos métodos presentan limitaciones, por ejemplo debido a la existencia ya sea de péptidos señales no canónicos (Bromley et al. 2004), de proteínas que aparentemente carecerían de estas señales. La ausencia de señales de localización mitocondrial a nivel peptídico, podría ser compensada con la existencia de señales de localización en el mensajero correspondiente.

Genoma mitocondrial

En la escala filogenética, el genoma de las mitocondrias se presenta en una gran diversidad estructural, yendo desde un único cromosoma circular hasta un arreglo complejo de múltiples cromosomas tanto lineales como circulares (Burger et al. 2003).

El genoma mitocondrial de los kinetoplástidos (ADNk) es distintivo del resto de los eucariotas por su tamaño y complejidad. La mitocondria contiene hasta el 25% del ADN celular total. El kinetoplasto está formado por dos tipos de moléculas de ADN circular concatenadas entre sí, llamados: maxi y mini círculos. Los maxicírculos (de 20 a 37 kb) contienen los genes que codifican para un pequeño número de proteínas y ARNr (Simpson 1987). Los minicírculos (de 0,5 a 2,8 kb) contienen las secuencias que codifican para los ARN guías (ARNg) necesarios para la modificación de los transcriptos mitocondriales (adición o delección de uridinas) en un proceso conocido como edición o "editing" (Shaw et al. 1988; Pollard et al. 1990;

Sturm & Simpson 1990). El resto de las proteínas mitocondriales se encuentran codificadas en el ADN nuclear del parásito.

La replicación del ADNk es un proceso complejo que requiere varias proteínas específicas que han sido mayormente caracterizadas en *T. brucei* y *Leishmania* (Liu et al. 2005). Varias proteínas asociadas al ADNk de *T. cruzi* han sido reportadas (Fragoso & Goldenberg 1992; Gómez et al. 2001; Zavala-Castro et al.; Duhagon et al. 2009). La replicación del kinetoplasto comienza sobre el final de la fase de síntesis del genoma nuclear (Elias et al. 2007). Los datos existentes también indican que el núcleo y el kinetoplasto segregan casi simultáneamente en *T. cruzi* (Elias et al. 2007). Este tipo de segregación se encuentra en *Crithidia* y *Leishmania* (Vanassel & Steinert 1971), mientras que en el estadio procíclico de *T. brucei* el kinetoplasto segrega cuando el núcleo se encuentra en la fase G2 del ciclo celular (Woodward & Gull 1990).

Los maxicírculos son funcionalmente equivalentes al ADN mitocondrial de otros eucariotas. Contienen únicamente los genes que codifican para los ARNr y proteínas mitocondriales hidrofóbicas, principalmente involucradas en la cadena de transporte de electrones (Simpson et al. 1987). A primera vista, el genoma de los maxicírculos parece carecer de varios genes que se encuentran conservados en otros genomas mitocondriales, además parece presentar genes que perdieron elementos esenciales para su traducción, como codones de inicio, o presentan marcos abiertos de lectura (ORFs) discontinuos. La inserción o delección postranscripcional de uridinas resuelve la mayoría de estos problemas creando codones de inicio (Shaw et al. 1988; Feagin, Shaw, et al. 1988), corrigiendo los marcos abiertos de lectura internos (ej.: agregado de 4 uridinas en el transcripto de COII) (Payne et al. 1985; Benne et al. 1986) y modificando extensivamente otros ARNm irreconocibles para generar ORFs enteros (van der Spek et al. 1988)(ej.: 547 uridinas insertadas y 41 eliminadas del transcripto que codifica para COIII de *T. brucei* (Feagin, Abraham, et al. 1988)). Por otro lado, se reportaron mutaciones en el genoma mitocondrial de *Trypanosoma vivax* recientemente ingresado en América, en donde muchos de los genes afectados no recobrarían su funcionalidad (Greif et al. 2015).

El ordenamiento y contexto de los genes presentes en los maxicírculos de *T. cruzi* se encuentra conservado con respecto a *T. brucei* y *Leishmania tarentolae*. En el caso de las secuencias no codificantes de *T. cruzi*, se observan regiones repetitivas cepa-específicas interrumpidas por un elemento de secuencia conservado que podría servir de origen de replicación (Westenberger et al. 2006). Se ha propuesto que estos elementos conservados podrían ser importantes para la replicación, transcripción y anclaje de los maxicírculos a la red del kinetoplasto.

El proceso de edición de los mensajeros incluye la hibridación de una región conocida como ancla del ARNg hacia el 3' del mensajero blanco. El

apareamiento secuencial de bases dirige la inserción y/o deleción en el mensajero. En este apareamiento se encuentran uniones no solo del tipo Watson-Crick sino que además, con frecuencia, se encuentran apareamientos del tipo G - U. Si bien el repositorio primario de ARNg se encuentra en los minicírculos, algunos ARNg se encuentran codificados a nivel de los maxicírculos (Thomas et al. 2007).

ANTECEDENTES DE PRODUCTOS NUCLEARES CON DESTINO MITOCONDRIAL

La mayoría de las proteínas mitocondriales que están codificadas a nivel nuclear son sintetizadas en el citoplasma y transportadas a la mitocondria de manera postraduccional (Figura 4). Muchos de estos polipéptidos tienen una extensión amino-terminal temporal denominada presecuencia, que es escindida una vez que estas proteínas son importadas a la mitocondria. Se describió que este proceso de importación requiere la existencia de una diferencia de potencial electroquímico a través de la membrana interna (Hay *et al*, 1984). Estudios realizados en varios eucariotas definieron que esta presecuencia (MTS) contendría la información no sólo sobre su destino mitocondrial sino también para su ubicación final en los compartimientos intramitocondriales en levaduras (van Loon et al. 1986).

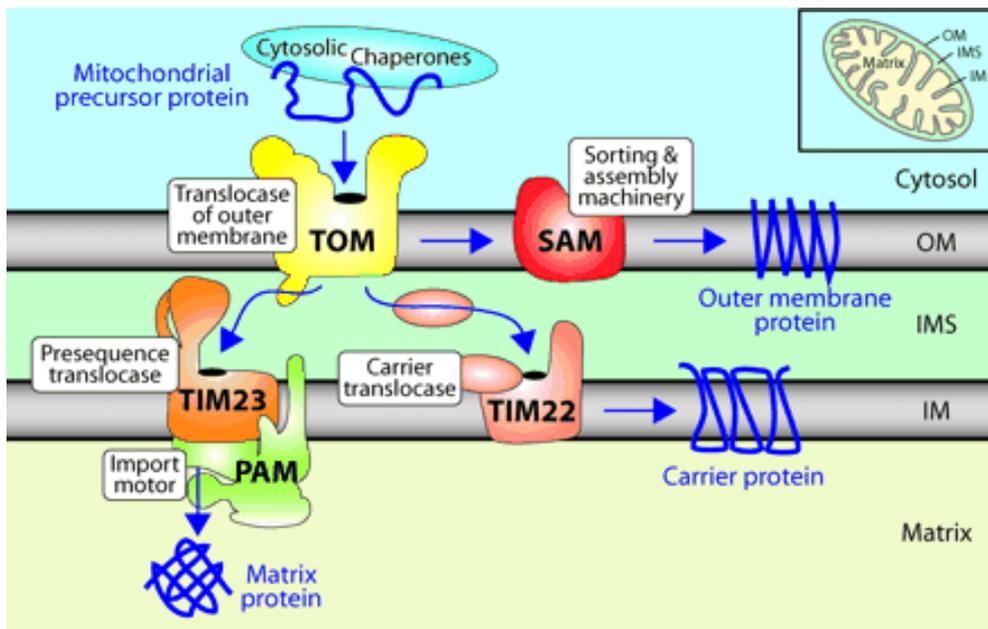


Figura 4. Vías de importación a la mitocondria. La mayoría de las proteínas mitocondriales son sintetizadas en el citoplasma. Estas precisan de la mediación de chaperonas citosólicas para ser transferidas al complejo TOM de donde finalmente se distribuye a su compartimiento final. Los precursores de las proteínas barril de la membrana externa de la mitocondria requieren de la participación del complejo SAM. Las proteínas destinadas a la matriz dependen de la traslocación mediada por la presecuencia y llevada a cabo por TIM23 y su motor de importación asociado PAM para su transporte a través de la membrana mitocondrial interna, con ayuda de la traslocasa TIM22. OM, membrana externa (*outer membrane*); IMS, Espacio intermembrana (*intermembrane space*); IM, membrana interna (*inner membrane*) (Wiedemann et al. 2004).

Si bien se sabe que los tripanosomátidos regulan la biogénesis mitocondrial a través de su ciclo de vida, el inventario de proteínas mitocondriales para *T. cruzi* permanece incierto. En cambio se realizaron estudios en *T. brucei* donde se describieron las proteínas que formaban parte de la mitocondria, mediante el estudio del proteoma mitocondrial (Panigrahi et al. 2009) y análisis bioinformáticos (Zhang et al. 2010). En estos estudios se encontraron o predijeron un total de 1.065 proteínas mitocondriales de las cuales, únicamente 18 estarían codificadas a nivel mitocondrial. Varias de estas proteínas si bien carecen de péptido señal, que guiaría su localización, son importadas a la mitocondria (Herrmann & Neupert 2000a; Priest et al. 1993). De las proteínas identificadas en el análisis bioinformático (Zhang et al. 2010), se demostró que los transcritos que codifican para 435 de ellas, cambian significativamente su expresión a lo largo del ciclo de vida del parásito, mientras que el resto permanece estable. En este estudio observaron eventos de procesamiento alternativo para 298 de los transcritos, donde un pequeño subgrupo podría emplear este mecanismo para la localización alternativa del mensajero en citoplasma y mitocondria (Zhang et al. 2010).

Particularmente han sido estudiadas algunas proteínas con gran relevancia a nivel de la biogénesis mitocondrial. Una de las más estudiadas es el citocromo c1 que corresponde a una subunidad del complejo citocromo c reductasa asociado a la membrana mitocondrial interna y expuesto hacia el espacio intermembrana. Esta proteína es codificada a nivel nuclear en todos los eucariotas estudiados, y se transcribe como un precursor con un MTS de 84 a 61 aminoácidos dependiendo de la especie (Hartl et al. 1989). Previo al ensamblado final en el complejo, el MTS es removido por proteasas para dar lugar al citocromo c1 maduro. En el caso de algunos tripanosomas, esta proteína carecería de un MTS convencional y, en el caso de *Crithidia fasciculata*, el procesamiento a proteína madura consistiría únicamente en la remoción de la metionina inicial. Este hecho implicaría que la citocromo c1 de kinetoplastos y de los euglenoides (Lill et al. 1992; Priest et al. 1993) sería dirigida e importada a la mitocondria por una vía no conservada.

Por otro lado, dentro del subgrupo de proteínas que se caracteriza por presentar doble localización, el caso más estudiado en levaduras es el de la fumarasa. Esta enzima que se encuentra cumpliendo su función tanto en la mitocondria como en el citoplasma, sería importada a la mitocondria cotraduccionalmente, su señal de localización mitocondrial sería escindida y luego sería nuevamente exportada al citoplasma (Sass et al. 2001). Sin embargo, análisis preliminares revelaron que *T. cruzi* posee dos genes independientes que codifican para la fumarasa de clase I, una citosólica (TcCLB.507257.60) y otra mitocondrial (TcCLB.507669.10) (Pádua 2014).

Otro grupo interesante es el de las aminoacil-ARNt-sintetasas (AAtSs), enzimas responsables de catalizar la aminoacilación de los ARNt con su

correspondiente aminoácido. En eucariotas, estas enzimas se requieren para la traducción tanto a nivel del citoplasma como de la mitocondria (Duchêne et al. 2009). En este contexto, se espera que los eucariotas posean dos sets de AAtSs (con excepción de las plantas que poseen 3 sets). Sin embargo, en algunos casos, varias de las enzimas son codificadas por un único gen. En tripanosomátidos esta situación es extrema ya que únicamente poseen 23 genes para AAtSs en su genoma (Berriman et al. 2005). En *T. brucei* se demostró que para la isoleucil-ARNt-sintetasa, el mecanismo de localización doble involucra el procesamiento por *trans-splicing* alternativo (Rettig et al. 2012). Previamente se había detectado procesamiento alternativo en por lo menos 5 AAtSs más (Asp-ARNt-S, Pro-ARNt-S, Gln-ARNt-S, Glu-ARNt-S, Ser-ARNt-S), sugiriéndose su vinculación con la localización múltiple de estas enzimas (Nilsson et al. 2010).

Los ribosomas mitocondriales parecen ser estructuralmente variables con respecto a otros organismos. A lo largo de su evolución sufrieron un proceso de remodelado exhaustivo donde se incluye pérdida de ARN y adquisición de proteínas para completar su estructura (Smits et al. 2007). Los ARNr, que son particularmente pequeños, se encuentran codificados junto con 18 proteínas en los maxicírculos. El ARNr 12S forma parte de la subunidad mayor del ribosoma mientras que el otro aún más pequeño 9S forma parte de la subunidad menor. La subunidad mayor (40S), la subunidad menor (30S) y el monosoma (50S) fueron purificadas de *L. tarentolae* y algunos de sus componentes identificados mediante análisis de espectrometría de masas (Maslov et al. 2006). Estudios llevados a cabo por el grupo de K. Stuart (Zíková et al. 2008), en *T. brucei*, permitieron identificar 133 proteínas de las cuales 38 presentan un grado variable de similitud con las proteínas ribosomales de bacteria u otras mitocondrias. Sin embargo, 56 proteínas no tienen homología por fuera de los kinetoplástidos. Se asignaron 77 proteínas a la subunidad mayor y 56 a la subunidad menor. La única proteína presente en los ribosomas mitocondriales que estaría codificada en el propio genoma del organelo, S12, no fue identificada mediante este abordaje. Con la excepción de este caso, las proteínas de los ribosomas mitocondriales de *T. brucei*, son significativamente más largas que su contraparte en bacteria. Como era de esperar, para la mayoría de las proteínas estudiadas se predijo la presencia de MTS en su secuencia (Zíková et al. 2008).

En eucariotas, el genoma mitocondrial además de codificar típicamente la mayoría de los ARNr, también codifica los ARNt necesarios para la síntesis de proteínas. Sin embargo, la importación hacia la mitocondria de ARNt codificados nuclearmente ha sido documentada en diversas especies (Schneider & Maréchal-drouard 2000). El número de ARNt a importar es altamente variable. El caso extremo lo encontramos precisamente en los kinetoplástidos, que deben importar el set completo de ARNt (Hancock et al. 1992; Simpson et al. 1989). El número de genes nucleares de ARNt en

estos parásitos es también variable siendo 82, 65 y 115 para *L. major*, *T. brucei* y *T. cruzi* respectivamente (Padilla-Mejía et al. 2009). Los ARNts se distribuyen entre la mitocondria y el citoplasma con diferente eficiencia, yendo desde 1 a 7,5% del ARNt total celular (Tan et al. 2002). Solamente dos ARNt se encuentran exclusivamente ubicados en el citoplasma: ARNt-Met(i) y ARNt-Sec, en cambio, no se detectó ningún ARNt exclusivamente mitocondrial (Tan et al. 2002). Aunque el rol de secuencias flanqueantes hacia el extremo 5' es bastante controversial (Tan et al. 2002), se ha demostrado la presencia de precursores de ARNt en la mitocondria de *T. brucei*. Este es el caso del precursor del ARNt-Ser que inicia 14nt corriente arriba del extremo 5' del ARNt maduro y se extiende hasta 59nt corriente abajo del gen (LeBlanc et al. 1999). En la misma línea, se evidenció la presencia de actividad ARNasa P en la mitocondria de estos parásitos, lo que permitiría el procesamiento del ARNt dentro de la mitocondria (Hancock et al. 1992). En el caso de *Leishmania* un determinante positivo de importación fue descubierto en el brazo D en el ARNt-Tyr (Mahapatra et al. 1998). Un motivo a nivel de secuencia (YGG(C/A)RRC), ubicado corriente arriba de genes que codifican para ARNt con destino mitocondrial de *T. brucei* podría estar implicado en la localización, siendo relevante en la distribución mitocondria/citoplasma de los ARNt endógenos del parásito, lo que le otorga un rol fundamental en el trafico intracelular (Sherrer et al. 2003) Al igual que para *T. brucei*, *T. cruzi* carece de genes que codifiquen para los ARNt mitocondriales a nivel del kinetoplasto, siendo los mismos codificados a nivel nuclear y teniendo que ser importados desde el citoplasma.

TRÁNSITO INTRACELULAR DE MENSAJEROS Y SUS ACTORES.

La distribución asimétrica del ARNm como mecanismo de regulación de la traducción se observa tanto en procariotas (Kannaiah & Amster-Choder 2014) como en eucariotas superiores (Jansen & Niessing 2012).

La mayoría de los estudios sobre traducción localizada se ha dado en células polarizadas (Sotelo-Silveira et al. 2006; Sotelo-Silveira et al. 2011; Perry et al. 2016; Fallini et al. 2016) o sistemas en desarrollo (Wreden et al. 1997; Lazzaretti et al. 2016). En los últimos años, se ha podido observar también traducción localizada cerca del núcleo, cloroplastos, peroxisomas y mitocondria (Weis et al. 2013; Zipor et al. 2009; Tian & Okita 2014; Zabezhinsky et al. 2016). Particularmente, en el caso de la ATP2p de levadura, que interviene en el control de la distribución asimétrica de su mensajero, se encontró que, además del proceso dependiente de la traducción, existen dos elementos de secuencia que participan en la localización subcelular de este ARNm. La estructura secundaria predicha a nivel de mensajero (Figura 5) se encuentra altamente conservada en al

menos 5 especies de levaduras (Garcia et al. 2010). Por otra parte, estudios previos habían revelado que la ocupación ribosomal y la tasa de traducción también influyen en la localización asimétrica de mensajeros (Paquin et al. 2007).

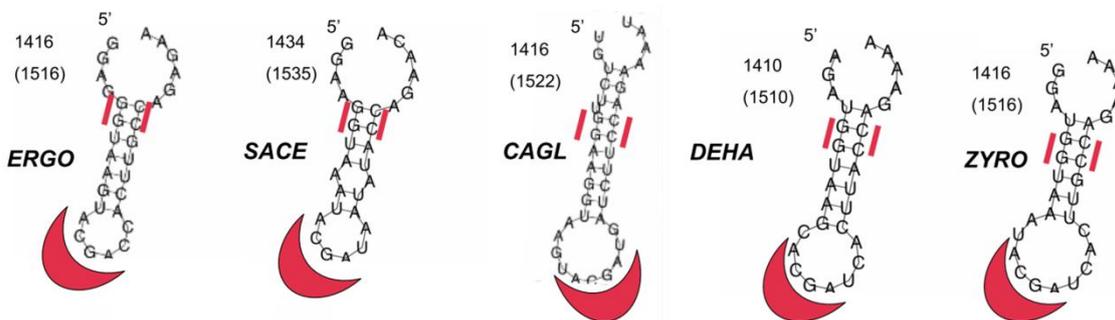


Figura 5. Estructura secundaria del ARN del motivo R2 que podría ser necesario para la localización asimétrica del ARNm. Se encontraron en el gen ATP2 de 5 especies de levaduras (<http://www.genolevures.org/concordance/>) CAGL = *Candida glabrata*; DEHA = *Debaryomyces hansenii*; ERGO = *Eremothecium gossypii*; SACE = *Saccharomyces cerevisiae* y ZYRO = *Zygosaccharomyces rouxii*. La posición de la primera base (5') se indica cerca de la posición de la última base, entre paréntesis. Se observó que todos los bucles contienen el motivo consenso ACGA (extraída de (Garcia et al. 2010)).

La localización de un ARNm en un entorno dado no necesariamente implica que el proceso de traducción sea llevado a cabo en ese lugar. Como ejemplo característico podemos citar la localización de mensajeros en *p-bodies*, donde estarían sometidos a un mecanismo de represión de la traducción (Decker & Parker 2012; Alves & Goldenberg 2016; Cassola 2011). Sin embargo, es razonable pensar que en aquellos casos en los cuales la localización del ARNm coincide con el destino subcelular de la proteína que codifica, la traducción sería llevada a cabo en ese sitio.

Si bien se conoce que el mecanismo de importación a la mitocondria de un gran número de proteínas codificadas a nivel nuclear se encuentra mediada por proteínas ubicadas en la membrana mitocondrial (Mani et al. 2015), no está completamente claro como las proteínas alcanzan esa localización (Figura 6). De los 780 genes nucleares que en levaduras codifican para proteínas mitocondriales 514 producen ARNm que son traducidos en el entorno mitocondrial (Saint-Georges et al. 2008). La asociación de ribosomas en el entorno de un organelo fue identificada por primera vez para el retículo endoplasmático, donde además se probó que el proceso de traducción se daba en el lugar (Palade 1955). En este caso se trata de la síntesis de proteínas de membrana o que tienen como destino ser secretadas. La síntesis de estas proteínas está acoplada a su transporte al lumen del retículo endoplasmático o a su membrana. También se observó que la localización de ciertos mensajeros y su traducción en el entorno del retículo endoplasmático podría ser una forma selectiva de respuesta a estrés por proteínas desplegadas (Reid et al. 2014).

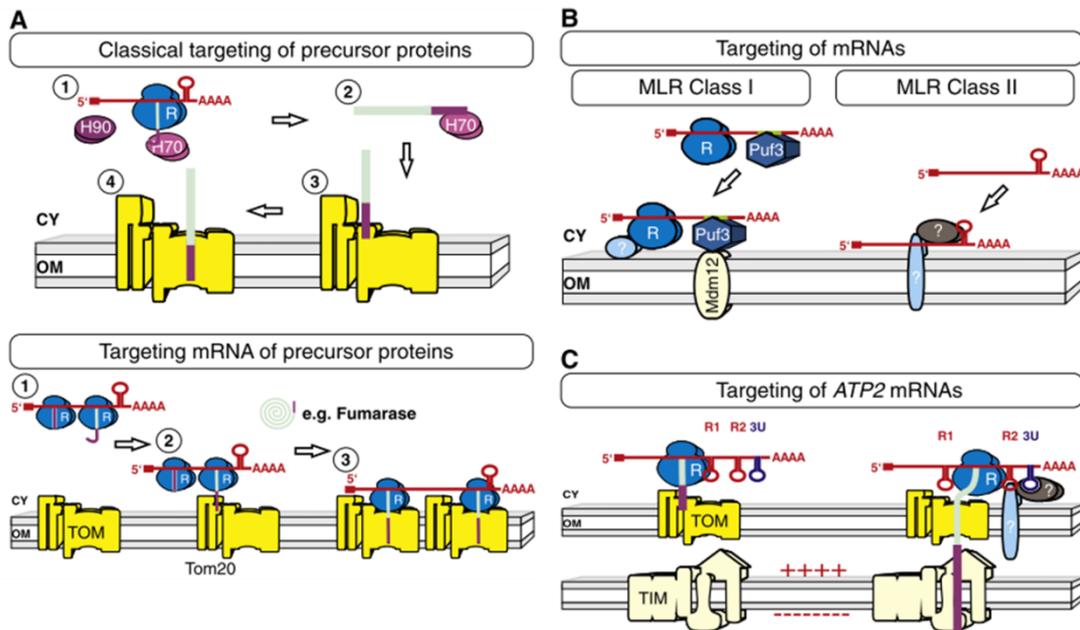


Figura 6. Mecanismo de ubicación de los mensajeros en el entorno mitocondrial. En A) Se observa la vía clásica de asociación mediante el péptido señal emergente. En este caso la proteína es sintetizada en el citoplasma (1), se asocia con chaperonas moleculares (2) que la obligan a permanecer en un estado desplegado y se importa de forma postraduccional vía el complejo TOM (3 y 4). Siguiendo esta misma línea existen casos en los cuales el precursor proteico es importado a la mitocondria donde es procesado y puede nuevamente ser exportado al citoplasma. En ambos casos el ribosoma asociado al ARNm es llevado al entorno mitocondrial vía la señal peptídica. Otra alternativa es la presentada en B) mediante la cual existe una asociación directa del ARNm con la superficie mitocondrial. Se ejemplifican dos vías para la localización diferencial del mensajero. En la clase I como ya se comentó, se encuentra la localización dependiente de Puf3 para los cuales los mensajeros que cuentan con una señal específica a nivel de la 3'UTR son reconocidos por esta proteína y llevados al entorno mitocondrial. Conectándose con Mmd12p. Por otro lado, la vía II es independiente de Puf3 y los protagonistas de este proceso se desconocen. Finalmente en C) se muestra el caso del ARNm de ATP2 en el entorno mitocondrial que corresponde a la vía 2 representada en B. Se vio que la asociación de este mensajero con la mitocondria se lleva a cabo mediante un mecanismo que involucra elementos en la región 3'UTR (3U) en la señal de localización mitocondrial (R1) y en la CDS (R2) (Cy, citoplasma OM, membrana externa; R, ribosoma). Extraído de (Weis et al. 2013).

Williams y colaboradores emplearon la técnica de perfilado ribosomal para aquellos ribosomas que estuvieran específicamente en la proximidad de la superficie mitocondrial de levaduras (Williams et al. 2014). A partir de ese estudio, encontraron que, en forma similar a lo que sucede en el retículo endoplasmático, la mayoría de las proteínas de la membrana interna fueron transportadas a la mitocondria de manera cotraduccional. De esta manera se confirmó que los ribosomas que se encuentran en el entorno mitocondrial, cerca de la membrana externa de la mitocondria, están activos y traduciendo proteínas mitocondriales (Williams et al. 2014). El mecanismo de traducción localizada en el entorno mitocondrial, ha sido estudiado en detalle (Williams et al. 2014; Lesnik et al. 2014; Gehrke et al. 2015).

Las proteínas mitocondriales se unen a la traslocasa en la membrana mitocondrial externa (complejo TOM) mediante la secuencias en la porción amino-terminal. Esto significa que el segmento inicial del polipéptido podría

unirse a la subunidad Tom20 en la membrana externa de la mitocondria, de modo de iniciar la importación cotraduccional y también impulsar el anclado de los ribosomas a la mitocondria (Garcia et al. 2010). El plegamiento de la proteína previene la importación de las mismas a través de las membranas mitocondriales, por lo que la importación de manera cotraduccional es energéticamente favorable ya que se evita el hecho de prevenir el plegamiento de la proteína (Lithgow 2000).

Por otra parte, las vías de exportación nuclear alternativas de mensajeros podrían ser avaladas por secuencias nucleotídica que necesitan ser traducidas para cumplir su función de señal de localización mitocondrial (Cenik et al. 2011). Esto sugirió que los mensajeros de aquellos genes que codifican para proteínas mitocondriales podrían compartir como característica particular, un mecanismo de exportación nuclear común. Vale la pena señalar que los transcritos para los cuales es atribuible un origen endosimbiótico, están enriquecidos entre aquellos que tienden a localizarse en el entorno mitocondrial (Lesnik et al. 2015). Por otra parte, análisis experimentales mostraron que la asociación entre los ARNm y la mitocondria es más leve cuando se agregó EDTA a la fracción mitocondrial apoyando la idea de que ribosomas (y por lo tanto, traducción activa) son importantes para la asociación (Lesnik et al. 2015).

OBJETIVO GENERAL

Contribuir a la caracterización de los procesos moleculares en *Trypanosoma cruzi* mediante la definición de señales de secuencia y/o estructurales específicas de transcritos de origen nuclear que codifiquen para proteínas mitocondriales.

OBJETIVOS ESPECÍFICOS

Construir una base de datos de genes nucleares que codifican proteínas con destino mitocondrial y estudiar sus características composicionales.

Agrupar estos genes en función de criterios vinculantes focalizados en la expresión diferencial y la funcionalidad.

- Determinar, en los agrupamientos, la existencia de señales de secuencia específicas.
- Determinar, en los agrupamientos, la existencia de señales conformacionales específicas.

MATERIALES Y MÉTODOS

En la Figura 7, se esquematiza la **estrategia** que se emplea en este trabajo con el fin de cumplir los objetivos planteados.

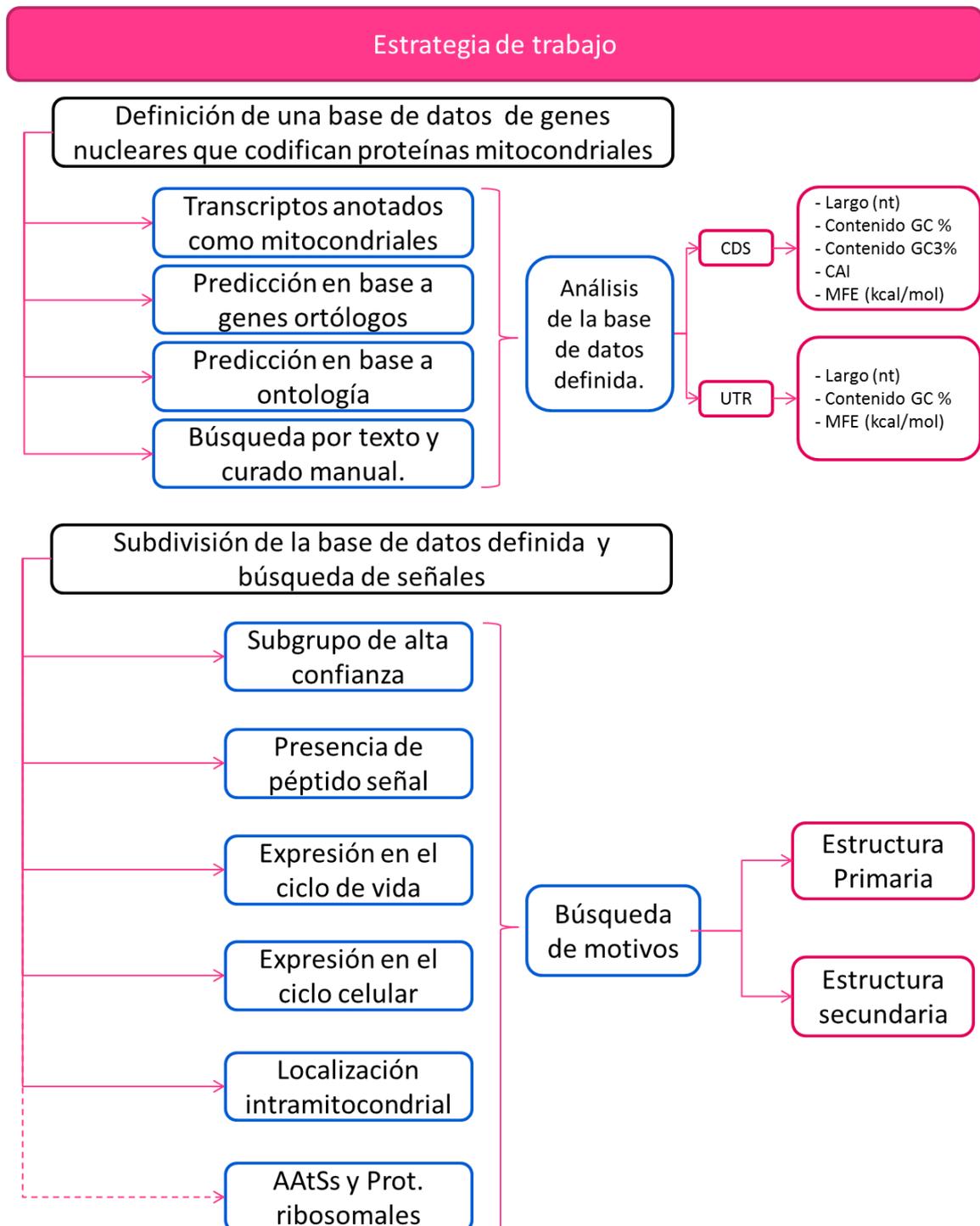


Figura 7. Esquema de la estrategia seguida en este trabajo. Contenido GC3: contenido GC en la tercera posición del codón, CAI: índice de adaptación de codones, MFE: mínimo de energía libre.

Bases de datos

Para la primer parte del trabajo se empleó la versión 4.0 de la base de datos de *Trypanosoma cruzi* disponible en Tritypdb.org (Aslett et al. 2010) (<http://tritypdb.org/common/downloads/release-4.0/Tcruzi/>). Se descargaron las bases de datos de CDSs y proteínas disponibles para el genoma *Esmeraldo-like*. Si bien se empleó la versión 4 para definir la base de datos de genes nucleares que codifican **proteínas con destino mitocondrial**, se actualizó a medida que cambió la versión. Por un lado se trabajó utilizando filtros de *bash* en línea de comando¹, con el fin de encontrar los genes anotados como tales, generar las listas y verificar la interacción entre subgrupos. Por otro lado se emplearon las herramientas disponibles en *TriTrypDB*. Este procedimiento se detalla en cada paso.

Para el análisis de las **UTRs** se emplearon los datos obtenidos a partir de experimentos de NGS por parte del grupo del Dr. El-Sayed (Li et al. 2016) para los estadios tripomastigota y amastigota y en nuestro laboratorio por parte del Mag. Santiago Chávez para el estadio epimastigota (Chávez 2016). Siempre que fue posible se utilizaron estas regiones UTR definidas a partir de los datos de NGS. Las mismas no fueron utilizadas en todos los casos debido a limitaciones de los programas empleados. En estos casos, las bases de datos tanto de las regiones 3'UTRs como las 5'UTRs se generaron empleando las herramientas disponibles en *TriTrypDB*, en función de los largos promedio reportados para las 5' y 3' UTR de *Trypanosoma brucei* (Campos et al. 2008) y de *T. cruzi* (Brandão & Jiang 2009). Se definieron las regiones 5'UTR como la secuencia desde 80 nt corriente arriba de la adenina del codón de inicio hasta 270 nt corriente debajo de la misma (350nt total), y la región 3'UTR de 350nt corriente abajo del codón de terminación anotado (Figura 8).

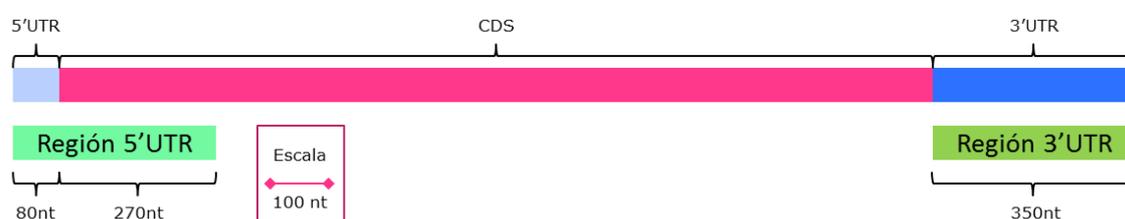


Figura 8. Representación de las regiones empleadas en los análisis de búsqueda de secuencia.

Por su parte para la **comparación de motivos lineales** encontrados en este trabajo con respecto a los presentes en otros organismos se empleó la base de datos de motivos de ARN descrita en (Ray et al. 2013).

¹ Todos los análisis computacionales se llevaron a cabo utilizando software de acceso libre disponible en la web y/o compilado para entorno LINUX (Ubuntu 14.04)

Métodos de análisis

En cuanto a la definición de **ortólogos** se empleó por un lado la herramienta disponible en la base de datos Tritypdb.org con sus parámetros por defecto y por otro lado en casos donde la búsqueda de ortólogos involucró organismos ajenos a la base de datos se recurrió al método computacional de “*Best Reciprocal BLAST Hits*” (Ward et al. 2014). Para ello se descargaron las secuencias de proteínas anotadas para los organismos a comparar y se llevó a cabo el análisis utilizando *scripts* en *bash* desarrollados para este fin.

Para determinar las **categorías GO** (*gene ontology*) se empleó la herramienta en línea DAVID (*Database for Annotation, Visualization and Integrated Discovery*, versión 6.7), diseñada para analizar datos de genómica funcional (Glynn Dennis 2003). Este programa permite agrupar genes según su ontología así como comprobar si alguna de las categorías encontradas se encuentra sobrerrepresentada con respecto a lo que se espera por azar teniendo en cuenta la proporción de genes que caen dentro de esta categoría en el genoma (o una lista de referencia arbitraria). La herramienta clasifica también las proteínas según los dominios presentes en la proteína según InterPro. Consideramos un grupo sobrerrepresentado si presenta un “Puntaje de Enriquecimiento” ES mayor a 1,3 (*Enrichment score*, definido como el menos logaritmo de la media geométrica de los valores p) (Huang et al. 2009). Los valores p son calculados utilizando una variante del test exacto de Fisher (*EASE score*). Para cada categoría ontológica dentro de la agrupación, se controló el efecto del testeo múltiple sobre los valores p obteniéndose el valor de FDR (*false discovery rate*) mediante el método de Benjamini (Benjamini et al. 2001), aceptándose valores menores a 0,05. Las listas usadas como referencia fueron los genes del haplotipo Esmeraldo de *T. cruzi* o el subconjunto de genes de origen nuclear que codifican para productos mitocondriales, según el caso. Dentro del paquete DAVID se utilizó la herramienta “Agrupamiento por Anotación Funcional” (*functional annotation clustering*) que se centra en el análisis de las categorías de GO que están enriquecidas en la lista de genes analizados y las agrupa en función de los genes que contienen en común. Esto disminuye sustancialmente la redundancia en los reportes de categorías individuales más clásicos (Huang et al. 2009).

Alternativamente se empleó la herramienta ofrecida por el *TriTrypDB* gracias a la cual se pueden analizar los subconjuntos de genes ingresados en búsqueda de enriquecimiento significativo en cuanto a **términos de ontología representados o rutas metabólicas**. Dentro de la base, en la sección “Análisis de resultados” (*analyze results*) se encuentran disponibles herramientas, dos de las cuales se emplearon en este trabajo: *Gene ontology enrichment* y *methabolic pathway enrichment*. En ambos casos se aceptaron valores de *p-value* menores que 0,01.

A la hora de buscar características distintivas para la base de transcritos nucleares que codifican para proteínas mitocondriales se analizaron varios aspectos. Entre ellos encontramos la fracción de **contenido GC** de los transcritos que fue calculada empleando la herramienta disponible en el paquete EMBOSS <http://www.bioinformatics.nl/cgi-bin/emboss/geecee/>. Por otro lado **el contenido de GC en la tercera posición del codón** (%GC3) para los análisis de las CDSs correspondientes se llevó a cabo mediante el programa INCA (*INteractive Codon usage Analysis*) (Supek & Vlahovicek 2004)

Por otra parte, dentro de los índices que se emplean para medir el sesgo en el uso de codones, empleamos el **índice de adaptación de codones** (*codon adaptation index*, CAI) que mide la desviación de un gen particular con respecto a un grupo de referencia (Sharp & Li 1987). Para esto se utiliza un grupo de proteínas de alta expresión como referencia, en nuestro caso las proteínas ribosomales, ya que se espera que estos presenten el mayor grado de optimización en el uso de los sinónimos.

Otra rasgo importante a la hora caracterizar nuestro subgrupo de interés es el grado de estructuración del transcripto. Para ello se suele emplear como parámetro el **mínimo de energía libre** (MFE) para la estructura secundaria de un transcripto o región del mismo. En este caso, empleando el algoritmo RNAfold del paquete Vienna RNA 2.0 (Hofacker 2003) se calculó el mínimo de energía libre para las regiones 5' y 3' UTR de los transcritos analizados así como la de las regiones codificantes. Por otra parte, como se detallará más adelante esta herramienta fue empleada para el cálculo del MFE de los motivos estructurales encontrados. Se emplearon los parámetros por defecto.

A la hora de la definición de señales dentro del subgrupo de transcritos que codifican para proteínas mitocondriales, se procedió a generar diferentes subgrupos con criterios racionales (ver estrategia en Figura 7). En el caso de la diferenciación entre transcritos que codifican una **señal de localización peptídica** y los que no, se emplearon 3 aproximaciones diferentes. En un caso se utilizó la herramienta MEME dentro del paquete MEME-suite (Bailey et al. 2009) para definir una señal primaria utilizando como grupo de entrada proteínas en cuya secuencia hay evidencias de la presencia de un péptido señal. Este motivo se optimizó como se detalla más adelante en la sección Resultados y discusión, empleando además la herramienta FIMO (Grant et al. 2011) del mismo paquete. Una vez obtenido el motivo optimizado se llevó a cabo la búsqueda propiamente dicha en la base de datos de transcritos que codifican proteínas mitocondriales empleando nuevamente la herramienta FIMO. Por otro lado, la búsqueda se amplió empleando herramientas en línea como TargetP (Emanuelsson et al. 2007), PredSL (Petsalaki et al. 2006) para las cuales no se modificaron los parámetros pre-definidos empleando como entrada a ambos programas un archivo multifasta conteniendo los primeros

151 aminoácidos de cada proteína mitocondriales codificadas a nivel nuclear. Para la discriminación final de aquellas proteínas que poseen o no señal peptídica de localización mitocondrial se empleó una estrategia semi-exclusiva mediante la cual solo se tuvieron en cuenta aquellas proteínas para las cuales al menos dos de los 3 métodos predijeron la presencia de péptido señal.

Otros subgrupos fueron generados teniendo en cuenta los datos de **expresión** tanto a lo largo del ciclo de vida (Minning et al. 2009) como del ciclo celular (Chávez 2016). Para esto se generaron 6 agrupamientos, en cada caso, según su perfil de expresión diferencial empleando la herramienta Java MeV (*Multiple Experiment Viewer* versión 4.9.0). Este programa además de soportar distintos formatos de entrada ofrece una interfaz gráfica para visualizar los agrupamientos y llevar a cabo análisis estadísticos pertinentes.

Una vez definidas las bases de datos de regiones UTR para cada agrupamiento, y para el resto del transcriptoma en general, se procedió a la definición de señales enriquecidas en los subgrupos con respecto al resto del transcriptoma. En primer lugar, para la búsqueda de **señales a nivel de secuencia (LiMs = *Linear Motifs*)** en ácidos nucleicos, se empleó el programa DREME (Bailey 2011), perteneciente al paquete MEME-Suite (Bailey et al. 2009), de forma local. Se buscaron motivos que estuvieran enriquecidos en el subgrupo analizado con respecto al subgrupo complementario (que se especifica caso a caso). Solamente se tuvieron en cuenta aquellos motivos que estuvieran presentes en la hebra transcripta y con un *E-value* mayor a 0,01. Los motivos obtenidos se compararon empleando la herramienta Tom-Tom (Gupta et al. 2007), del mismo paquete (MEME-suite) con el fin de comparar cada motivo contra una base de datos de motivos conocidos. En este caso empleamos la base de datos proporcionada por Ray y colaboradores, (Ray et al. 2013), que permite además asociar los motivos a proteínas de unión en diversos organismos. En este caso el *E-value* que se empleó como límite, se aclara en cada caso. En los casos en que se consideró pertinente se intentó definir la ubicación de los motivos empleando la herramienta FIMO contra la base de datos que generó al motivo.

Por otra parte, también empleando las bases de datos de regiones UTRs definidas, se definieron **señales a nivel de estructura secundaria (StMs = *Structural Motifs*)**. Los elementos de estructura fueron identificados usando CMfinder 0.2 en su versión local (Yao et al. 2006) (<http://bio.cs.washington.edu/yzizhen/CMfinder/>). Los parámetros empleados en las búsquedas fueron *-s1* (número de tallos por estructura), *-f 0.6* (fracción de secuencias conteniendo el motivo), y *-c10* (número de estructuras candidatas por secuencia) al igual que los utilizados por el Dr. De Gaudenzi y colaboradores (De Gaudenzi et al. 2013) que tienen como finalidad identificar únicamente los motivos más confiables. Los elementos

encontrados dentro de cada agrupamiento fueron empleados como entrada del programa *Infernal* (Nawrocki et al. 2009). *Infernal* permite la implementación de un modelo de covarianza (CM) que puntúa la combinación entre secuencia consenso y la estructura secundaria consenso permitiendo de esta forma la identificación de conservación a nivel de la estructura secundaria del ARN. Para ello se construyó un modelo de covarianza a partir de la salida de *cmfinder* empleando la herramienta *cmbuild* para luego calibrar los parámetros para el modelo empleando *cmcalibrate*. Los modelos obtenidos fueron usados como interrogantes en búsquedas más refinadas empleando el algoritmo *cmsearch*. Para validar el uso de estos modelos se evaluó: la cobertura a nivel de subgrupos, la sensibilidad de la búsqueda, la especificidad del motivo encontrado así como la tasa de falsos positivos. De esta forma, se definió la cobertura (Figura 9) del motivo como la fracción de secuencias de un subgrupo que aportan al motivo sobre la cantidad de secuencias del subgrupo. La sensibilidad fue calculada considerando que fracción de las secuencias que aportan al motivo fueron luego levantadas en una búsqueda recíproca empleando *CMsearch* (m_Grupo 1 vs Grupo 1, donde m_Grupo 1 es el motivo al que contribuyen miembros del Grupo 1). La tasa de falsos positivos y por ende la especificidad del modelo fue calculada realizando la búsqueda del motivo en 5 x n grupos al azar de los transcriptos que no se encuentran en la base de datos de interés, y del mismo tamaño que el grupo a partir del cual se generó el motivo analizado. De esta forma, la fracción de secuencias identificadas en los grupos al azar sobre el total fue tomada como falsos positivos, mientras que la especificidad fue calculada como "1 -Tasa de falsos positivos". Como criterios de aceptación de los motivos nos propusimos que sólo serían válidos aquellos que en primer lugar su especificidad fuera mayor o igual al 90%, la sensibilidad superase el valor de los falsos positivos más dos desvíos estándar y a su vez la estabilidad de la estructura, reflejada mediante el cálculo del mínimo de energía libre de formación de la misma fuera menor que el promedio del conjunto de estructuras generadas en la búsqueda. Cabe resaltar que tanto para el modelado de la estructura como para el cálculo del MFE se empleó la herramienta *RNAfold* en línea de comando, forzando el cálculo de energía libre de la estructura consenso encontrada ("*-C --enforceconstraint*").

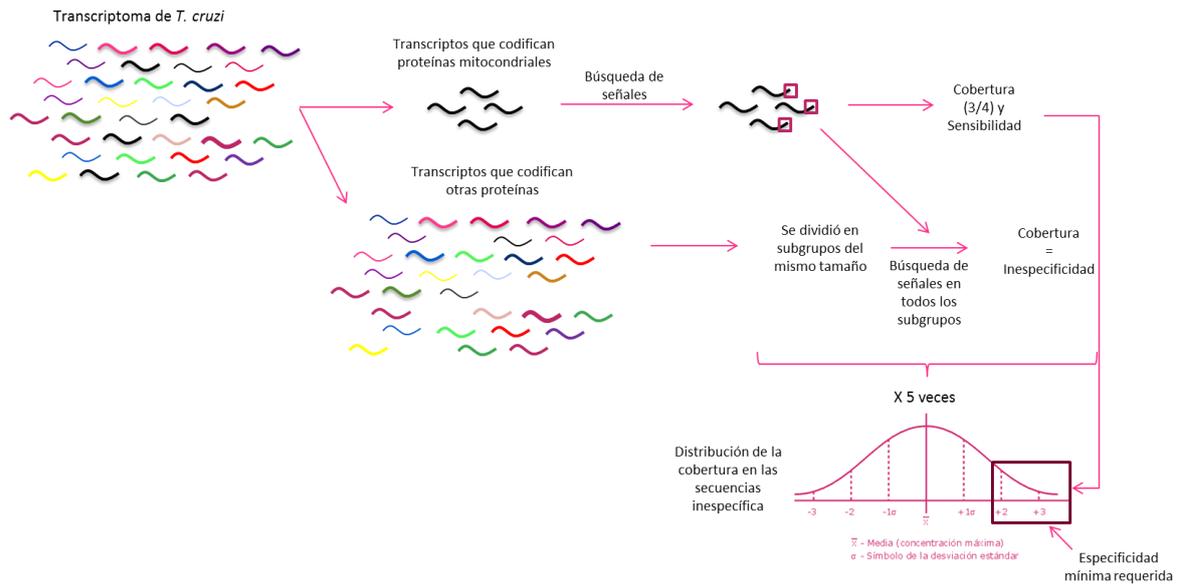


Figura 9. Estrategia de validación de las señales de estructura secundaria encontradas. A partir de los subgrupos de mensajeros nucleares que codifican proteínas mitocondriales se lleva a cabo la búsqueda de señales. El número de mensajeros que contiene esta señal se compara con el número de transcritos del subgrupo analizado. De esta forma se informa la cobertura de la señal (en el ejemplo, $\frac{3}{4}$). Por otro lado la señal se busca nuevamente en el subgrupo a partir del cual se generó. El resultado de esta búsqueda se compara con la cobertura. De esta forma, si se encuentra el motivo en todos los transcritos que aportan al mismo, la sensibilidad es del 100% y si en caso contrario, el motivo se encuentra en menos transcritos la sensibilidad será menor. Por otro lado el análisis de especificidad se lleva a cabo empleando el resto de los transcritos de *T. cruzi*. En este caso se generan n grupos al azar conteniendo el mismo número de transcritos que el subgrupo a partir del cual se generó la señal a analizar. De esta forma se busca la señal en cada uno de los n subgrupos. En cada caso se analiza la cobertura, que en este caso corresponde a la inespecificidad del motivo. Este procedimiento se repite 5 veces, generándose una distribución de inespecificidad. Para que el motivo sea aceptado por este criterio, debe ocurrir que la sensibilidad calculada para el mismo, supere al promedio más dos desvíos estándar en la distribución de inespecificidad.

Finalmente, a la hora de la presentación de las **secuencias consenso** encontradas para las estructuras primarias como secundarias, en caso de ser necesario se empleó la herramienta Weblogo (Crooks et al. 2004) disponible en línea.

Análisis estadístico

Los **análisis estadísticos** empleados a lo largo del trabajo, *test de chi-cuadrado* y *one way anova*, se llevaron a cabo en **R**.

RESULTADOS Y DISCUSIÓN

IDENTIFICACIÓN Y CARACTERIZACIÓN DE GENES NUCLEARES QUE CODIFICAN PARA PRODUCTOS MITOCONDRIALES.

Selección y predicción de genes nucleares codificantes de productos mitocondriales.

Con el objetivo de construir una **base de datos** de genes nucleares que codifican para productos mitocondriales en *T. cruzi* se buscaron en primer lugar los genes cuya anotación indica que codifican proteínas mitocondriales. Para ello, se trabajó con el genoma de *T. cruzi* *CL Brener Esmeraldo like* (Minning, Weatherly et al. 2009) (<http://tritrypdb.org/common/downloads/release-4.0/Tcruzi/>). Al momento de resumir los resultados, los mismos se actualizaron empleando la versión 9.0 (diciembre de 2014). De los 10597 genes que se encuentran en la base de datos para esta cepa el 51% (5362) se encuentran anotados como genes hipotéticos, de los cuales el 94% codificarían para proteínas. Utilizando el archivo de anotación y trabajando en línea de comando en *Bash* se generó un primer grupo de genes correspondiendo a aquellos que codifican para proteínas mitocondriales y están **anotados** como tales (105 genes). Estos genes se hayan distribuidos a lo largo de los cromosomas del parásito no presentando una distribución diferencial aparente.

Dado que el abordaje basado en la anotación rindió resultados sumamente exigüos, a modo de completar la base de datos de genes nucleares que codifican proteínas mitocondriales se buscaron diferentes estrategias (Figura 10). En primer lugar se llevó a cabo una búsqueda por **ontología** empleando las herramientas disponibles en Tritrypdb.org (*Cellular Component* = "*Mitochondrial*") que permitió identificar 292 proteínas que incluían a las 105 identificadas previamente.

Teniendo en cuenta que en *Trypanosoma brucei* se cuenta con datos que indican un número mayor de genes nucleares codificantes de productos con destino mitocondrial, aproximadamente 1.065 genes (Zhang et al. 2010), se decidió realizar la búsqueda de los ortólogos correspondientes en *T. cruzi*. Para ello, basado en el método de *Best Reciprocal BLAST Hit*, se llevó a cabo un BLAST recíproco. Utilizando como base las proteínas mitocondriales de *T. brucei* codificadas a nivel nuclear, para las cuales existe una evidencia sólida de su destino celular, agrupadas en la base de datos que se denominó "**MitoCarta**", (Zhang et al. 2010), se procedió a definir una lista de genes de *T. cruzi* ortólogos a las mismas. Se encontró que de los 479 genes codificantes para proteínas mitocondriales de *T. brucei*, 339 poseen ortólogos en *T. cruzi*. De estas, 47 coincidían con los genes seleccionados por anotación, 86 con los genes seleccionados por ontología, y 32 genes eran comunes a las tres estrategias hasta el

momento realizadas. De esta forma se llegó a una base de datos de 527 genes.

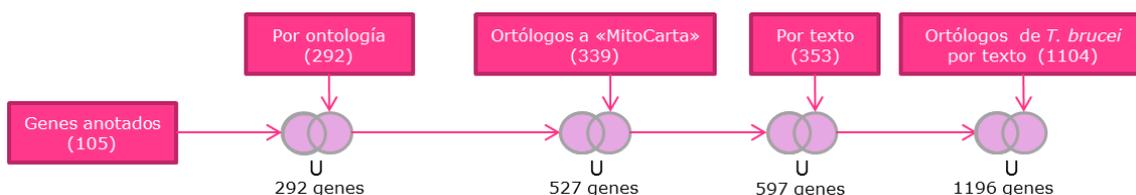


Figura 10. Estrategia empleada para la generación de la base de datos que nominamos MiNT.

Se describe paso a paso la estrategia empleada en la búsqueda de transcritos de proteínas mitocondriales codificadas a nivel del genoma nuclear.

Otra estrategia alternativa y complementaria para detectar genes nucleares que codifican productos mitocondriales se basó en realizar una búsqueda por **“texto”** (en TritypDB.org) sin acotar los campos. De esta forma se encontraron 357 genes en *Trypanosoma cruzi* relacionados a mitocondria. El resultado de esta búsqueda fue curado manualmente y de esta forma 4 de los transcritos fueron desestimados, resultando entonces en un total de 353 genes. Un subgrupo de 283 genes coincide con los 527 de la base previamente generada. Los 70 genes restantes, detectados exclusivamente por esta aproximación se analizaron empleando el método de agrupación de función DAVID. Estos genes mayoritariamente se agrupan en varios términos entre los que se encuentran: respiración celular, procesos asociados al ciclo de Krebs, etc. Este resultado apoya la inclusión de estos genes en la base de datos.

Por último, se llevó a cabo una búsqueda de los genes de *T. brucei* **Treu927** que estuvieran anotados como mitocondriales, usando el mismo criterio de búsqueda poco acotado que para *T. cruzi*. De esta manera, se encontraron 1.242 genes mitocondriales en *T. brucei* de los cuales 1.104 presentarían **ortólogos** en *T. cruzi*. Esta alternativa permitió la detección de 599 genes que no habían sido incluidos en la base de datos siguiendo las aproximaciones anteriores. De forma similar a lo realizado anteriormente, se estudió la ontología de los componentes de este subgrupo nuevo, encontrándose un enriquecimiento en términos asociados con el balance redox, síntesis de aminoacil-ARNt, síntesis de ácidos nucleicos y producción de energía. Nuevamente, este resultado apoya la inclusión de estos genes en la base de datos.

De esta manera y gracias al aporte de las diferentes estrategias, se completó la base de datos que denominamos **MiNT** (*Mitochondrial Nuclear Transcripts*) conteniendo un total de 1.196 genes nucleares de un total 10.597, que codificarían para proteínas mitocondriales. Este número de genes resulta significativamente comparable con el hallado por el grupo de Ochsenreiter (Zhang et al. 2010) para *T. brucei*. En ese caso se predijeron 1.065 genes que codifican para proteínas mitocondriales de un total de 12.065.

De acuerdo con lo esperado ($p\text{-value} < 0,002$), la base MiNT tiene un 54% (652/1.196) genes que codifican **proteínas hipotéticas**, de las cuales, por encima de lo esperado ($p\text{-value} < 0,0001$), el 82% (535/652), presenta ortólogos tanto en *T. brucei* como en *L. major* (Kissinger 2006). Por otro lado con respecto a MiNT, existen evidencias de expresión, *expressed sequence tags* (ESTs), para el 39% (470/1.196) de los genes. Sin embargo datos de RNAseq (Li et al. 2016; Chávez 2016) y micro-arreglos (Minning et al. 2009) empleados en este trabajo, avalarían la expresión de al menos el 98% (1.167/1.196) de los genes de la base de datos. El 2% restante (Tabla 1) codifica para productos con funciones asociadas a transporte a través de la membrana mitocondrial, respuesta a estrés oxidativo, etc. La falta de detección de su expresión puede deberse a problemas metodológicos, niveles de expresión reducidos, o expresión sujeta a condiciones particulares. Resulta conveniente notar que de los 29 genes para los cuales no contamos con evidencia de expresión únicamente 5 codifican proteínas hipotéticas. Por ello, de todas formas, estos genes serán incluidos en los análisis globales que se detallan a continuación.

Tabla 1. Términos de ontología enriquecidos en el subgrupo de genes para los cuales no se cuenta con evidencia de expresión.

ID	Término	Fold enrichment
GO:0042127	<i>regulation of cell proliferation</i>	98,10
GO:0006952	<i>defense response</i>	49,05
GO:0006839	<i>mitochondrial transport</i>	39,24
GO:0006801	<i>superoxide metabolic process</i>	28,03
GO:0006378	<i>mRNA polyadenylation</i>	24,52
GO:0006811	<i>ion transport</i>	7,64
GO:0006468	<i>protein phosphorylation</i>	6,23
GO:0006793	<i>phosphorus metabolic process</i>	4,63
GO:0036211	<i>protein modification process</i>	3,74
GO:0043412	<i>macromolecule modification</i>	3,18

Se construyó la base de datos MiNT que contiene 1196 genes nucleares que codifican proteínas con destino mitocondrial. Este número es comparable con lo reportado en organismos relacionados.

Análisis composicional de transcritos de MiNT

A modo de caracterizar la base de datos MiNT, aquí definida, se llevó a cabo un estudio de los transcritos correspondientes. Para esto contamos con los datos de experimentos de NGS para los estadios amastigota, tripomastigota, (Li et al. 2016) y epimastigota (Chávez 2016) del parásito. El estudio se realizó tanto para las **regiones codificantes** (CDSs) como para las **transcritas no traducidas** (UTRs). En el caso de estas últimas regiones, el análisis inicial global mostró una amplia distribución en el largo de las UTRs (Figura 11), por lo que se decidió trabajar con aquellas UTRs que se encuentren entre los percentiles 2,5% y el 97,5% eliminando regiones posiblemente mal determinadas.

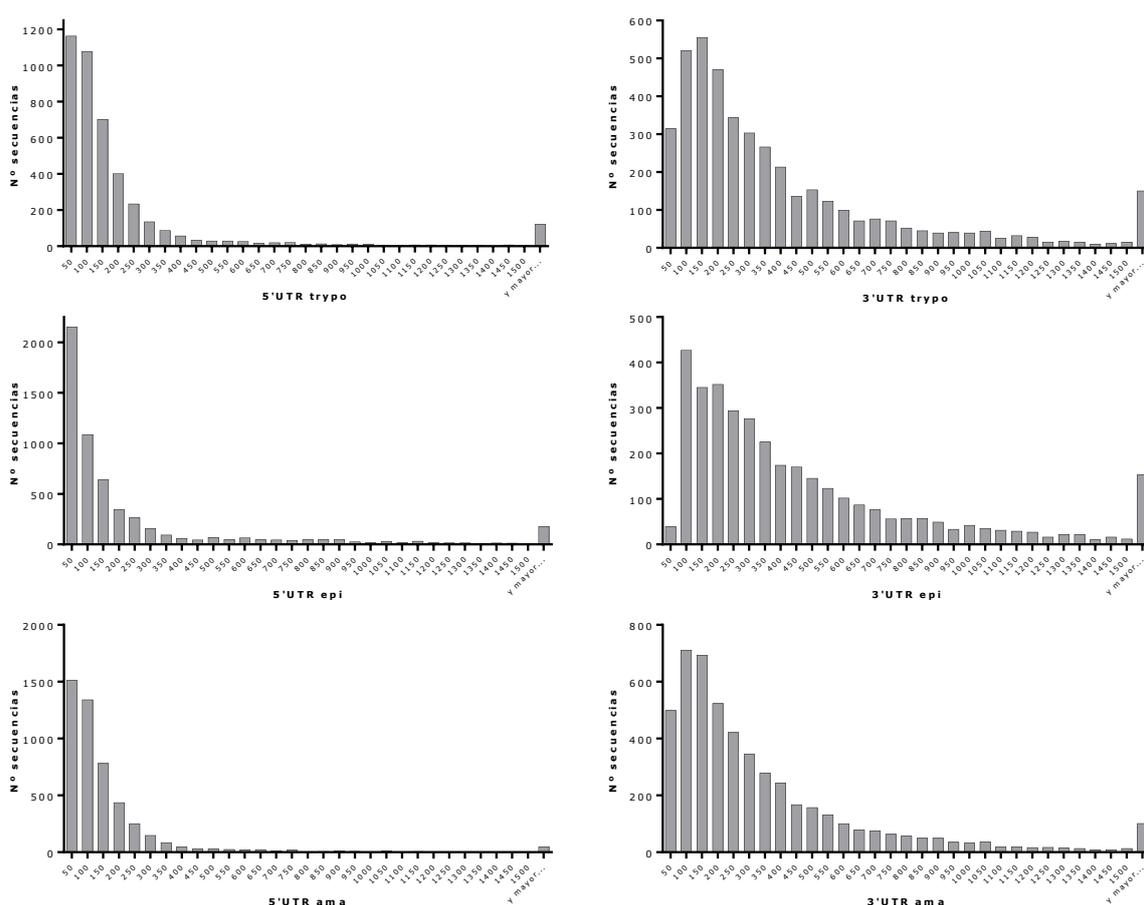


Figura 11. Distribución de largos de las regiones no traducidas de los transcritos en general para los distintos estadios del ciclo de vida del parásito.

Para todas las regiones se estudiaron las siguientes **características**:

- Largo (pb) de las regiones no traducidas
- Contenido GC
- Nivel de estructuración

Con respecto a las regiones codificante se analizó además:

- Contenido GC3
- Índice de adaptación de codones (CAI)

Se comparó mediante métodos estadísticos los **largos** obtenidos para las **UTRs** empleados en este trabajo en comparación con los datos obtenidos por el grupo de Brandão (Brandão & Jiang 2009) donde, a pesar de las diferencias encontradas en el tamaño de la muestra, los valores obtenidos son comparables (Tabla 2). La mediana del largo de las 5'UTRs empleados en este trabajo varía entre 94, 82 y 84 nucleótidos en el estadio tripomastigota, amastigota y epimastigota respectivamente, mientras que para el caso de la 3'UTR varía entre 241, 207 y 302, respectivamente (Figura 12).

Tabla 2. Comparación del largo de las regiones UTR determinadas experimentalmente. *

Valores se refieren al promedio de los largos encontrados para los 3 estadios por separado.

Fuente	(Brandão & Jiang 2009)		(Chávez 2016; Li et al. 2016)	
Largo (pb)	5'UTR	3'UTR	5'UTR*	3'UTR*
Mínimo	11	17	10	32
Máximo	526	2847	1250	1889
Promedio	81	334	156	360
Mediana	55	289	86	250
1er cuartil	36	135	42	132
3er cuartil	104	419	174	472
20 percentil	33	108	37	111
80 percentil	118	453	210	546
Número de secuencias	173	179	5177	4053

En el caso de los genes codificantes para proteínas mitocondriales se encontró que el largo de las UTRs varía entre 88, 77 y 81 para la 5'UTR de tripomastigota, amastigota y epimastigota mientras que la 3'UTR varía entre 221, 201 y 258 respectivamente.

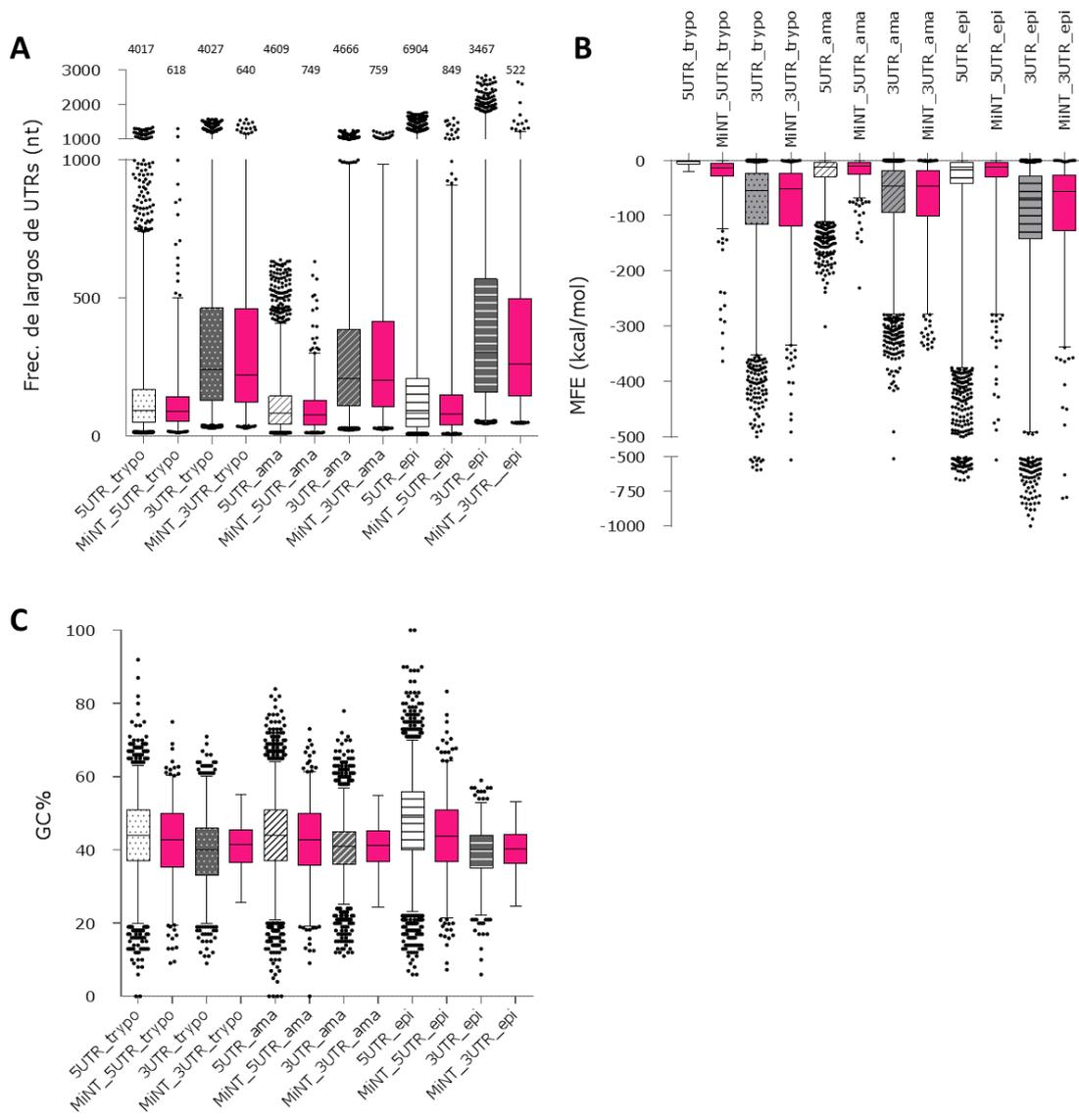


Figura 12. Comparación de las características de las regiones UTR de los transcritos de MiNT vs el transcriptoma completo de *T. cruzi*. A) Boxplot comparativo del largo de 3' y 5' UTRs para los transcritos nucleares obtenidos mediante experimentos de NGS para los estadios amastigota (5UTR_ama, 3UTR_ama), epimastigota (5UTR_epi, 3UTR_epi) y tripomastigota (5UTR_trypo, 3UTR_trypo) entre los transcritos de *T. cruzi* en general y aquellos que pertenecen a MiNT (prefijo: MiNT). B) Comparación del grado de estructuración de las regiones no traducidas de los mensajeros entre ambos subgrupos. C) Comparación del porcentaje de GC dentro de cada región no traducida para los distintos estadios, entre ambos subgrupos.

Mediante este análisis no se encontraron características particulares para los transcritos de MiNT en cuanto al **largo** o **grado de estructuración** de las regiones no traducidas (Figura 12). Sin embargo, si se observa una diferencia significativa ($p\text{-value} < 0,0001$) en cuanto al **contenido GC%** de los transcritos que pertenecen a MiNT y la tendencia general de los transcritos de *T. cruzi* para el estadio epimastigota. Por otro lado si bien el alto contenido en GC% podría sugerir en un principio mayor grado de estructuración, esta diferencia no se ve directamente reflejada en cuanto al

comportamiento general al comparar los valores de energía libre. Cabe resaltar que la dispersión de los datos con los que contamos para el estadio epimastigota es mayor cuando se lo compara con el resto de los estadios. La diferencia en cuanto a la dispersión de los datos puede deberse a la diferencia en las metodologías empleadas para la obtención de los mismos. Esta dispersión podría ser en parte responsable de la diferencia observada para el contenido GC%.

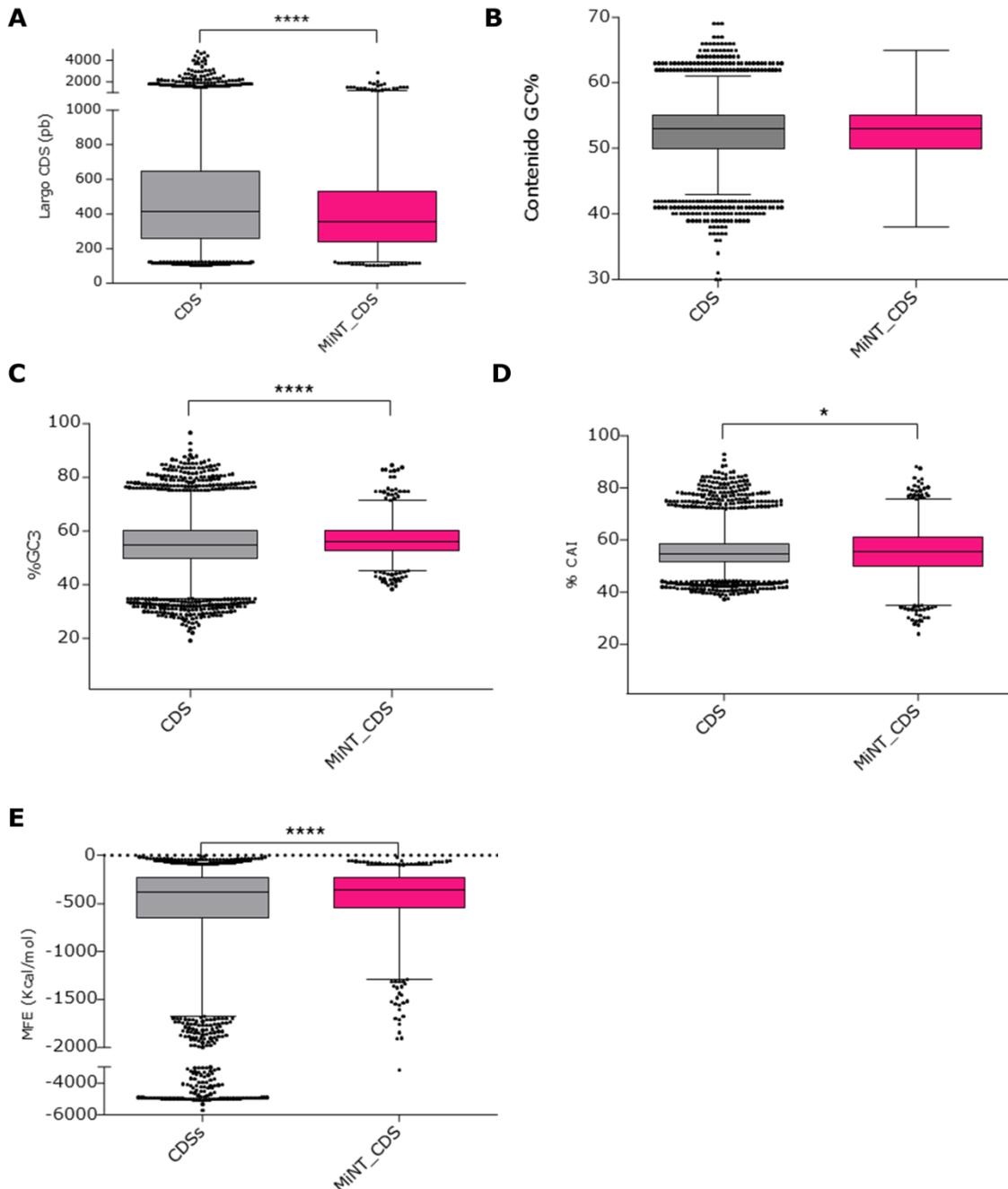


Figura 13. Comparación de las características de las CDSs de los transcritos de MiNT vs el transcriptoma completo de *T. cruzi*. Boxplots comparativo del A) largo (p -value < 0,0001), B) contenido GC% (p -value = 0,2316) C) %GC3 (p -value < 0,0001) D) %CAI (p -value = 0,0189) y E) Mínimo de energía libre de estructuración (p -value = 0,0009) de las CDSs, entre ambos subgrupos.

Con respecto a las **CDSs** (Figura 13), se observó que la diferencia es significativa en mayor o menor medida para la mayoría los parámetros estudiados. Por un lado teniendo en cuenta el **origen procariótico** de la mitocondria esperamos ver que se conserven alguno de los parámetros que evolutivamente se han mostrado distintos. Con respecto al largo de los genes se ha visto que en promedio el largo de los genes eucariotas es mayor que el de los procariotas, tanto para las proteínas que poseen ortólogos en ambos organismos como en general (Xu et al. 2006). En este sentido se observa que el **largo** de las CDSs es significativamente menor para aquellos que codifican proteínas mitocondriales consistente con la hipótesis de que estas proteínas se encontraban originalmente codificadas a nivel del genoma mitocondrial. Mientras que, a pesar de que el **contenido GC%** no difiere entre los dos grupos, el **contenido de G+C en la tercera posición del codón** y el **índice de adaptación de codones** es significativamente mayor para las CDSs de transcritos pertenecientes a MiNT. Por un lado tanto un alto contenido en GC3 como un alto CAI se asocia en general con alta tasa de producción de proteínas. Por otra parte se ha encontrado que en *T. cruzi* a diferencia de lo que sucede en otros tripanosomátidos, *T. vivax* y *T. brucei*, si bien si se cumple para aquellos genes que se encuentran conservados, el contenido en la tercera posición del codón en general no refleja lo esperado (Rodríguez 2013). En este caso al ser un subgrupo acotado de proteínas que son necesarias para el correcto funcionamiento del organelo, esperamos que la mayoría de las mismas estén siendo producidas con una tasa relativamente alta. Con respecto al **grado de estructuración** se observa menor complejidad a lo largo de las CDSs de transcritos pertenecientes a MiNT. En estudios realizados sobre transcritos de procariotas con respecto a eucariotas en general se había observado una diferencia significativa en cuanto al plegamiento de los transcritos con respecto a la misma secuencia al azar concluyéndose que la selección natural estaría actuando en pro de mantener estas estructuras secundarias. Sin embargo para eucariotas no sucedería lo mismo (Solaimuthu 2010) apoyando el resultado obtenido.

En resumen se observó que las CDSs de los transcritos de MiNT poseen menor tamaño y menor grado de estructuración que resto de los transcritos de T. cruzi. Esto es similar a lo observado cuando se comparan genes de origen procariota y eucariota lo cual puede estar reflejando el posible origen mitocondrial de los genes de MiNT. El contenido GC3 y el CAI sugieren que estos transcritos presentarían una alta tasa traduccional.

ANÁLISIS DE LA BASE MINT POR CATEGORÍAS

Considerando que señales compartidas por subgrupos de genes podrían no ser detectadas en un análisis global, debido a limitaciones de los algoritmos disponibles, se decidió proceder a realizar agrupamientos dentro de la base MiNT siguiendo diferentes criterios de categorización y analizar posteriormente los grupos generados en cada categoría.

En primer lugar nos planteamos dos categorías que definen los genes nucleares que codifican productos mitocondriales. Estas son: por un lado, la clasificación como tales en todas las aproximaciones planteadas (**MiNT-core**) y por otro la **presencia de señal peptídica de localización mitocondrial (MiNT-MTS)**. En segundo lugar, nos planteamos la categorización en función de la expresión a lo largo del **ciclo de vida (MiNT-DiExTra-LC, *differentially expressed transcripts throughout life cycle*)** y del **ciclo proliferativo (MiNT-DiExTra-CC, *differentially expressed transcripts throughout cellular cycle*)**. En tercer lugar, nos planteamos una categorización en función de su **localización intramitocondrial (MiNT-Intra)**. Por último, nos propusimos analizar **subgrupos de interés particular (AAtS, aminoacil ARNt sintetetas y PRiMito, proteínas ribosomales mitocondriales)** (Figura 7).

MiNT-core

Dada la estrategia inclusiva que seguimos a la hora de generar MiNT, es inevitable la presencia de genes no específicos que enmascaren resultados. Una forma de disminuir este posible efecto es analizar de a subgrupos racionales la base de datos. En este caso seleccionamos el subgrupo de genes que fueron incluidos en MiNT con mayor confianza. Por ende se entiende por MiNT-core a los **32 genes** que fueron identificados como genes nucleares que codifican proteínas mitocondriales en forma coincidente en todos los abordajes empleados para completar la base de datos MiNT.

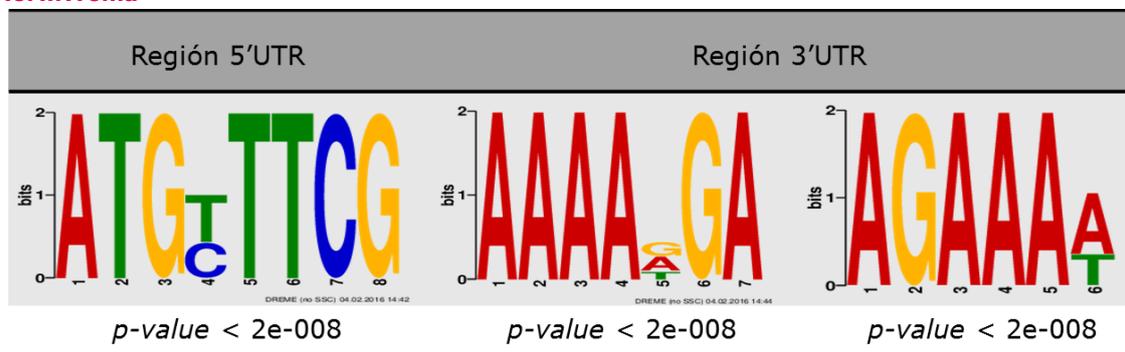
Búsqueda de motivos lineales (LiMs) en las regiones UTRs

Se realizó una búsqueda de LiMs enriquecidos en los transcritos de MiNT-core con respecto al resto de los transcritos del parásito (NoMiNToma = Transcriptoma de *T. cruzi* una vez excluidos los transcritos integrantes de la base MiNT). Se encontraron enriquecidos, con un, $p\text{-value} < 2e-8$, un LiM asociado a la región 5'UTR y dos LiMs asociados a la región 3'UTR. Estos motivos se especifican en la Tabla 3.

Para los tres casos se buscaron proteínas asociadas a los motivos empleando el programa Tom-Tom del paquete *meme-suite*. Se encontró que la **proteína de unión a poli(A) (PABP)** de *T. brucei* (Tbg972.9.5210)

reconocería un motivo significativamente similar al motivo "AGAAAW". La PABP posee un ortólogo en *T. cruzi* (TcCLB.506885.70). Estas PABP presentan 4 dominios RRM típicos hacia el extremo N-terminal que se detallan en la Figura 14, donde además se compara la proteína codificada por el gen ortólogo en *Drosophila melanogaster* por motivos que se detallan más adelante. Se sabe que los primeros dos dominios RRM son suficientes para la unión específica a poli(A) y el tercer y cuarto dominio RRM podrían mediar el reconocimiento de secuencias no poli(A). La región C terminal posee un motivo denominado PABC (*Poly(A)-binding protein C-terminal domain*) el cual no uniría ARN pero sería responsable de la interacción de la nucleasa PAN con la poli(A) (da Costa Lima et al. 2010). Se ha propuesto que esta proteína en *T. brucei* tendría un rol en los **procesos de estabilización** postranscripcional (Coller & Wickens 2007). Trabajos previos también sugieren un rol desestabilizador en acción conjunta con la proteína de unión a elementos ricos en AU (TcUBP1) en el estadio tripomastigota (D'Orso & Frasch 2002).

Tabla 3. LiMs enriquecidos en las regiones 5' y 3'UTR de MiNT-core con respecto a NoMiNToma



No se encontraron proteínas asociadas para los otros motivos. Sin embargo, resulta interesante señalar que el motivo encontrado en la región 5', se ubica en todos los casos al **comienzo de las CDSs** y codifica para la metionina inicial seguida de fenilalanina o leucina y finalmente arginina (M [F o L] R). La vinculación de este motivo con la **señal de localización peptídica** se discute más adelante.

Los **LiMs** encontrados para el subgrupo MiNT-core se buscaron a lo largo de **MiNT**. Para ello se empleó el programa FIMO (MEME-suite). Se encontró que el motivo "ATGYTTCG" está presente en 76 de los transcritos de MiNT de los cuales, únicamente 1 posee el motivo en la región 5'UTR propiamente dicha. En este caso el transcripto (TcCLB.506753.150) si bien posee la señal en la 5'UTR, esta señal está en fase con la CDS que se encuentra anotado. Sin embargo, proponemos que la metionina que se encuentra señalada como principio del marco abierto de lectura no estaría indicada correctamente ya que 75 nucleótidos corriente arriba habría un ATG en fase con la proteína que permitiría que la señal MLR se exprese a 10 aá del sitio de inicio de la proteína (Figura 15). La proteína presenta un dominio de la

familia de las **rodanasas**. Por lo que a pesar de encontrarse anotada como una proteína hipotética, sugerimos también su anotación como **tiosulfato X-transferasa** ya que si bien no se tiene evidencia en qué etapa puntual actuaría, por poseer este dominio se sabe que formaría parte de la vía de detoxificación del cianuro, mediante su conversión en tiocianato en la mitocondria (<http://pfam.xfam.org/family/rhodanese>).



Figura 14. Estudio de los dominios de unión a ARN presentes en las proteínas que reconocerían los LiMs encontrados. Se muestra el alineamiento de los dominios en Pfam (A) y secuencias de las proteínas de unión a ARN (B) que reconocerían los motivos GAAAR, AAARAW y ARAAA en *T. brucei* y/o en *D. melanogaster* y la proteína codificada por su gen ortólogo de *T. cruzi*. Se señala con recuadros verdes cada motivo RRM descrito en A y con recuadro rojo el dominio PABC.

Búsqueda de motivos de estructura (StMs) en las regiones no traducidas de mensajeros.

La búsqueda de StMs en este subgrupo fue llevada a cabo en colaboración con el Dr. Javier De Gaudenzi (UNSAM-CONICET). Los parámetros empleados fueron los descritos en la sección Materiales y métodos, y luego se utilizaron los mismos en las subsiguientes búsquedas de StMs.

En este caso se encontró una señal que denominamos **MiSTres_{core}** (motivo estructural derivado de la región 3'UTR de la base de datos MiNT-core) enriquecida significativamente a nivel de la 3'UTR de MiNT-core. Esta señal presentó un 72% de cobertura por estar presente en 23 de los 32 transcritos (Figura 17).

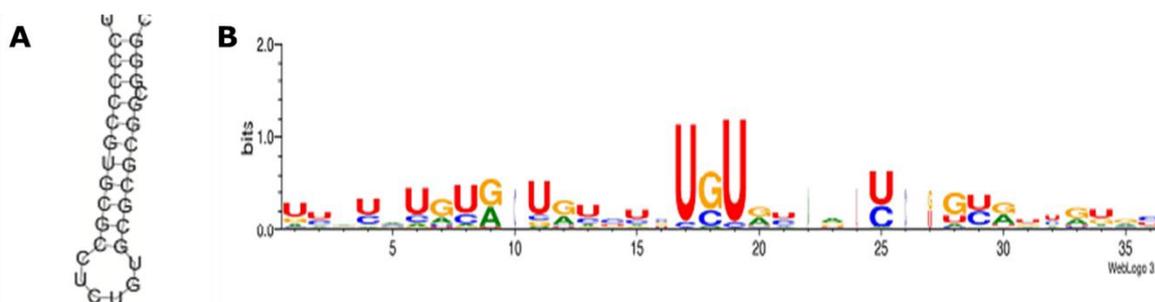


Figura 17. STM encontrado en la región 3'UTR de los transcritos componentes de MiNT-core. Se muestra la estructura esperada(A) y la secuencia consenso (B) para la misma.

Dado que el análisis que permitió la identificación de la señal se realizó en base a UTRs predichas por métodos *in silico*, se investigó la presencia de esta señal en las regiones definidas experimentalmente en nuestro laboratorio por estudios de transcriptómica para el estadio epimastigota (Chávez 2016), y por el grupo del Dr. Najib El Sayed, para los estadios amastigota y tripomastigota (Li et al. 2016). Se encontró que el 87% (20) de los transcritos contienen la señal MiSTres_{core} en su secuencia 3'UTR. Por otro lado, se constató que en muchos casos los transcritos que teóricamente no contarían con la señal en la región empleada para el análisis, sus UTRs son más largas (>350 nt) que las empleadas para la búsqueda. Esto implica que el número de transcritos que poseen la señal podría ser mayor.

Siguiendo la estrategia estándar de búsqueda de StMs empleando la herramienta CMsearch, se alimentó el programa con MiSTres_{core} y se buscó a lo largo de todo el **transcriptoma de *T. cruzi***. En este caso se encontró que 655 transcritos de *T. cruzi* poseen el motivo MiSTres_{core} de los cuales, 109 pertenecen a MiNT. El análisis de ontología de estos transcritos (Tabla 4) muestra que los mismos estarían a grandes rasgos asociados a la replicación del kinetoplasto, a la importación de proteínas desde el citoplasma a la mitocondria y de la homeostasis de metales en la mitocondria. Se observa además que, los 546 transcritos que poseen la

señal y que no forman parte de MiNT se encuentran enriquecidos en la **vía del metabolismo de las riboflavinas y la vía fotosintética**. En el caso de esta última obviamente se debe a una agrupación errónea de proteínas relacionadas con alguna vía de transporte de electrones y/u obtención de energía, por lo que posiblemente estos transcritos codifiquen para proteínas mitocondriales. Sin embargo para este trabajo no fueron incorporados a MiNT.

Tabla 4. Análisis de ontología asociados a los transcritos que poseen el motivo MiSTres_core en su 3'UTR. Panel izquierdo: Términos de ontología asociados a los 109 transcritos de MiNT que poseen el motivo. Panel superior derecho: Términos de ontología enriquecidos para los 546 transcritos del NoMiNToma de *T. cruzi* que poseen el motivo. Panel inferior derecho: vías metabólicas (KEGG PATHWAY) que se encuentran enriquecidas en estos 546 transcritos.

ID	Término	Fold enrichment
GO:0055072	<i>iron ion homeostasis</i>	58,43
GO:0006264	<i>mitochondrial DNA replication</i>	29,22
GO:0016226	<i>iron-sulfur cluster assembly</i>	25,04
GO:0055065	<i>metal ion homeostasis</i>	23,37
GO:0000002	<i>mitochondrial genome maintenance</i>	19,48
GO:0006873	<i>cellular ion homeostasis</i>	16,7
GO:0006839	<i>mitochondrial transport</i>	10,62
GO:0006626	<i>protein targeting to mitochondrion</i>	10,31
GO:0007005	<i>mitochondrion organization</i>	9,74
GO:0006260	<i>DNA replication</i>	7,72
GO:0017038	<i>protein import</i>	7,62
GO:0009451	<i>RNA modification</i>	4,61
GO:0006259	<i>DNA metabolic process</i>	3,63
GO:0090304	<i>nucleic acid metabolic process</i>	2,19

ID	Término	Fold enrichment
GO:0005515	<i>protein binding</i>	1,45

ID	Término	Fold enrichment
ec00740	<i>Riboflavin metabolism</i>	36,97
ec00195	<i>Photosynthesis</i>	16,31

Se estudió el conjunto de genes con mayor grado de confianza dentro de MiNT que cuenta con 32 genes (MiNT-core). El análisis de este subgrupo permitió la definición de 3 señales específicas. En la región 5'UTR se observa un LiM que codifica el péptido M [F o L] R y que claramente corresponde al inicio de la señal peptídica de localización mitocondrial y en la región 3'UTR aparece el motivo "AGAAAW". A nivel de estructura secundaria se encuentra una señal en la 3'UTR que denominamos MiSTres_core. Esta señal está enriquecida entre los transcritos de MiNT. Por otro lado, se destaca que la búsqueda de las señales mencionadas ha permitido detectar y corregir errores en la anotación tanto a nivel de función como la definición de la metionina inicial.

Otra de las categorías que se consideró válida para proceder a la agrupación de genes en MiNT se basó en la discriminación entre aquellos transcritos de MiNT que codifican o no **péptido señal de localización mitocondrial**. Para ello se emplearon distintas herramientas disponibles en la red o locales (ver Materiales y métodos).

La definición del subgrupo de MiNT que posee péptido señal se implementó en primer lugar en la sub-base MiNT₁₀₅, que se encuentra constituida por los genes que se encuentran anotados como mitocondriales, con el fin de probar la estrategia para luego escalarla a MiNT. Para esto, siguiendo la estrategia descrita en la Figura 18, se alimentó el programa con un conjunto de secuencias correspondientes a péptidos de localización mitocondrial de tripanosomátidos previamente reportados (Häusler et al. 1997) y otros obtenidos mediante comunicación personal con la Dra. Noreen Williams (Department of Microbiology and Immunology, University at Buffalo). De esta forma se generó un motivo cuya secuencia consenso se muestra en la Figura 18 (A). Con el fin de mejorarlo, sesgándolo específicamente hacia el motivo presente en las proteínas de *T. cruzi*, se analizaron las proteínas codificadas por los genes de **MiNT₁₀₅**. De las 105 proteínas anotadas como mitocondriales, 17 portan la señal (p -value < 0,0001) de localización mitocondrial a nivel peptídico. Empleando las secuencias de estas 17 proteínas en conjunto con las proteínas que se emplearon para generar la expresión regular inicial se generó el motivo **TryM-TaPe** (*Trypanosomal Mitochondrial Targeting Peptide*) (Figura 18 (B)). Una vez definido TryM-TaPe, se buscó el motivo nuevamente en MiNT₁₀₅ donde se observó que 38 (p -value < 0,0001) de las 105 proteínas (o sea el 36%) poseerían a TryM-TaPe, mientras que las 67 restantes carecerían de la misma.

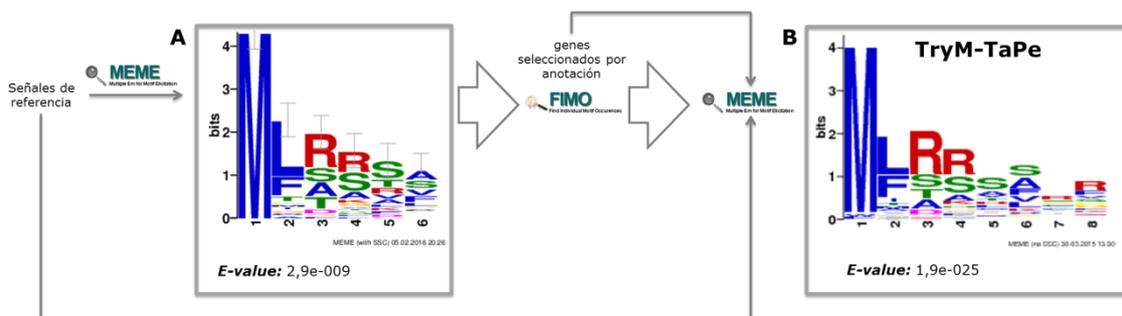


Figura 18. Estrategia empleada para la definición de TryM-TaPe. A partir de las secuencias disponibles en (Häusler et al. 1997) y otras comunicadas personalmente, se generó la primer señal (A). Con este motivo y empleando la herramienta FIMO (MEME-suite) se buscó en la base MiNT₁₀₅ aquellas proteínas codificadas que presentan esta señal en su secuencia. El conjunto de secuencias que resultan positivas de esta búsqueda se suman a las señales usadas como referencia para luego recalculer el motivo señal (B) al que se denomina TryM-TaPe.

Con este abordaje, se logró mejorar el motivo para poder definir con mayor confianza aquellas proteínas cuya localización subcelular estuviera mediada por la presencia de un péptido específico localizado en **el extremo N-terminal**.

Alternativamente, empleando la herramienta **TargetP** se predijo que 58 de las proteínas codificadas en MiNT₁₀₅ (55%) poseerían péptido señal de localización mitocondrial. De estas, 31 (53%) coinciden con las halladas mediante la búsqueda anterior.

Por otro lado, utilizando el programa **PredSL**, se encontró que el subconjunto de genes de MiNT₁₀₅ que codificarían señal peptídica de localización mitocondrial, sería de 43 proteínas, lo que equivaldría a 41% del total. De estos, 23 (53%) fueron también reconocidos en la primer búsqueda y 41 (95%) en la segunda búsqueda.

Finalmente, para definir el subgrupo de genes de MiNT₁₀₅ que codifican señal peptídica de localización (**MiNT-MTS**) se propuso como criterio de inclusión la detección en por lo menos 2 de los 3 métodos empleados para predecir la presencia de señal peptídica (Figura 19). De esta forma, de los 105 transcritos nucleares que codifican para proteínas mitocondriales, 51, esto es casi el 50%, codificarían señal peptídica de localización mitocondrial. En el 63% de ellos, la señal TryM-TaPe se encuentra con localización N-terminal.

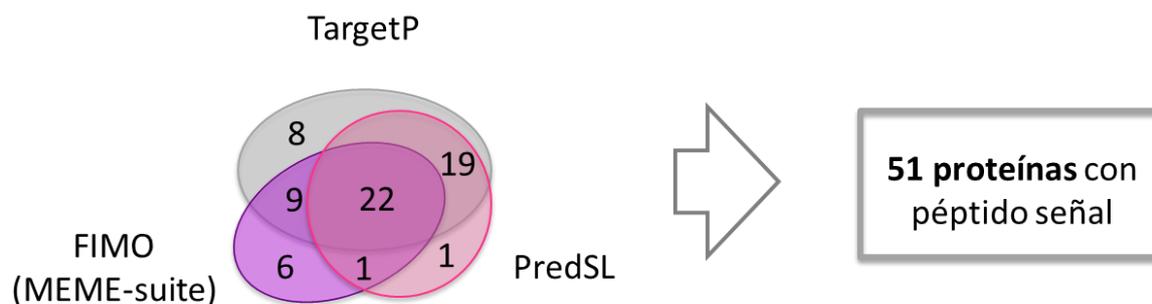


Figura 19. Definición del subgrupo de proteínas con péptido señal codificadas en MiNT₁₀₅. Se muestra esquemáticamente los resultados obtenidos empleando las distintas estrategias.

Posteriormente, siguiendo los mismos criterios de búsqueda que para MiNT₁₀₅, la búsqueda se hizo extensiva a la totalidad de **MiNT** (Figura 20). En primer lugar se empleó el programa FIMO, alimentado con el motivo **TryM-TaPe**. De esta forma se encontró que 731 (61%) de las proteínas codificadas por MiNT poseerían el motivo buscado ($p\text{-value} < 0,001$). En cambio, empleando el programa **PredSL**, se predijo que 483 de las proteínas codificadas por MiNT (40%) tendrían péptido de localización mitocondrial, sin embargo solo coincidieron con el primer subgrupo 362 proteínas. El tercer método elegido para el análisis, **TargetP**, determinó

que 614 proteínas (51%) poseerían péptido de localización mitocondrial. Siguiendo los mismos criterios anteriormente establecidos, se aceptaron como positivos aquellos genes cuyos productos proteicos clasificaran en 2 de los 3 métodos como tales. De esta forma se llegó a un total de 580 proteínas (49%) con señal peptídica de localización mitocondrial, de las cuales el 83% (485) contienen el motivo consenso TryM-TaPe.

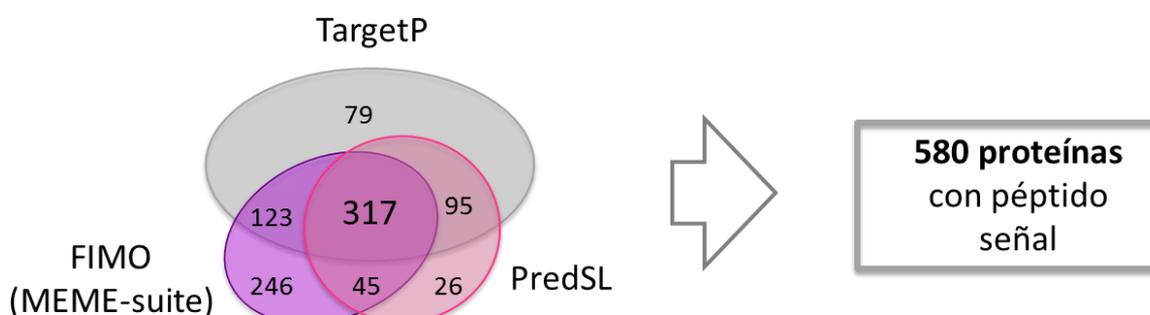


Figura 20. Definición del subgrupo de proteínas con péptido señal codificadas en MiNT. Se muestra esquemáticamente los resultados obtenidos empleando las distintas estrategias.

Posiblemente debido a que cada uno de estos programas emplea distintas estrategias y condiciones, es que las listas de genes que codifican para proteínas con péptido señal derivadas de los distintos análisis, es sólo parcialmente convergente. De hecho, los programas TargetP y PredSL limitan el número de aminoácidos a usar para alimentar el análisis predictivo. Teniendo en cuenta esta limitación, para estas búsquedas se empleó únicamente la porción amino-terminal de las proteínas. Sin embargo, en el caso de TryM-TaPe encontramos que el péptido señal presenta una ubicación heterogénea encontrándose incluso más allá de los primeros 150 aminoácidos incluidos en las búsquedas empleando TargetP y PredSL. El análisis de la distribución muestra que en el 57% de las proteínas que poseen a TryM-TaPe en su secuencia, el motivo se ubica en los primeros 100 aminoácidos (Figura 21). Pero sólo un tercio de las proteínas comienzan con el motivo. La ubicación corriente abajo podía ser atribuible a errores en la anotación de los genes. Este tipo de problema ha resultado bastante frecuente y la anotación del genoma de *T. cruzi* se encuentra en continua revisión. Independientemente de que esta parece ser la situación evidente en varios casos no se puede descartar la posibilidad de **sitios de inicio alternativos** o de **localización interna de la señal peptídica**.

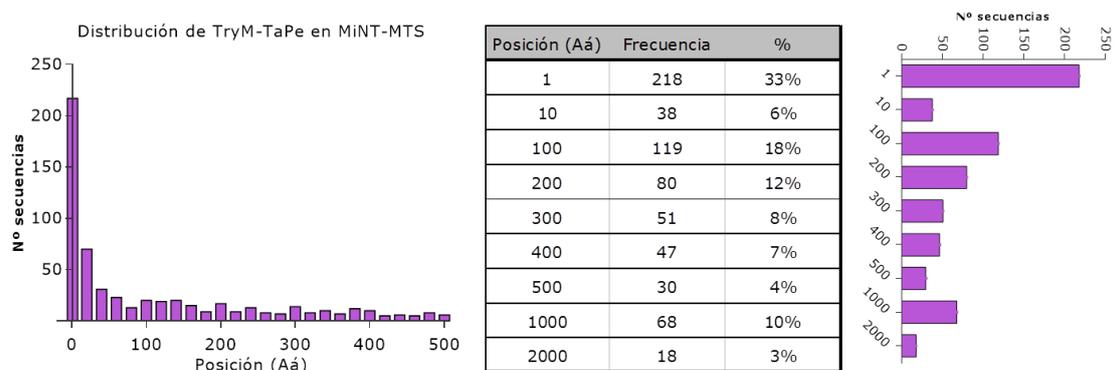


Figura 21. Distribución de TryM-TaPe a lo largo de las proteínas que lo contienen en su secuencia. En A) se muestra el histograma de la distribución de cerca del 90% de los motivos TryM-TaPe a lo largo de la secuencia peptídica de las proteínas de MiNT. En la tabla, B) se especifican los valores cuando se tienen en cuenta ciertos intervalos y los porcentajes del total a los que corresponden estos valores que se representan gráficamente en C).

En suma, mediante este abordaje se definió la base **MiNT-MTS** que contiene a los 580 genes que codifican para proteínas que según nuestro análisis contienen péptido señal de localización mitocondrial. Por otro lado, se denominó **MiNT-noMTS** al conjunto de 616 genes que no poseen péptido señal y se procedió a realizar un análisis comparativo global de los transcritos que codifican y los que no codifican péptido señal a nivel composicional y estructural.

Por un lado, el largo de las CDSs de MiNT-MTS es significativamente mayor cuando se lo compara con los presentes en MiNT-noMTS (Figura 22 A). Otra diferencia significativa se observa en el contenido GC de las CDSs de estos transcritos, siendo significativamente mayor en el caso de MiNT-MTS (Figura 22 B). En ambos casos es posible que el aporte significativo este dado por la presencia de la señal peptídica de localización mitocondrial ya sea TryM-TaPe u otra. En concordancia con el origen común de ambas poblaciones de transcritos el contenido GC en la tercera posición del codón y el índice de adaptación de codones no muestran diferencias significativas entre estos dos grupos (Figura 22 C y D).

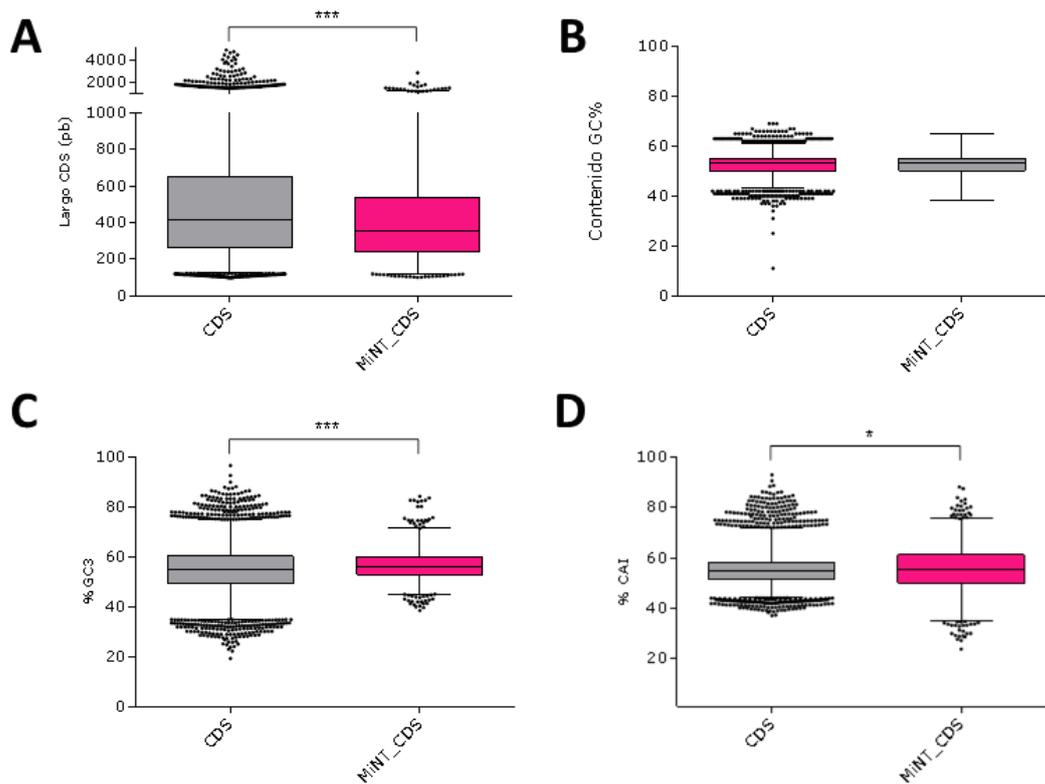


Figura 22. Análisis comparativo de las CDSs de MiNT-MTS vs MiNT-noMTS. Se analizaron distintas características de las CDSs dentro de cada grupo. A) En el caso del largo de las CDSs de cada subgrupo se observa una diferencia significativa (**** p-value < 0.0001) así como en B) el contenido GC (*** p-value = 0.0007), en cambio para C) el contenido GC en la tercera posición de los codones (GC3) así como con respecto al D) índice de adaptación de codones no se evidenciaron diferencias significativas.

En conclusión, entre las características estudiadas, los transcritos de MiNT-MTS y los de MiNT-noMTS se diferencian en el **largo** y en la proporción del **contenido GC** en las CDSs. No llama la atención que a pesar de que a nivel de las regiones UTR de los transcritos no se observaron diferencias significativas en cuanto al largo, no sucede lo mismo para la región codificante de los transcritos que codifican para proteínas con MTS ya que esta señal necesariamente adiciona bases a esta región. En cambio, la variación en el contenido de GC no era esperado y el significado no resulta claro. El contenido en GC de las CDSs ha sido en general asociado a niveles de expresión. Sin embargo, el estudio de la **expresión global** para los genes de MiNT-MTS y MiNT-noMTS no mostró diferencias significativas ni aun cuando se compararon la expresión dentro de cada estadio específico (Figura 23).

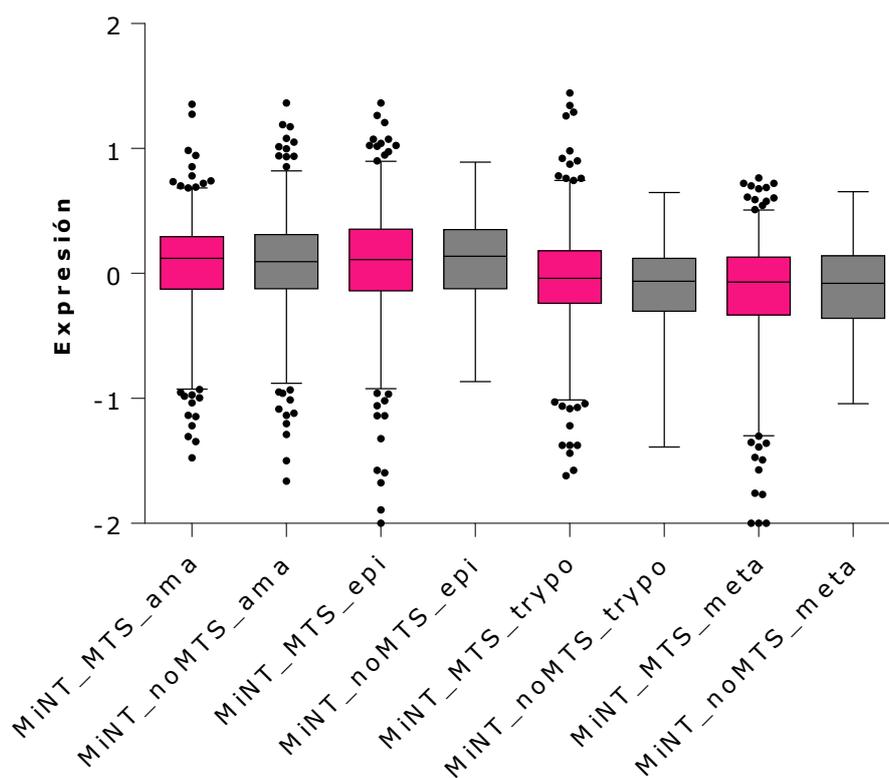


Figura 23. Niveles de expresión estadio específico para los transcritos de MiINT (MiINT-MTS vs MiINT-noMTS). Se muestra *boxplot* comparativo de los niveles de expresión para los transcritos de los distintos subgrupos empleando los datos de (Minning et al. 2009).

Búsqueda de LiMs en las regiones UTRs

Siguiendo la estrategia descrita anteriormente se buscaron los LiMs que se encuentren enriquecidos en cada subgrupo (MiINT-MTS / MiINT-noMTS). En primer lugar, se buscaron aquellos LiMs que se encuentren enriquecidos en la región 5'UTR de los transcritos de **MiINT-MTS** con respecto a los transcritos de MiINT-noMTS (Figura 24). Se encontraron **4 motivos** enriquecidos con un *p-value* < 5 e-3: "ATGYTBCG", "CAGCGS", "CGHCGC", y "GDTGCGT". El motivo "**ATGYTBCG**" es similar al que se describió en el análisis de MiINT-core. En este caso el 76% de los transcritos que lo poseen (114 / 150) lo tienen ubicado en la posición +1 de la CDS. Solamente el 7% (11/150) de los motivos se encuentran fuera de fase y únicamente uno se encuentra en la 5'UTR propiamente dicho sin embargo el péptido codificado por este motivo, que nuevamente sería M [L o F] R, se encuentra también representado en el inicio de la CDS de este transcripto.

Únicamente en el caso del motivo "CAGCGS" se encontró una proteína de humano que podría reconocer el motivo especificado (Figura 24). Sin embargo, no se encontraron genes ortólogos para esta proteína en *T. cruzi*. Si bien el motivo se encuentra en la mayoría de los casos dentro de la CDS del transcripto (93%), cabe señalar que en el 54% que corresponde a 158 transcritos, el motivo se encuentra en fase, mientras que en el restante

46% (182 transcritos) el motivo está fuera de fase. Esto indicaría que la señal es netamente nucleotídica y no requiere de su traducción.

Para el motivo "CGHCGC" en el 96% (437) de los casos se encuentra ubicado en la región de la CDS, siendo que la mayoría de los casos, el 63%, se encuentra fuera de fase. Del mismo modo, para "GDTGCGT" el 67% (37) de los casos lo encontramos en la CDS de los cuales el 70% (26) se encuentra fuera de fase. La falta de restricción para una localización en fase en estos casos, apoya la funcionalidad de la **señal a nivel del ARN**. La búsqueda del resto de los motivos dentro de la base de datos que los generó no fue posible debido que el programa de búsqueda no arrojó resultados significativos para estos motivos. En el caso de las regiones 3'UTR de los transcritos que codifican para proteínas que presentan péptido señal, no se encontraron motivos significativamente enriquecidos.

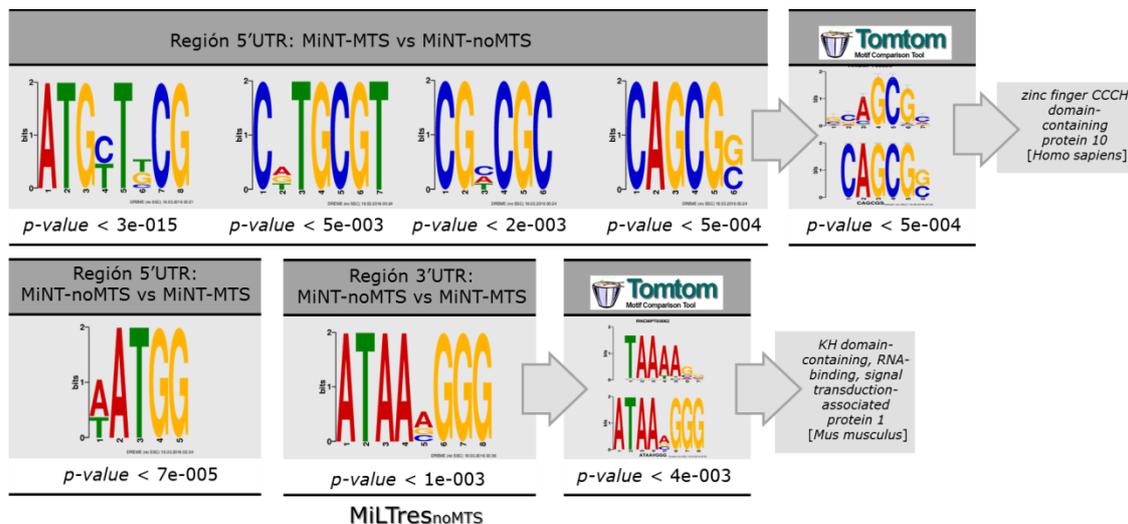


Figura 24. LiMs encontrados al analizar las regiones no traducidas de los transcritos pertenecientes a MiNT-MTS y MiNT-noMTS. En el panel superior se muestran los motivos encontrados enriquecidos en la base MiNT-MTS con respecto a MiNT-noMTS. Se muestra también el resultado obtenido mediante el programa Tom-Tom. En el panel inferior por el contrario se muestran los motivos que se encontraron enriquecidos en el análisis recíproco resaltando un motivo a nivel de la región 3'UTR de los transcritos de MiNT-noMTS al que denominamos MiLTres_{noMTS}.

El análisis recíproco, para el cual se definieron señales enriquecidas a nivel de los mensajeros de **MiNT-noMTS**, permitió definir una señal en la 3'UTR de los mismos que denominamos **MiLTres_{noMTS}** (motivo lineal en la región 3'UTR de transcritos asociados a **MiNT-noMTS**). Aquellos mensajeros que no tienen una señal a nivel peptídica de localización mitocondrial, poseen este motivo en su región 3'UTR del mensajero que, a falta de señal peptídica, podría jugar un rol crucial en la **distribución subcelular**. Además, para este motivo se encontró que podría ser reconocido por una proteína de *Mus musculus*, sin embargo el gen que codifica esta proteína no posee ortólogo en *T. cruzi*.

Búsqueda de StMs en las regiones UTRs

Además de los motivos de secuencia se buscaron motivos de estructura que pudieran contribuir con la localización mitocondrial de los productos génicos, en forma independiente (MiNT-noMTS) o complementando a la señal peptídica (MiNT-MTS). Por tal motivo se buscaron StMs a nivel de 5' y 3'UTR para los transcritos presentes en MiNT-MTS y en MiNT-noMTS.

Se encontraron 4 motivos que superaron el umbral de **especificidad/sensibilidad** propuesto. En todos los casos estos se corresponden con motivos presentes en la región 5'UTR (Tabla 5).

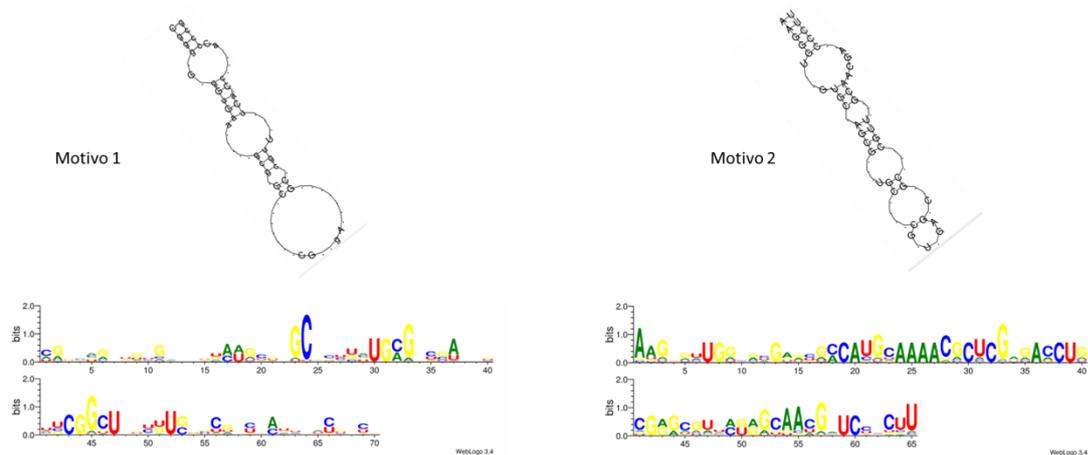
Tabla 5. Búsqueda de motivos estructurales en los subgrupos MiNT-MTS y MiNT-noMTS.

Motivo	Genes aporte	Largo (nt)	DE Largo (nt)	Cobertura	Sensibilidad CM search	Especificidad	ΔG^* kcal/mol
m1_3UTR_MiNT_MTS *	517	30	6	89%	72%	0%	0,00
m2_3UTR_MiNT_MTS *	510	37	8	88%	11%	0%	0,00
m3_3UTR_MiNT_MTS *	510	28	3	88%	74%	7%	0,00
m1_5UTR_MiNT_MTS	531	35	4	92%	19%	98%	-6,90
m2_5UTR_MiNT_MTS	68	30	8	12%	56%	98%	-8,80
m3_5UTR_MiNT_MTS *	421	59	10	73%	39%	37%	0,00
m1_3UTR_MiNT_noMTS *	338	27	9	55%	50%	71%	0,00
m2_3UTR_MiNT_noMTS *	277	41	17	45%	62%	72%	0,00
m3_3UTR_MiNT_noMTS *	277	34	5	45%	63%	71%	-7,40
m1_5UTR_MiNT_noMTS *	377	29	8	61%	26%	70%	0,00
m2_5UTR_MiNT_noMTS	538	37	5	87%	15%	98%	-18,70
m3_5UTR_MiNT_noMTS	538	41	5	87%	24%	97%	-2,70

Se señala con * los motivos que no superaron el criterio de especificidad/sensibilidad propuesto para este tipo de búsquedas. Los círculos verdes señalan aquellos agrupamientos cuya especificidad supera el 90% mientras que el rojo a los menores del 80% de especificidad. Se señala además en verde aquellas estructuras cuyo ΔG de formación a 28°C es menor al promedio de ΔG encontrados.

En la Figura 25 se muestran las estructuras para los motivos predichos. Los elementos estructurales tienen aproximadamente un largo de entre 30 y 40nt (Tabla 5). Los 4 motivos muestran una estructura del tipo tallo y horquilla cuya secuencia consenso muestra en algunos casos varios nucleótidos conservados.

Región 5' de MiNT-MTS



Región 5' de MiNT-noMTS

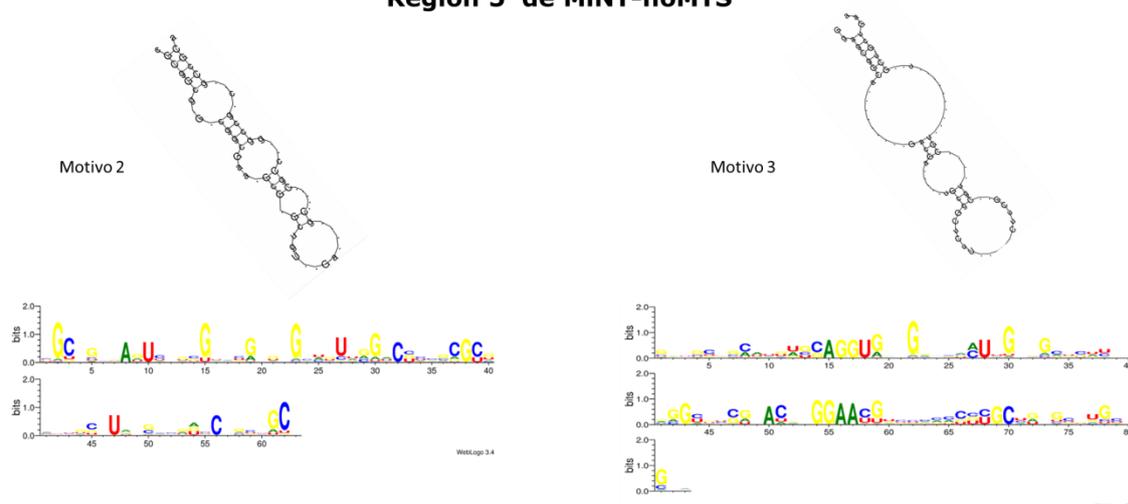


Figura 25. Estructura conservada en los motivos específicos encontrados para la región 5'UTR de los transcritos de MiNT-MTS y MiNT-NoMTS y sus respectivas secuencias consenso.

Al analizar la posición de inicio de cada uno de estos motivos encontramos que en el caso de MiNT-MTS se ubican por completo dentro de la CDS (Figura 26, m1_5UTR_MTS) o presentan su inicio preferencialmente en el comienzo de la CDS (Figura 26, m2_5UTR_MTS). En el caso de los motivos que se encontraron en MiNT-noMTS si bien se ubican preferencialmente dentro de la CDS, también encontramos varios casos con el motivo ubicado en la 5'UTR propiamente dicho.

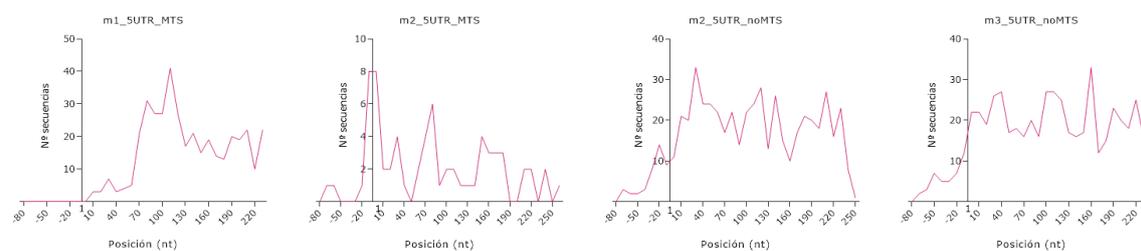


Figura 26. Frecuencia de sitios de inicio de las estructuras encontradas para la región 5'. La posición 1 corresponde a la posición 1 del AUG anotado para cada transcripto.

Por otro lado, se analizó la posición de cada motivo transcripto a transcripto para evaluar su presencia en función del largo real de las UTRs (Tabla 6). Encontramos que, para cada motivo definido, apenas 1% de los transcriptos no poseerían el motivo, ya que su UTR no abarcaría la región donde se encontró el motivo.

Tabla 6. Síntesis de total de UTRs que no presentarían el motivo o para las cuales no se cuenta con el largo real de las mismas.* ND: largo del UTR no disponible para ningún estadio

Motivo	Total	UTRs ND*		Motivo no presente	
		UTRs ND*	%	UTRs ND*	%
m1_5UTR_MiNT_MTS	542	43	8%	2	0%
m2_5UTR_MiNT_MTS	68	5	7%	1	1%
m2_5UTR_MiNT_noMTS	538	133	25%	3	1%
m3_5UTR_MiNT_noMTS	538	107	20%	0	0%

*Se establece una secuencia aminoacídica consenso que define el péptido señal en *T. cruzi* y que denominamos TryM-TaPe. El motivo se empleó para definir dos sub-bases: MiNT-MTS con 580 transcriptos y MiNT-NoMTS con los restantes 616 transcriptos de MiNT. Estas bases presentan tanto características composicionales como señales distintivas. Se encontró un motivo lineal enriquecido en aquellos transcriptos que no poseen péptido señal de localización mitocondrial, constituyendo una señal prometedora para cumplir una función compensatoria de la señal peptídica característica de estos sistemas de localización subcelular de proteínas. Este motivo fue denominado MiLTres_{NoMTS}.*

MiNT-DiExTra-LC

Empleando los datos de expresión diferencial se llevó a cabo una nueva subdivisión de MiNT. Se utilizó el programa MeV para analizar los datos disponibles a partir de **experimentos de microarreglos** (Minning et al. 2009).

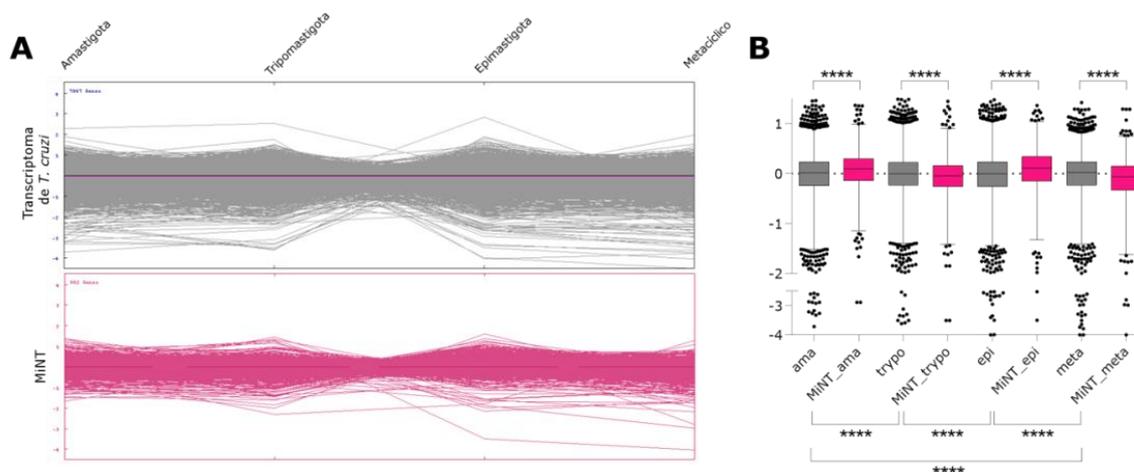


Figura 27. Panorama general de la expresión a lo largo del ciclo de vida del parásito. A) En el panel superior se muestra el perfil de expresión de los genes del transcriptoma de *T. cruzi* completo, mientras que en el panel inferior únicamente se analizan aquellos genes que pertenecen a MiNT. B) *Boxplot* comparativo del nivel de expresión por estadios para *T. cruzi* (gris) y para MiNT (fucsia) en particular.

El análisis global del **transcriptoma** permite observar que presenta una distribución homogénea a lo largo de los estadios (Figura 27). Sin embargo, no sucede lo mismo con respecto a MiNT. En este caso, la expresión en los **estadios replicativos** es significativamente mayor con respecto al transcriptoma, en general, y con respecto a la expresión de los genes que componen MiNT en los estadios no replicativos. Este comportamiento se explica en base a la progresión del estadio tripomastigota sanguíneo a la forma epimastigota, al igual que el pasaje a la forma amastigota. Es en estas etapas de diferenciación donde no sólo se observan cambios a nivel de morfología sino que además, el cambio de ambiente motiva la adaptación y aumento de complejidad de la mitocondria y, entre otros, a la expresión de componentes clave de la cadena de transporte de electrones y chaperonas mitocondriales (Tyler et al. 2002). También el complejo III (citocromo reductasa) se activa en la transición entre tripomastigotas y amastigotas.

Por otro lado, únicamente para 234 genes de los 1196 genes pertenecientes a MiNT, no habría datos de expresión. Estos genes que no se encuentran en los datos de microarreglos se asocian con diversos procesos (Tabla 7), encontrándose particular enriquecimiento con respecto a la ontología, en términos asociados a regulación de ARNm mitocondrial, metabolismo de ácidos grasos, metabolismo de nucleótidos y transporte a través de la

membrana mitocondrial. Por lo que su ausencia en los datos de transcriptoma sería atribuible a la sensibilidad del método ya que el 35% de los genes (84) poseen un nivel de expresión, según los datos de (Minning et al. 2009), por debajo de la mediana de los valores de expresión obtenidos en nuestro laboratorio para los mismos genes en el estadio epimastigota a lo largo del ciclo celular. Por otro lado 18% (36) de los mismos tampoco se encuentran en este abordaje del transcriptoma.

Tabla 7. Términos de ontología enriquecidos en el subgrupo de transcritos no representados en los datos de expresión a lo largo del ciclo de vida.

ID	Término	Fold enrichment
GO:0019395	<i>fatty acid oxidation</i>	35,67
GO:0016554	<i>cytidine to uridine editing</i>	22,30
GO:0043488	<i>regulation of mRNA stability</i>	17,84
GO:0000959	<i>mitochondrial RNA metabolic process</i>	13,38
GO:0006839	<i>mitochondrial transport</i>	10,70
GO:0007005	<i>mitochondrion organization</i>	5,95

Se procedió a generar subgrupos de genes de acuerdo a su perfil de expresión diferencial. Se obtuvieron 6 agrupamientos con un promedio de (160 ± 31) genes en cada subgrupo (Tabla 8). Se analizaron los genes dentro de cada uno de los 6 perfiles diferenciales en busca de enriquecimiento a nivel de ontología. Mediante el análisis empleando DAVID se obtuvieron categorías enriquecidas para los agrupamientos 2 y 5. Los transcritos presentes en el agrupamiento 2 poseen una expresión intermedia en el estadio amastigota para luego presentar un máximo de expresión en el estadio tripomastigota. En las formas del parásito presentes en el vector hematófago, estos transcritos se encuentran subexpresados. Este subgrupo de mensajeros se destaca por poseer un claro enriquecimiento en la categoría ontológica relacionada a **complejos ribonucleoproteicos y la traducción** (*Enrichment score* = 1,64). Si bien se ha reportado que el estadio tripomastigota metacíclico presenta una maquinaria traduccional reducida (Smircich et al. 2015), el enriquecimiento observado se justifica por la especialización funcional.

Por otra parte, el agrupamiento 5 muestra un perfil donde en los estadios presentes en vertebrado, amastigota y tripomastigota, se encuentra subexpresado, mientras que aumenta su expresión en el estadio epimastigota y se mantiene con algún descenso para el estadio tripomastigota metacíclico. En este caso se encuentra enriquecida la categoría de **unión a iones metálicos** (*Enrichment score* = 1,70). En esta categoría se encuentran ubicadas proteínas de la cadena de transporte de electrones, para las cuales esperamos una sobreexpresión en este estadio del parásito así como las proteínas encargadas de la protección al ataque mediante especies reactivas del oxígeno y nitrógeno, como la superóxido dismutasa (SOD) (Piacenza et al. 2009).

Búsqueda de LiMs en las regiones UTRs

Como ya se describió en la estrategia general mostrada en la Figura 7, en primer lugar, llevamos a cabo la búsqueda de señales a nivel de secuencia tanto dentro de la 3'UTR como de la región 5'UTR. Empleando la herramienta proporcionada por MEME-Suite, DREME, se buscaron LiMs enriquecidos en cada uno de los agrupamientos con respecto a NoMiNToma. De esta forma se obtuvieron en el entorno de 5 motivos por conjunto de secuencias donde el *p-value* de los mismos nunca superó el $2,5 \times 10^{-7}$. Todos los motivos generados se usaron como entrada para la búsqueda de posibles proteínas de unión a ARN que tuvieran como blanco la secuencia consenso obtenida. Los resultados se resumen en la Tabla 8. Una vez analizado con Tom-Tom, programa también disponible en MEME-Suite, se encontró que de los 57 motivos totales (entre los hallados para los 6 subgrupos tanto en la región 5' como en la 3'UTR), solamente 16 mostraron coincidencias a distintos niveles con motivos reconocidos por proteínas de unión al ARN (Ray et al. 2013), y al menos 10 de los motivos serían reconocidos por alguna proteína de unión a ARN con mayor significancia. En este caso sólo 2 de ellos serían reconocidos por proteínas que presentan un gen ortólogo en *T. cruzi*. El motivo "GAAGGAKT" que se encuentra enriquecido en la región 3'UTR del agrupamiento 2 es reconocido por una proteína de *Physcomitrella patens* que presenta un gen ortólogo en el parásito. Esta proteína, codificada por TcCLB.510143.80, presenta dos dominios RRM1 que en este caso no muestran similitud significativa con los presentes en la proteína de *P. patens*. En este caso, es muy probable que la proteína de *T. cruzi* no este reconociendo el mismo motivo que su ortóloga en *P. patens*. Por otro lado, el motivo "AGGGK" es similar al motivo reconocido por la proteína EIE75345.1 de *Rhizopus delemar*. La búsqueda de genes ortólogos para el que codifica esta proteína sugirió que la proteína RBP40 de *T. cruzi*, que según la bibliografía se encuentra restringida a reservosomas del parásito, sería codificada por un ortólogo del gen que codifica la proteína de *R. delemar*. En este caso presentan incompatibilidades en cuanto a sus **dominios de unión al ARN**. Cabe resaltar que la proteína RBP40 presenta un único dominio de unión a ARN (RRM1) mientras que la proteína de *R. delemar* posee 2, y a pesar de que el dominio presente en la primera se asemeja a uno de los dos dominios encontrados en la proteína del hongo, no se cuenta con información al respecto de la interacción domino-motivo, por lo que es posible aseverar que la RBP40 reconocería el motivo "AGGGK".

Tabla 8. Motivos de secuencia generados para cada uno de los agrupamientos. Se muestran aquellos motivos para los cuales se encontró una proteína heteróloga que podría reconocerlo y en aquellos casos de que no, se muestra el motivo con menor *p*-value de los generados LiMs enriquecidos dentro de cada agrupamiento con respecto a NoMiNToma.

	A T E M		Motivo enriquecidos	Proteína de unión	Motivo reconocido	Proteína ortóloga en <i>T. cruzi</i>
Agrupamiento 1 (172 genes)		5'UTR	p-value 1.0e-7 	XP_662150.1 (E-value < 0,1) <i>Aspergillus nidulans</i>		No presenta
		3'UTR	p-value 1.3e-9 	NP_649552.1 (E-value < 0,1) <i>Drosophila melanogaster</i>		No presenta
Agrupamiento 2 (137 genes)		5'UTR	p-value 2,6e-8 	No se encontraron coincidencias (E-value < 1)	No presenta	No presenta
		3'UTR	p-value 1,7e-8 	NP_995835 (E-value < 1) <i>Drosophila melanogaster</i>		No presenta
		3'UTR	p-value 2,4e-7 	XP_001764808.1 (E-value < 1) <i>Physcomitrella patens</i>		TcCLB.510143.80 <i>RNA-binding protein, putative</i>
Agrupamiento 3 (182 genes)		5'UTR	p-value 1,6e-7 	NP_495121.1 (E-value < 0,01) <i>Caenorhabditis elegans</i>		No presenta
		3'UTR	p-value 4,2e-8 	No se encontraron coincidencias (E-value < 1)	No presenta	No presenta
Agrupamiento 4 (108 genes)		5'UTR	p-value 1,9e-8 	No se encontraron coincidencias (E-value < 1)	No presenta	No presenta
		3'UTR	p-value 6,8e-10 	XP_001314386.1 (E-value < 0,1) <i>Trichomonas vaginalis</i>		No presenta
Agrupamiento 5 (207 genes)		5'UTR	p-value 1,6e-9 	NP_599124.1 (E-value < 0,1) <i>Drosophila melanogaster</i>		No presenta
		3'UTR	p-value 1,2e-8 	NM_024939.2 (E-value < 0,1) <i>Homo sapiens</i>		No presenta
Agrupamiento 6 (156 genes)		5'UTR	p-value 9,6e-8 	NM_018834.5 (E-value < 0,1) <i>Homo sapiens</i>		No presenta
		3'UTR	p-value 1,4e-9 	EIE75345.1 (E-value < 1) <i>Rhizopus delemar</i>		TcCLB.506565.12 <i>RNA-binding protein, putative</i>

Búsqueda de StMs en las regiones UTRs

Continuando con la estrategia de trabajo se buscaron señales a nivel de estructura enriquecidas en las regiones UTR de los transcritos que forman parte de MiNT-DiExTra-LC. Los tres motivos encontrados en cada región (5' o 3' UTR) para cada agrupamiento se analizaron para determinar su especificidad en cuanto a NoMiNToma (Tabla 9), así como su cobertura con respecto al agrupamiento específico (Figura 28). De los **36 motivos** encontrados, 14 superan los límites de sensibilidad/especificidad y energía propuestos (Figura 29 y Figura 31). Para cada una de las estructuras mostradas en la Figura 29 se analizó la posición que ocupan a lo largo de la región 5'UTR seleccionada para el análisis, y se muestra en la Figura 30. Se observa en la mayoría de los casos que los motivos se encuentran principalmente en la porción codificante de la región 5' empleada o entorno al inicio del mismo. En varios casos, Agrup_2_5UTR_m2, Agrup_2_5UTR_m3, la distribución es bastante homogénea. Mientras que en el caso de Agrup_4_5UTR_m3 el motivo parece encontrarse enriquecido en la porción 5'UTR propiamente dicha.

Tabla 9. Búsqueda de motivos estructurales en los agrupamientos según el perfil de expresión a lo largo del ciclo de vida del parásito.

Motivo	Genes aporte	Largo (nt)	DE Largo (nt)	Cobertura	Sensibilidad CM search	Especificidad	ΔG kcal/mol	
Agrup_1_3UTR_1	85	32	6	49%	51%	●	85%	9,80
Agrup_1_3UTR_2	122	28	1	71%	38%	●	81%	2,90
Agrup_1_3UTR_3	141	31	8	82%	9%	●	98%	0,00
Agrup_1_5UTR_1	58	35	6	34%	62%	●	99%	-6,80
Agrup_1_5UTR_2	162	32	3	94%	17%	●	90%	2,10
Agrup_1_5UTR_3	166	28	4	97%	12%	●	99%	-10,30
Agrup_2_3UTR_1	109	31	2	80%	37%	●	87%	-1,20
Agrup_2_3UTR_2	40	29	8	29%	45%	●	100%	-23,00
Agrup_2_3UTR_3	7	37	9	5%	100%	●	100%	-24,50
Agrup_2_5UTR_1*	79	29	3	58%	6%	●	98%	0,00
Agrup_2_5UTR_2	116	33	5	85%	13%	●	99%	-5,60
Agrup_2_5UTR_3	39	40	5	28%	59%	●	98%	-8,30
Agrup_3_3UTR_1*	66	31	5	36%	65%	●	81%	8,20
Agrup_3_3UTR_2	162	30	8	89%	27%	●	99%	-7,80
Agrup_3_3UTR_3*	127	39	5	70%	10%	●	80%	0,00
Agrup_3_5UTR_1	139	33	1	76%	10%	●	99%	-6,40
Agrup_3_5UTR_2*	116	36	1	64%	25%	●	84%	0,00
Agrup_3_5UTR_3	172	32	3	95%	10%	●	99%	-9,90
Agrup_4_3UTR_1*	80	31	2	74%	55%	●	79%	4,60
Agrup_4_3UTR_2*	97	39	5	90%	73%	●	39%	0,00
Agrup_4_3UTR_3*	75	38	4	69%	63%	●	78%	-8,64
Agrup_4_5UTR_1	91	34	3	84%	41%	●	100%	-14,70
Agrup_4_5UTR_2*	83	33	6	77%	10%	●	95%	2,00
Agrup_4_5UTR_3	95	41	2	88%	26%	●	99%	-12,84
Agrup_5_3UTR_1	155	30	9	75%	46%	●	75%	0,60
Agrup_5_3UTR_2	110	33	13	53%	50%	●	81%	-2,60
Agrup_5_3UTR_3*	115	30	4	56%	33%	●	84%	13,40
Agrup_5_5UTR_1	148	33	4	71%	7%	●	100%	-9,10
Agrup_5_5UTR_2	197	34	11	95%	12%	●	95%	-7,60
Agrup_5_5UTR_3	156	36	3	75%	13%	●	93%	-2,00
Agrup_6_3UTR_1*	86	30	7	55%	53%	●	80%	9,50
Agrup_6_3UTR_2*	84	29	10	54%	38%	●	84%	1,00
Agrup_6_3UTR_3*	147	33	4	94%	34%	●	72%	-4,40
Agrup_6_5UTR_1	63	33	16	40%	38%	●	100%	-6,80
Agrup_6_5UTR_2	145	40	7	93%	31%	●	100%	-2,69
Agrup_6_5UTR_3	31	39	8	20%	84%	●	99%	3,50

Se señala con * los motivos que no superaron el criterio de especificidad/sensibilidad propuesto para este tipo de búsquedas. Los círculos verdes señalan aquellos agrupamientos cuya especificidad supera el 90% mientras que el amarillo corresponde a aquellos que están entre el 80 y 90% y el rojo a los menores del 80% de especificidad. Se señala además en verde aquellas estructuras cuyo ΔG de formación a 28°C es menor al promedio de ΔG encontrados.

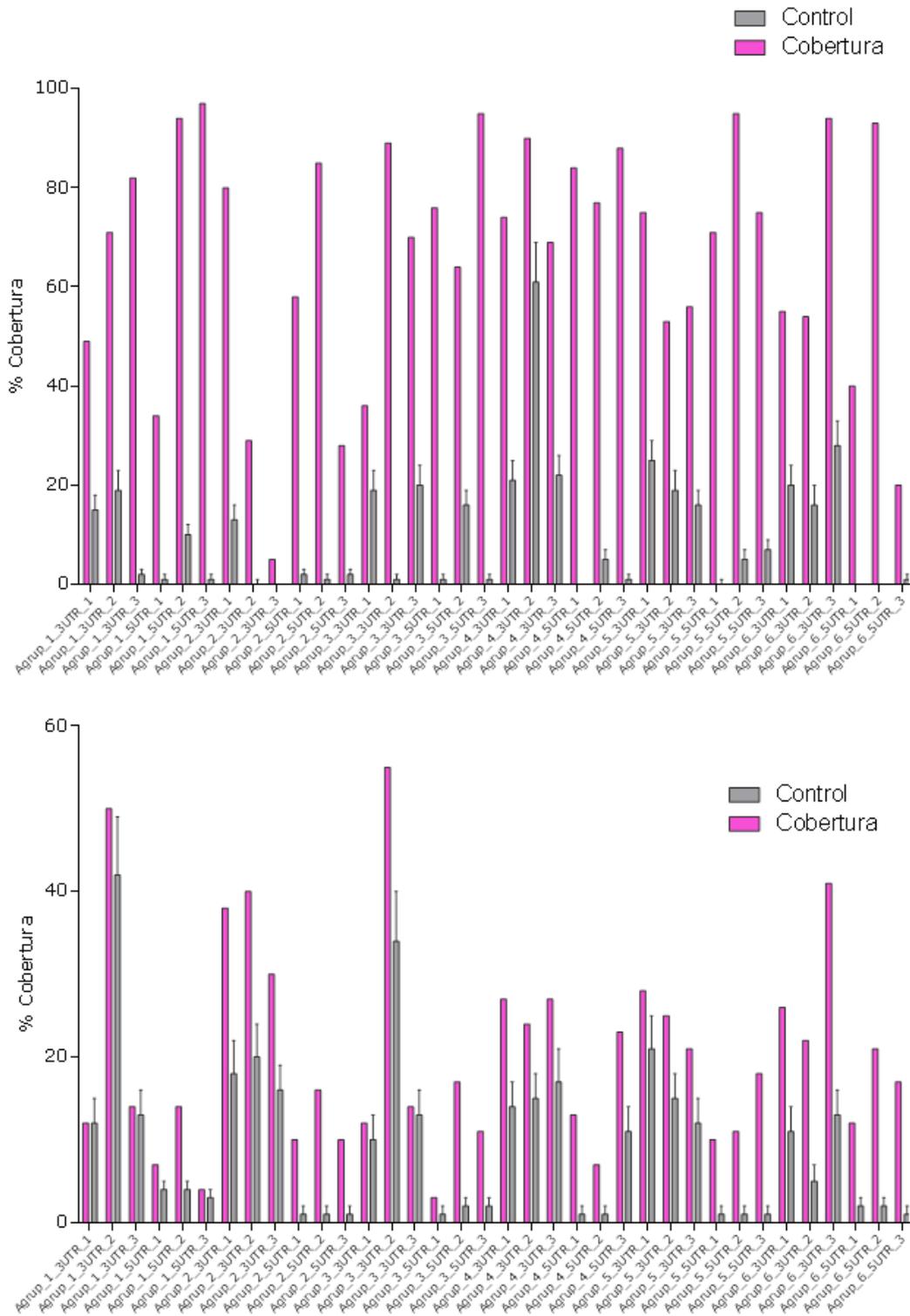


Figura 28. Cobertura y especificidad de los motivos estructurales encontrados para los distintos agrupamientos del ciclo celular. En A) se presentan los resultados obtenidos a partir del análisis de la cobertura en cuanto a las secuencias que aportan a la definición del motivo estructural en comparación de la cobertura que presentan los grupos de transcritos al azar cuando se utiliza el motivo como entrada. En B) se representan los resultados de cobertura obtenidos cuando se lleva a cabo la búsqueda del motivo contra el grupo que lo generó, en comparación con la cobertura que presentan los grupos de transcritos al azar cuando se utiliza el motivo como entrada.

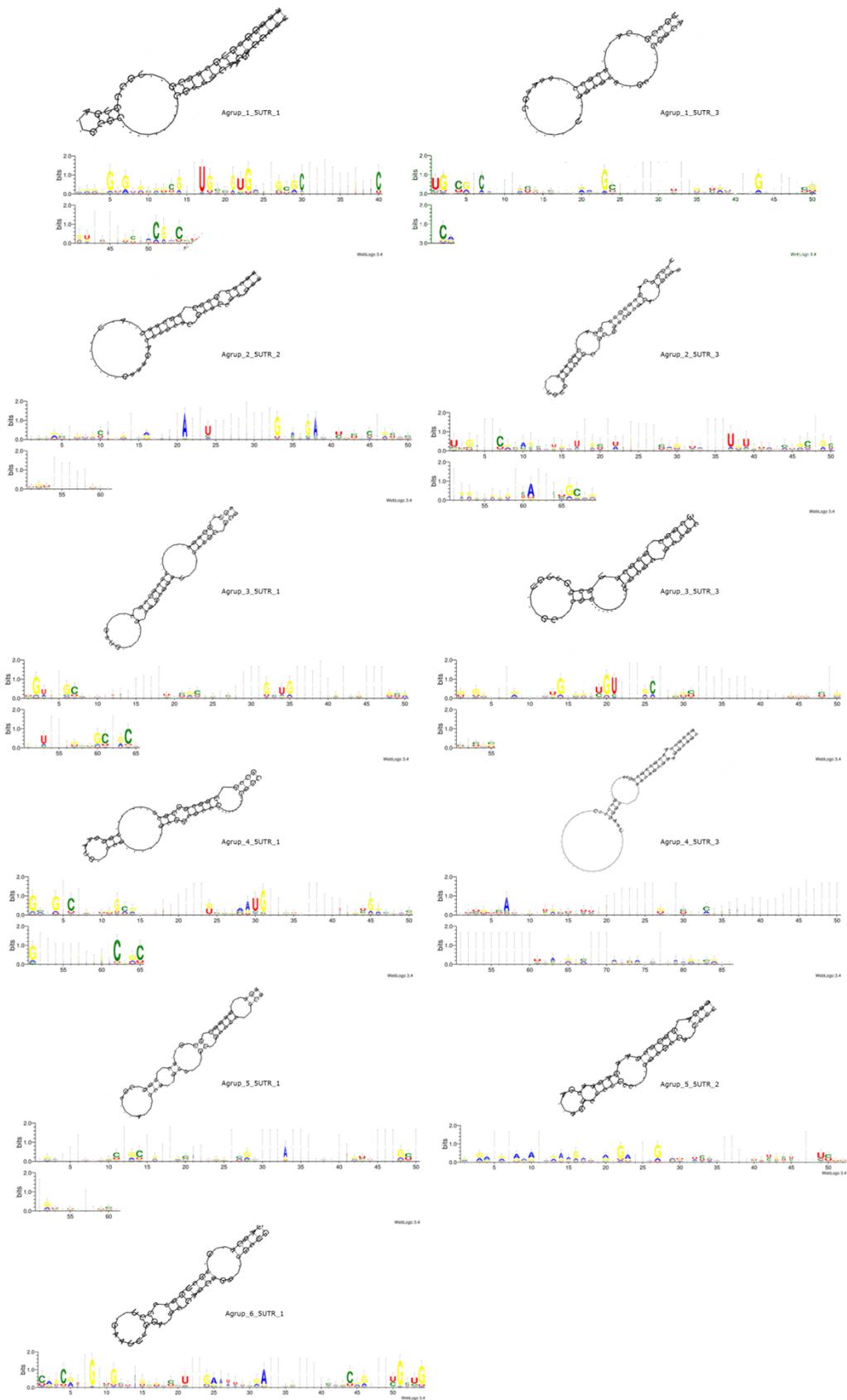


Figura 29. Estructuras que superaron los criterios propuestos para la región 5'UTR. Se observan las 11 estructuras modeladas mediante RNAfold así como el logo de la secuencia consenso para cada una. En cada caso se señala el agrupamiento del cual provienen, la región específica, en este caso 5'UTR y el número del motivo (1, 2, o 3).

En cada caso se buscaron enriquecimientos significativos a nivel de **ontología** para los conjuntos de transcritos que poseen una señal específica (Tabla 10). Aquellos términos que se encontraron repetidos de un subgrupo a otro se eliminaron y sólo tuvimos en cuenta aquellos específicos para cada subgrupo. En este caso se observó que los **procesos biosintéticos** así como la **organización del organelo** en general abarcan la mayoría de los procesos enriquecidos. En el caso del Agrup_2_5UTR_m2 sobresale la categoría de modificación del ARNm y participación en la expresión génica. Por otro lado, el subgrupo de transcritos que presentarían el Agrup_4_5UTR_m3, el cual vimos que mostraba una distribución preferencial hacia la 5'UTR propiamente dicho, cuando se lo compara con la distribución para el resto de los motivos, no presenta enriquecimiento específico con respecto a términos de ontología asociados.

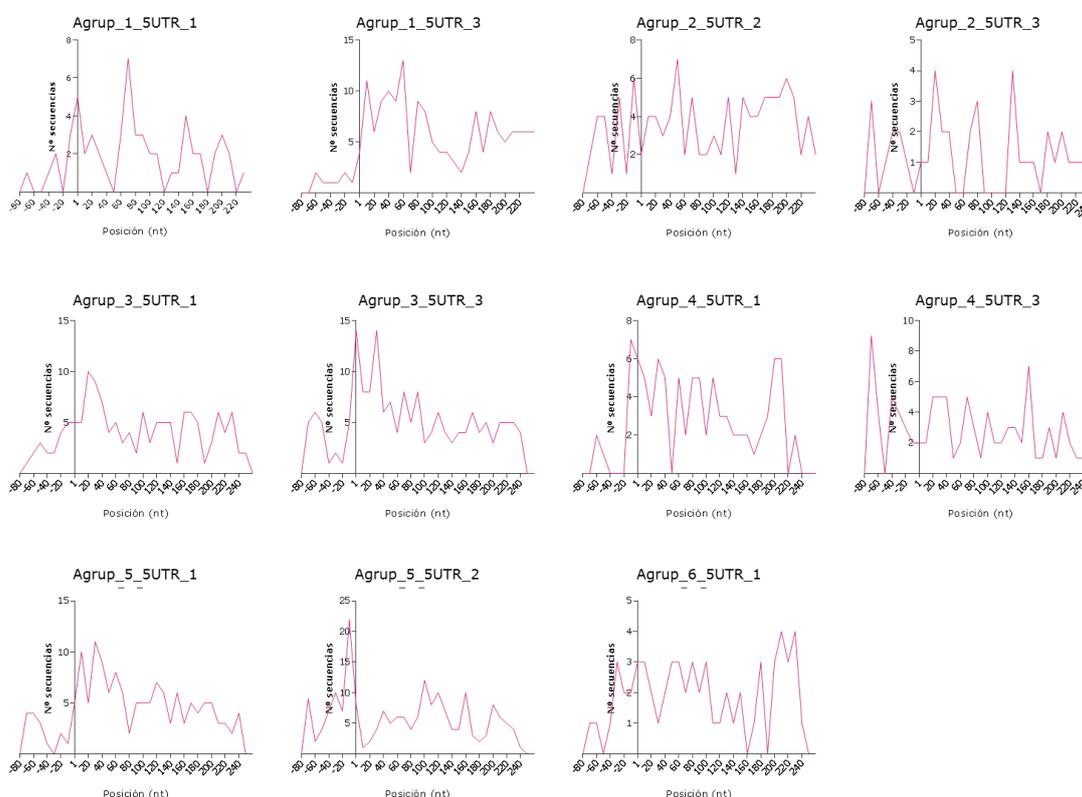


Figura 30. Distribución de las estructuras que se encontraron significativamente enriquecidas a nivel de la región 5'UTR. Se señala con 1 el inicio de la región codificante.

En el caso de la región **3'UTR** se encontraron únicamente 3 motivos que cumplieron con los criterios de especificidad/sensibilidad propuestos. Los motivos cuya estructura y secuencia consenso se muestran en la Figura 31, se encontraron en el agrupamiento 2 y el agrupamiento 3. Las secuencias

consenso para estas estructuras son ricas en G, C y U, y en general exponen secuencias **ricas en GC**, salvo por la estructura encontrada en el agrupamiento 3 para en cual no se encontró una secuencia consenso tan marcada como en los otros casos.

Tabla 10. Términos de ontología enriquecidos diferencialmente para cada subgrupo de transcritos que aporta a cada uno de los motivos.

Agrup_1_5UTR_1			Agrup_1_5UTR_3		
ID	Término	Fold enrichment	ID	Término	Fold enrichment
GO:0033865	<i>nucleoside bisphosphate metabolic process</i>	35,35	GO:0006695	<i>cholesterol biosynthetic process</i>	38,10
GO:0015936	<i>coenzyme A metabolic process</i>	35,35	GO:0006696	<i>ergosterol biosynthetic process</i>	25,40
			GO:0044108	<i>cellular alcohol biosynthetic process</i>	25,40
			GO:0042559	<i>pteridine-containing compound biosynthetic process</i>	21,77
			GO:0030150	<i>protein import into mitochondrial matrix</i>	14,29
			GO:0009108	<i>coenzyme biosynthetic process</i>	9,94
			GO:0044271	<i>cellular nitrogen compound biosynthetic process</i>	3,65
			GO:0006520	<i>cellular amino acid metabolic process</i>	3,63
			GO:0044249	<i>cellular biosynthetic process</i>	2,02
			GO:0009058	<i>biosynthetic process</i>	1,99
Agrup_2_5UTR_2			Agrup_2_5UTR_3		
ID	Término	Fold enrichment	ID	Término	Fold enrichment
GO:0009451	<i>RNA modification</i>	7,28	GO:0022603	<i>regulation of anatomical structure morphogenesis</i>	206,53
GO:0010467	<i>gene expression</i>	2,44	GO:0046168	<i>glycerol-3-phosphate catabolic process</i>	206,53
Agrup_3_5UTR_1			Agrup_3_5UTR_3		
ID	Término	Fold enrichment	ID	Término	Fold enrichment
GO:0009107	<i>lipoate biosynthetic process</i>	53,03	GO:0006284	<i>base-excision repair</i>	27,83
GO:0006633	<i>fatty acid biosynthetic process</i>	11,36	GO:0045039	<i>protein import into mitochondrial inner membrane</i>	27,83
			GO:0061024	<i>membrane organization</i>	13,91
Agrup_4_5UTR_1					
ID	Término	Fold enrichment			
GO:0031163	<i>metallo-sulfur cluster assembly</i>	36,50			
GO:0016226	<i>iron-sulfur cluster assembly</i>	36,50			
Agrup_5_5UTR_1			Agrup_5_5UTR_2		
ID	Término	Fold enrichment	ID	Término	Fold enrichment
GO:0034654	<i>nucleobase-containing compound biosynthetic process</i>	4,69	GO:0006102	<i>isocitrate metabolic process</i>	39,24
			GO:0006744	<i>ubiquinone biosynthetic process</i>	26,16
			GO:0006733	<i>oxidoreduction coenzyme metabolic process</i>	8,41
			GO:0006091	<i>generation of precursor metabolites and energy</i>	7,01
			GO:0016043	<i>cellular component organization</i>	3,00
			GO:0071842	<i>cellular component organization at cellular level</i>	3,00
			GO:0006996	<i>organelle organization</i>	2,78
			GO:0071840	<i>cellular component organization or biogenesis</i>	2,73

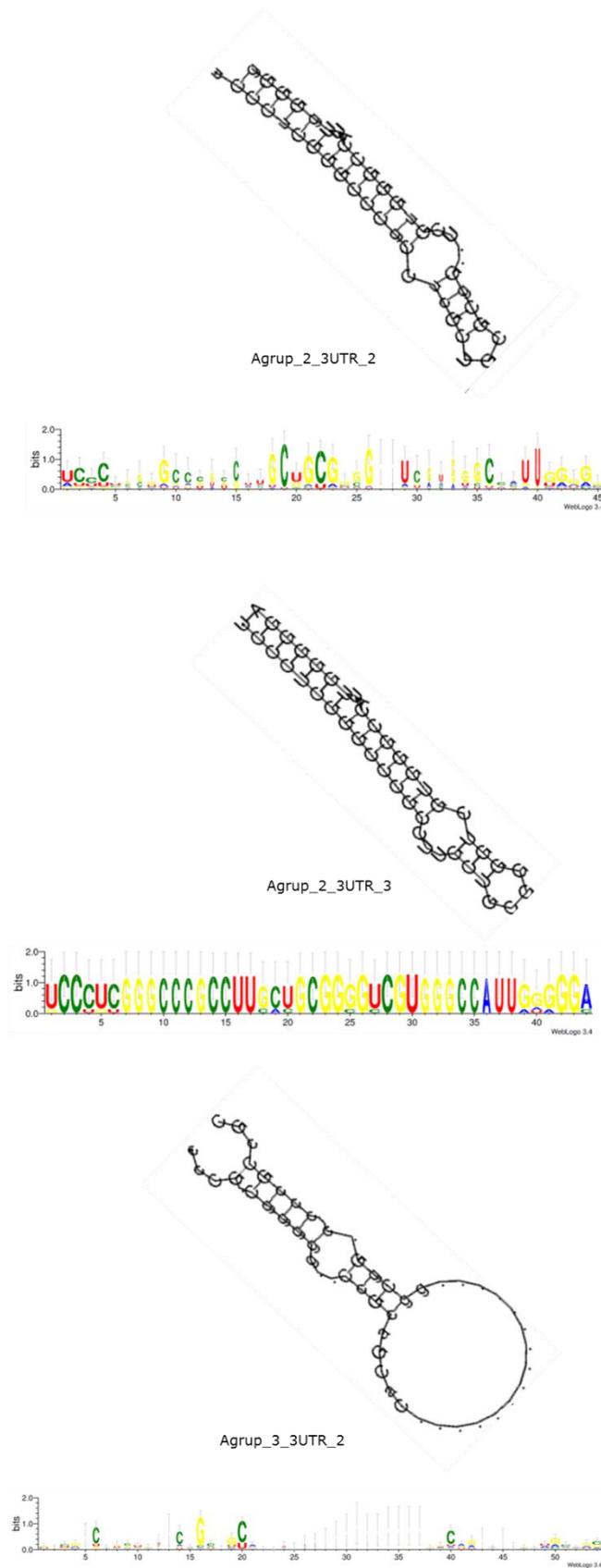


Figura 31. Estructuras que superaron los criterios propuestos para la región 3'UTR. Se observan las 3 estructuras modeladas mediante RNAfold así como el logo de la secuencia consenso para cada una. En cada caso se señala el agrupamiento del cual provienen, la región específica, en este caso 3'UTR y el número del motivo (2 o 3).

La distribución de estos motivos a lo largo de la región 3'UTR no evidenciaría una ubicación preferencial de los mismos Figura 32. El análisis con respecto a las secuencias de las UTRs reales reveló que para el motivo Agrup_2_3UTR_m2, se encontraría presente en el 28% (11/40) de los transcritos que aportan al motivo mientras que para un 25% (10/40) no contamos con información acerca de largo de la 3'UTR. De los 7 transcritos que contiene la estructura Agrup_2_3UTR_m3 en el 71% (5/7) de los casos no podemos predecir la presencia o no del motivo mientras que en el restante 29% (2/7) no se encontraría presente. Finalmente, el motivo presente en el agrupamiento 3, Agrup_3_3UTR_m2, en un 35% (14/40) de los transcritos estaría presente, en un 48% (19/40) no, y para el resto (7/40) no es posible verificarlo con los datos empleados.

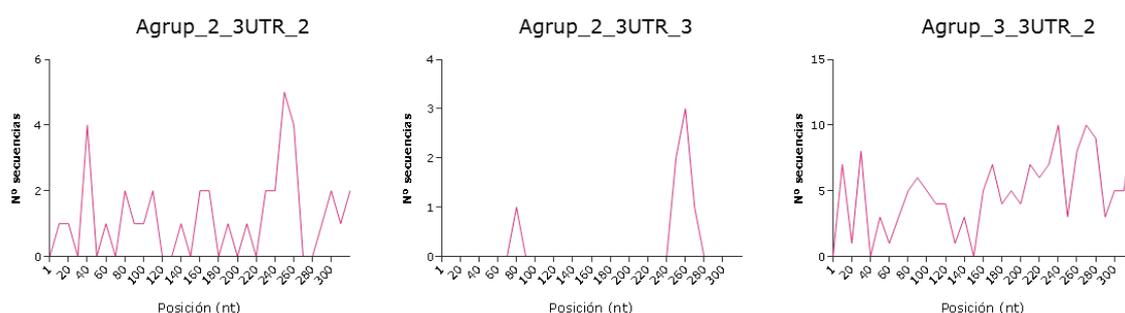


Figura 32. Distribución de las estructuras que se encontraron significativamente enriquecidas a nivel de la región 3'UTR

Se estudió el enriquecimiento en cuanto a **ontología** de los transcritos que poseen cada señal (Tabla 11). Si bien para el motivo tres localizado en unos pocos mensajeros del agrupamiento 2 (Agrup_2_3UTR_m3) no se encontraron términos enriquecidos, los transcritos del Agrupamiento 2 que poseen el motivo 2 (Agrup_2_3UTR_2) presentan un claro enriquecimiento en proteínas asociadas a la **morfogénesis y a la organización del citoesqueleto** dejando ver que las proteínas que presentan este motivo podrían ser responsables de la asociación/interacción del organelo con esta estructura. Por otro lado, es importante resaltar la presencia significativa de subgrupos específicos de transcritos en aquellos mensajeros del agrupamiento 3 que poseen la señal 2 (Agrup_3_3UTR_m2). Más allá del proceso en el que se encuentre involucrado el motivo de estructura encontrado, ya sugiere un mecanismo de regulación del **tipo regulón**, función específica, y no un mecanismo general para los transcritos de MiNT.

Tabla 11. Términos de ontología enriquecidos para cada subgrupo de transcriptos que aporta a cada uno de los motivos encontrados en la región 3'UTR.

Agrup_2_3UTR_2			Agrup_3_3UTR_2		
ID	Término	Fold enrichment	ID	Término	Fold enrichment
GO:0022603	regulation of anatomical structure morphogenesis	218,00	GO:0009107	lipote biosynthetic process	44,59
GO:0007010	cytoskeleton organization	19,82	GO:0045039	protein import into mitochondrial inner membrane	29,73
			GO:0007007	inner mitochondrial membrane organization	26,75
			GO:0006633	fatty acid biosynthetic process	9,56
			GO:0006626	protein targeting to mitochondrion	8,92
			GO:0042180	cellular ketone metabolic process	4,01
			GO:0055114	oxidation-reduction process	3,60

Del estudio del perfil de expresión a lo largo del ciclo de vida del parásito de los genes de MiNT, MiNT-DiExTra-LC, se conformaron 6 agrupamientos. A nivel de secuencia, de las 57 señales encontradas enriquecidas en los agrupamientos con respecto a NoMiNToma, varias son composicionalmente semejantes, y 16 resultaron significativamente comparables con los motivos reconocidos por proteínas de unión a ARN listados en la base de datos disponible a la fecha. Por otra parte, la búsqueda de señales a nivel de estructura para cada agrupamiento permitió definir 11 señales con alta confianza para la región 5'UTR, así como 3 señales en la región 3'UTR. Puntualmente, el motivo Agrup_2_3UTR_m2 parece ser específico para transcriptos que codifican proteínas que permiten mantener la morfología y organización de la mitocondria así como su anclaje al citoesqueleto.

A partir de datos obtenidos en el laboratorio por el Mag. Santiago Chávez (Chávez 2016), de transcriptomas derivados de cultivo de epimastigotas de *T. cruzi* sincronizados, se llevó a cabo un análisis de la expresión a lo largo del **ciclo celular**. Se observó que, de la totalidad de genes (10.597) se obtuvieron datos para 7860, de los cuales el 14,4% son genes de MiNT siendo que estos genes corresponden a un 11,3% del genoma completo. Por lo tanto, de los 1196 genes que componen la base de datos, 1132, o sea más del 94% de los genes de MiNT estarían presentes en el estadio epimastigota. Este hecho da peso a la hipótesis de que el genoma de estos parásitos se transcribe en su totalidad y permite mantener la postura de que la regulación tiene una mayor importancia a nivel postranscripcional.

Para los **1132 genes** de MiNT contamos con los niveles de expresión para las distintas etapas del ciclo celular permitiendo diferenciar aquellos genes que se expresan preferencialmente en la etapa de crecimiento, G1, en la etapa de síntesis de ADN, S, o en G2 donde se da la replicación del kinetoplasto propiamente dicha y la preparación para la división celular. Empleando el programa MeV se evaluaron distintos perfiles de expresión de los transcritos de MiNT. Se generaron **6 agrupamientos** según el comportamiento observado (Figura 33).

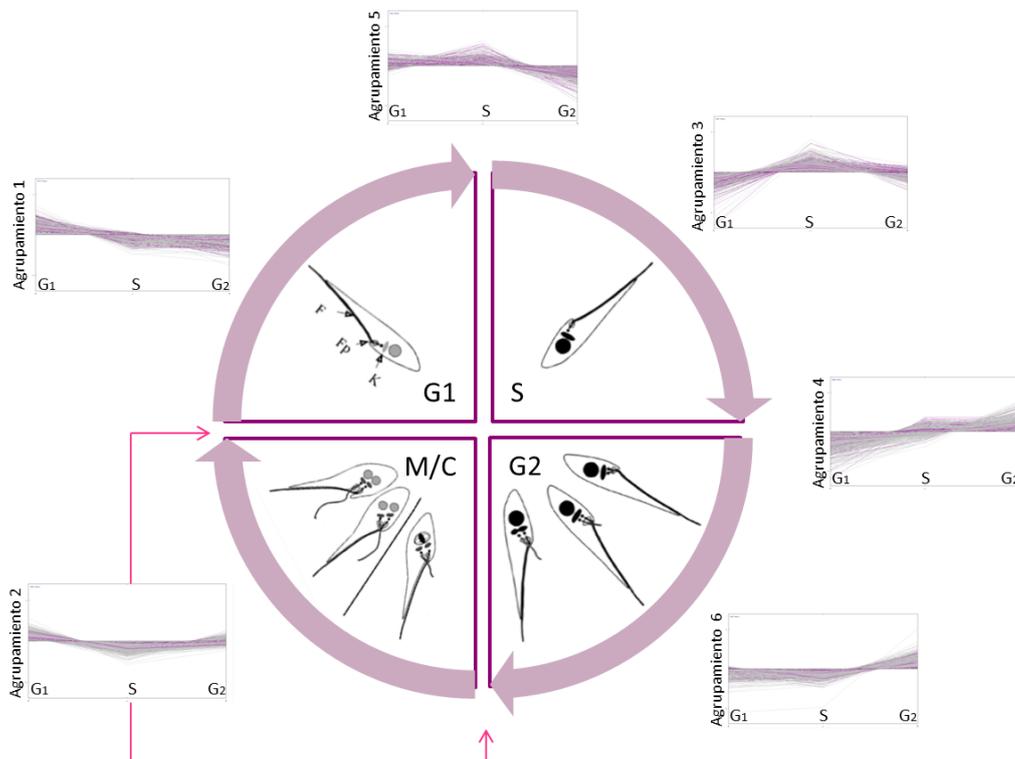


Figura 33. Cambios morfológicos a lo largo del ciclo celular. Se señala con flechas F- flagelo, Fp- Bolsillo Flagelar, K- kinetoplasto. Se muestran los distintos agrupamientos obtenidos a partir de los datos de expresión derivados de cultivos de epimastigotas de *T. cruzi* sincronizados a través de las distintas etapas del ciclo (Chávez 2016) (G1, S-síntesis, G2, M-Mitosis, C- Citocinesis). Adaptado de (Elias et al. 2007).

Se analizaron los genes de cada etapa en busca de enriquecimiento a nivel de ontología empleando la herramienta ofrecida por DAVID. Se obtuvieron categorías ontológicas enriquecidas significativamente para los agrupamientos 4 y 6 (*Enrichment score* > 1,3). En el caso del **agrupamiento 4** cuyo perfil deja ver un aumento en la expresión de mensajeros en la etapa S y G2 del ciclo, el grupo que presentó mayor índice de enriquecimiento (*Enrichment score* = 1,58) fue el que se asocia con unión a desoxiribonucleótidos y a ribonucleótidos. Para el **agrupamiento 6** el perfil describe un aumento de mensajeros hacia G2 mientras que se encuentran subexpresados en el resto de las etapas. Este agrupamiento presentó enriquecimiento significativo en términos asociados a biosíntesis de nucleótidos y de ácidos nucleicos así como con cascadas de señalización y receptores (*Enrichment score* = 2,08 y 1,87 respectivamente). Un enriquecimiento en estas funciones sugiere que el organelo se encuentra en un estadio preparativo para el proceso de **replicación del kinetoplasto**. Este resultado acompaña lo reportado previamente por el grupo de Shenkman (Elias et al. 2007) que propuso que la replicación del kinetoplasto se daría en la etapa G2 del ciclo celular (Figura 33). Teniendo en cuenta al parásito en general, las etapas S y G2 están condicionadas por la disponibilidad energética y la comunicación con el entorno, para lo cual son cruciales los grupos de proteínas que se muestran enriquecidos.

Búsqueda de LiMs en las regiones UTRs

Nos propusimos encontrar motivos de secuencia a nivel tanto de la 3'UTR como de la 5'UTR. Empleando el programa DREME componente de MEME-Suite, se buscaron motivos enriquecidos en los distintos agrupamientos con respecto a la totalidad de los genes de *T. cruzi*, una vez excluidos los que componen la base MiNT (NoMiNToma). Los resultados se resumen en la Tabla 12.

Para cada motivo encontrado, se analizó la posible asociación con proteínas de unión a ARN, empleando el programa Tom-Tom (MEME-suite) que permite comparar con la base de datos de motivos reconocidos por proteínas de unión a ARN (Ray et al. 2013). Las proteínas identificadas son candidatos a ser "reguladores y/o transportadores" del set de mensajeros que presenten el motivo. Para las proteínas identificadas, se buscaron ortólogos en *T. cruzi*. El resultado fue exitoso para 4 de ellos, particularmente dos de estos se encuentran en los *contigs* correspondientes al haplotipo Esmeraldo-like mientras que otros dos únicamente presentaban genes ortólogos en *T. cruzi* haplotipo No Esmeraldo (Tabla 12).

Tabla 12. Motivos lineales encontrados para cada agrupamiento del ciclo celular. En la tabla se resumen los motivos encontrados, así como el resultado de la comparación de estos con el programa Tom-Tom. La proteína que se encontró responsable del reconocimiento se describe y en el caso de que lo hubiere se muestra el resultado de la búsqueda de genes ortólogos para los que codifican estas proteínas en *T. cruzi*.

Agrup. 1 3UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	AAGRAA	2,20E-08	NP_005769.1	RNA-binding protein 5 [<i>Homo sapiens</i>]	No presenta
AMAAGGT	5,50E-08	NP_035447.3	KH domain-containing, RNA-binding, signal transduction-associated protein 1 [<i>Mus musculus</i>]	No presenta	
Agrup. 1 5UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	ATGCGDCG	4,10E-10	No presenta		
	TDGGATCC	1,90E-08	No presenta		
	AAGAGAGG	6,10E-08	NP_473357.1	serine/arginine-rich splicing factor 10 isoform 2 [<i>Homo sapiens</i>]	No Presenta
CRTTT	9,20E-08	No Presenta			
Agrup. 2 3UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	GAAGAY	1,60E-08	AAC39596.1	poly(A) binding protein II [<i>Homo sapiens</i>]	
	TTTWITT	6,00E-08	AAA28922.1 AAB41913.1- XP_001766363.1 EAX81576.1 NP_001261808.1	Sx1 [<i>Drosophila melanogaster</i>] HuR RNA binding protein [<i>Homo sapiens</i>] Pp_0228 RRM [<i>Physcomitrella patens</i>] hypothetical protein [<i>Trichomonas vaginalis</i>] bruno-3, isoform O [<i>Drosophila melanogaster</i>]	No Presenta No Presenta No Presenta No Presenta No Presenta
	AAAG	1,00E-07	No presenta		
Agrup. 2 5UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	CSCCATAT	8,40E-09	No Presenta		
	ATGCKATCG	1,60E-08	NP_650473.1	CG5213 [<i>Drosophila melanogaster</i>]	No presenta
	TTGGCD	4,30E-08	No presenta		
TWTC	1,10E-07	NP_957486.1	Pcbp2 protein [<i>Danio rerio</i>]	No Presenta	
Agrup. 3 3UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	GAAAR	1,90E-10	NP_476667.1	pAbp, isoform A [<i>Drosophila melanogaster</i>]	TcCLB.506885.70
	MTTA	1,20E-09	No presenta		
ARAWG	3,20E-09	NP_731830.2	rasputin, isoform G [<i>Drosophila melanogaster</i>]	No presenta	
Agrup. 3 5UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	ATGCKKCG	3,70E-14	No presenta		
	TWTCG	5,10E-11	No presenta		
	TAGCTTY	5,50E-09	No presenta		
	TYAA	5,00E-09	No presenta		
	AYCCTTGG	1,50E-08	No presenta		
	GGGCGYA	9,90E-08	NP_001027084.1	SRA stem-loop interacting RNA binding protein 1, isoform A [<i>Drosophila melanogaster</i>]	No presenta
	CGCYTT	1,50E-07	No presenta		
GMTTGGTG	1,50E-07	No presenta			
Agrup. 4 3UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	AAARAW	2,60E-09	Tbg972.9.5210	polyadenylate-binding protein 1, putative, PABP1 [<i>Trypanosoma brucei gambiense</i>]	TcCLB.506885.70
	CATTTVGG	4,30E-09	No presenta		
	RGGAAA	5,40E-08	No presenta		
	ARGGGG	9,60E-08	No presenta		
GWTA	1,10E-07	No presenta			
Agrup. 4 5UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	ATGYTTCG	7,40E-20	No presenta		
	GGCMTTGG	1,50E-07	No presenta		
TITSAG	1,60E-07	No presenta			
Agrup. 5 3UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	ARAAG	8,10E-12	AAC39596.1 Tbg972.3.3970	poly(A) binding protein II [<i>Homo sapiens</i>] Triple RNA binding domain protein 3 [<i>Trypanosoma brucei gambiense</i>]	No Presenta TcCLB.510149.140
	ARAAA	9,50E-09	Tbg972.9.5210	polyadenylate-binding protein 1, putative, PABP1 [<i>Trypanosoma brucei gambiense</i>]	TcCLB.506885.70
	CATAGW	5,80E-08	No presenta		
GVA AA	1,50E-07	No presenta			
Agrup. 5 5UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	ATGCKTCG	8,80E-10	NP_650473.1	CG5213 [<i>Drosophila melanogaster</i>]	No presenta
	CGSCTT	9,80E-08	No presenta		
	ATGTTT	1,50E-07	No presenta		
ACATCGVG	2,70E-08	NP_004550.2 NP_731510.1	YBX1 [<i>Homo sapiens</i>] RNA-binding protein 1, isoform D [<i>Drosophila melanogaster</i>]	No presenta No presenta	
Agrup. 6 3UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
	AGAWGG	4,90E-08	Tbg972.3.3970 NP_650107.1	Triple RNA binding domain protein 3 [<i>Trypanosoma brucei gambiense</i>] uncharacterized protein Dmel_CG14718 [<i>Drosophila melanogaster</i>]	TcCLB.510149.140 TcCLB.509733.110*
Agrup. 6 5UTR	P-value		TomTom	RBP [organismo]	Proteína Ortóloga en <i>T. cruzi</i>
GTTGTAGG	3,60E-09	No presenta			

* Solamente presenta un ortólogo en la cepa No Esmeraldo.

Se encontró que la **PABP1**, que se describió en detalle en la sección MiNT-core, reconocería los motivos "GAAAR", "AAARAW" y "ARAAA" que se encuentran enriquecidos en la región 3'UTR de los agrupamientos 3, 4 y 5 respectivamente. En este caso, en particular para el primer motivo, el programa Tom-Tom asignó con un $e\text{-value} < 1$ la proteína de *Drosophila melanogaster* NP_476667.1. Con el mismo $e\text{-value}$ fue asignada la proteína de unión codificada por Tbg972.9.5210 de *T. brucei* al segundo motivo, mientras que para el último motivo, la misma proteína fue asignada con un $e\text{-value} < 0,1$. Estas proteínas presentan una clara conservación a nivel de sus dominios de unión a ARN. En *T. cruzi* corresponde un único gen ortólogo **TcCLB.506885.70** para los genes que las codifican (Figura 14).

La otra proteína de interés que reconocería un motivo similar a los enriquecidos en los agrupamientos 5, "ARAAG", y 6 "AGAWGG" corresponde a la proteína **TRRM3**, *Triple RNA binding domain protein 3* (Tbg972.3.3970) de *Trypanosoma brucei gambiense*. El ortólogo en *T. cruzi* (**TcCLB.510149.140**, TRRM3, Figura 34) presenta una identidad del 75% en el dominio de unión al ARN. La proteína de *T. brucei* (Tb927.3.3670) presenta un 100% de identidad de secuencia con el de la proteína de *T. b. gambiense* (Wurst et al. 2009). Estudios de espectrometría de masas en las fracciones mitocondriales del estadio procíclico ubican a esta proteína en la mitocondria o **asociada a la membrana mitocondrial** (Niemann et al. 2013). Se observó además que esta proteína se encuentra regulada por la presencia de modificaciones postraduccionales, mediante la **metilación de argininas** en *T. brucei* (Lott et al. 2013).



Figura 34. Alineamiento de los motivos (A) y secuencias de las proteínas de unión a ARN (B) que reconocerían los motivos ARAAG y AGAWGG de *T. brucei gambiense* y sus ortólogos en *T. cruzi* y *T. brucei*. Se señala con recuadro verde el dominio RRM1.

Para ambas proteínas codificadas por **TcCLB.506885.70** y **TcCLB.510149.140**, se analizó el perfil de expresión de sus mensajeros a lo largo del ciclo celular usando los datos del Mag. Chávez (**Figura 35**).

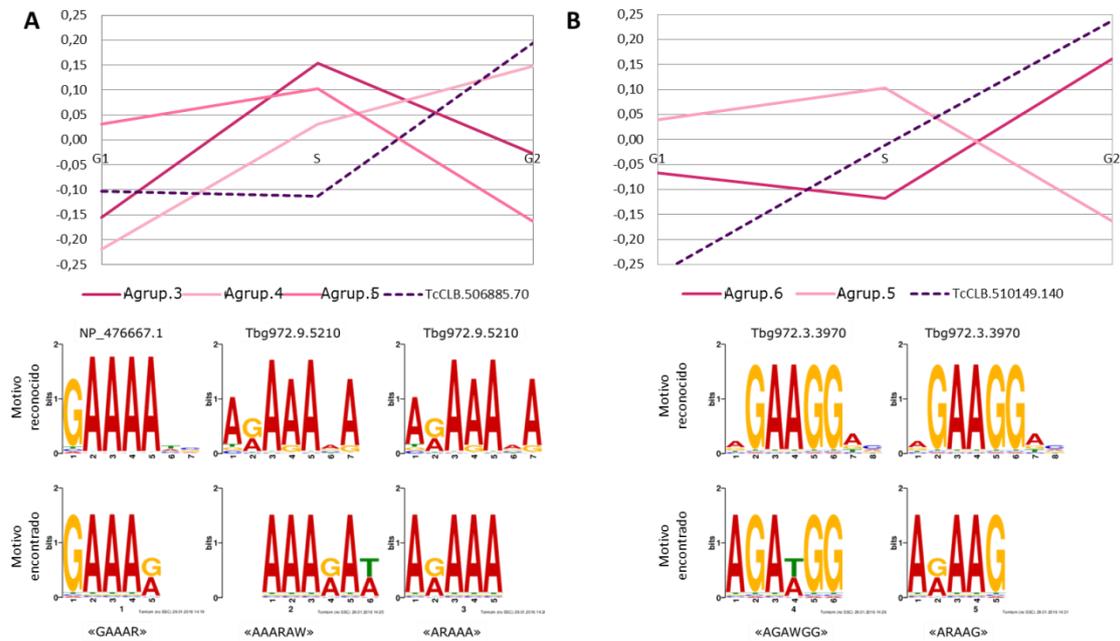


Figura 35. Perfil de expresión a lo largo del ciclo celular de los mensajeros de las proteínas TcCLB.506885.70 (A) y TcCLB.510149.140 (B) posiblemente involucradas en el reconocimiento de los motivos enriquecidos en los distintos agrupamientos. Se indica el perfil de expresión consenso de cada agrupamiento. Se muestra además las secuencias reconocidas por cada RBP comparada con cada motivo encontrado en los distintos agrupamientos. En el caso del motivo "GAAAR" la proteína de reconocimiento fue asignada con un $p\text{-value} = 3,56e-03$, para "AAARAW" $p\text{-value} = 2,80e-03$, para "ARAAA" $p\text{-value} = 1,11e-04$, para "AGAWGG" $p\text{-value} = 6,96e-04$ y para "ARAAG" $p\text{-value} = 6,34e-03$.

Para el caso del transcripto de la proteína PABP (**TcCLB.506885.70**) se observa que el mensajero de esta proteína de unión a ARN presenta una sobreexpresión en G2. Esta proteína que reconocería los motivos "GAAAR", "AAARAW" y "ARAAA" enriquecidos en la región 3'UTR de los agrupamientos 3, 4 y 5 respectivamente, presenta un perfil prácticamente espejudo a nivel de mensajeros con los agrupamientos 3 y 5. En cambio, el perfil de expresión acompaña al correspondiente en promedio al agrupamiento 4 (Figura 35A). En el caso del transcripto que codifica la proteína TRRM3 (**TcCLB.510149.140**) se aprecia un aumento en la expresión a lo largo del ciclo, comenzando por valores bajos en G1 y aumentando paulatinamente hasta alcanzar un máximo de expresión en G2. Sin embargo, los grupos (5 y 6) que poseen un enriquecimiento en las señales "ARAAG" y "AGAWGG" que serían reconocidos por esta proteína presentan un comportamiento diametralmente opuesto entre ellas. Esto implicaría que la proteína podría estar reconociendo a sus mensajeros blanco indistintamente de la fase del ciclo en la que se encuentre y no parece estar regulando, al menos *a priori*, su expresión.

Ya que los transcritos reconocidos por estas proteínas no parecen presentar un perfil de expresión diferencial que los caracterice, se intentó encontrar algún criterio que los uniera como subgrupo. Para esto, se analizó por ontología los transcritos con la señal que reconocería cada una de las proteínas. En el caso de los transcritos que poseen el motivo "GAAAR" se observa (Tabla 13) predominancia de términos asociados a **estabilidad y procesamiento** (edición) de mensajeros. Para aquellos que contienen al motivo "AAARAW" (Tabla 13) en su 3'UTR, se observan términos asociados plenamente al metabolismo mitocondrial incluyendo la **fosforilación oxidativa y el metabolismo redox**. También se observa un claro enriquecimiento en el metabolismo de proteínas. En estos casos, ambos subgrupos comparten además de los términos genéricos, un conjunto de términos asociados al proceso de localización mitocondrial de proteínas y en particular a la importación de proteínas a la matriz del organelo. Los procesos que se observan no parecen ser típicos de ninguna etapa en particular en el ciclo celular del parásito por lo que no es posible generar una asociación directa entre expresión diferencial y localización mitocondrial. El último conjunto de transcritos (Tabla 13) es el más distante de los tres, observándose un enriquecimiento en términos que hacen a la **replicación del kinetoplasto** propiamente dicha (motivo "ARAAA"). Estos transcritos presentan un aumento en la expresión de sus mensajeros hacia la etapa S del ciclo y disminuye en la etapa G2 donde efectivamente sus productos proteicos cumplirían su función (Figura 35 A).

Tabla 13. Términos de ontología en los que se encuentran enriquecidos los transcritos que contienen da motivo. Se muestran los términos más específicos dentro de los cuales el enriquecimiento fue significativo. En los tres grupos se descartaron los subgrupos asociados a términos genéricos como procesos metabólicos, metabolismo mitocondrial, etc.

"GAAAR"	ID	Término	Fold enrichment
	GO:0016554	<i>cytidine to uridine editing</i>	21,91
	GO:0048255	<i>mRNA stabilization</i>	17,98
	GO:0042559	<i>pteridine-containing compound biosynthetic process</i>	16,7
	GO:0008299	<i>isoprenoid biosynthetic process</i>	14,61
	GO:0030150	<i>protein import into mitochondrial matrix</i>	14,61
	GO:0043488	<i>regulation of mRNA stability</i>	14,61
	GO:0072655	<i>establishment of protein localization in mitochondrion</i>	9,74
	GO:0043038	<i>amino acid activation</i>	9,42
	GO:0010608	<i>posttranscriptional regulation of gene expression</i>	9,35
	GO:0006091	<i>generation of precursor metabolites and energy</i>	7,44
	GO:0034660	<i>ncRNA metabolic process</i>	3,81
	GO:0044281	<i>small molecule metabolic process</i>	2,96
"AAARAW"	ID	Término	Fold enrichment
	GO:0032049	<i>cardiolipin biosynthetic process</i>	35,53
	GO:0097033	<i>mitochondrial respiratory chain complex III biogenesis</i>	35,53
	GO:0000963	<i>mitochondrial RNA processing</i>	26,65
	GO:0033108	<i>mitochondrial respiratory chain complex assembly</i>	26,65
	GO:0097034	<i>mitochondrial respiratory chain complex IV biogenesis</i>	26,65
	GO:0008535	<i>respiratory chain complex IV assembly</i>	23,69
	GO:0000959	<i>mitochondrial RNA metabolic process</i>	17,76
	GO:0030150	<i>protein import into mitochondrial matrix</i>	17,76
	GO:0007005	<i>mitochondrion organization</i>	13,82
	GO:0072655	<i>establishment of protein localization in mitochondrion</i>	11,84
	GO:0009063	<i>cellular amino acid catabolic process</i>	9,47
	GO:0070271	<i>protein complex biogenesis</i>	8,61
	GO:0046395	<i>carboxylic acid catabolic process</i>	8,36
	GO:0065003	<i>macromolecular complex assembly</i>	5,47
	GO:0006732	<i>coenzyme metabolic process</i>	4,1
	GO:0034622	<i>cellular macromolecular complex assembly</i>	4,02
	GO:0034621	<i>cellular macromolecular complex subunit organization</i>	4,01
GO:0006811	<i>ion transport</i>	3,85	
GO:0055114	<i>oxidation-reduction process</i>	3,74	
"AAAAA"	ID	Término	Fold enrichment
	GO:0032042	<i>mitochondrial DNA metabolic process</i>	18,5
	GO:0006260	<i>DNA replication</i>	7,33
	GO:0006259	<i>DNA metabolic process</i>	2,76

El motivo "AGAWGG" (Tabla 14) por su parte está presente en transcritos asociados a las **cascadas de señalización intracelular**. Como ya se mencionó este grupo de transcritos, así como TRRM3 (Figura 35 B) que

estaría encargada del reconocimiento del motivo descrito, se encuentran sobreexpresados en G2. En el caso de la señalización mediante AMPc, el 100% de los transcriptos asociados a este término poseen el motivo en cuestión, por lo que podríamos al menos asegurar que esta función estaría asociada a la etapa pre-mitosis del ciclo.

Tabla 14. Términos de ontología en los que se encuentran enriquecidos los transcriptos que contienen cada motivo. Se muestran los términos más específicos dentro de los cuales el enriquecimiento fue significativo. Se descartaron los subgrupos asociados a términos genéricos como procesos metabólicos, metabolismo mitocondrial, etc.

	ID		Término	Fold enrichment
"AGAWGG"	GO:0019933		<i>cAMP-mediated signaling</i>	67,29
	GO:0009168		<i>purine ribonucleoside monophosphate biosynthetic process</i>	29,91
	GO:0009127		<i>purine nucleoside monophosphate biosynthetic process</i>	29,91
	GO:0019932		<i>second-messenger-mediated signaling</i>	28,84
	GO:0009190		<i>cyclic nucleotide biosynthetic process</i>	24,47
	GO:0044242		<i>cellular lipid catabolic process</i>	16,82
	GO:0015980		<i>energy derivation by oxidation of organic compounds</i>	8,41
	GO:0045333		<i>cellular respiration</i>	8,41
	GO:0034654		<i>nucleobase-containing compound biosynthetic process</i>	7,27
	GO:0055114		<i>oxidation-reduction process</i>	4,19
	GO:0044281		<i>small molecule metabolic process</i>	3,19
"ARAAG"	GO:0006260		<i>DNA replication</i>	6,38
	GO:0055114		<i>oxidation-reduction process</i>	3,00
	GO:0019752		<i>carboxylic acid metabolic process</i>	2,76
	GO:0043436		<i>oxoacid metabolic process</i>	2,76

Los transcriptos asociados al motivo "ARAAG" (Tabla 14) por su parte se encuentran enriquecidos, además de en términos comunes a la mayoría de los agrupamientos como son los **procesos redox** y el **metabolismo de ácidos carboxílicos, en la replicación de ADN** que se sucede en la etapa G2 del ciclo. Al igual que sucede en el caso de los transcriptos conteniendo el motivo "ARAAA", por pertenecer al mismo agrupamiento, se encuentran subexpresados en la etapa en la que se daría la replicación del kinetoplasto mientras que **TRRM3** se encuentra sobreexpresada en G2 con respecto al resto de las etapas del ciclo celular. Claramente los transcriptos que poseen estos motivos no se encuentran asociados a una función en particular ni poseen un perfil diferencial lo largo del ciclo celular. Es posible que la señal encontrada para estos dos subgrupos en particular sea reconocida por TRRM3 y sea esta proteína finalmente la responsable de acarrear estos mensajeros al entorno mitocondrial. El motivo reconocido por la TRRM3 según el trabajo de Ray y colaboradores (Ray et al. 2013) fue "RGAAGGDM". Al analizar los resultados anteriores con respecto a las búsquedas de LiMs, encontramos que en la mayoría de los casos se observa

representado este motivo en las regiones 3'UTR de los distintos agrupamientos analizados (Figura 36 A), incluyendo el motivo encontrado enriquecido en la región 3'UTR de MiNT-NoMTS (MiLTres_{NoMTS}). Se alinearon dichos motivos así como los que se describen en los apartados siguientes que cumplen con el criterio. Se generó una secuencia consenso y se buscó empleando el programa Tom-Tom si este motivo era reconocido por la proteína de interés. Se encontró que la TRRM3 de *T. b. gambiense* reconocería el motivo consenso con un ($p\text{-value} = 2,27e-2$) (Figura 36 B).

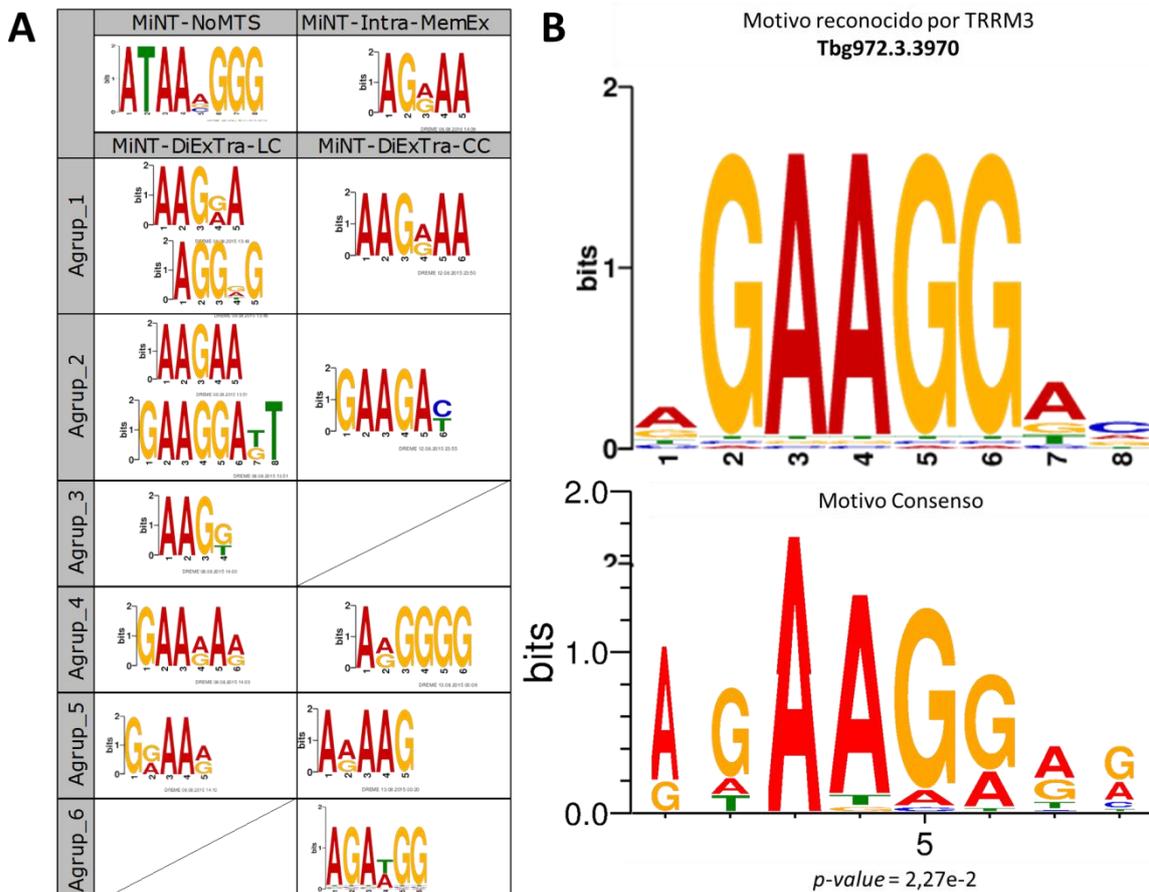


Figura 36. Secuencia consenso derivada de LiMs encontrados en la región 3'UTR a lo largo de este trabajo. En A) Se muestra una compilación de los LiMs encontrados en los distintos agrupamientos detallando en que subgrupo se encontró enriquecido. B) Alineamiento de la secuencia consenso generada a partir de los LiMs mostrados en A) y el motivo lineal reconocido por la proteína TRRM3.

Es claro que el motivo aquí encontrado se encuentra ampliamente distribuido en los transcritos de MiNT. El hecho de que este motivo pueda ser reconocido por una proteína en *T. brucei* que además fue encontrada en el entorno mitocondrial mediante análisis del proteoma de la membrana mitocondrial, deja la puerta abierta a pensar que esta proteína podría ser responsable del transporte de los mensajeros que codifican para proteínas mitocondriales mediante el reconocimiento de esta señal a nivel de la 3'UTR.

Por otra parte, se estudió en detalle el motivo "ATGCKTCG", dado que se encontraron motivos similares enriquecidos en la región 5'UTR de 5 de los 6

agrupamientos generados. Para ello se generó un nuevo motivo consenso entre los motivos de características similares obtenidos (ver Tabla 12). Se buscó el motivo generado (Figura 37 A) en las regiones 5' de los 6 agrupamientos definidos. Cabe resaltar que si bien estas secuencias contienen parte de la región codificante las secuencias que se usan como control también la poseen, por lo que el enriquecimiento de todas formas es válido. De la totalidad de los motivos que se encuentran en transcritos de MiNT se observa que el 1% de los motivos estaría ubicado en la 5'UTR propiamente dicho, mientras que del 99% restante, el 80% se encuentra en fase con el marco abierto de lectura ubicándose la mayoría de los casos (55%) en la metionina inicial, el 4% en el segundo aminoácido de la proteína y el restante 21% interno en la CDS. Es llamativo que en el 20% de los casos el motivo se encuentre en otro marco de lectura (Tabla 15). Para 14 de los 229 transcritos que poseen la señal se encontró el motivo duplicado. En 7 de estos casos, uno de los motivos se encuentra a nivel de la metionina inicial y el otro interno en la CDS, en fase con el marco abierto de lectura. Particularmente, hay 5 transcritos que presentan dos veces el motivo en su secuencia de las cuales una vez se encuentra en fase y la otra no, mientras que los dos restantes cuentan con el motivo por duplicado en su secuencia ambas veces fuera de fase.

Tabla 15. Distribución del motivo "ATGCKDCG" a lo largo de la región 5' del transcripto, en cada uno de los agrupamientos generados para el ciclo celular. Cabe resaltar que el agrupamiento 6 a pesar de no encontrarse enriquecido el motivo, se encontraron algunos transcritos que lo incluyen en su secuencia.

	UTR	CDS			
		En fase			Fuera de fase
		Met Inicial	2º AA	Interno	
A1	2% (1)	47% (22)	2% (1)	21% (10)	29% (14)
A2	0% (0)	43% (18)	7% (3)	33% (14)	17% (7)
A3	2% (1)	55% (30)	5% (3)	17% (9)	20% (11)
A4	3% (1)	67% (26)	5% (2)	15% (6)	10% (4)
A5	0% (0)	52% (22)	5% (2)	21% (9)	21% (9)
A6	0% (0)	70% (14)	0% (0)	15% (3)	15% (3)

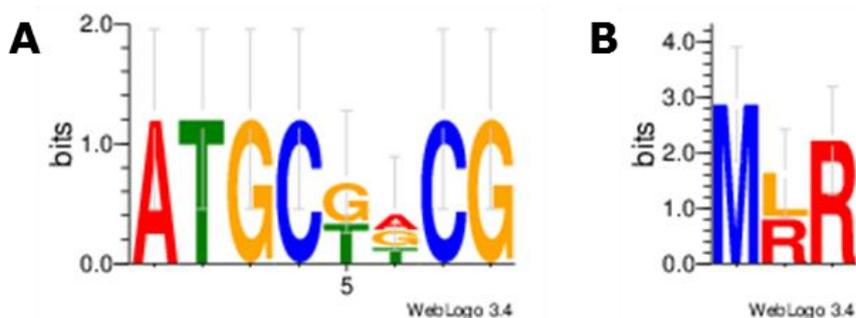


Figura 37. Secuencia nucleotídica (A) y aminoacídica (B) del motivo encontrado enriquecido en la región 5' de los transcritos pertenecientes a agrupamientos en base al ciclo celular de MiNT.

Al ubicarse en la mayoría de los casos en la **región codificante del gen**, en fase con el marco abierto de lectura hace pensar que más que tratarse de un motivo con relevancia a nivel del transcripto podría codificar para un motivo a nivel proteico. Teniendo en cuenta la secuencia nucleotídica se generó además el motivo proteico derivado de la misma (Figura 37B). Debe tenerse en cuenta que los aminoácidos que son codificados por estos arreglos de nucleótidos podrían ser codificados por 3 arreglos más en el caso de la leucina y arginina en la segunda posición y por 2 arreglos más en el caso de la leucina en la tercera posición. El motivo peptídico que encontramos se asemeja al motivo encontrado en la sección MiNT-core así como a TryM-TaPe definido en la sección "MiNT-MTS".

Se compararon los subgrupos conteniendo el motivo **TryM-TaPe** y aquellos transcritos para los que se encontró la señal a nivel nucleotídica. De los 188 transcritos, que poseen esta señal incluida en la CDS y en fase, 130 se encuentran dentro del subgrupo definido con péptido señal (69%). Sin embargo, el resto de los transcritos se encuentran dentro de MiNT-NoMTS. Por otro lado, los tres transcritos (TcCLB.511215.110, TcCLB.506925.14 y TcCLB.506753.150) que según el análisis contarían con la señal en la secuencia 5'UTR propiamente dicha, codifican para proteínas hipotéticas conservadas, una de las cuales fue analizada en la sección: MiNT-core (TcCLB.506753.150). Para TcCLB.506925.14, se observa que la señal se encuentra efectivamente fuera de fase y no habría en principio ningún indicio de que estuviera mal anotada, una vez analizado el gen ortólogo en *T. brucei* (Tb927.11.600). Finalmente, si bien el motivo fue encontrado con mayor puntaje en la 5'UTR en el caso de TcCLB.511215.110, la proteína también comienza con la misma secuencia aminoacídica, MRL, pero codificada por el arreglo "ATGCGGCT". Cabe señalar que esta proteína ya se encuentra formando parte de la base de datos que contiene péptido señal.

Búsqueda de StMs en las regiones UTRs

La búsqueda de señales a nivel de mensajeros se extendió empleando las herramientas disponibles en el paquete INFERNAL y *CMfinder*. De esta forma, se analizaron los mensajeros específicos de cada agrupamiento, tanto a nivel de 3'UTR como de la región 5'UTR definidas. En todos los casos se generaron 3 motivos para cada búsqueda (Tabla 16). Estos motivos se analizan con el fin de determinar su especificidad en cuanto a MiNT, así como su cobertura con respecto al agrupamiento específico (Figura 38).

Tabla 16. Búsqueda de motivos estructurales en los agrupamientos según el perfil de expresión a lo largo del ciclo celular

Motivo	Genes aporte	Largo (nt)	DE Largo (nt)	Cobertura	Sensibilidad CM search	Especificidad	ΔG kcal/mol	
Agrup_1_3UTR_1*	111	25	7	48%	25%	●	88%	-0,20
Agrup_1_3UTR_2*	222	60	44	97%	51%	●	58%	-10,16
Agrup_1_3UTR_3*	115	34	8	50%	29%	●	87%	-8,22
Agrup_1_5UTR_1	174	33	0	76%	9%	●	96%	0,00
Agrup_1_5UTR_2	219	37	6	96%	15%	●	96%	-4,37
Agrup_1_5UTR_3*	168	32	0	73%	6%	●	97%	0,00
Agrup_2_3UTR_1	100	30	4	63%	61%	●	82%	0,00
Agrup_2_3UTR_2	106	36	3	67%	59%	●	80%	-7,42
Agrup_2_3UTR_3	140	39	7	88%	34%	●	84%	-7,04
Agrup_2_5UTR_1	132	27	6	83%	12%	●	99%	-8,72
Agrup_2_5UTR_2	153	35	6	96%	16%	●	99%	-12,48
Agrup_2_5UTR_3	70	38	7	44%	23%	●	99%	-5,12
Agrup_3_3UTR_1*	123	31	4	57%	21%	●	90%	-4,65
Agrup_3_3UTR_2	188	38	5	88%	63%	●	66%	-1,19
Agrup_3_3UTR_3*	126	42	12	59%	25%	●	87%	-10,37
Agrup_3_5UTR_1	183	25	10	86%	4%	●	99%	-6,20
Agrup_3_5UTR_2	213	29	5	100%	17%	●	98%	-6,98
Agrup_3_5UTR_3	207	29	7	97%	12%	●	98%	-24,91
Agrup_4_3UTR_1	131	28	6	63%	43%	●	86%	-2,49
Agrup_4_3UTR_2	124	26	3	60%	40%	●	85%	-3,78
Agrup_4_3UTR_3	140	30	6	67%	40%	●	83%	-6,68
Agrup_4_5UTR_1	204	29	3	98%	13%	●	99%	-4,25
Agrup_4_5UTR_2	170	34	7	82%	8%	●	99%	-18,96
Agrup_4_5UTR_3	118	41	2	57%	41%	●	89%	-1,52
Agrup_5_3UTR_1*	151	27	6	76%	36%	●	79%	-3,81
Agrup_5_3UTR_2	92	39	9	46%	54%	●	85%	-4,19
Agrup_5_3UTR_3	101	37	11	51%	41%	●	88%	-5,03
Agrup_5_5UTR_1	180	28	3	90%	11%	●	99%	0,00
Agrup_5_5UTR_2	168	38	3	84%	13%	●	99%	-7,79
Agrup_5_5UTR_3	190	35	5	95%	18%	●	99%	-12,48
Agrup_6_3UTR_1	66	29	4	54%	48%	●	89%	0,00
Agrup_6_3UTR_2	116	41	5	94%	23%	●	95%	0,00
Agrup_6_3UTR_3	120	43	4	98%	43%	●	87%	-13,10
Agrup_6_5UTR_1	109	32	6	89%	14%	●	98%	-1,86
Agrup_6_5UTR_2	112	29	8	91%	23%	●	98%	-7,63
Agrup_6_5UTR_3	84	40	7	68%	25%	●	99%	-15,50

Se señala con * los motivos que no superaron el criterio de especificidad/sensibilidad propuesto para este tipo de búsquedas. Los círculos verdes señalan aquellos agrupamientos cuya especificidad supera el 90% mientras que el amarillo corresponde a aquellos que están entre el 80 y 90% y el rojo a los menores del 80% de especificidad. Se señala además en verde aquellas estructuras cuyo ΔG de formación a 28°C es menor al promedio de ΔG encontrados.

Una vez obtenidos los motivos se filtraron los resultados para seguir trabajando con aquellas estructuras que resultasen más significativas. Se seleccionaron las estructuras teniendo en cuenta su especificidad y ΔG de formación a 28°C. Las estructuras que pasaron los criterios de selección se representan en la Figura 39.

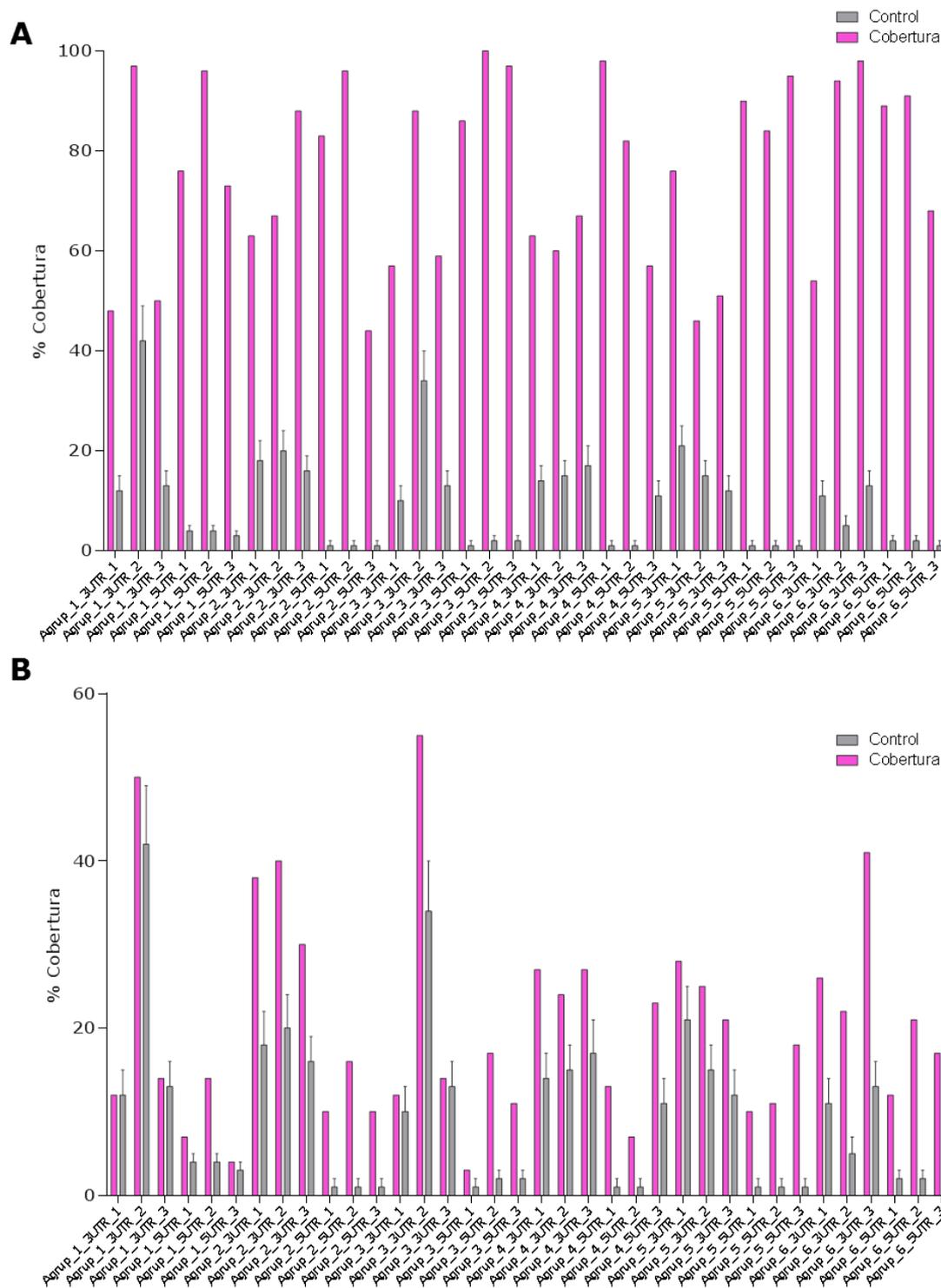


Figura 38. Cobertura y especificidad de los motivos estructurales encontrados para los distintos agrupamientos del ciclo celular. En A) se presentan los resultados obtenidos a partir del análisis de la cobertura en cuanto a las secuencias que aportan a la definición del motivo estructural en comparación de la cobertura que presentan los grupos de transcritos al azar cuando se utiliza el motivo como entrada. En B) se representan los resultados de cobertura obtenidos cuando se lleva a cabo la búsqueda del motivo contra el grupo que lo generó, en comparación con la cobertura que presentan los grupos de transcritos al azar cuando se utiliza el motivo como entrada.

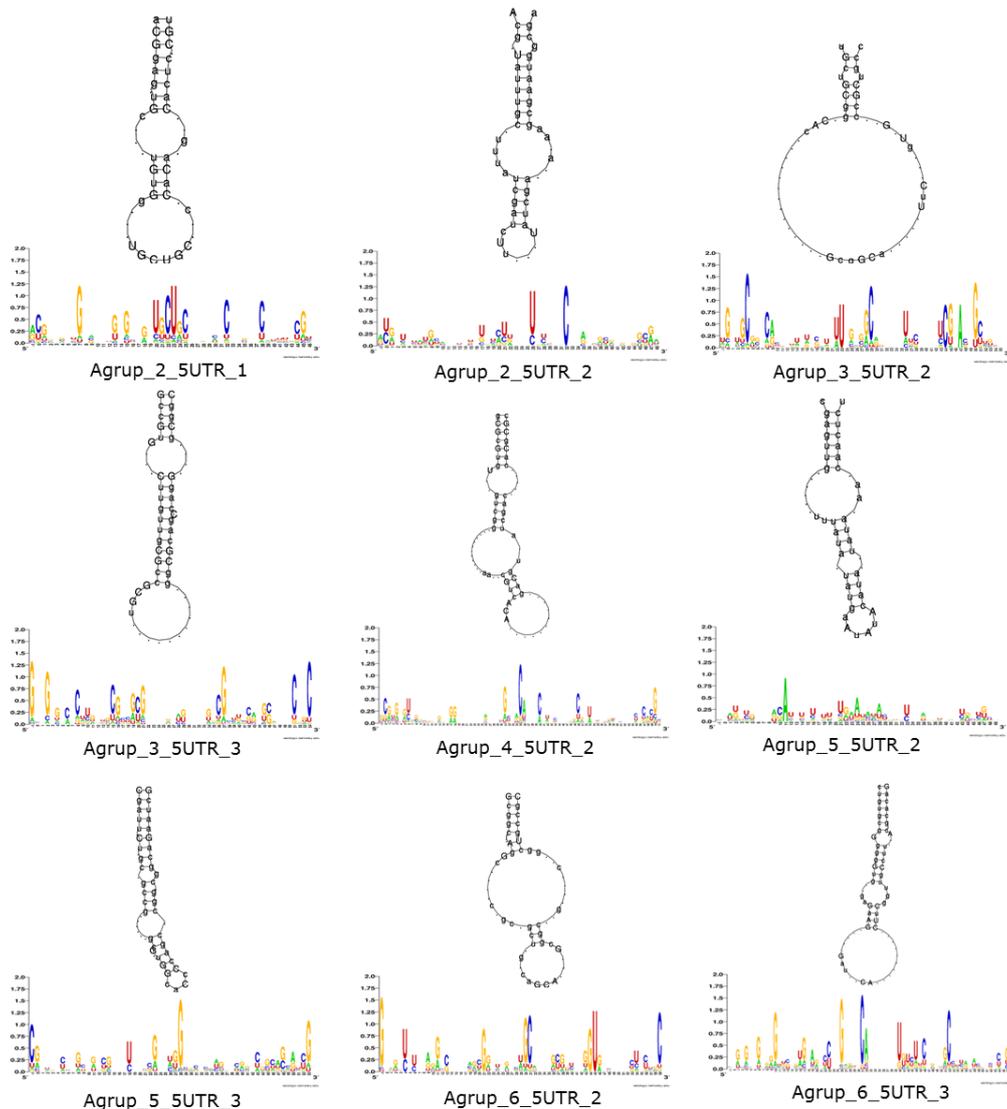


Figura 39. Estructura y secuencia consenso para los motivos seleccionados. Las estructuras con RNAfold los logos con weblogo.

Se estudió la ontología ($e\text{-value} < 1e\text{-}3$) de los subgrupos de transcritos asociados a alguna de estas señales, empleando como contexto todos los genes de *T. cruzi*. Para ellos se observa que:

- › Transcritos del **agrupamiento 2** que contienen alguno de los motivos definidos para este agrupamiento (Agrup_2_5UTR_m1 ó Agrup_2_5UTR_m2) se encuentran asociados a procesos de oxidación-reducción, GO:0055114.
- › Transcritos del **agrupamiento 3** que contienen alguno de los motivos definidos para este agrupamiento (Agrup_3_5UTR_m2 Agrup_3_5UTR_m3) se ven mayormente enriquecidos motivos asociados a procesos metabólicos (GO:0008152; GO:0019752;GO:0016054) y de localización proteica en la mitocondria (GO:0006626; GO:0033365; GO:0070585; GO:0072655).
- › Transcritos del **agrupamiento 4** que contienen el motivo Agrup_4_5UTR_m2 presentan un enriquecimiento significativo en

- términos asociados a procesos metabólicos de ARN (GO:0043489; GO:0048255; GO:0043488; GO:0043487; GO:0009451; GO:0016070)
- › Transcritos del **agrupamiento 5** que contienen alguno de los motivos definidos para este agrupamiento (Agrup_5_5UTR_m2 ó Agrup_5_5UTR_m3) se encuentran: replicación del ADN (GO:0006260) y procesos metabólicos en general (GO:0008152)
 - › Finalmente, para los motivos encontrados en el **agrupamiento 6** (Agrup_6_5UTR_m2 ó Agrup_6_5UTR_m3) principalmente se encuentran transcritos asociados al metabolismo y biosíntesis de nucleótidos y ribonucleótidos (GO:0009123; GO:0009124; GO:0009126; GO:0009127; GO:0009161; GO:0009167; GO:0009168).

Se observa que los subgrupos de transcritos que poseen cada motivo se encuentran enriquecidos prácticamente en los mismos términos que el subgrupo que los contiene, dando cuenta de la cobertura que alcanzan estos motivos. Sin embargo, es claro que no se puede atribuir un motivo estructural a una función dada.

En cuanto a la posición de los motivos estructurales encontrados, corresponden a la región usada para la búsqueda de señales en la 5'UTR. Sin embargo, dado el método empleado en la búsqueda de los motivos, resulta que los mismos se distribuyen desde 80 nucleótidos corriente arriba del sitio de inicio de la traducción hasta 270 nt corriente abajo del mismo. En la Figura 40 se muestra la distribución a lo largo de esta región con más detalle. Se observa que el motivo 2 correspondiente al agrupamiento 2 (Agrup_2_5UTR_m2) así como el motivo 2 correspondiente al agrupamiento 5 (Agrup_5_5UTR_m2) presentan un enriquecimiento a nivel de la 5'UTR propiamente dicho. En el resto de los casos la distribución del motivo en la región es pareja o se encuentra enriquecida en la región codificante.

Caso a caso y para cada motivo se analizó el largo de la 5'UTR obtenido de forma experimental para cualquiera de los 3 estadios indistintamente. De esta forma se buscó evaluar en cada caso si el motivo definido, estaría o no presente en el transcripto real (Tabla 17). Se encontró que los motivos definidos estarían presentes en el (98 ± 2) % de los casos para los que se cuenta con información de los largos de las regiones no traducidas. Solamente en un (17 ± 3) % no se tiene información al respecto de los largos de las UTRs.

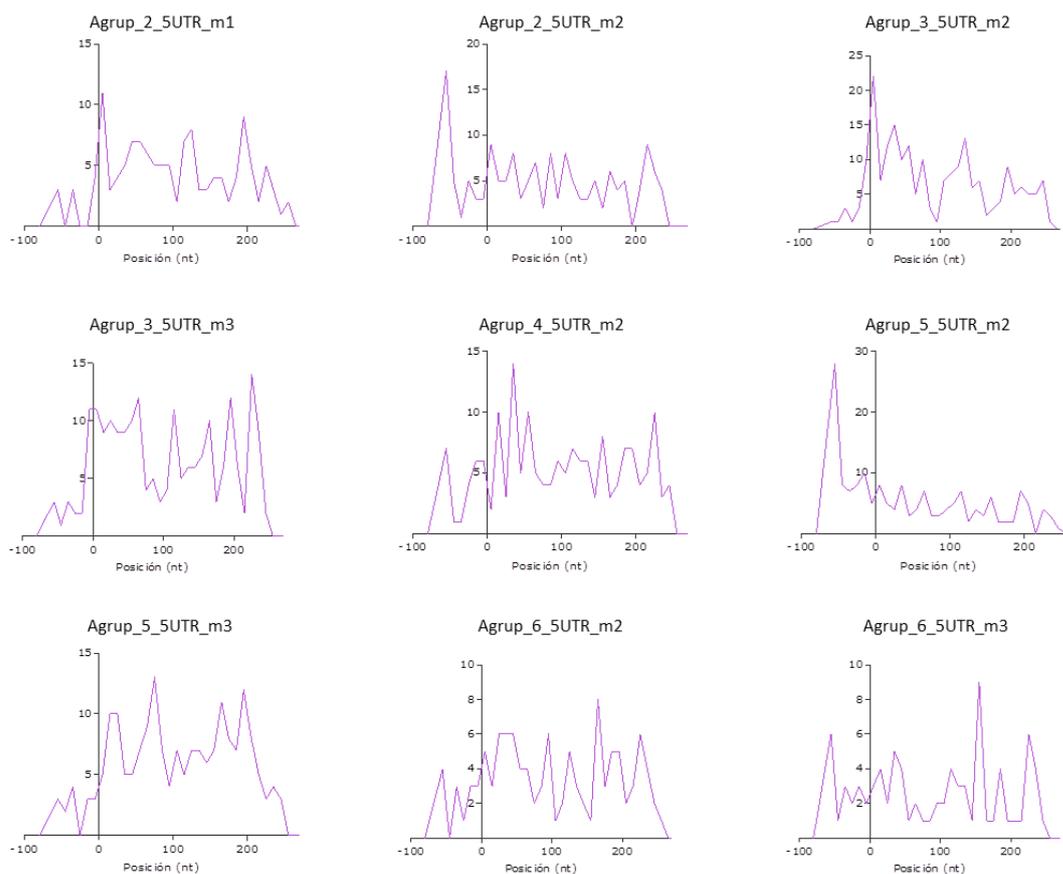


Figura 40. Frecuencia de sitios de inicio de las estructuras encontradas para la región 5'. +1 corresponde a la posición del AUG anotado.

Los agrupamientos para los cuales se definió más de un motivo relevante fueron analizados en cuanto a la **distribución** de los distintos motivos a través de la población de mensajeros que los contienen. Para ello, se revisó la **coincidencia/exclusividad** de un motivo en particular transcrito a transcritos (Tabla 18).

Tabla 17. Síntesis de total de UTRs que no presentarían el motivo o para las cuales no se cuenta con el largo.

Motivo	Total	UTRs ND*		Motivo no presente	
Agrup_2_5UTR_m1	132	20	15%	1	1%
Agrup_2_5UTR_m2	153	28	18%	5	3%
Agrup_3_5UTR_m2	213	28	13%	0	0%
Agrup_3_5UTR_m3	207	27	13%	2	1%
Agrup_4_5UTR_m2	170	26	15%	2	1%
Agrup_5_5UTR_m2	168	28	17%	13	8%
Agrup_5_5UTR_m3	190	31	16%	3	2%
Agrup_6_5UTR_m2	112	24	21%	0	0%
Agrup_6_5UTR_m3	84	17	20%	1	1%

*ND: largo de UTR no disponible para ningún estadio

Tabla 18. Resumen de posiciones relativas de motivos pertenecientes a un mismo agrupamiento.

Comparación	Exclusivos		Total de aportes	Solapados			Posición Relativa	
	m_x	m_y		m_x m_y	m_x m_y	m_x m_y	m_x	m_y
	Corriente Arriba	Superpuesto		Corriente abajo	m_x	m_y	m_x	m_y
Agrup2_5UTR_m1 vs. Agrup2_5UTR_m2	5% (6)	18% (27)	159	9% (11)	10% (12)	6% (8)	35% (44)	40% (51)
Agrup3_5UTR_m2 vs. Agrup3_5UTR_m3	3% (7)	0% (1)	215	12% (24)	9% (18)	13% (26)	39% (81)	28% (58)
Agrup5_5UTR_m2 vs. Agrup5_5UTR_m3	5% (8)	16% (30)	198	9% (15)	3% (4)	6% (9)	60% (96)	23% (36)
Agrup6_5UTR_m2 vs. Agrup6_5UTR_m3	32% (36)	10% (8)	119	7% (5)	13% (10)	9% (7)	31% (23)	40% (30)

Por lo tanto, se observó que la presencia de un único motivo de manera exclusiva se da como máximo en el 32% de los transcriptos siendo el promedio de $(11 \pm 10) \%$. El índice de solapamiento ya sea total o parcial se da en un promedio de $(9 \pm 3) \%$ de los transcriptos que poseen los dos motivos. Cuando se tienen en cuenta la presencia de ambos motivos en un único transcripto, la posición relativa de los mismos se acerca a proporciones que permitirían concluir que no hay una preferencia particular por la posición relativa. Sin embargo, es llamativo el caso del Agrup5_5UTR_m2 el cual en un 69% de los casos se ubica corriente arriba con respecto al Agrup5_5UTR_m3.

Para los 9 motivos encontrados se hizo una búsqueda contra MiNT. Si bien debido a la baja sensibilidad del método de búsqueda en cuanto a los motivos analizados, no se obtienen resultados confiables, comparativamente es claro que los motivos se encuentran enriquecidos en los agrupamientos que lo generaron (Tabla 19) por lo que serían a su vez **agrupamiento específicos**. De todas formas, se analizó el enriquecimiento en cuanto a términos de ontología para ver si era posible agruparlos según este criterio, sin embargo, como era de esperarse los términos que se encuentran enriquecidos son a grandes rasgos los mismos.

Tabla 19. Análisis de los motivos contra MiNT. En la tabla se resume el número de transcriptos que posee cada motivo en cada agrupamiento por separado. Se representan gráficamente con barras rosadas los porcentajes del número de motivos con respecto al total de aciertos. Se encuentra entre paréntesis el número de motivos.

	Agrup_2_5UTR		Agrup_3_5UTR		Agrup_4_5UTR	Agrup_5_5UTR		Agrup_6_5UTR	
	motivo 1	motivo 2	motivo 2	motivo 3	motivo 2	motivo 2	motivo 3	motivo 2	motivo 3
Agrup 1	7% (2)	9% (3)	7% (4)	11% (5)	4% (1)	3% (1)	4% (2)	5% (2)	0% (0)
Agrup 2	53% (16)	71% (25)	5% (3)	4% (2)	21% (5)	3% (1)	4% (2)	5% (2)	13% (4)
Agrup 3	0% (0)	0% (0)	63% (37)	53% (24)	8% (2)	0% (0)	6% (3)	12% (5)	0% (0)
Agrup 4	13% (4)	9% (3)	12% (7)	2% (1)	58% (14)	7% (2)	4% (2)	9% (4)	16% (5)
Agrup 5	17% (5)	6% (2)	2% (1)	18% (8)	0% (0)	74% (22)	75% (36)	7% (3)	3% (1)
Agrup 6	10% (3)	6% (2)	12% (7)	11% (5)	8% (2)	10% (3)	6% (3)	63% (27)	68% (21)
Total	30	35	59	45	24	29	48	43	31

Del estudio del perfil de expresión diferencial a lo largo del ciclo celular del estadio epimastigota del parásito de los genes de MiNT, MiNT-DiExTra-CC, se definieron seis agrupamientos. Las categorías enriquecidas en cada agrupamiento corroboran el perfil de expresión génica esperado para cada etapa del ciclo celular. De interés particular encontramos la señal que

denominamos $MiLTre_{TRRM3}$. Esta señal, es significativamente similar a la reconocida por la proteína TRRM3 de *T. b. gambiense*. La proteína codificada por el gen ortólogo en *T. brucei* se identificó en la fracción enriquecida en membrana mitocondrial. Por esto, y habiendo encontrado el gen ortólogo en *T. cruzi*, proponemos que la proteína TRRM3 podría localizar los transcritos de MiNT reconociendo el motivo antes mencionado. Por otra parte, se observa que en este caso los motivos de estructura son altamente específicos de los agrupamientos que los generaron.

MiNT-intra

Se generaron distintos subgrupos en base al destino intra-organelo que presentan las proteínas de MiNT. En este caso, se llevó a cabo una clasificación con respecto a la ontología mediante el cual se identificaron las proteínas que formarían parte de la membrana externa (**MemEx**), del espacio intermembrana (**EspIM**), de la membrana interna (**MemIn**) o que se alojaran en la matriz mitocondrial (**Matrix**). En la Tabla 20 se muestran aquellos genes que se encontraron mediante la discriminación por ontología empleando la herramienta de análisis disponible en la base de datos *tritypdb.org*.

Tabla 20. Distribución intramitocondrial. En la tabla se muestran los transcritos que se encontraron agrupados bajo los términos de ontología que se indican en los encabezados.

GO:0005741 : mitochondrial outer membrane		GO:0005743 : mitochondrial inner membrane	
ID	Producto	ID	Producto
TcCLB.507711.300	Archaic translocase of outer membrane 11 kDa subunit, putative (ATOM11)	TcCLB.511867.140	Acyl carrier protein, mitochondrial, putative (ACP)
TcCLB.504057.145	cytochrome b5-like Heme/Steroid binding domain containing protein, putative (POMP29)	TcCLB.506399.89	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (fragment) (SDH1-2)
TcCLB.508347.70	Engulfment and cell motility domain-containing protein (POMP3)	TcCLB.509233.180	ATP synthase F1, beta subunit, putative (ATPB)
TcCLB.503525.40	FAD dependent oxidoreductase, putative (POMP20)	TcCLB.510293.54	hypothetical protein, conserved
TcCLB.511733.40	FUN14 family, putative (POMP32)	TcCLB.510173.70	MIX protein, putative
TcCLB.511071.100	Glycerophosphoryl diester phosphodiesterase family, putative (POMP42)	TcCLB.509769.60	succinate dehydrogenase iron-sulfur subunit (SDH2N)
TcCLB.506399.70	hypothetical protein, conserved	TcCLB.504089.40	long-chain-fatty-acid-CoA ligase, putative
TcCLB.506399.80	hypothetical protein, conserved	TcCLB.510609.40	ATP synthase F1, alpha subunit, putative (pseudogene)
TcCLB.505807.150	Leucine-rich repeat, ribonuclease inhibitor subtype domain-containing protein (POMP30)	TcCLB.509267.60	NADH-ubiquinone oxidoreductase complex I subunit, putative (TAP151)
TcCLB.506467.29	methyltransferase domain containing protein, putative (POMP1)	TcCLB.510149.150	NADH-ubiquinone oxidoreductase complex I subunit, putative
TcCLB.506885.340	Mitochondrial import receptor subunit ATOM40, putative (ATOM40)	TcCLB.503745.30	ascorbate peroxidase (APX)
TcCLB.511687.10	Mitochondrial outer membrane protein porin, putative (VDAC)	TcCLB.510759.120	rieske iron-sulfur protein, mitochondrial precursor, putative (RISP)
TcCLB.509141.40	Mitochondrial outer membrane protein porin, putative (VDAC)	TcCLB.506211.160	Mitochondrial ADP/ATP carrier protein 5, putative (MCP5)
TcCLB.509129.10	Mitochondrial outer membrane protein porin, putative (VDAC)	TcCLB.437545.10	3-oxo-5-alpha-steroid 4-dehydrogenase, putative
TcCLB.507093.260	multidrug resistance protein, mitochondrial, putative (ABCT)	TcCLB.506529.360	cytochrome oxidase subunit IV, putative (COXIV)
TcCLB.504057.40	NAD dependent epimerase/dehydratase family, putative (POMP27)	TcCLB.506649.50	cytochrome c oxidase assembly factor, putative
TcCLB.508083.10	Present in the outer mitochondrial membrane proteome 10 (POMP10)	TcCLB.511419.60	NADH-ubiquinone oxidoreductase complex I subunit, putative
TcCLB.504153.250	Present in the outer mitochondrial membrane proteome 11 (POMP11)		
TcCLB.510431.180	Present in the outer mitochondrial membrane proteome 13 (POMP13)		
TcCLB.508647.110	Present in the outer mitochondrial membrane proteome 14 (ATOM14)		
TcCLB.508813.60	Present in the outer mitochondrial membrane proteome 16 (ATOM46)		
TcCLB.511559.10	Present in the outer mitochondrial membrane proteome 17 (fragment)		
TcCLB.511557.50	Present in the outer mitochondrial membrane proteome 17 (POMP17)		
TcCLB.506529.530	Present in the outer mitochondrial membrane proteome 18 (POMP18)		
TcCLB.508785.10	Present in the outer mitochondrial membrane proteome 19 (POMP19)		
TcCLB.508347.80	Present in the outer mitochondrial membrane proteome 2 (POMP2)		
TcCLB.503975.100	Present in the outer mitochondrial membrane proteome 22 (POMP22)		
TcCLB.506697.80	Present in the outer mitochondrial membrane proteome 24 (POMP24)		
TcCLB.510317.7	Present in the outer mitochondrial membrane proteome 24 (POMP24)		
TcCLB.504431.30	Present in the outer mitochondrial membrane proteome 25 (POMP25)		
TcCLB.504057.80	Present in the outer mitochondrial membrane proteome 28		
TcCLB.506941.220	Present in the outer mitochondrial membrane proteome 31 (POMP31)		
TcCLB.510525.130	Present in the outer mitochondrial membrane proteome 33 (POMP33)		
TcCLB.507817.60	Present in the outer mitochondrial membrane proteome 34 (POMP34)		
TcCLB.509153.80	Present in the outer mitochondrial membrane proteome 36 (POMP36)		
TcCLB.508383.10	Present in the outer mitochondrial membrane proteome 37 (fragment)		
TcCLB.510611.79	Present in the outer mitochondrial membrane proteome 37 (fragment)		
TcCLB.508973.90	Present in the outer mitochondrial membrane proteome 38 (POMP38)		
TcCLB.506865.50	Present in the outer mitochondrial membrane proteome 40 (ATOM12)		
TcCLB.511217.20	Present in the outer mitochondrial membrane proteome 41 (POMP41)		
TcCLB.510429.40	Present in the outer mitochondrial membrane proteome 5 (POMP5)		
TcCLB.506147.80	Present in the outer mitochondrial membrane proteome 6 (POMP6)		
TcCLB.510741.150	Present in the outer mitochondrial membrane proteome 7 (POMP7)		
TcCLB.503557.40	Present in the outer mitochondrial membrane proteome 8 (POMP8)		
TcCLB.511803.40	Present in the outer mitochondrial membrane proteome 9 (ATOM69)		
TcCLB.509455.114	Protein of unknown function (DUF1295), putative (POMP23)		
TcCLB.508369.20	squalene synthase, putative		
TcCLB.509379.20	Tab55, putative		
TcCLB.504225.20	voltage-dependent anion-selective channel, putative (fragment) (VDAC)		
TcCLB.511687.19	voltage-dependent anion-selective channel, putative (fragment) (VDAC)		
TcCLB.509149.20	zinc finger protein family member, putative (POMP35)		
		GO:0005758 : mitochondrial intermembrane space	
		ID	Producto
		TcCLB.506691.64	cytochrome c oxidase copper chaperone, putative
		TcCLB.430605.40	ERV/ALR sulfhydryl oxidase domain-containing protein (POMP4)
		TcCLB.505843.49	ERV/ALR sulfhydryl oxidase domain-containing protein (POMP4)
		TcCLB.510089.120	hypothetical protein, conserved
		TcCLB.503789.34	hypothetical protein, conserved
		TcCLB.510507.29	hypothetical protein, conserved
		TcCLB.509775.34	hypothetical protein, conserved
		GO:0005759 : mitochondrial matrix	
		ID	Producto
		TcCLB.511867.140	Acyl carrier protein, mitochondrial, putative (ACP)
		TcCLB.506635.100	ATP dependent DEAD-box helicase, putative (RH)
		TcCLB.506739.50	ATP-dependent protease ATPase subunit HsIU1, putative (HsIU1)
		TcCLB.510719.260	ATP-dependent protease subunit HsIV, putative (HsIV)
		TcCLB.508127.9	hypothetical protein, conserved
		TcCLB.506495.6	hypothetical protein, conserved
		TcCLB.510615.10	hypothetical protein, conserved
		TcCLB.511589.220	hypothetical protein, conserved
		TcCLB.508731.40	hypothetical protein, conserved
		TcCLB.508323.120	hypothetical protein, conserved
		TcCLB.511423.100	kinetoplast ribosomal PPR-repeat containing protein 3, putative
		TcCLB.511529.80	kinetoplast-associated protein 3 (KAP3)
		TcCLB.503831.40	mitochondrial DNA primase, putative
		TcCLB.507603.50	Mitochondrial Ribosomal Protein of Small Subunit 29, putative
		TcCLB.507547.30	Mitochondrial-processing peptidase subunit alpha (MPPA)
		TcCLB.511585.80	mitochondrial-processing peptidase subunit beta, putative (MPPB)
		TcCLB.509965.290	p22 protein precursor, putative
		TcCLB.506945.170	thiosulfate sulfurtransferase, mitochondrial, putative

Búsqueda de LiMs en las regiones UTRs

En este caso se emplearon los cuatro subgrupos de transcritos definidos en la sección anterior. En todos los casos se buscaron motivos enriquecidos con respecto a NoMiNToma. Los motivos encontrados se reportan en la

Tabla 21. Para la región 3' UTR del subgrupo EspIM y la región 5'UTR del subgrupo Matrix, no se encontró ningún motivo enriquecido significativamente. Empleando el programa Tom-Tom con la base de datos de (Ray et al. 2013), en todos los casos se buscaron semejanzas con motivos reportados, reconocidos por factores proteicos. Únicamente 3 de estos motivos se asemejan a los presentes en otros organismos, para los que además se conoce la proteína involucrada en el reconocimiento. De estos, el motivo 5 "AGRAA" presente en Tabla 21 MemEx 3UTR, es el mismo motivo que se encontró enriquecido en la región 3'UTR del agrupamiento 1 del ciclo celular y en el ciclo de vida. Como ya se observó en aquella instancia, el gen que codifica para la proteína que reconoce el motivo en *Homo sapiens* no presentaría un ortólogo en *T. cruzi*. Por otro lado los motivos 1 "CTTTGTGA" y 3 "WGGCGACG" presentes en Tabla 21 MemEx 5UTR, podrían ser reconocidos por las proteínas SUP (*SUPressor*, NP_508674.1) de *Caenorhabditis elegans* y SARF7 (*serine/arginine-rich splicing factor 7 isoform 1*, NP_001026854.1) de *Homo Sapiens*, respectivamente. Sin embargo ninguno de los genes que codifican para estas proteínas presentan ortólogo en *T. cruzi*.

Tabla 21. Motivos lineales encontrados para cada agrupamiento de MiNT-intra. En la tabla se resumen los motivos encontrados

<p>EspIM 5UTR</p> <p>Motivo 1 bits 1 2 3 4 5 6 7 8 T G G A A C A T</p> <p>Motivo 2 bits 1 2 3 4 5 6 7 T G T - T A G</p>	<p>Matrix 3UTR</p> <p>Motivo 1 bits 1 2 3 4 5 6 7 A T C T A T</p>
<p>MemEx 5UTR</p> <p>Motivo 1 bits 1 2 3 4 5 6 7 8 C T T T G T G A</p> <p>Motivo 2 bits 1 2 3 4 5 6 7 8 T C - C T A</p> <p>Motivo 3 bits 1 2 3 4 5 6 7 8 G G C G A C G</p> <p>Motivo 4 bits 1 2 3 4 5 6 7 8 T T G C A A T G</p> <p>Motivo 5 bits 1 2 3 4 5 6 7 8 G A A G C T G G</p> <p>Motivo 6 bits 1 2 3 4 5 6 7 8 C A A G T C C C</p> <p>Motivo 7 bits 1 2 3 4 5 6 7 8 T T G A G C C C</p> <p>Motivo 8 bits 1 2 3 4 5 6 7 8 T - A T A T T T</p>	<p>MemEx 3UTR</p> <p>Motivo 1 bits 1 2 3 4 5 6 7 8 A - G A T C G A</p> <p>Motivo 2 bits 1 2 3 4 5 6 7 8 G G - A C G T C</p> <p>Motivo 3 bits 1 2 3 4 5 6 7 8 G A C C T T A G</p> <p>Motivo 4 bits 1 2 3 4 5 6 7 8 T - A C A G G A</p> <p>Motivo 5 bits 1 2 3 4 5 A G - A A</p>
<p>MemIn 5UTR</p> <p>Motivo 1 bits 1 2 3 4 5 6 7 T A - G C A A</p> <p>Motivo 2 bits 1 2 3 4 5 6 7 A - C G T T T</p>	<p>MemIn 3UTR</p> <p>Motivo 1 bits 1 2 3 4 5 6 7 8 T - T G T C A A</p>

En general cuando se analizan contra **MiNT** se observa que de los 11 motivos presentes en la región 5'UTR, en su mayoría (TGGAACAT (EspIM); AGGCGACG, TTGAGCCC, CAAGTCCC, GAAGCTGG, TTGCAATG (MemEx); TAAGCAA (MemIn)) las señales se encuentran mayormente en la región codificante. De estas, en CAAGTCCC, GAAGCTGG, TTGCAATG (MemEx) no están en fase con el marco de lectura, mientras que en el resto de los casos más del 85% de los motivos se encontraron en fase. Por otro lado el motivo

TGATATTT (MemEx) es el único que parece estar diferencialmente ubicado en la región 5'UTR propiamente dicha. El resto de los motivos no presentarían distribución preferencial. Con respecto a los motivos ubicados en la región 3'UTR no se evidencia una localización preferencial de los mismos, encontrándose distribuidos casi uniformemente a lo largo de la región no traducida.

Si bien los transcritos que poseen cada señal se encuentran enriquecidos en el términos asociados al componente celular que las deriva, se encuentran varios genes que no fue posible definir su **ubicación intramitocondrial** al menos mediante términos de ontología.

Búsqueda de StMs en las regiones no traducidas de mensajeros.

Siguiendo la misma estrategia que para el resto de las subdivisiones, se buscaron señales a nivel de estructura que pudieran estar enriquecidas con respecto al NoMiNToma. Las estructuras encontradas se analizaron en cuanto a cobertura, sensibilidad y especificidad así como en cuanto a la probabilidad de formación de las mismas calculando la variación de energía libre de Gibbs para la estructura predicha. En la Tabla 22 se resumen los parámetros obtenidos para las estas estructuras.

En este caso, únicamente 3 de las estructuras cumplieron nuestro criterio de selección. Las tres estructuras se muestran en la Figura 41. Si bien se analizó la posición en la que se encuentran a lo largo de la región 5' no es posible concluir si se localizan en alguna posición preferencial dado el bajo número de transcritos que aportan en cada caso.

Tabla 22. Búsqueda de motivos estructurales en los agrupamientos según el destino intramitocondrial de las proteínas codificadas.

Motivo	Genes aporte	Largo (nt)	DE Largo (nt)	Cobertura	Sensibilidad	Especificidad	ΔG^* kcal/mol
m1_3UTR_MiNT_MemIn	16	31	8	94%	75%	91%	3,20
m2_3UTR_MiNT_MemIn*	15	37	10	88%	67%	85%	-16,00
m3_3UTR_MiNT_MemIn*	15	49	6	88%	47%	79%	N.C.
m1_5UTR_MiNT_MemIn	16	32	3	94%	81%	98%	-9,60
m2_5UTR_MiNT_MemIn	13	38	7	76%	100%	97%	0,00
m3_5UTR_MiNT_MemIn	13	34	15	76%	85%	98%	-13,60
m1_3UTR_MiNT_MemEx	49	33	4	96%	59%	97%	N.C.
m2_3UTR_MiNT_MemEx	37	9	16	73%	49%	90%	-7,50
m3_3UTR_MiNT_MemEx	44	34	22	86%	61%	78%	-9,44
m1_5UTR_MiNT_MemEx	11	29	7	22%	91%	94%	-5,20
m2_5UTR_MiNT_MemEx	42	32	9	82%	79%	95%	-2,20
m3_5UTR_MiNT_MemEx	5	39	5	10%	80%	98%	-15,10
m1_3UTR_MiNT_EspIM	5	33	5	71%	80%	86%	-23,80
m2_3UTR_MiNT_EspIM	5	33	3	71%	100%	81%	-5,90
m3_3UTR_MiNT_EspIM	5	35	5	71%	120%	81%	-13,20
m1_5UTR_MiNT_EspIM	5	33	2	71%	100%	85%	-8,90
m2_5UTR_MiNT_EspIM	5	31	3	71%	80%	85%	-5,60
m3_5UTR_MiNT_EspIM	4	31	6	57%	125%	89%	-14,00
m1_3UTR_MiNT_Matrix	16	34	5	89%	94%	90%	-3,20
m2_3UTR_MiNT_Matrix	17	43	5	94%	94%	79%	-27,70
m3_3UTR_MiNT_Matrix	12	46	5	67%	100%	81%	-30,90
m1_5UTR_MiNT_Matrix	17	30	2	94%	94%	87%	-7,90
m2_5UTR_MiNT_Matrix	14	39	5	78%	93%	91%	-25,20
m3_5UTR_MiNT_Matrix	11	40	4	61%	82%	88%	-15,50

Se señala con * los motivos que no superaron el criterio de especificidad/sensibilidad propuesto para este tipo de búsquedas. Los círculos verdes señalan aquellos agrupamientos cuya especificidad supera el 90% mientras que el amarillo corresponde a aquellos que están entre el 80 y 90% y el rojo a los menores de 80% de especificidad. Se señala además en verde aquellas estructuras cuyo ΔG de formación a 28°C es menor al promedio de ΔG encontrado. Los valores señalados como N.C. implican que no fue posible calcularlos con el programa empleado.

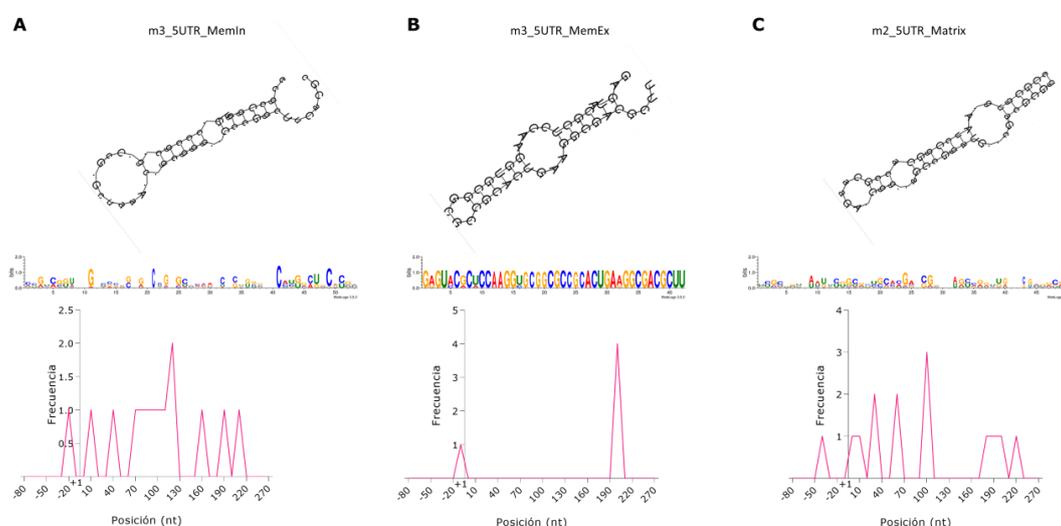


Figura 41. Estructuras que superaron los criterios propuestos y su distribución a lo largo de la región 5'UTR. En el panel superior se observan las 3 estructuras que se encontraron significativamente enriquecidas en las búsquedas, modeladas mediante RNAfold, así como el logo de la secuencia consenso para cada una de ellas. En cada caso se señala el agrupamiento del cual provienen, la región específica, en este caso 5UTR, y el número del motivo (1,2 o 3). En el panel inferior se observa la distribución de las estructuras que se muestran en el panel superior.

De todas formas, por estar presente mayormente en la región de la CDS de los transcriptos que los contienen en la mayoría de los casos el motivo efectivamente se encontraría en el ARNm. Sin embargo, en dos casos, uno para m3_5UTR_MemEx y otro para m2_5UTR_Matrix, en los cuales el motivo se encuentra propiamente en la región 5'UTR, no contamos con evidencia del largo de esta región no traducida.

Al igual que sucede en el caso de los motivos lineales, la búsqueda de estas señales a lo largo de la región 5'UTR de los transcriptos de MiNT, arrojan resultados, aunque confiables, escasos. La baja sensibilidad del método, así como la anotación deficiente del genoma de *T. cruzi* no permiten llegar a una conclusión precisa. Sin embargo, la búsqueda confirma la **especificidad** del motivo en cuanto al subgrupo que la define. Lamentablemente en la mayoría de los nuevos transcriptos identificados para cada señal no se conoce el **destino intramitocondrial** de las proteínas codificadas.

Si bien escapa al objetivo de este trabajo, esta estrategia de búsqueda empleando los motivos específicos generados, podría emplearse para predecir el destino intramitocondrial de las proteínas codificadas.

Para el estudio del destino intramitocondrial de las proteínas codificadas por los genes de MiNT (MiNT-intra) se definieron 4 agrupamientos. Se encontró el motivo "AGRAA" enriquecido en la región 3'UTR de los transcriptos de MemEx, que ya fue incluido en la definición del motivo consenso MiLTreSTRM3. Este análisis permitió además establecer motivos específicos no sólo con respecto a NoMiNToma sino también entre los agrupamientos

Estudio de casos particulares

Para continuar con el análisis se estudiaron en particular subgrupos de genes específicos:

- Proteínas Ribosomales (PRiMito)
- Aminoacil-ARNt-sintetasas (AAtSs)

PRiMito

Contrariamente a lo que sucede con respecto a las proteínas ribosomales asociadas a ribosomas citoplasmáticos, las proteínas ribosomales mitocondriales no se encuentran definidas en *T. cruzi*. Por ello se definió la base de datos **PRiMito** (Proteínas Ribosomales Mitocondriales) a partir de la búsqueda de ortólogos a las proteínas ribosomales mitocondriales halladas con alto nivel de confianza en *T. brucei* (Zíková et al. 2008). De las 77 proteínas de *T. brucei* de la subunidad mayor (SuMa) se encontraron en *T. cruzi* genes ortólogos para 70 de ellas (Tabla 23). Para el caso de las 56 proteínas que se corresponden con la subunidad menor (SuMe) de en *T. cruzi* se encontraron 47 genes ortólogos (Tabla 24).

Tabla 23. Búsqueda de proteínas de la subunidad mayor de los ribosomas mitocondriales de *T. cruzi*. Se muestran los genes que codifican para las proteínas de la subunidad mayor de ribosomas mitocondriales en *T. brucei* y sus ortólogos en *T. cruzi*. *señala el nombre propuesto para las proteínas de *T. brucei* en (Zíková et al. 2008)

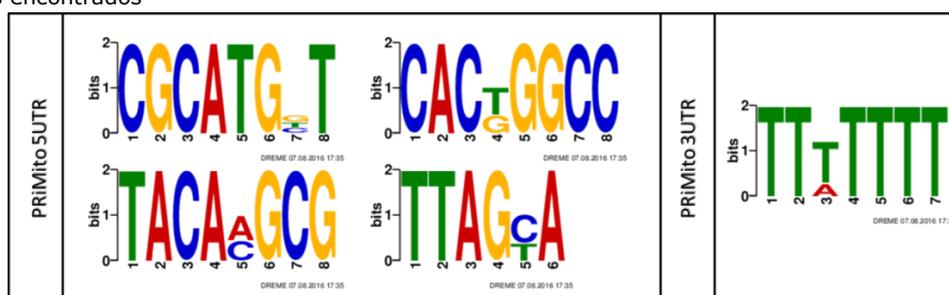
GeneID	Descripción del producto	Nombre sugerido*	ortólogo Gene ID	Descripción del producto
Tb927.7.4710	39S mitochondrial ribosomal protein L46, putative	TbMRPL46	TcCLB.507521.120	hypothetical protein, conserved
Tb927.4.1070	50S ribosomal protein L13, putative	TbMRPL13	TcCLB.506405.110	ribosomal protein L13, putative
Tb927.4.930	50S ribosomal protein L14, putative	TbMRPL14	TcCLB.469435.4	ribosomal protein L14, putative
Tb927.7.3960	50S ribosomal protein L16, putative	TbMRPL16	TcCLB.506657.79	50S ribosomal protein L16, putative
Tb927.8.5860	50S ribosomal protein L17, putative	TbMRPL17	TcCLB.506743.30	50S ribosomal protein L17, putative
Tb927.8.2760	50S ribosome-binding GTPase, putative	---	TcCLB.510509.20	hypothetical protein, conserved
Tb927.7.2630	50S ribosome-binding GTPase, putative	---	TcCLB.506859.160	hypothetical protein, conserved
Tb927.11.16990	50S ribosome-binding GTPase, putative	---	TcCLB.511537.60	hypothetical protein, conserved
Tb927.7.4550	60S ribosomal protein-like	TbMRPL7/12	TcCLB.506605.150	60S ribosomal protein, putative
Tb927.7.6800	Alpha/beta hydrolase family, putative	---	TcCLB.503967.10	hypothetical protein, conserved
Tb927.4.2720	ATP dependent DEAD-box helicase (RH)	---	TcCLB.506635.100	ATP dependent DEAD-box helicase, putative (RH)
Tb927.6.2480	chaperone protein DNAJ, putative	---	TcCLB.509965.229	heat shock protein (pseudogene), putative
Tb927.11.9450	cyclophilin type peptidyl-prolyl cis-trans isomerase, putative (PPIase)	---	TcCLB.511589.50	cyclophilin-like protein, putative
Tb927.7.3430	cyclophilin-type peptidyl-prolyl cis-trans isomerase, putative (PPIase)	---	TcCLB.508323.94	cyclophilin, putative
Tb927.9.8290	hypothetical protein, conserved	---	TcCLB.504255.10	hypothetical protein, conserved
Tb927.9.3640	hypothetical protein, conserved	---	TcCLB.506895.30	hypothetical protein, conserved
Tb927.8.3300	hypothetical protein, conserved	---	TcCLB.509769.120	hypothetical protein, conserved
Tb927.8.3170	hypothetical protein, conserved	---	TcCLB.510517.30	hypothetical protein, conserved
Tb927.7.7010	hypothetical protein, conserved	---	TcCLB.508387.40	hypothetical protein, conserved
Tb927.7.3510	hypothetical protein, conserved	---	TcCLB.508265.40	hypothetical protein, conserved
Tb927.7.3460	hypothetical protein, conserved	---	TcCLB.508323.110	hypothetical protein, conserved
Tb927.7.2990	hypothetical protein, conserved	---	TcCLB.510529.10	hypothetical protein, conserved
Tb927.6.4200	hypothetical protein, conserved	---	TcCLB.511733.20	hypothetical protein, conserved
Tb927.6.4080	hypothetical protein, conserved	---	TcCLB.511735.20	hypothetical protein, conserved
Tb927.6.4040	hypothetical protein, conserved	TbMRPL28	TcCLB.511735.50	hypothetical protein, conserved
Tb927.6.3930	hypothetical protein, conserved	---	TcCLB.511739.20	hypothetical protein, conserved
Tb927.6.3600	hypothetical protein, conserved	---	TcCLB.506941.120	hypothetical protein, conserved
Tb927.6.1440	hypothetical protein, conserved	---	TcCLB.505807.50	hypothetical protein, conserved
Tb927.5.4120	hypothetical protein, conserved	---	TcCLB.507941.80	hypothetical protein, conserved
Tb927.5.3980	hypothetical protein, conserved	TbMRPL15	TcCLB.506947.100	hypothetical protein, conserved
Tb927.5.3870	hypothetical protein, conserved	---	TcCLB.510603.150	hypothetical protein, conserved
Tb927.5.3410	hypothetical protein, conserved	TbMRPL9	TcCLB.509777.80	hypothetical protein, conserved
Tb927.5.3110	hypothetical protein, conserved	TbMRPL49	TcCLB.503955.60	hypothetical protein, conserved
Tb927.5.2070	hypothetical protein, conserved	---	TcCLB.506519.120	hypothetical protein, conserved
Tb927.4.4610	hypothetical protein, conserved	---	TcCLB.407477.60	hypothetical protein, conserved
Tb927.4.1810	hypothetical protein, conserved	TbMRPL33	TcCLB.510879.90	hypothetical protein, conserved
Tb927.3.820	hypothetical protein, conserved	---	TcCLB.511867.100	hypothetical protein, conserved
Tb927.11.9830	hypothetical protein, conserved	---	TcCLB.507163.90	hypothetical protein, conserved
Tb927.11.870	hypothetical protein, conserved	---	TcCLB.506925.280	hypothetical protein, conserved
Tb927.11.5990	hypothetical protein, conserved	---	TcCLB.504149.140	hypothetical protein, conserved
Tb927.11.5880	hypothetical protein, conserved	---	TcCLB.508515.80	hypothetical protein, conserved
Tb927.11.5530	hypothetical protein, conserved	---	TcCLB.508647.190	hypothetical protein, conserved
Tb927.11.4650	hypothetical protein, conserved	---	TcCLB.506227.110	hypothetical protein, conserved
Tb927.11.4650	hypothetical protein, conserved	---	TcCLB.506227.120	hypothetical protein, conserved
Tb927.11.3640	hypothetical protein, conserved	---	TcCLB.511051.20	hypothetical protein, conserved
Tb927.11.1630	hypothetical protein, conserved	---	TcCLB.506367.140	hypothetical protein, conserved
Tb927.11.15500	hypothetical protein, conserved	---	TcCLB.509269.10	hypothetical protein, conserved
Tb927.11.11630	hypothetical protein, conserved	---	TcCLB.509205.70	hypothetical protein, conserved
Tb927.11.10570	hypothetical protein, conserved	---	TcCLB.510285.110	hypothetical protein, conserved
Tb927.11.10170	hypothetical protein, conserved	---	TcCLB.510719.170	hypothetical protein, conserved
Tb927.11.10080	hypothetical protein, conserved	---	TcCLB.510719.74	hypothetical protein, conserved
Tb927.11.10050	hypothetical protein, conserved	---	TcCLB.510719.50	hypothetical protein, conserved
Tb927.10.12050	hypothetical protein, conserved	---	TcCLB.510123.10	hypothetical protein, conserved
Tb927.10.11350	hypothetical protein, conserved	---	TcCLB.507709.110	hypothetical protein, conserved
Tb927.10.11050	hypothetical protein, conserved	---	TcCLB.507711.280	hypothetical protein, conserved
Tb927.1.1160	kinetoplast ribosomal PPR-repeat containing protein 3 (KRIPP3)	---	TcCLB.511423.100	kinetoplast ribosomal PPR-repeat containing protein 3, putative (KRIPP3)
Tb927.9.7170	Mitochondrial 39-S ribosomal protein L47 (MRP-L47), putative	---	TcCLB.508127.9	hypothetical protein, conserved
Tb927.4.4600	Mitochondrial ribosomal protein L51 / S25 / C1-B8 domain containing protein, putative	TbMRPL43	TcCLB.407477.69	hypothetical protein, conserved (fragment)
Tb927.10.380	mitochondrial RNA binding complex 1 subunit (PPR5)	---	TcCLB.506869.40	mitochondrial RNA binding complex 1 subunit, putative
Tb927.11.14980	phosphatidylethanolamine-binding protein, putative	---	TcCLB.511715.80	hypothetical protein, conserved
Tb927.9.3350	pseudouridylyl synthase, putative	---	TcCLB.511577.90	pseudouridylyl synthase-like protein, putative
Tb927.7.1640	ras-like small GTPase, putative (TbEAR)	---	TcCLB.508207.200	small GTP-binding protein, putative
Tb927.2.4890	ribosomal protein L11, putative	TbMRPL11	TcCLB.511621.200	ribosomal protein L11, putative
Tb927.2.4890	ribosomal protein L11, putative	TbMRPL11	TcCLB.511621.70	ribosomal protein L11, putative
Tb927.7.4140	ribosomal protein L21, putative	TbMRPL21	TcCLB.510989.40	hypothetical protein, conserved
Tb927.7.2760	ribosomal protein L22p/L17e, putative	TbMRPL22	TcCLB.510525.20	hypothetical protein, conserved
Tb927.3.5610	ribosomal protein L3 mitochondrial, putative	TbMRPL3	TcCLB.507795.100	mitochondrial ribosomal protein L3, putative
Tb927.11.6000	ribosomal protein L4/L1 family, putative (MRPL4)	---	TcCLB.504149.130	hypothetical protein, conserved
Tb927.9.12850	SpoU rRNA Methylase family, putative	---	TcCLB.510745.20	hypothetical protein, conserved
Tb927.10.6090	tRNA pseudouridine synthase A, putative	---	TcCLB.507639.50	tRNA pseudouridine synthase A, putative

Tabla 24. Búsqueda de proteínas de la subunidad menor de los ribosomas mitocondriales de *T. cruzi*. Se muestran los genes que codifican para las proteínas de la subunidad menor de ribosomas mitocondriales en *T. brucei* y sus ortólogos en *T. cruzi*. * señala el nombre propuesto para las proteínas de *T. brucei* en (Ziková et al. 2008).

GeneID	Descripción del producto	Nombre sugerido*	ortólogo Gene ID	Descripción del producto
Tb927.10.13300	30S ribosomal protein S8, putative	---	TcCLB.506357.110	hypothetical protein, conserved
Tb927.10.13820	Domain of unknown function (DUF4379), putative	---	TcCLB.507993.33	hypothetical protein, conserved
Tb927.11.10150	Enoyl-CoA hydratase/isomerase family, putative	---	TcCLB.510719.150	hypothetical protein, conserved
Tb927.9.6510	hypothetical protein, conserved	---	TcCLB.504881.40	hypothetical protein, conserved
Tb927.9.13780	hypothetical protein, conserved	---	TcCLB.510755.80	hypothetical protein, conserved
Tb927.9.11880	hypothetical protein, conserved	---	TcCLB.507011.90	hypothetical protein, conserved
Tb927.8.5280	hypothetical protein, conserved	TbMRPS34	TcCLB.509979.230	hypothetical protein, conserved
Tb927.8.5200	hypothetical protein, conserved	---	TcCLB.511217.90	hypothetical protein, conserved
Tb927.8.4860	hypothetical protein, conserved	---	TcCLB.508837.160	hypothetical protein, conserved
Tb927.8.4550	hypothetical protein, conserved	---	TcCLB.511409.20	hypothetical protein, conserved
Tb927.7.3240	hypothetical protein, conserved	---	TcCLB.510533.100	hypothetical protein, conserved
Tb927.7.3050	hypothetical protein, conserved	---	TcCLB.510529.30	hypothetical protein, conserved
Tb927.7.2620	hypothetical protein, conserved	---	TcCLB.506859.130	hypothetical protein, conserved
Tb927.6.4580	hypothetical protein, conserved	---	TcCLB.511731.70	hypothetical protein, conserved
Tb927.6.4560	hypothetical protein, conserved	---	TcCLB.511731.80	hypothetical protein, conserved
Tb927.6.4080	hypothetical protein, conserved	---	TcCLB.511735.20	hypothetical protein, conserved
Tb927.6.2180	hypothetical protein, conserved	---	TcCLB.507927.90	hypothetical protein, conserved
Tb927.5.4040	hypothetical protein, conserved	---	TcCLB.507941.10	hypothetical protein, conserved
Tb927.5.3640	hypothetical protein, conserved	---	TcCLB.509779.40	hypothetical protein, conserved
Tb927.5.1790	hypothetical protein, conserved	---	TcCLB.505555.30	hypothetical protein, conserved
Tb927.5.1510	hypothetical protein, conserved	---	TcCLB.506739.40	hypothetical protein, conserved
Tb927.3.770	hypothetical protein, conserved	---	TcCLB.511867.40	hypothetical protein, conserved
Tb927.3.5240	hypothetical protein, conserved	---	TcCLB.508075.7	hypothetical protein, conserved
Tb927.3.5240	hypothetical protein, conserved	---	TcCLB.510143.129	hypothetical protein, conserved
Tb927.2.4400	hypothetical protein, conserved	---	TcCLB.509099.80	hypothetical protein, conserved
Tb927.11.7790	hypothetical protein, conserved	---	TcCLB.508813.70	hypothetical protein, conserved
Tb927.11.11470	hypothetical protein, conserved	---	TcCLB.511803.50	hypothetical protein, conserved
Tb927.11.10400	hypothetical protein, conserved	---	TcCLB.510283.100	hypothetical protein, conserved
Tb927.10.3580	hypothetical protein, conserved	---	TcCLB.510311.160	hypothetical protein, conserved
Tb927.10.3250	hypothetical protein, conserved	---	TcCLB.503575.54	hypothetical protein, conserved
Tb927.10.16090	hypothetical protein, conserved	---	TcCLB.511353.20	hypothetical protein, conserved
Tb927.10.11820	hypothetical protein, conserved	---	TcCLB.504009.20	hypothetical protein, conserved
Tb927.10.11260	hypothetical protein, conserved	---	TcCLB.507711.50	hypothetical protein, conserved
Tb927.10.10400	hypothetical protein, conserved	---	TcCLB.504507.10	hypothetical protein, conserved
Tb927.4.3690	Iron/manganese superoxide dismutases, C-terminal domain containing protein, putative	---	TcCLB.504129.30	hypothetical protein, conserved
Tb927.11.15850	kinteoplast poly(A) polymerase complex 1 subunit	---	TcCLB.511545.70	kinteoplast poly(A) polymerase complex 1 subunit, putative
Tb927.6.1250	Mitochondrial Ribosomal Protein of Small Subunit 29 (MRPS29)	TbMRPS29	TcCLB.507603.50	Mitochondrial Ribosomal Protein of Small Subunit 29, putative (MRPS29)
Tb927.11.5060	Mitochondrial small ribosomal subunit Rsm22,	---	TcCLB.507667.20	hypothetical protein, conserved
Tb927.11.2530	Mitochondrial SSU ribosomal protein, putative	---	TcCLB.504153.190	mitochondrial edited mRNA stability factor 1 subunit
Tb927.11.1250	Mitochondrial SSU ribosomal protein, putative	---	TcCLB.511445.130	Mitochondrial SSU ribosomal protein, putative
Tb927.11.5500	Mitochondrial SSU ribosomal protein, putative	---	TcCLB.508647.220	mitochondrial RNA binding protein 1
Tb927.3.970	NAD binding domain of 6-phosphogluconate dehydrogenase/NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase, putative	---	TcCLB.511871.40	hypothetical protein, conserved
Tb927.1.2990	PPR repeat family, putative	---	TcCLB.506363.50	hypothetical protein, conserved
Tb927.10.6300	Ribosomal protein S5, C-terminal domain containing protein, putative	---	TcCLB.507641.150	hypothetical protein, conserved
Tb927.8.3110	ribosomal protein S9/S16, putative	TbMRPS9	TcCLB.510515.110	hypothetical protein, conserved
Tb927.6.4930	thiosulfate sulfurtransferase, mitochondrial,	---	TcCLB.506945.170	thiosulfate sulfurtransferase, mitochondrial, putative
Tb927.9.5280	unspecified product	---	TcCLB.508277.200	hypothetical protein, conserved

Se llevó a cabo la búsqueda de motivos lineales en la base de datos PRiMito tanto a nivel de la región 5' como 3'UTR encontrándose los motivos que se muestran en la Tabla 25.

Tabla 25. Motivos lineales encontrados para la base de datos PRiMito. En la tabla se resumen los motivos encontrados



Empleando el programa Tom-Tom se buscaron motivos similares en la base de datos. Para el motivo "CACKGGCC" que se observó enriquecido en la

región 5'UTR de la base PRiMito, se encontró un motivo similar reconocido por la proteína Vts1p de *Saccharomyces cerevisiae* (NP_015004.3) y *protein Smaug homolog 1 isoform 1* de *Homo sapiens* (NP_056404.4). Los genes que codifican estas proteínas, involucradas en la estimulación de la **deadenilación** mediada por CCR4-NOT-POP2p en *S. cerevisiae* (Rendl et al. 2008) y en la represión de la traducción en *H. sapiens* (Baez & Boccaccio 2005), no presentan ortólogos en *T. cruzi*. Por otro lado, la señal encontrada a nivel de la región 3'UTR se encontró significativamente enriquecida en los transcritos correspondientes al agrupamiento 2 de MiNT-DiExTra-CC. Estos transcritos presentan la peculiaridad de verse sobrepresados en las fases G1 y G2 del ciclo en concordancia con la demanda celular de **proteínas ribosomales**.

Con respecto a los motivos estructurales que se encontraron para las regiones no traducidas de los transcritos de PRiMito únicamente uno de los 6 motivos analizados (Tabla 26) (3 de la región 5' y 3 de la región 3') cumplió con los criterios establecidos.

Tabla 26. Búsqueda de motivos estructurales en los transcritos de PRiMito con respecto a NoMiNToma.

Motivo	Genes aporte	Largo (nt)	DE Largo (nt)	Cobertura	Sensibilidad	Especificidad	ΔG^* kcal/mol
m1_3UTR_PRiMito*	108	87	46	93%	62%	56%	0,00
m2_3UTR_PRiMito	115	29	3	99%	25%	89%	-2,80
m3_3UTR_PRiMito	102	38	6	88%	15%	94%	0,00
m1_5UTR_PRiMito	113	30	11	97%	30%	98%	N.C.
m2_5UTR_PRiMito	48	37	6	41%	58%	98%	1,00
m3_5UTR_PRiMito	46	38	7	40%	70%	98%	-18,80

Se señala con * los motivos que no superaron el criterio de especificidad/sensibilidad propuesto para este tipo de búsquedas. Los círculos verdes señalan aquellos agrupamientos cuya especificidad supera el 90% mientras que el amarillo corresponde a aquellos que están entre el 80 y 90% y el rojo a los menores de 80% de especificidad. Se señala además en verde aquellas estructuras cuyo ΔG de formación a 28°C es menor al promedio de ΔG encontrado. Los valores señalados como N.C. implican que no fue posible calcularlos con el programa empleado.

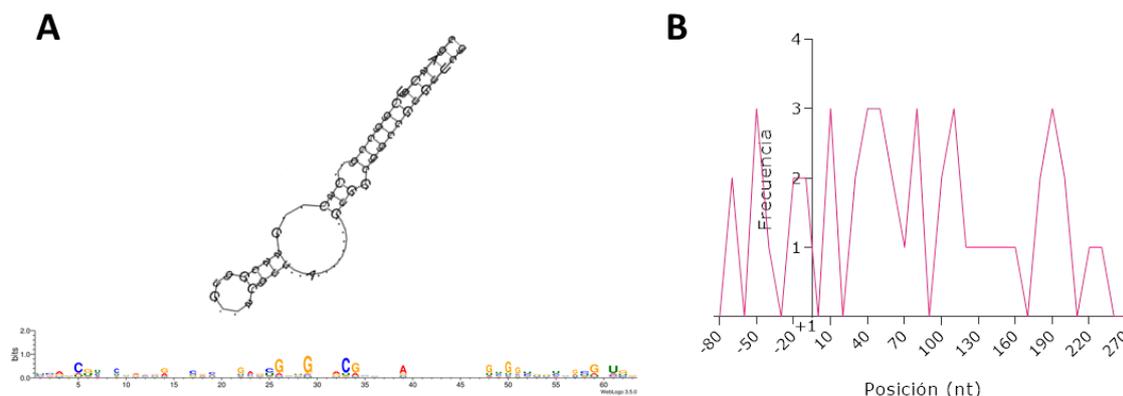


Figura 42. Estructura que superó los criterios propuestos y su distribución a lo largo de la región 5'UTR. En A) se observa la estructura que se encontró significativamente enriquecida en las búsquedas, modelada mediante el algoritmo RNAFold, así como el logo de su secuencia consenso. En B) se observa la distribución de la estructura a lo largo de la región 5'UTR que se muestran en el panel superior.

De las 46 transcritos que poseen este motivo, 25, se encuentran dentro de los 46 pertenecientes a PRiMito-SuMa (transcritos que codifican proteínas de la subunidad mayor del ribosoma mitocondrial). El 88% (22/25) de los transcritos que poseen el motivo se encuentran anotados como producto hipotético conservado. Los restantes transcritos pertenecen a PRiMito-SuMe (transcritos que codifican proteínas de la subunidad menor del ribosoma mitocondrial). De igual manera el 81% (17/21) de los transcritos su producto se encuentra anotado como **hipotético conservado**. Esto no llama la atención para estos subgrupos de transcritos, donde el porcentaje de productos hipotéticos es de 75% (53/70) en el caso de PRiMito-SuMa y 88% en el caso de PRiMito-SuMe.

Cuando el motivo se compara contra MiNT, se encuentran en total 136 transcritos que lo poseen para los cuales en 34 están presentes y para los 102 restantes no se cuenta con el dato de la longitud de la 5'UTR real. En este caso, 18 transcritos corresponden a la PRiMito-SuMa y 21 a PRiMito-SuMe. Las diferencias encontradas entre estos datos y los presentados para la cobertura se debe a la baja sensibilidad del método de búsqueda. Por otro lado, los restantes transcritos que poseen esta señal se encuentran vinculados mayormente a la aminoacilación de ARNt y al metabolismo del malato. También es importante señalar que habiendo repetido la búsqueda del motivo esta vez, únicamente contra las **proteínas ribosomales citoplasmáticas**, sólo para 16 de las 132 se encontró el motivo con un $p\text{-value} < 1$.

En este caso se definió la base de datos de los genes que codifican para las proteínas de ribosomas mitocondriales (PRiMito). Se encontraron 70 genes asociados a la subunidad mayor y 56 genes para la menor. Las señales encontradas son específicas de los transcritos que codifican proteínas mitocondriales cuando se las compara con los que codifican proteínas ribosomales citoplasmáticas.

AAtSs

Las aminoacil-ARNt-sintetasas presentan la particularidad de ser un subgrupo de proteínas necesarias para la traducción tanto a nivel citoplasmático como mitocondrial. Cabe resaltar que el ADNk no codifica ninguna de las al menos 20 especies de estas proteínas. Como ya se dijo, si bien se esperaría que a nivel del genoma nuclear se encontraran 2 juegos completos de estas proteínas, una con destino mitocondrial y otra con destino citoplasmático, tanto en *T. cruzi* como en *T. brucei* pocas son las codificadas con destino específico.

Dado que los ARNt mitocondriales son adquiridos desde el citosol y derivan del mismo set nuclear, también es de esperarse que sean aminoacilados por las mismas enzimas. El estudio de los mecanismos que permiten la localización dual de los ARNt implica un entendimiento mayor del metabolismo mitocondrial en estos parásitos y escapa a los objetivos de este trabajo. Sin embargo, una primera aproximación al problema se muestra en el Apéndice 1. En coherencia con este hecho la mayoría de las AAtSs son codificadas por un único gen a nivel nuclear. Sin embargo, se vio que en el genoma de *T. brucei* se encuentran dos genes diferentes para la Trp-ARNt-sintetasa. Donde el uso de la segunda enzima podría emplearse en un evento de edición mitocondrial que requiere el reasignado de UGA a triptófano.

El grupo de Nilsson (Nilsson et al. 2010) propuso un mecanismo de localización subcelular para 25 AAtSs codificadas a nivel del genoma nuclear de *T. brucei*. Para estos 25 genes buscamos ortólogos en *T. cruzi*. A pesar de encontrar 21 genes ortólogos no se logra completar con estos el set de aminoacil-ARNt-sintetasas necesarias para el correcto funcionamiento de la célula. Se buscó completar el pool haciendo una búsqueda por texto en el transcriptoma de *T. cruzi*. Se encontraron 7 genes más que permitieron mejorar el set (Tabla 27). Sin embargo, no se encontró ningún gen que codifique para la Arg-ARNt-S ni para la His-ARNt-S.

Las AAtSs que se encuentran codificadas en *T. cruzi* podrían poseer mecanismos diferenciales de localización subcelular. Encontramos 3 situaciones *a priori*:

1) AAtSs codificadas por un único gen cuyo transcripto no se encuentre en MiNT:

La Asn-ARNt-S, Gln-ARNt-S, Glu-ARNt-S, Ile-ARNt-S y Leu-ARNt-S se encuentran codificadas por un único gen y no fueron predichas de ser mitocondriales. Sin embargo, es claro que estas enzimas son esenciales para la traducción mitocondrial. En este caso, se espera observar un mecanismo de localización dual de la proteína.

2) AAtSs codificadas por un único gen cuyo transcripto se encuentre en MiNT:

La Val-ARnt-S, Thr-ARnt-S, Ser-ARnt-S, Pro-ARnt-S, Met-ARnt-S, Lys-ARnt-S, Cys-ARnt-S así como la Ala-ARnt-S son codificadas por un único gen cuyo transcripto se encuentra dentro de MiNT. De estas, la Ala-ARnt-S, Cys-ARnt-S, Lys-ARnt-S, Met-ARnt-S, Ser-ARnt-S Thr-ARnt-S y Val-ARnt-S, o sea todas menos la Pro-ARnt-S, se encuentran dentro del grupo de MiNT-MTS, contando por lo tanto con péptido señal de localización mitocondrial. Es de esperar en general que este subgrupo recurra a la estrategia de localización dual mediada por la escisión del péptido señal en la mitocondria para luego distribuirse diferencialmente entre la mitocondria y el citoplasma en un mecanismo similar al descrito para la fumarasa de levadura (Sass et al. 2001).

3) AAtSs codificadas por más de un gen:

En el caso de *T. cruzi*, son 5 los genes que presentan más de una isoforma, Asp-ARnt-S (2), Gly-ARnt-S (2), Trp-ARnt-S (5, de los cuales 1 es un fragmento y otros dos son pseudogenes), Tyr-ARnt-S (3, de las cuales 1 es un pseudogén y la otra se encuentra anotada como proteína hipotética conservada) y Phe-ARnt-S (2). Encontramos transcriptos en MiNT tanto para la Asp-ARnt-S, la Trp-ARnt-S, como para la Gly-ARnt-S. Únicamente uno de los pseudogenes Trp-ARnt-S y la Gly-ARnt-S, poseen péptido de localización mitocondrial encontrándose dentro del subgrupo MiNT-MTS. Ni la Phe-ARnt-S ni la Tyr-ARnt-S se encuentran codificadas en los transcriptos de MiNT.

Tabla 27. Aminoacil-ARNt-sintetasas de *T. cruzi*. Se señala además aquellas que se encuentran dentro de MiNT-MTS, así como el gen ortólogo en *T. brucei*.

Aminoácido	MTS	<i>T. cruzi</i>	Producto	Gen ortólogo en <i>T. brucei</i>	
Ala	*	TcCLB.506863.10	<i>alanyl-tRNA synthetase, putative</i>	Tb927.6.700	
Arg				Tb927.11.1990	
Asn		TcCLB.503899.90	<i>asparagine synthetase a, putative</i>		
				Tb927.4.2310	
Asp		TcCLB.510777.20	<i>aspartyl-tRNA synthetase, putative</i>	Tb927.6.1880	
		TcCLB.509967.30	<i>aspartyl-tRNA synthetase, putative</i>	Tb927.10.1260	
Cys	*	TcCLB.507537.30	<i>cysteinyl-tRNA synthetase, putative</i>	Tb927.6.950	
Gln		TcCLB.508277.160	<i>glutaminyl-tRNA synthetase, putative</i>	Tb927.9.5210	
Glu		TcCLB.511731.60	<i>glutamyl-tRNA synthetase, putative</i>	Tb927.6.4590	
Gly	*	TcCLB.504017.79	<i>glycyl-tRNA synthetase, putative</i>	Tb927.11.9640	
		TcCLB.511589.260	<i>glycyl-tRNA synthetase, putative</i>		
His				Tb927.6.2060	
Ile		TcCLB.504137.10	<i>isoleucyl-tRNA synthetase, putative</i>		
				Tb927.10.9190	
Leu		TcCLB.511049.40	<i>leucyl-tRNA synthetase, putative</i>		
Lys	*	TcCLB.505807.120	<i>lysyl-tRNA synthetase, putative</i>	Tb927.6.1510	
				Tb927.8.1600	
Met	*	TcCLB.509247.50	<i>methionyl-tRNA synthetase, putative</i>	Tb927.10.1500	
Phe		TcCLB.506127.30	<i>phenylalanyl-tRNA synthetase, putative</i>	Tb927.11.2360	
		TcCLB.508271.10	<i>phenylalanyl-tRNA synthetase, putative</i>	Tb927.11.2360	
				Tb927.11.14120	
Pro		TcCLB.509805.130	<i>prolyl-tRNA synthetase, putative</i>	Tb927.10.12890	
Ser	*	TcCLB.506777.80	<i>seryl-tRNA synthetase, putative</i>	Tb927.11.7170	
Thr	*	TcCLB.511181.80	<i>threonyl-tRNA synthetase, putative</i>	Tb927.5.1090	
Trp		TcCLB.506375.10	<i>tryptophanyl-tRNA synthetase, putative</i>	Tb927.3.5580	Tb927.8.2240
		TcCLB.506375.60	<i>tryptophanyl-tRNA synthetase, putative</i>	Tb927.3.5580	Tb927.8.2240
	*	TcCLB.507795.60	<i>tryptophanyl-tRNA synthetase (pseudogene), putative</i>	Tb927.3.5580	Tb927.8.2240
		TcCLB.509667.20	<i>tryptophanyl-tRNA synthetase (pseudogene), putative</i>		
		TcCLB.466823.20	<i>tryptophanyl-tRNA synthetase, putative (fragment)</i>		
Tyr		TcCLB.510407.60	<i>tyrosyl-tRNA synthetase (pseudogene), putative</i>	Tb927.7.3620	
		TcCLB.511217.200	<i>tyrosyl or methionyl-tRNA synthetase, putative</i>	Tb927.8.5330	Tb927.7.2400
		TcCLB.508321.50	<i>hypothetical protein, conserved</i>	Tb927.7.2400	Tb927.8.5330
Val	*	TcCLB.509167.100	<i>valyl-tRNA synthetase, putative</i>	Tb927.6.4480	

Se analizó la secuencia de los transcritos anotados como **pseudogenes** o fragmentos con el fin de confirmar su participación en el subgrupo. En el caso de las Trp-ARNt-S anotadas como pseudogén (TcCLB.5007795.60, TcCLB.509667.20) se analizó la secuencia proteica que codifica, así como la secuencia genómica propiamente dicha (Figura 43). En este caso encontramos que la proteína predicha presenta solamente parte del dominio *tRNA synthetases class I (W and Y)* (Figura 43 A). Sin embargo, al analizar la secuencia genómica se encontró que el motivo parece estar completo, aunque en dos marcos de lectura distintos (Figura 43 B). Nuevamente, recalando que el genoma de *T. cruzi* necesita aún ser pulido, esto podría ser únicamente un error de ensamblaje del mismo. Por otro lado, este transcripto está presente en los datos de microarreglos evidenciándose su expresión. Por ello de todas formas será tenida en cuenta en el análisis.

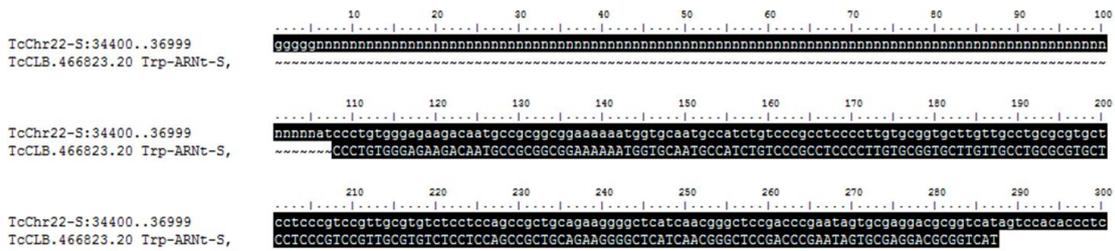


Figura 45. Alineamiento de la secuencia del fragmento de Trp-ARNt-S con la región genómica que lo contiene. Se observa la presencia de un tracto de Ns corriente arriba de lo que en la base de datos se encuentra definido como la región codificante.

Otro pseudogén presente en la lista lo constituye el TcCLB.510407.60 que corresponde con la Tyr-ARNt-S. Para este se llevó a cabo el mismo análisis que para los anteriores encontrándose a nivel de proteína y de la secuencia nucleotídica el resultado mostrado en la Figura 46. Se encontró el dominio tRNA-synt_1b completo, aunque la proteína representa únicamente el dominio sin otro contexto. Por otro lado, llama la atención que en el marco abierto de lectura 3+, a continuación de donde se encontraría el dominio existe un gran marco abierto que no posee sitios de terminación de la traducción interrumpiéndolo, sino hasta el final de la región codificante. Es de esperar que este defecto se deba al igual que sucede en el caso del Trp-ARNt-S (Figura 43) a una **inserción/delección** o simplemente a un problema de anotación o de ensamblaje del mismo. Así mismo, este transcripto está presente en los datos de microarreglos evidenciándose su expresión. En este caso seguiremos teniendo en cuenta el transcripto para la búsqueda de señales específicas.

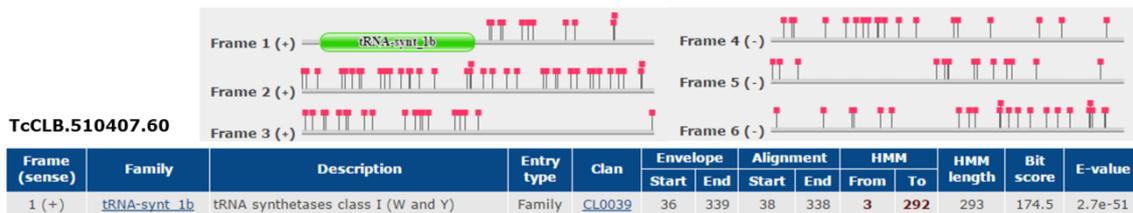


Figura 46. Esquema del dominio encontrado para el pseudogén de la Tyr-ARNt-S codificada por TcCLB.510407.60. Se esquematiza el dominio encontrado cuando se analizó el pseudogén mencionado.

Se llevó a cabo la búsqueda de señales a nivel de secuencia, que se encontraran en las regiones 5' y 3' UTR de los transcriptos de las AATs. Ninguno de los análisis empleando DREME encontró resultados significativos. Por otro lado, se analizaron las regiones no traducidas en búsqueda de motivos de estructura, cuyos resultados se resumen en la Tabla 28. Para estas regiones se encontraron dos motivos que cumplieron con los parámetros propuestos. Uno ellos ubicado en la región 5'UTR (Figura 47 A) y se encuentra presente en la porción de la CDS empleada en la búsqueda. Los transcriptos que aportan a este motivo se corresponden con la Val-ARNt-S (TcCLB.509167.100) y la Gly-ARNt-S (TcCLB.504017.79) que además de formar parte de MiNT también se encuentran en MiNT-MTS. Por

otro lado, también contienen el motivo, el transcripto de la Tyr-ARNt-S (TcCLB.510407.60) y ambos transcriptos de la Phe-ARNt-S.

Tabla 28. Búsqueda de motivos estructurales en los transcriptos que codifican para las AAtSs

Motivo	Genes aporte	Largo (nt)	DE Largo (nt)	Cobertura	Sensibilidad	Especificidad	ΔG^* kcal/mol
m1_3UTR_AAtSs*	4	32	7	15%	100%	93%	-9,50
m2_3UTR_AAtSs	19	35	7	70%	89%	93%	-11,80
m3_3UTR_AAtSs	7	30	5	26%	100%	96%	-13,30
m1_5UTR_AAtSs	22	27	6	81%	95%	96%	-10,10
m2_5UTR_AAtSs	20	35	6	74%	95%	81%	-12,70
m3_5UTR_AAtSs	5	46	6	19%	100%	96%	-16,80

Se señala con * los motivos que no superaron el criterio de especificidad/sensibilidad propuesto para este tipo de búsquedas. Los círculos verdes señalan aquellos agrupamientos cuya especificidad supera el 90% mientras que el amarillo corresponde a aquellos que están entre el 80 y 90% de especificidad. Se señala además en verde aquellas estructuras cuyo ΔG de formación a 28°C es menor al promedio de ΔG encontrado.

Por su parte, en la región 3'UTR encontramos otro motivo que superó los parámetros propuestos (Figura 47 B). Si bien sólo 7 aportaron a la construcción del motivo, la búsqueda para el análisis de la sensibilidad del método encontró que 10 de los 27 transcriptos de las AAtSs lo poseen en sus 3'UTR. En este caso los motivos se distribuyen preferencialmente hacia la región 3' de la 3'UTR. Al comparar la posición de los motivos con el largo de las UTRs reales, se vio que 4 de los 10 la poseerían en su secuencia, 2 de los restantes 6 presentan un largo menor y para el resto no se cuenta con los datos.

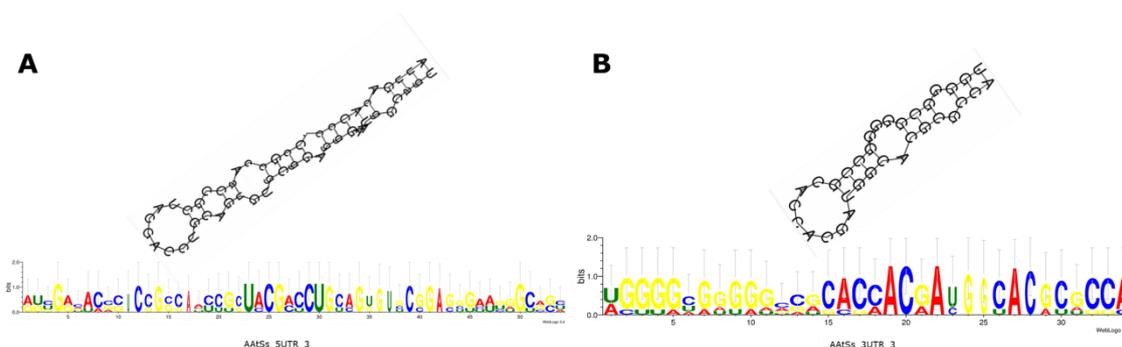


Figura 47. Estructura y secuencia consenso encontrada para los motivos que se observaron enriquecidos en las regiones 5' y 3'UTR de los transcriptos de las AAtSs.

De los 10 transcriptos que poseen el motivo, 4 pertenecen a MiNT: la Val-ARNt-S (TcCLB.509167.100), que posee además el motivo descrito en la región 5'; Trp-ARNt-S (TcCLB.506375.10, TcCLB.506375.60) y la Cys-ARNt-S (TcCLB.507537.30). De estos solamente la Val-ARNt-S y la Cys-ARNt-S poseen péptido de localización mitocondrial. El resto: la Tyr-ARNt-S (TcCLB.510407.60), el pseudogén de la Trp-ARNt-S (TcCLB.509667.20), la Phe-ARNt-S (TcCLB.508271.10), la Leu-ARNt-S (TcCLB.511049.40), la Gln-ARNt-S (TcCLB.508277.160) y la Asp-ARNt-S (TcCLB.510777.20), ni poseerían señal peptídica de localización mitocondrial, ni se encuentran en MiNT.

Los motivos encontrados para estas enzimas, cuya distribución se detalla en la Tabla 29, representan un par de señales prometedoras para ahondar en su estudio.

Tabla 29. Resumen de las señales presentes para cada transcrito

Aminoácido	<i>T. cruzi</i>	Producto	MiNT	MTS	5UTR	3UTR
Ala	TcCLB.506863.10	<i>alanyl-tRNA synthetase, putative</i>	*	*		
Asn	TcCLB.503899.90	<i>asparagine synthetase a, putative</i>				
Asp	TcCLB.510777.20	<i>aspartyl-tRNA synthetase, putative</i>				*
	TcCLB.509967.30	<i>aspartyl-tRNA synthetase, putative</i>	*			
Cys	TcCLB.507537.30	<i>cysteinyl-tRNA synthetase, putative</i>	*	*		*
Gln	TcCLB.508277.160	<i>glutaminyl-tRNA synthetase, putative</i>				*
Glu	TcCLB.511731.60	<i>glutamyl-tRNA synthetase, putative</i>				
Gly	TcCLB.504017.79	<i>glycyl-tRNA synthetase, putative</i>	*	*	*	
	TcCLB.511589.260	<i>glycyl-tRNA synthetase, putative</i>				
Ile	TcCLB.504137.10	<i>isoleucyl-tRNA synthetase, putative</i>				
Leu	TcCLB.511049.40	<i>leucyl-tRNA synthetase, putative</i>				*
Lys	TcCLB.505807.120	<i>lysyl-tRNA synthetase, putative</i>	*	*		
Met	TcCLB.509247.50	<i>methionyl-tRNA synthetase, putative</i>	*	*		
Phe	TcCLB.506127.30	<i>phenylalanyl-tRNA synthetase, putative</i>			*	
	TcCLB.508271.10	<i>phenylalanyl-tRNA synthetase, putative</i>			*	
Pro	TcCLB.509805.130	<i>prolyl-tRNA synthetase, putative</i>	*			
Ser	TcCLB.506777.80	<i>seryl-tRNA synthetase, putative</i>	*	*		
Thr	TcCLB.511181.80	<i>threonyl-tRNA synthetase, putative</i>	*	*		
Trp	TcCLB.506375.10	<i>tryptophanyl-tRNA synthetase, putative</i>	*			*
	TcCLB.506375.60	<i>tryptophanyl-tRNA synthetase, putative</i>	*			*
	TcCLB.509667.20	<i>tryptophanyl-tRNA synthetase (pseudogene), putative</i>				*
Tyr	TcCLB.510407.60	<i>tyrosyl-tRNA synthetase (pseudogene), putative</i>	*		*	*
	TcCLB.511217.200	<i>tyrosyl or methionyl-tRNA synthetase, putative</i>				
	TcCLB.508321.50	<i>hypothetical protein, conserved</i>				
Val	TcCLB.509167.100	<i>valyl-tRNA synthetase, putative</i>	*	*	*	

Por un lado, encontramos aquellas enzimas cuyos transcritos se encuentran en MiNT y que a pesar de estar codificado por un **único gen** poseen señal peptídica de localización subcelular. Tal es el caso de las enzimas que catalizan la adición de Ala, Cys, Lys, Met, Ser, Thr y Val a sus respectivos ARNt. Para estas proteínas se podría sugerir la existencia de transcritos de diferentes largos derivados de *trans-splicing* alternativo. Teniendo en cuenta el antecedente de la fumarasa (Sass et al. 2001), estas proteínas podrían ser sintetizadas en el entorno mitocondrial, ingresar a la mitocondria mediante la interacción entre la señal peptídica y los complejos de importación de la membrana mitocondrial, y una vez que el MTS sea escindido, retornar en parte al citoplasma.

Otro subgrupo de transcritos, poseen al menos **dos genes distintos** para cada enzima. En particular la Gly-ARNt-S se encuentra codificada por dos

genes. En uno de los casos parecería claro que el destino final de la proteína codificada sería la mitocondria por poseer MTS, mientras que, el otro transcripto, aunque no se encuentra anotado completamente, es probable que codifique a la Gly-ARNt-S citoplasmática.

También encontramos aquellas enzimas cuyos transcriptos se encuentran en MiNT y sin embargo no tienen MTS. Este es el caso de una de las Asp-ARNt-S, la Pro-ARNt-S, dos de las Trp-ARNt-S y una de las Tyr-ARNt-S. Por un lado, los transcriptos que codifican para la Asp-ARNt-S y para la Pro-ARNt-S no poseen ninguno de los motivos descritos en esta sección y, por lo tanto, debería existir una señal alternativa. Por otro lado, la Tyr-ARNt-S cuyo transcripto forma parte de MiNT posee los dos motivos estructurales propuestos en esta sección. Cabe resaltar que, si bien encontramos 2 genes más que codifican para esta enzima, uno de ellos se encuentra anotado como proteína hipotética conservada y solamente posee un dominio de unión a ARNt y no el dominio característico de las AAtSs. Por otra parte, el otro gen podría codificar para una Met-ARNt-S. En el caso de las Trp-ARNt-S los tres genes que lo codifican poseen la señal que describimos en la región 3'UTR. Sin embargo, según el largo real de las UTRs con las que contamos, en este caso el transcripto TcCLB.509667.20 no poseería la señal.

Finalmente, entre aquellos transcriptos que no fueron agrupados en MiNT, encontramos que los dos transcriptos que codifican para la Phe-ARNt-S presentan la señal estructural descrita en la región 5'. Contrariamente, los transcriptos que codifican para la Leu-ARNt-S, Gln-ARNt-S y Asp-ARNt-S solamente poseen la señal en la región 3'UTR y finalmente para Ile-ARNt-S la región 5'UTR no se encuentra anotada por lo que la búsqueda de señal a este nivel fue imposible. No contamos con evidencia de señales específicas para los transcriptos de Glu-ARNt-S y la Asn-ARNt-S.

Para las AAtSs se definieron señales conservadas. Extender la búsqueda a MiNT serviría como una primera aproximación para el estudio de su vinculación con el destino subcelular. Mediante ese estudio, encontramos que el motivo ubicado en la 5'UTR se presenta solamente en 60 transcriptos dentro de MiNT que se caracterizan por participar en el metabolismo de ácidos carboxílicos (GO:0019752, *fold enrichment* = 6,47) y ser componentes de las vías de síntesis de aminoacil-ARNts (ec00970, *fold enrichment* =26,96). Los motivos se distribuyen principalmente en la porción de CDS empleada en la búsqueda (Figura 48 A). No se puede predecir destino dual para estos transcriptos.

Cuando se enfoca la atención sobre el motivo que se definió para la región 3'UTR, encontramos que 58 transcriptos de MiNT lo presentan. En todos los casos, la ubicación del motivo se encuentra mayormente hacia el final de la 3'UTR hipotética (Figura 48 B), por lo que, se esperaría que fuera más probable encontrarlo en aquellos transcriptos que presenten 3'UTRs

mayores a la media. Su presencia se confirma en la mayoría (48/58) de los transcritos para los que contamos con, al menos, una 3'UTR definida. En este caso encontramos que, si bien el número de transcritos que poseen el motivo es similar, las funciones a las que se encuentran asociados son más variadas (Tabla 30).

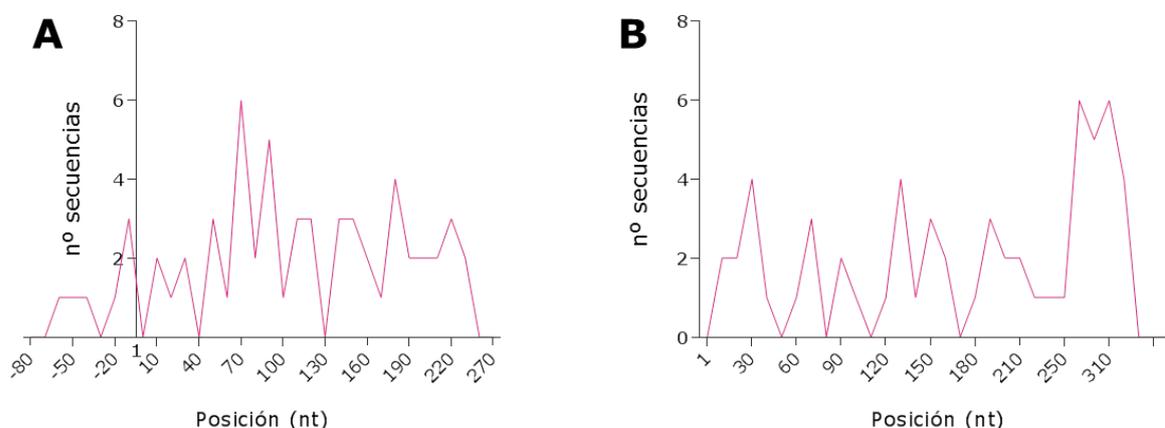


Figura 48. Frecuencia de sitios de inicio de las estructuras encontradas para A) la región 5' donde +1 corresponde a la posición del AUG anotado y para B) la región 3'UTR.

Tabla 30. Términos de ontología que encontramos enriquecidos cuando analizamos los transcritos de MINT que poseen la señal ubicada en la región 3'UTR.

ID	Término	Fold enrichment
GO:0043039	<i>tRNA aminoacylation</i>	21,33
GO:0006626	<i>protein targeting to mitochondrion</i>	16,35
GO:0070585	<i>protein localization in mitochondrion</i>	15,33
GO:0034660	<i>ncRNA metabolic process</i>	7,32
GO:0019752	<i>carboxylic acid metabolic process</i>	5,06
GO:0055114	<i>oxidation-reduction process</i>	3,72

Se definió la base de datos AAtSs conteniendo los genes que codifican para la mayor parte de las aminoacil-ARNt-sintetasas de *T. cruzi*. Para varios de los genes incluidos se actualizó su anotación dada la presencia de errores en la misma. Se definieron dos motivos de estructuras, uno en la región 5'UTR y otro en la región 3'UTR, enriquecidos en las AAtSs.

CONCLUSIONES

- › Se construyó una base de datos que contiene 1196 genes nucleares que codifican proteínas con destino mitocondrial (MiNT).
- › Las características de largo de CDS y grado de estructuración de los transcritos de MiNT, dan evidencia de su posible origen procariótico. El contenido GC3 y el CAI sugieren que estos transcritos presentarían una alta tasa traduccional.
- › El análisis de MiNT mediante la estrategia de agrupamientos racionales derivó en la definición de señales prometedoras para su estudio.
- › El estudio de MiNT-core llevo a la identificación de MiSTres_{core}, una señal a nivel de estructura secundaria enriquecida en la región 3'UTR de los mensajeros de MiNT.
- › Se encontró una señal peptídica conservada que denominamos TryM-TaPe y que seguramente constituya el péptido que dirija la localización mitocondrial de la proteína que lo contiene.
- › Se halló un motivo lineal enriquecido en los transcritos que no poseen péptido señal MiLTres_{NoMTS}, constituyendo una señal que podría cumplir una función compensatoria de localización subcelular de proteínas.
- › Se descubrió, la señal que denominamos MiLTres_{TRRM3}, representada en varias de las búsquedas llevadas a cabo a lo largo de este trabajo que probablemente sea reconocida por la proteína TRRM3 de *T. cruzi*. Los antecedentes de TRRM3 permiten plantearnos la hipótesis de su participación en la localización de los mensajeros blanco en el entorno mitocondrial.
- › Se construyó la base de datos de gene que codifican proteínas ribosomales mitocondriales, PRiMito, compuesta por 70 genes asociados a la subunidad mayor y 56 a la subunidad menor.
- › Se definió la base de datos AAtSs conteniendo los genes que codifican para la mayor parte de las aminoacil-ARNt-sintetasas de *T. cruzi*. Se encontraron dos motivos estructurales, uno en la región 5'UTR y otro en la 3'UTR

En primer lugar nos planteamos poner a punto un protocolo de fraccionamiento subcelular que permita obtener una fracción enriquecida en mitocondria de *T. cruzi*. El análisis a nivel proteómico de esta fracción permitirá validar la base de datos MiNT que hemos construido en este trabajo.

Basado en el análisis *in silico* de esta base de datos, hemos propuesto señales compartidas entre genes cuyos productos tienen un destino mitocondrial. La función de esas señales en el metabolismo de los mensajeros es desconocida. En algunos casos, los antecedentes permiten sospechar que estarían involucradas en mecanismos de localización diferencial. De esta forma, surge la necesidad de validarlos mediante un abordaje experimental.

El análisis transcriptómico de la fracción mitocondrial permitirá estudiar la distribución asimétrica de los mensajeros y validar la presencia enriquecida de aquellos que contengan las señales predichas.

Nos proponemos estudiar el papel de las señales descritas, tanto a nivel peptídico como nucleotídico, en la localización subcelular. Para ello, usaremos genes reporteros fusionados a estas señales y analizaremos la localización de mensajeros y productos proteicos, mediante el estudio por fraccionamiento subcelular y eventualmente *in situ*. No descartamos recurrir al análisis de mutaciones para verificar estos alcances.

Por otro lado, podría resultar de interés identificar los factores en *trans* que reconozcan las señales propuestas. Para ello, emplearíamos estrategias basadas en afinidad por el motivo específico.

Particularmente, nos interesa estudiar en profundidad la proteína TRRM3 de *T. cruzi*. En este sentido, nos planteamos producir la proteína de forma recombinante y validar la unión a la señal de interés mediante ensayos de retardo en gel. Complementariamente, validaremos los resultados obtenidos en el análisis bioinformático mediante CLIP-seq (*cross-linking immunoprecipitation & sequencing*), lo que además permitirá ampliar el espectro de mensajeros blanco.

Los resultados obtenidos en este trabajo, junto con las perspectivas que se abren a partir de ellos, constituyen un aporte a la caracterización de la regulación de la expresión génica en *Trypanosoma cruzi*.

BIBLIOGRAFÍA

- Almeida, I.C. & Nakayasu, E.S., 2010. Subcellular Proteomics and Global Analysis of Posttranslational Modifications to Study Functional Roles of *Trypanosoma cruzi* Molecules. , pp.167–177.
- Álvarez, G. et al., 2013. 1,2,4-thiadiazol-5(4H)-ones: a new class of selective inhibitors of *Trypanosoma cruzi* triosephosphate isomerase. Study of the mechanism of inhibition. *Journal of enzyme inhibition and medicinal chemistry*, 28(5), pp.981–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22803666> [Accessed March 2, 2016].
- Alves, L.R. & Goldenberg, S., 2016. RNA-binding proteins related to stress response and differentiation in protozoa. *World journal of biological chemistry*, 7(1), pp.78–87. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26981197>.
- Araújo, P.R. & Teixeira, S.M., 2011. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review. *Memórias do Instituto Oswaldo Cruz*, 106(3), pp.257–266. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762011000300002&lng=en&nrm=iso&tlng=en [Accessed August 22, 2016].
- Archer, S.K. et al., 2009. *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS pathogens*, 5(8), p.e1000565. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2727004&tool=pmcentrez&rendertype=abstract> [Accessed May 17, 2014].
- Aslett, M. et al., 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Research*, 38(Database), pp.D457–D462. Available at: <http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkp851> [Accessed July 19, 2016].
- Ayub, M.J. et al., 2009. Proteomic analysis of the *Trypanosoma cruzi* ribosomal proteins. *Biochemical and biophysical research communications*, 382(1), pp.30–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19245787> [Accessed March 1, 2015].
- Baez, M. V. & Boccaccio, G.L., 2005. Mammalian Smaug Is a Translational Repressor That Forms Cytoplasmic Foci Similar to Stress Granules. *Journal of Biological Chemistry*, 280(52), pp.43131–43140. Available at: <http://www.jbc.org/cgi/doi/10.1074/jbc.M508374200> [Accessed August 7, 2016].
- Bailey, T.L., 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics (Oxford, England)*, 27(12), pp.1653–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21543442> [Accessed July 12, 2016].
- Bailey, T.L. et al., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, 37(Web Server issue), pp.W202–8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2703892&tool=pmcentrez&rendertype=abstract> [Accessed July 9, 2014].
- Bayer-Santos, E. et al., 2012. Regulatory elements in the 3' untranslated region of the GP82 glycoprotein are responsible for its stage-specific expression in *Trypanosoma cruzi* metacyclic trypomastigotes. *Acta Tropica*, 123, pp.230–233.
- Benjamini, Y. et al., 2001. Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, 125, pp.279–284. Available at: www.elsevier.com/locate/bbr [Accessed August 13, 2016].
- Benne, R. et al., 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell*, 46(6), pp.819–26. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3019552> [Accessed February 11, 2016].
- Berriman, M. et al., 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science (New York, N.Y.)*, 309(5733), pp.416–22. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16020726> [Accessed May 1, 2014].
- Bolognani, F., Contente-Cuomo, T. & Perrone-Bizzozero, N.I., 2010. Novel recognition motifs and biological functions of the RNA-binding protein HuD revealed by genome-wide identification of its targets. *Nucleic acids research*, 38(1), pp.117–30. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2800223&tool=pmcentrez&rendertype=abstract> [Accessed May 17, 2014].
- Brandão, A. & Jiang, T., 2009. The composition of untranslated regions in *Trypanosoma cruzi* genes. *Parasitology international*, 58(3), pp.215–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19505588> [Accessed August 9, 2013].

- Bromley, E. V. et al., 2004. The amino terminal domain of a novel WD repeat protein from *Trypanosoma cruzi* contains a non-canonical mitochondrial targeting signal. *International Journal for Parasitology*, 34(1), pp.63–71. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0020751903002789> [Accessed December 16, 2014].
- Buckner, F.S. & Navabi, N., 2010. Advances in Chagas disease drug development: 2009–2010. *Current opinion in infectious diseases*, 23(6), pp.609–16. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3603362&tool=pmcentrez&rendertype=abstract> [Accessed February 6, 2016].
- Burger, G., Gray, M.W. & Lang, B.F., 2003. Mitochondrial genomes: Anything goes. *Trends in Genetics*, 19(12), pp.709–716.
- Buscaglia, C.A. et al., 2006. *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature Reviews Microbiology*, 4(3), pp.229–236. Available at: <http://www.nature.com/doi/10.1038/nrmicro1351> [Accessed July 15, 2016].
- Campos, P.C. et al., 2008. Sequences involved in mRNA processing in *Trypanosoma cruzi*. *International journal for parasitology*, 38(12), pp.1383–9.
- Cassola, A., 2011. RNA Granules Living a Post-transcriptional Life: the Trypanosomes' Case. *Current chemical biology*, 5(2), pp.108–117. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21949551> [Accessed August 22, 2016].
- Cassola, A., De Gaudenzi, J.G. & Frasch, A.C., 2007. Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. *Molecular microbiology*, 65(3), pp.655–70. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17635187> [Accessed August 22, 2016].
- Cenik, C. et al., 2011. Genome analysis reveals interplay between 5'UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS genetics*, 7(4), p.e1001366. Available at: <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1001366> [Accessed February 9, 2016].
- Chagas, C., 1909. Nova tripanozomíaze humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n.gen. n.sp., agente etiológico de nova entidade morbida do homem. *Mem Inst Oswaldo Cruz*, 1, pp.159–218.
- Charrière, F., 2006. Mitochondrial tRNA import in the parasitic protozoon *Trypanosoma brucei* and its consequences on mitochondrial translation. , (1544).
- Chávez, S., 2016. *Identificación de transcriptos periódicamente expresados y sus múltiples niveles de control mediante el análisis transcriptómico del ciclo proliferativo de Trypanosoma cruzi*. Tesis de Maestría en Ciencias Biológicas. Facultad de ciencias, Universidad de la República.
- Chen, D.H., Shi, X. & Suyama, Y., 1994. In vivo expression and mitochondrial import of normal and mutated tRNA(thr) in *Leishmania*. *Molecular and biochemical parasitology*, 64(1), pp.121–33. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7521523> [Accessed August 7, 2016].
- Cho, P.F. et al., 2006. Cap-dependent translational inhibition establishes two opposing morphogen gradients in *Drosophila* embryos. *Current biology*: CB, 16(20), pp.2035–41. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2238800&tool=pmcentrez&rendertype=abstract> [Accessed February 18, 2016].
- Choi, J. & El-Sayed, N.M., 2012. Functional genomics of trypanosomatids. *Parasite immunology*, 34(2–3), pp.72–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22132795> [Accessed December 16, 2014].
- Clayton, C. & Shapira, M., 2007. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Molecular and biochemical parasitology*, 156(2), pp.93–101. Available at: <http://www.sciencedirect.com/science/article/pii/S0166685107002101> [Accessed May 8, 2014].
- Clayton, C.E., 2002. Life without transcriptional control? From fly to man and back again. *The EMBO journal*, 21(8), pp.1881–8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=125970&tool=pmcentrez&rendertype=abstract> [Accessed February 9, 2016].
- Coller, J. & Wickens, M., 2007. Tethered function assays: an adaptable approach to study RNA regulatory proteins. *Methods in enzymology*, 429(7), pp.299–321. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17913629> [Accessed December 16, 2014].
- Cordingley, J.S., 1985. Nucleotide sequence of the 5S ribosomal RNA gene repeat of *Trypanosoma brucei*. *Molecular and biochemical parasitology*, 17(3), pp.321–30. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/4079966> [Accessed February 10,

- 2016].
- da Costa Lima, T.D. et al., 2010. Functional characterization of three Leishmania poly(A) binding protein homologues with distinct binding properties to RNA and protein partners. *Eukaryotic Cell*, 9(10), pp.1484–1494.
- Coughlin, B.C. et al., 2000. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein. *The Journal of biological chemistry*, 275(16), pp.12051–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10766837> [Accessed May 17, 2014].
- Crausaz Esseiva, A. et al., 2004. The T-stem determines the cytosolic or mitochondrial localization of trypanosomal tRNAsMet. *Molecular biology of the cell*, 15(6), pp.2750–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15064351> [Accessed August 3, 2016].
- Crooks, G.E. et al., 2004. WebLogo: A Sequence Logo Generator. *Genome research*, 14(14), pp.1188–1190. Available at: <http://www.genome.org/cgi/doi/10.1101/> [Accessed August 7, 2016].
- D'Orso, I. & Frasch, A.C.C., 2002. TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and Poly(A)-binding proteins forming a ribonucleoprotein complex. *The Journal of biological chemistry*, 277(52), pp.50520–8. Available at: <http://www.jbc.org/content/277/52/50520.full> [Accessed January 27, 2016].
- Dallagiovanna, B. et al., 2008. Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein from *Trypanosoma cruzi*. *The Journal of biological chemistry*, 283(13), pp.8266–73. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2276385&tool=pmcentrez&rendertype=abstract> [Accessed May 17, 2014].
- Das, A. & Bellofatto, V., 2009. The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in *Trypanosoma brucei*. *PLoS one*, 4(9), p.e6959. Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006959> [Accessed February 9, 2016].
- Decker, C.J. & Parker, R., 2012. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harbor perspectives in biology*, 4(9), p.a012286. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3428773&tool=pmcentrez&rendertype=abstract> [Accessed February 16, 2016].
- Ding, D., Dave, K.R. & Bhattacharya, S.K., On Message Ribonucleic Acids Targeting to Mitochondria. , pp.71–82.
- Docampo, R. et al., 2005. Acidocalcisomes - conserved from bacteria to man. *Nature reviews. Microbiology*, 3(3), pp.251–61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15738951> [Accessed February 12, 2016].
- Duchêne, A.-M., Pujol, C. & Maréchal-Drouard, L., 2009. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Current genetics*, 55(1), pp.1–18. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19083240> [Accessed February 19, 2016].
- Duhagon, M.A. et al., 2009. The *Trypanosoma cruzi* nucleic acid binding protein Tc38 presents changes in the intramitochondrial distribution during the cell cycle. *BMC microbiology*, 9(1), p.34. Available at: <http://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-9-34> [Accessed May 7, 2015].
- Ekanayake, D. & Sabatini, R., 2011. Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi*: modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. *Eukaryotic cell*, 10(11), pp.1465–72. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3209055&tool=pmcentrez&rendertype=abstract> [Accessed February 12, 2016].
- Elias, M.C. et al., 2007. Morphological events during the *Trypanosoma cruzi* cell cycle. *Protist*, 158(2), pp.147–57. Available at: <http://www.sciencedirect.com/science/article/pii/S1434461006001052> [Accessed September 2, 2015].
- El-Sayed, N.M. et al., 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science (New York, N.Y.)*, 309(5733), pp.409–15. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16020725> [Accessed May 7, 2014].
- Emanuelsson, O. et al., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, 2(4), pp.953–71. Available at: <http://dx.doi.org/10.1038/nprot.2007.131> [Accessed July 11, 2014].
- Emanuelsson, O., von Heijne, G. & Schneider, G., 2001. Analysis and prediction of

- mitochondrial targeting peptides. *Methods in cell biology*, 65, pp.175–187. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11381593> [Accessed May 17, 2014].
- Fallini, C. et al., 2016. Deficiency of the Survival of Motor Neuron Protein Impairs mRNA Localization and Local Translation in the Growth Cone of Motor Neurons. *The Journal of neuroscience*: the official journal of the Society for Neuroscience, 36(13), pp.3811–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27030765> [Accessed August 13, 2016].
- Feagin, J.E., Shaw, J.M., et al., 1988. Creation of AUG initiation codons by addition of uridines within cytochrome b transcripts of kinetoplastids. *Proceedings of the National Academy of Sciences of the United States of America*, 85(2), pp.539–43. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=279586&tool=pmcentrez&rendertype=abstract> [Accessed February 11, 2016].
- Feagin, J.E., Abraham, J.M. & Stuart, K., 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell*, 53(3), pp.413–22. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2452697> [Accessed February 11, 2016].
- Fragoso, S.P. & Goldenberg, S., 1992. Cloning and characterization of the gene encoding *Trypanosoma cruzi* DNA topoisomerase II. *Molecular and biochemical parasitology*, 55(1–2), pp.127–34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1331785> [Accessed February 10, 2016].
- Garcia, M. et al., 2010. Mitochondrial presequence and open reading frame mediate asymmetric localization of messenger RNA. *EMBO reports*, 11(4), pp.285–91. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20224577> [Accessed August 9, 2013].
- De Gaudenzi, J.G. et al., 2013. Genome-wide analysis of 3'-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes. *PeerJ*, 1, p.e118. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3728762&tool=pmcentrez&rendertype=abstract> [Accessed August 8, 2013].
- Gehrke, S. et al., 2015. PINK1 and Parkin control localized translation of respiratory chain component mRNAs on mitochondria outer membrane. *Cell metabolism*, 21(1), pp.95–108. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4455944&tool=pmcentrez&rendertype=abstract> [Accessed February 16, 2016].
- Gerber, A.P. et al., 2006. Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12), pp.4487–92. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1400586&tool=pmcentrez&rendertype=abstract>.
- Glynn Dennis, J.B.T.S.D.A.H.J.Y.W.G.H.C.L.R.A.L., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, 4(9), p.R60.
- Gómez, E.B. et al., 2001. Characterization of the *Trypanosoma cruzi* Cdc2p-related protein kinase 1 and identification of three novel associating cyclins. *Molecular and biochemical parasitology*, 113(1), pp.97–108. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11254958> [Accessed February 10, 2016].
- Grant, C.E., Bailey, T.L. & Noble, W.S., 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*, 27(7), pp.1017–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21330290> [Accessed July 12, 2016].
- Greif, G. et al., 2015. Kinetoplast adaptations in American strains from *Trypanosoma vivax*. *Mutation research*, 773, pp.69–82. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25847423> [Accessed March 3, 2016].
- Gu, W. et al., 2004. A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. *Genes & development*, 18(12), pp.1452–65. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=423195&tool=pmcentrez&rendertype=abstract> [Accessed February 18, 2016].
- Guda, C., Fahy, E. & Subramaniam, S., 2004. MITOPRED: a genome-scale method for prediction of nucleus-encoded mitochondrial proteins. *Bioinformatics (Oxford, England)*, 20(11), pp.1785–94. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15037509> [Accessed August 9, 2013].
- Guerra-Slompo, E.P. et al., 2012. Molecular characterization of the *Trypanosoma cruzi* specific RNA binding protein TcRBP40 and its associated mRNAs. *Biochemical and biophysical research communications*, 420(2), pp.302–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22425988> [Accessed August 22, 2016].
- Gupta, S. et al., 2007. Quantifying similarity between motifs. *Genome biology*, 8(2), p.R24.

- Available at: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2007-8-2-r24> [Accessed April 6, 2016].
- Haile, S. & Papadopoulou, B., 2007. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Current opinion in microbiology*, 10(6), pp.569–77. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18177626> [Accessed May 17, 2014].
- Hancock, K. et al., 1992. Identification of nuclear encoded precursor tRNAs within the mitochondrion of *Trypanosoma brucei*. *The Journal of biological chemistry*, 267(33), pp.23963–71. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1385429> [Accessed February 26, 2016].
- Hartl, F.U. et al., 1989. Mitochondrial protein import. *Biochimica et biophysica acta*, 988(1), pp.1–45. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2642391> [Accessed February 22, 2016].
- Häusler, T. et al., 1997. Conservation of mitochondrial targeting sequence function in mitochondrial and hydrogenosomal proteins from the early-branching eukaryotes *Crithidia*, *Trypanosoma* and *Trichomonas*. *European journal of cell biology*, 73(3), pp.240–51. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9243185> [Accessed July 12, 2016].
- Hay, R., Böhni, P. & Gasser, S., 1984. How mitochondria import proteins. *Biochimica et biophysica acta*, 779(1), pp.65–87. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6318829> [Accessed February 20, 2016].
- Herrmann, J.M. & Neupert, W., 2000a. Protein transport into mitochondria. *Current opinion in microbiology*, 3(2), pp.210–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10744987> [Accessed May 17, 2014].
- Herrmann, J.M. & Neupert, W., 2000b. What fuels polypeptide translocation? An energetical view on mitochondrial protein sorting. *Biochimica et biophysica acta*, 1459(2–3), pp.331–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11004448> [Accessed May 17, 2014].
- Hofacker, I.L., 2003. Vienna RNA secondary structure server. *Nucleic acids research*, 31(13), pp.3429–31. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=169005&tool=pmcentrez&endertype=abstract> [Accessed February 8, 2015].
- Holetz, F.B. et al., 2007. Evidence of P-body-like structures in *Trypanosoma cruzi*. *Biochemical and biophysical research communications*, 356(4), pp.1062–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17399688> [Accessed August 22, 2016].
- Horton, P. et al., 2007. WoLF PSORT: protein localization predictor. *Nucleic acids research*, 35(Web Server issue), pp.W585–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17517783> [Accessed August 31, 2016].
- Hotz, H.R. et al., 1997. Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3'-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic acids research*, 25(15), pp.3017–26. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=146859&tool=pmcentrez&endertype=abstract> [Accessed May 17, 2014].
- Huang, D.W., Sherman, B.T. & Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1), pp.44–57. Available at: <http://www.nature.com/doi/10.1038/nprot.2008.211> [Accessed July 9, 2014].
- Ivens, A.C. et al., 2005. The genome of the kinetoplastid parasite, *Leishmania major*. *Science (New York, N.Y.)*, 309(5733), pp.436–42. Available at: <http://science.sciencemag.org/content/309/5733/436.abstract> [Accessed February 7, 2016].
- Jansen, R.-P. & Niessing, D., 2012. Assembly of mRNA-protein complexes for directional mRNA transport in eukaryotes--an overview. *Current protein & peptide science*, 13(4), pp.284–93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22708485> [Accessed August 13, 2016].
- Jiang, H., Guan, W. & Gu, Z., 2010. Tinkering evolution of post-transcriptional RNA regulons: puf3p in fungi as an example. *PLoS genetics*, 6(7), p.e1001030. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2908677&tool=pmcentrez&rendertype=abstract> [Accessed August 9, 2013].
- Kaneko, T. et al., 2003. Wobble modification differences and subcellular localization of tRNAs in *Leishmania tarentolae*: implication for tRNA sorting mechanism. *The EMBO journal*, 22(3), pp.657–67. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12554666> [Accessed August 7, 2016].

- Kannaiah, S. & Amster-Choder, O., 2014. Protein targeting via mRNA in bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(8), pp.1457–1465.
- Kapushoc, S.T. et al., 2000. End Processing Precedes Mitochondrial Importation and Editing of tRNAs in *Leishmania tarentolae**.
- Keene, J.D., 2007. RNA regulons: coordination of post-transcriptional events. *Nature reviews. Genetics*, 8, pp.533–543.
- Kelly, S. et al., 2012. Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes. *Open biology*, 2(4), p.120033. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3376733&tool=pmcentrez&rendertype=abstract>.
- Kissinger, J.C., 2006. A tale of three genomes: the kinetoplastids have arrived. , 22(6), pp.240–243.
- Kramer, S., 2013. RNA in development: how ribonucleoprotein granules regulate the life cycles of pathogenic protozoa. *Wiley interdisciplinary reviews. RNA*, 5(2), pp.263–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24339376> [Accessed August 22, 2016].
- Lazzaretti, D. et al., 2016. The bicoid mRNA localization factor Exuperantia is an RNA-binding pseudonuclease. *Nature structural & molecular biology*, 23(8), pp.705–13. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27376588> [Accessed August 13, 2016].
- LeBlanc, A.J., Yermovsky-Kammerer, A.E. & Hajduk, S.L., 1999. A nuclear encoded and mitochondrial imported dicistronic tRNA precursor in *Trypanosoma brucei*. *The Journal of biological chemistry*, 274(30), pp.21071–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10409659> [Accessed February 26, 2016].
- Lee, C.M. et al., 1999. The DNA helicase, Hmi1p, is transported into mitochondria by a C-terminal cleavable targeting signal. *The Journal of biological chemistry*, 274(30), pp.20937–42. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10409639> [Accessed March 3, 2016].
- Lesnik, C. et al., 2014. OM14 is a mitochondrial receptor for cytosolic ribosomes that supports co-translational import into mitochondria. *Nature communications*, 5, p.5711. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4268710&tool=pmcentrez&rendertype=abstract> [Accessed February 16, 2016].
- Lesnik, C., Golani-Armon, A. & Arava, Y., 2015. Localized translation near the mitochondrial outer membrane: An update. *RNA Biology*, 6286(July), pp.00–00. Available at: <http://www.tandfonline.com/doi/full/10.1080/15476286.2015.1058686>.
- Li, Y. et al., 2016. Transcriptome Remodeling in *Trypanosoma cruzi* and Human Cells during Intracellular Infection. *PLoS pathogens*, 12(4), p.e1005511. Available at: <http://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1005511> [Accessed April 6, 2016].
- Li, Z.-H. et al., 2012. A 43-nucleotide U-rich element in 3'-untranslated region of large number of *Trypanosoma cruzi* transcripts is important for mRNA abundance in intracellular amastigotes. *The Journal of biological chemistry*, 287(23), pp.19058–69. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22500021> [Accessed August 22, 2016].
- Lill, R. et al., 1992. Import of cytochrome c heme lyase into mitochondria: a novel pathway into the intermembrane space. *The EMBO journal*, 11(2), pp.449–56. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=556474&tool=pmcentrez&rendertype=abstract> [Accessed March 4, 2016].
- Lima, B.D. & Simpson, L., 1996. Sequence-dependent in vivo importation of tRNAs into the mitochondrion of *Leishmania tarentolae*. *RNA (New York, N.Y.)*, 2(5), pp.429–40. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8665410> [Accessed August 3, 2016].
- Lithgow, T., 2000. Targeting of proteins to mitochondria. *FEBS letters*, 476(1–2), pp.22–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10878243> [Accessed February 18, 2016].
- Liu, B. et al., 2005. Fellowship of the rings: the replication of kinetoplast DNA. *Trends in parasitology*, 21(8), pp.363–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15967722> [Accessed February 8, 2016].
- van Loon, a P., Brändli, a W. & Schatz, G., 1986. The presequences of two imported mitochondrial proteins contain information for intracellular and intramitochondrial sorting. *Cell*, 44(5), pp.801–12. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3004746>.
- Lott, K. et al., 2013. Global proteomic analysis in trypanosomes reveals unique proteins and

- conserved cellular processes impacted by arginine methylation. *Journal of proteomics*, 91, pp.210–25. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23872088> [Accessed August 22, 2016].
- Machado-Silva, J.R., Neves, R.H. & Rodrigues-Silva, R., 2010. Do schistosomes grow old? A confocal laser scanning microscopy study. *Journal of helminthology*, 84(3), pp.305–11. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19941681> [Accessed May 17, 2014].
- Mahapatra, S. et al., 1998. The D arm of tRNATyr is necessary and sufficient for import into Leishmania mitochondria in vitro. *Nucleic acids research*, 26(9), pp.2037–41. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9547256> [Accessed August 3, 2016].
- Mair, G. et al., 2000. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA (New York, N.Y.)*, 6(2), pp.163–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1369902&tool=pmcentrez&rendertype=abstract> [Accessed February 10, 2016].
- Mani, J. et al., 2015. Mitochondrial protein import receptors in Kinetoplastids reveal convergent evolution over large phylogenetic distances. *Nature communications*, 6, p.6646. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4389251&tool=pmcentrez&rendertype=abstract> [Accessed September 8, 2015].
- Manne-Goehler, J., Reich, M.R. & Wirtz, V.J., 2015. Access to care for Chagas disease in the United States: a health systems analysis. *The American journal of tropical medicine and hygiene*, 93(1), pp.108–13. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25986581> [Accessed March 2, 2016].
- Maria, T.A., Tafuri, W. & Brener, Z., 1972. The fine structure of different bloodstream forms of *Trypanosoma cruzi*. *Annals of tropical medicine and parasitology*, 66(4), pp.423–31. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/4632495> [Accessed February 10, 2016].
- Martinez-Calvillo, S. et al., 2010. Gene expression in trypanosomatid parasites. *Journal of Biomedicine and Biotechnology*, 2010.
- Maslov, D.A. et al., 2006. Isolation and characterization of mitochondrial ribosomes and ribosomal subunits from *Leishmania tarentolae*. *Molecular and Biochemical Parasitology*, 148(1), pp.69–78.
- Michels, P.A.M. et al., 2006. Metabolic functions of glycosomes in trypanosomatids. *Biochimica et biophysica acta*, 1763(12), pp.1463–77. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17023066> [Accessed November 25, 2015].
- Minning, T.A. et al., 2009. The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. *BMC genomics*, 10, p.370. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2907688&tool=pmcentrez&rendertype=abstract> [Accessed May 16, 2014].
- Monnerat, S. et al., 2004. Genomic organization and gene expression in a chromosomal region of *Leishmania major*. *Molecular and biochemical parasitology*, 134(2), pp.233–43. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15003843> [Accessed February 10, 2016].
- Mortara, R.A. et al., 2005. Mammalian cell invasion and intracellular trafficking by *Trypanosoma cruzi* infective forms. *Anais da Academia Brasileira de Ciencias*, 77, pp.77–94.
- Nawrocki, E.P., Kolbe, D.L. & Eddy, S.R., 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics (Oxford, England)*, 25(10), pp.1335–7.
- Newberry, L.B. & Paulin, J.J., 1989. Reconstruction of the chondriome of the amastigote form of *Trypanosoma cruzi*. *The Journal of parasitology*, 75(4), pp.649–52. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2668492> [Accessed February 10, 2016].
- Niemann, M. et al., 2013. Mitochondrial outer membrane proteome of *Trypanosoma brucei* reveals novel factors required to maintain mitochondrial morphology. *Molecular & cellular proteomics*: MCP, 12(2), pp.515–28. Available at: <http://www.mcponline.org/content/12/2/515.long> [Accessed September 8, 2015].
- Nilsson, D. et al., 2010. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS pathogens*, 6(8), p.e1001037. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2916883&tool=pmcentrez&rendertype=abstract> [Accessed December 16, 2014].
- Nóbrega, A. et al., 2009. Oral transmission of Chagas disease by consumption of açai palm fruit, Brazil. *Headache*. Available at: <http://wwwnc.cdc.gov/eid/article/15/4/pdfs/08-1450.pdf> [Accessed February 9, 2016].
- Noé, G., De Gaudenzi, J.G. & Frasch, A.C., 2008. Functionally related transcripts have

- common RNA motifs for specific RNA-binding proteins in trypanosomes. *BMC molecular biology*, 9, p.107. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2637893&tool=pmcentrez&rendertype=abstract> [Accessed December 16, 2014].
- Opperdoes, F.R. & Michels, P.A.M., 2008. Complex I of Trypanosomatidae: does it exist? *Trends in parasitology*, 24(7), pp.310–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18534909> [Accessed February 10, 2016].
- Padilla-Mejía, N.E. et al., 2009. Gene organization and sequence analyses of transfer RNA genes in Trypanosomatid parasites. *BMC genomics*, 10(1), p.232. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2695483&tool=pmcentrez&rendertype=abstract> [Accessed February 9, 2016].
- Pádua, R.A.P. de, 2014. *Structural and functional characterization of Trypanosoma cruzi fumarate hydratase isoforms*. University of São Paulo, Ribeirão Preto.
- Palade, G.E., 1955. A small particulate component of the cytoplasm. *The Journal of biophysical and biochemical cytology*, 1(1), pp.59–68. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2223592&tool=pmcentrez&rendertype=abstract> [Accessed October 22, 2015].
- Panigrahi, A.K. et al., 2009. A comprehensive analysis of Trypanosoma brucei mitochondrial proteome. *Proteomics*, 9(2), pp.434–50. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2869593&tool=pmcentrez&rendertype=abstract> [Accessed September 7, 2015].
- Papadopoulou, M. V et al., 2015. 3-Nitrotriazole-based piperazides as potent antitrypanosomal agents. *European journal of medicinal chemistry*, 103, pp.325–34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26363868> [Accessed February 12, 2016].
- Paquin, N. et al., 2007. Local activation of yeast ASH1 mRNA translation through phosphorylation of Khd1p by the casein kinase Yck1p. *Molecular cell*, 26(6), pp.795–809. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17588515> [Accessed February 20, 2016].
- Park, K.-J. & Kanehisa, M., 2003. Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. *Bioinformatics (Oxford, England)*, 19(13), pp.1656–63. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12967962> [Accessed February 10, 2016].
- Pastro, L. et al., 2013. Implication of CA repeated tracts on post-transcriptional regulation in Trypanosoma cruzi. *Experimental parasitology*, 134(4), pp.511–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23631879> [Accessed December 16, 2014].
- Payne, M. et al., 1985. Identification of mitochondrial genes in Trypanosoma brucei and homology to cytochrome c oxidase II in two different reading frames. *Molecular and biochemical parasitology*, 15(2), pp.159–70. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2989684> [Accessed February 11, 2016].
- Pérez-Díaz, L. et al., 2013. Evidence for a negative feedback control mediated by the 3' untranslated region assuring the low expression level of the RNA binding protein TcRBP19 in T. cruzi epimastigotes. *Biochemical and biophysical research communications*, 436(2), pp.295–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23743203>.
- Perry, R.B.-T. et al., 2016. Nucleolin-Mediated RNA Localization Regulates Neuron Growth and Cycling Cell Size. *Cell reports*, 16(6), pp.1664–76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27477284> [Accessed August 13, 2016].
- Petsalaki, E.I. et al., 2006. PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. *Genomics, proteomics & bioinformatics*, 4(1), pp.48–55. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16689702> [Accessed May 15, 2015].
- Piacenza, L. et al., 2009. Enzymes of the antioxidant network as novel determiners of Trypanosoma cruzi virulence. *International journal for parasitology*, 39(13), pp.1455–64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19505468> [Accessed June 16, 2016].
- Pollard, V.W. et al., 1990. Organization of minicircle genes for guide RNAs in trypanosoma brucei. *Cell*, 63(4), pp.783–790. Available at: <http://www.cell.com/article/0092867490901444/fulltext> [Accessed February 10, 2016].
- Priest, J.W., Wood, Z. a. & Hajduk, S.L., 1993. Cytochromes c1 of kinetoplastid protozoa lack mitochondrial targeting presequences. *Biochimica et biophysica acta*, 1144(2), pp.229–231. Available at: <http://linkinghub.elsevier.com/retrieve/pii/000527289390178I>.
- Quenault, T., Lithgow, T. & Traven, A., 2011. PUF proteins: repression, activation and mRNA localization. *Trends in cell biology*, 21(2), pp.104–12. Available at:

- <http://www.ncbi.nlm.nih.gov/pubmed/21115348> [Accessed August 9, 2013].
- Raghavan, A. et al., 2004. Patterns of coordinate down-regulation of ARE-containing transcripts following immune cell activation. *Genomics*, 84(6), pp.1002–1013.
- Ray, D. et al., 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, 499(7457), pp.172–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3929597&tool=pmcentrez&rendertype=abstract> [Accessed July 10, 2014].
- Reid, D.W. et al., 2014. The unfolded protein response triggers selective mRNA release from the endoplasmic reticulum. *Cell*, 158(6), pp.1362–74. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4163055&tool=pmcentrez&rendertype=abstract> [Accessed February 16, 2016].
- Rendl, L.M., Bieman, M.A. & Smibert, C.A., 2008. *S. cerevisiae* Vts1p induces deadenylation-dependent transcript degradation and interacts with the Ccr4p-Pop2p-Not deadenylase complex. *RNA (New York, N.Y.)*, 14(7), pp.1328–36. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18469165> [Accessed August 7, 2016].
- Respuela, P. et al., 2008. Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. *The Journal of biological chemistry*, 283(23), pp.15884–92. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3259629&tool=pmcentrez&rendertype=abstract> [Accessed March 2, 2016].
- Rettig, J. et al., 2012. Dual targeting of isoleucyl-tRNA synthetase in *Trypanosoma brucei* is mediated through alternative trans-splicing. *Nucleic acids research*, 40(3), pp.1299–306. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3273800&tool=pmcentrez&rendertype=abstract> [Accessed May 17, 2014].
- Reynolds, D. et al., 2016. Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of siRNA Loci in *Trypanosoma brucei*. *PLoS genetics*, 12(1), p.e1005758. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4721609&tool=pmcentrez&rendertype=abstract> [Accessed January 25, 2016].
- Riou, G. & Delain, E., 1969. Abnormal circular DNA molecules induced by ethidium bromide in the kinetoplast of *Trypanosoma cruzi*. *Proceedings of the National Academy of Sciences of the United States of America*, 64(2), pp.618–25. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=223389&tool=pmcentrez&rendertype=abstract> [Accessed February 10, 2016].
- Riou, G.F. & Yot, P., 1977. Heterogeneity of the kinetoplast DNA molecules of *Trypanosoma cruzi*. *Biochemistry*, 16(11), pp.2390–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/324518> [Accessed February 10, 2016].
- Rodríguez, M., 2013. *Genómica comparativa y biología molecular de tripanosomas analizando datos de RNAseq Resultados y discusión*. Facultad de Ciencias, Universidad de la República.
- Saint-Georges, Y. et al., 2008. Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS one*, 3(6), p.e2293. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2387061&tool=pmcentrez&rendertype=abstract> [Accessed August 9, 2013].
- Sass, E. et al., 2001. Mitochondrial and cytosolic isoforms of yeast fumarase are derivatives of a single translation product and have identical amino termini. *The Journal of biological chemistry*, 276(49), pp.46111–7. Available at: <http://www.jbc.org/content/276/49/46111.full> [Accessed February 25, 2016].
- Schneider, A. & Maréchal-drouard, L., 2000. Mitochondrial tRNA import: are there distinct mechanisms? , 8924(0), pp.509–513.
- Shapiro, T.A. & Englund, P.T., 1995. The structure and replication of kinetoplast DNA. *Annual review of microbiology*, 49, pp.117–43. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8561456> [Accessed February 12, 2016].
- Sharp, P.M. & Li, W.H., 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research*, 15(3), pp.1281–95. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3547335> [Accessed August 14, 2016].
- Shaw, J.M. et al., 1988. Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. *Cell*, 53(3), pp.401–11. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2452696> [Accessed February 10, 2016].
- Shepard, K. a et al., 2003. Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA microarray analysis. *Proceedings of the*

- National Academy of Sciences of the United States of America*, 100(20), pp.11429–34. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=208774&tool=pmcentrez&rendertype=abstract>.
- Sherrer, R.L. et al., 2003. A Sequence Motif within Trypanosome Precursor tRNAs Influences Abundance and Mitochondrial Localization. *Molecular and Cellular Biology*, 23(24), pp.9061–9072. Available at: <http://mcb.asm.org/cgi/doi/10.1128/MCB.23.24.9061-9072.2003> [Accessed August 3, 2016].
- da Silva Augusto, L. et al., 2015. A membrane-bound eIF2 alpha kinase located in endosomes is regulated by heme and controls differentiation and ROS levels in *Trypanosoma cruzi*. *PLoS pathogens*, 11(2), p.e1004618. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25658109> [Accessed August 22, 2016].
- Simpson, A.M. et al., 1989. Kinetoplastid mitochondria contain functional tRNAs which are encoded in nuclear DNA and also contain small minicircle and maxicircle transcripts of unknown function. *Nucleic acids research*, 17(14), pp.5427–45. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=318168&tool=pmcentrez&rendertype=abstract> [Accessed February 26, 2016].
- Simpson, L. et al., 1987. Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence. *The Journal of biological chemistry*, 262(13), pp.6182–96. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3032958> [Accessed February 10, 2016].
- Simpson, L., 1987. The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. *Annual review of microbiology*, 41, pp.363–382.
- Smircich, P. et al., 2013. Genomic analysis of sequence-dependent DNA curvature in *Leishmania*. *PLoS one*, 8(4), p.e63068. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3639952&tool=pmcentrez&rendertype=abstract> [Accessed March 2, 2016].
- Smircich, P. et al., 2015. Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. *BMC genomics*, 16, p.443. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4460968&tool=pmcentrez&rendertype=abstract> [Accessed March 2, 2016].
- Smith, D.F. & Parsons, M., 1996. *Molecular biology of parasitic protozoa*, IRL Press at Oxford University Press.
- Smits, P. et al., 2007. Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic acids research*, 35(14), pp.4686–703. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1950548&tool=pmcentrez&rendertype=abstract> [Accessed February 24, 2016].
- Solaimuthu, S., 2010. *Natural selection on mRNA secondary structure and its correlation with protein functional groups*. Faculty of New Jersey Institute of Technology.
- Sotelo-Silveira, J.R. et al., 2011. Localization of mRNA in vertebrate axonal compartments by in situ hybridization. *Methods in molecular biology (Clifton, N.J.)*, 714, pp.125–38. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21431738> [Accessed February 29, 2016].
- Sotelo-Silveira, J.R. et al., 2006. RNA trafficking in axons. *Traffic (Copenhagen, Denmark)*, 7(5), pp.508–15. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16643274> [Accessed March 5, 2016].
- de Souza, W., Attias, M. & Rodrigues, J.C.F., 2009. Particularities of mitochondrial structure in parasitic protists (Apicomplexa and Kinetoplastida). *The international journal of biochemistry & cell biology*, 41(10), pp.2069–80. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19379828> [Accessed February 10, 2016].
- van der Spek, H. et al., 1988. Transcripts from the frameshifted MURF3 gene from *Crithidia fasciculata* are edited by U insertion at multiple sites. *The EMBO journal*, 7(8), pp.2509–14. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=457121&tool=pmcentrez&rendertype=abstract> [Accessed February 11, 2016].
- Sturm, N.R. & Simpson, L., 1990. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell*, 61(5), pp.879–884. Available at: <http://www.cell.com/article/009286749090198N/fulltext> [Accessed February 10, 2016].
- Supek, F. & Vlahovicek, K., 2004. INCA: synonymous codon usage analysis and clustering by means of self-organizing map. *Bioinformatics (Oxford, England)*, 20(14), pp.2329–30. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15059815> [Accessed September 1, 2016].

- Tan, T.H.P. et al., 2002. tRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import. *Molecular and cellular biology*, 22(11), pp.3707–17. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=133840&tool=pmcentrez&rendertype=abstract> [Accessed February 26, 2016].
- Thomas, S. et al., 2007. A population study of the minicircles in *Trypanosoma cruzi*: predicting guide RNAs in the absence of empirical RNA editing. *BMC genomics*, 8, p.133. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1892023&tool=pmcentrez&rendertype=abstract> [Accessed August 9, 2013].
- Tian, L. & Okita, T.W., 2014. mRNA-based protein targeting to the endoplasmic reticulum and chloroplasts in plant cells. *Current opinion in plant biology*, 22, pp.77–85. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25282588> [Accessed February 16, 2016].
- Tomás, A.M. & Castro, H., 2013. Redox metabolism in mitochondria of trypanosomatids. *Antioxidants & redox signaling*, 19(7), pp.696–707. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3739956&tool=pmcentrez&rendertype=abstract> [Accessed February 10, 2016].
- Tyler, K., Olson, C. & Engman, D., 2002. The Life Cycle Of *Trypanosoma Cruzi*. *American Trypanosomiasis*, 7, pp.1–11. Available at: <http://www.springerlink.com/index/V201GR75056M2457.pdf>.
- Vanassel, S. & Steinert, M., 1971. Nuclear and kinetoplasmic DNA replication cycles in normal and synchronously dividing. *Experimental Cell Research*, 65(2), pp.353–358. Available at: <http://www.sciencedirect.com/science/article/pii/0014482771900139> [Accessed February 11, 2016].
- Vieira, M. et al., 2005. Role for a P-type H⁺-ATPase in the acidification of the endocytic pathway of *Trypanosoma cruzi*. *The Biochemical journal*, 392(Pt 3), pp.467–74. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1316285&tool=pmcentrez&rendertype=abstract> [Accessed February 12, 2016].
- Ward, N. et al., 2014. Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss? V. de Crécy-Lagard, ed. *PLoS ONE*, 9(7), p.e101850. Available at: <http://dx.plos.org/10.1371/journal.pone.0101850> [Accessed July 19, 2016].
- Weis, B.L., Schleiff, E. & Zerges, W., 2013. Protein targeting to subcellular organelles via mRNA localization. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833(2), pp.260–273. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S016748891200095X> [Accessed August 9, 2013].
- Westenberger, S.J. et al., 2006. *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC genomics*, 7, p.60. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1559615&tool=pmcentrez&rendertype=abstract> [Accessed August 7, 2013].
- Wharton, R.P. & Aggarwal, A.K., 2006. mRNA regulation by Puf domain proteins. *Science's STKE*: *signal transduction knowledge environment*, 2006(354), p.pe37. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17003467> [Accessed February 18, 2016].
- WHO, 2015. Chagas disease (American trypanosomiasis). *Media centre*, 340. Available at: <http://www.who.int/mediacentre/factsheets/fs340/en/>.
- Wickens, M. et al., 2002. A PUF family portrait: 3'UTR regulation as a way of life. *Trends in genetics*: *TIG*, 18(3), pp.150–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11858839> [Accessed January 21, 2016].
- Wiedemann, N., Frazier, A.E. & Pfanner, N., 2004. The Protein Import Machinery of Mitochondria. *Journal of Biological Chemistry*, 279(15), pp.14473–14476.
- Williams, C.C., Jan, C.H. & Weissman, J.S., 2014. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science (New York, N.Y.)*, 346(6210), pp.748–51. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4263316&tool=pmcentrez&rendertype=abstract>.
- Woese, C.R., Kandler, O. & Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, 87(12), pp.4576–4579. Available at: <http://www.pnas.org/cgi/doi/10.1073/pnas.87.12.4576> [Accessed August 31, 2016].
- Woodward, R. & Gull, K., 1990. Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *Journal of cell science*, 95

- (Pt 1), pp.49–57. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2190996> [Accessed February 11, 2016].
- Wreden, C. et al., 1997. Nanos and pumilio establish embryonic polarity in *Drosophila* by promoting posterior deadenylation of hunchback mRNA. *Development (Cambridge, England)*, 124(15), pp.3015–23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9247343> [Accessed February 18, 2016].
- Wurst, M. et al., 2009. Molecular & Biochemical Parasitology An RNAi screen of the RRM-domain proteins of *Trypanosoma brucei*. *Molecular & Biochemical Parasitology*, 163, pp.61–65.
- Xu, L. et al., 2006. Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Molecular Biology and Evolution*, 23(6), pp.1107–1108.
- Yao, Z., Weinberg, Z. & Ruzzo, W.L., 2006. CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics (Oxford, England)*, 22(4), pp.445–52. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16357030> [Accessed October 3, 2014].
- Zabehinsky, D. et al., 2016. An Essential Role for COPI in mRNA Localization to Mitochondria and Mitochondrial Function. *Cell Reports*, 15(3), pp.540–549. Available at: <http://dx.doi.org/10.1016/j.celrep.2016.03.053>.
- Zavala-Castro, J.E. et al., Kinetoplast DNA-binding protein profile in the epimastigote form of *Trypanosoma cruzi*. *Archives of medical research*, 33(3), pp.250–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12031629> [Accessed February 10, 2016].
- Zhang, X. et al., 2010. The *Trypanosoma brucei* MitoCarta and its regulation and splicing pattern during development. *Nucleic acids research*, 38(21), pp.7378–87. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2995047&tool=pmcentrez&rendertype=abstract> [Accessed August 9, 2013].
- Zíková, A. et al., 2008. *Trypanosoma brucei* mitochondrial ribosomes: affinity purification and component identification by mass spectrometry. *Molecular & cellular proteomics*: MCP, 7(7), pp.1286–96. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2493383&tool=pmcentrez&rendertype=abstract> [Accessed December 16, 2014].
- Zipor, G. et al., 2009. Localization of mRNAs coding for peroxisomal proteins in the yeast, *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 106(47), pp.19848–53. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2785255&tool=pmcentrez&rendertype=abstract> [Accessed February 16, 2016].

ANÁLISIS DE ARN DE TRANSFERENCIA

Como ya se mencionó, en una sección anterior realizamos la búsqueda preliminar de posibles señales presentes en los ARNt de *T. cruzi*. Dado que la mitocondria de *T. cruzi* no posee ARNt codificados a nivel de su genoma estos deben ser importados desde el citoplasma. Si bien para *T. cruzi* no hay evidencia al respecto, existen estudios en *T. brucei* que permiten afirmar que el ARNt-Met iniciador se encuentra presente exclusivamente en citoplasma (Tan et al. 2002). Por su parte, el ARNt-Met elongador es el que se importa a la mitocondria para hacer las veces de iniciador, en el caso de que sea formilado, o de elongador en caso contrario. La enzima que se encarga de la reacción de formilación es la metionil-ARNt-formiltransferasa (MTF). El IF-2 mitocondrial reconoce el ARNt-Formil-Met, pero no su contraparte no formilada. Además se sugirió que el ARNt-Gln (CUG) sea también citoplasma específico.

Por otro lado, si bien varios trabajos atribuyeron la localización diferencial de los ARNt a distintas regiones en el mismo ARNt, en *T. brucei* se encontraron algunos ejemplos de ARNt con extensiones en la región 5'(Hancock et al. 1992) y en un caso se probó que era efectivamente esta la región responsable de la distribución del mensajero en la mitocondria (Tan et al. 2002).

Tabla 31. Señales de importación de los ARNt. Extraída y modificada de(Charrière 2006)

Organismo	ARNt	Localización de la señal	Evidencia	Referencia
<i>Trypanosoma brucei</i>	Met - ARNt (CAU #e)	Brazo T	mutación puntual	(Crausaz Esseiva et al. 2004)
	Leu - ARNt (CAA)	Región 5' - pre-ARNt	delección, mutación puntual	(Sherrer et al. 2003)
<i>Leishmania spp.</i>	Tyr - ARNt (GUA)	Brazo D	Fragmentación, mutación puntual	(Mahapatra et al. 1998)
	Ile - ARNt (UAU)	Brazo F	cambio de dominio	(Lima & Simpson 1996)
		brazo V-T		(Ding et al. n.d.)
	Trp - ARNt (CCA)	Anticodón	Fraccionamiento	(Kapushoc et al. 2000).
	Thr - ARNt (AGU)	bucle V	C--U Edición	(Chen et al. 1994)
Glu - ARNt (UUC)	Anticodón	inserción de bases, modificación de bases	(Kaneko et al. 2003)	

En este trabajo, se empleó la estrategia desarrollada anteriormente con el fin de buscar secuencias enriquecidas en la región corriente arriba del sitio de inicio del ARNt maduro. Para ello, se construyó una base de datos contando con las secuencias de los ARNt codificados en el genoma nuclear de *T. cruzi*. Estos ARNt presentan los anticodones que se describen en la

Tabla 32 y se distribuyen a lo largo de algunos cromosomas de *T. cruzi* como se muestra en la Figura 49.

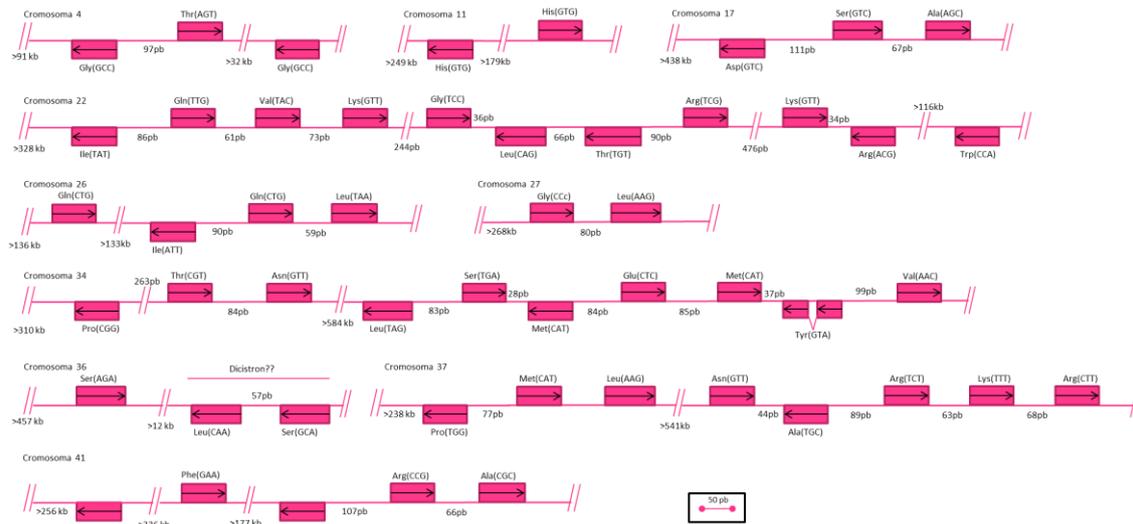


Figura 49. Organización genómica de los genes de ARNt en *T. cruzi*. Aproximadamente 13 agrupamientos y alguno de los genes dispersos encontrados en los mismos cromosomas que hacen un total de 52 genes, se representaron a escala. Los números indican el largo de las regiones intergénicas. Los genes de los 3 ARNt que se encuentran dispersos en los cromosomas 8, 9 y 23 de *T. cruzi* no se representaron.

Tabla 32. Distribución de anticodones en el set de ARNt de *Trypanosoma cruzi*.

Producto	Número de genes	Anticodones representados
tRNA-Ala	3	TGC CGC AGC
tRNA-Arg	6	CCT TCT TCG CCG ACG (2)
tRNA-Asn	2	GTT (2)
tRNA-Asp	1	GTC
tRNA-Cys	1	GCA
tRNA-Gln	3	TTG CTG (2)
tRNA-Glu	2	CTC (2)
tRNA-Gly	4	TCC CCC GCC (2)
tRNA-His	2	GTG (2)
tRNA-Ile	3	TAT ATT (2)
tRNA-Leu	6	TAG CAG CAA TAA AAG (2)
tRNA-Lys	3	TTT GTT (2)
tRNA-Met	3	CAT (3)
tRNA-Phe	2	GAA (2)
tRNA-Pro	3	CGG AGG TGG
tRNA-Ser	4	TGA CGA AGA GCT
tRNA-Thr	3	CGT TGT AGT
tRNA-Trp	1	CCA
tRNA-Tyr	1	GTA
tRNA-Val	2	AAC TAC

Para los 55 genes se obtuvieron las secuencias genómicas correspondientes a los 30 nucleótidos corriente arriba del inicio de la secuencia del ARNt maduro. La señal de localización mitocondrial se buscó en estas secuencias a partir de la expresión regular reportada en la literatura (Sherrer et al. 2003) (YGG(C/A)RRC) donde Y es una pirimidina (C/U) y R es una purina (A/G). Esta búsqueda no arrojó resultados significativos.

Por otra parte, se buscaron señales que estuvieran presentes en este grupo de genes. Se comenzó con la búsqueda de un motivo a nivel de secuencia, empleando el programa MEME de la plataforma MEME-Suite. Se ingresaron como parámetros que el largo de la secuencia encontrada fuera de entre 5 y 10 nucleótidos de largo, y que la búsqueda se llevara a cabo únicamente en la hebra proporcionada. Se encontró el motivo lineal descrito en la Figura 50. A este motivo contribuyen 32 de las 55 secuencias correspondientes a la región corriente arriba de los genes ARNt. Se llevó a cabo un control de especificidad realizando una búsqueda de este motivo en secuencias de 30 nucleótidos corriente arriba del inicio del gen o de las CDSs de 50 genes al azar. La búsqueda no encontró la señal en el grupo de secuencias control. Sin embargo a la hora de ver la sensibilidad del método (realizando una búsqueda en las mismas condiciones que el control, pero contra las secuencias que generaron el motivo) únicamente un 6% de las secuencias fueron recuperadas con un $e\text{-value} < 0,09$. Por otra parte, cuando se comparó el motivo contra la base de datos de motivos de ARN (Ray et al. 2013), empleando el programa Tom-Tom, no se encontraron motivos significativamente similares.

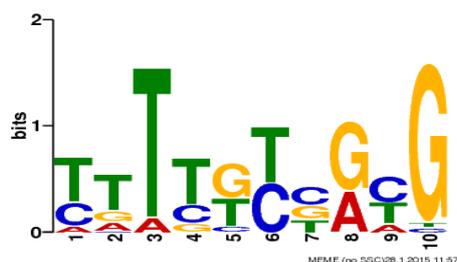


Figura 50. Logo correspondiente a la búsqueda de una secuencia conservada en la región -30 corriente arriba de los genes que codifican para ARNt en *Trypanosoma cruzi*. Esta secuencia se correspondería con la expresión regular “[TC]TT[TC][GT][TC][CGT][GA][CT]G” y tendría un $E\text{-value}$ de $1,9 \times 10^{-1}$.

En otra línea, se llevó a cabo la búsqueda de un motivo estructural en las mismas secuencias. En este caso se tomaron secuencias desde -60 nucleótidos corriente arriba del sitio de inicio del ARN maduro hasta +30 nucleótidos corriente abajo del sitio del mismo. Se encontraron 3 motivos para los cuales inicialmente se estudió su distribución a lo largo de la región empleada. En base a esto el motivo 1 fue descartado por encontrarse en el ARNt maduro (Figura 51 y Tabla 33).

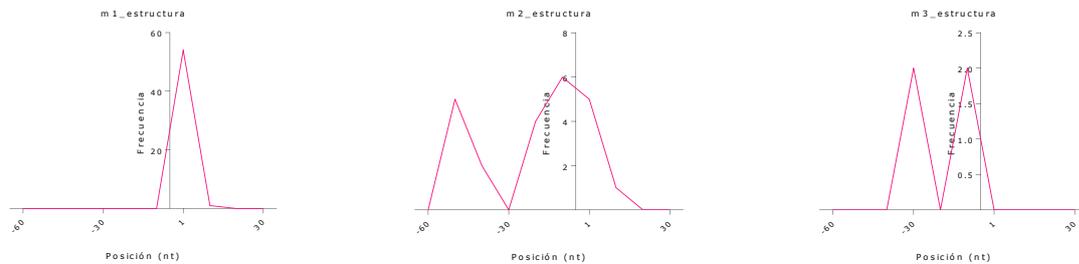


Figura 51. Distribución de las estructuras encontradas significativamente enriquecidas para la región flanqueante corriente abajo de la región codificante de los ARNts

Tabla 33. Búsqueda de motivos estructurales presentes en la región flanqueante de los ARNts.

Motivo	Genes aporte	Largo (nt)	DE Largo (nt)	Cobertura	Sensibilidad	Especificidad	ΔG^* kcal/mol
m2_ARNt	23	28	3	42%	78%	● 98%	-17,8
m3_ARNt	4	40	2	7%	175%	● 97%	-12

Se señala con * los motivos que no superaron el criterio de especificidad/sensibilidad propuesto para este tipo de búsquedas. Los círculos verdes señalan aquellos agrupamientos cuya especificidad supera el 90% mientras que el amarillo corresponde a aquellos que están entre el 80 y 90% y el rojo a los menores de 80% de especificidad.

Las estructuras encontradas se modelaron y se representan en la Figura 52.

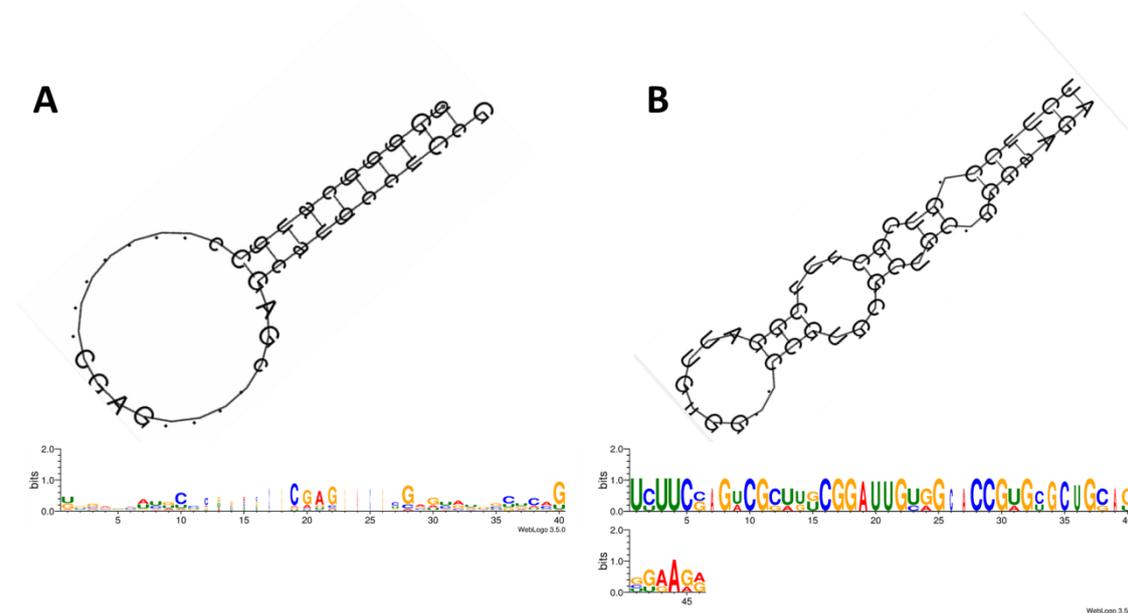


Figura 52. Estructuras que superaron los criterios propuestos en la región corriente arriba del inicio del transcrito maduro. Se observan las 2 estructuras que se encontraron significativamente enriquecidas en las búsquedas, modeladas mediante RNAFold, así como el logo de la secuencia consenso para cada una de ellas.

Debido a que no profundizamos en la definición de cuales de los ARNts son requeridos en la mitocondria no podemos definir si los motivos encontrados están presentes en los ARNts que deben ser importados al organelo. Por lo tanto, en este caso, se requiere ahondar en el estudio del mecanismo traduccional que tiene lugar a nivel de la mitocondria de estos parásitos para poder discutir estos resultados con mayor profundidad.

AGRADECIMIENTOS

La galaxia conspiró para que fueran parte de esta tesis sin darles elección, mi **familia**, que son muchos para nombrarlos a todos, mis **amig@s** que no me abandonaron a pesar del número "lick tablet" de veces que rechacé invitaciones o no me hice presente por quedarme inmersa en el mundo de las señales y que ayudaron en la terapia de superación de frustración casi en cada paso de análisis y en cada experimento. Mis **compañeros** de baile y canto, quienes entendieron las innumerables faltas a ensayos por los mismos motivos presentados anteriormente. Mi tutora, **Beatriz** y mi co-tutor **Pablo** que si bien se encuentran incluidos en alguno de los subgrupos anteriores, es importante resaltar la paciencia que me tuvieron, tienen, y espero tendrán para recorrer este camino que recién comienza. El **chongo** y mis dos hermosas hijas que a esta altura ya podrían defender mi trabajo incluso mejor que yo de tantas veces que participaron en la lectura, preparación y presentaciones del mismo

Agradecimientos especiales a la **CAP, PEDECIBA, la ANII, CSIC, Dr. De Gaudenzi** y no menos importante a **Alexandra Elbakyan** sin quien este trabajo hubiera sido casi impagable. Finalmente agradecer a los miembros del tribunal por aceptar criticar este trabajo y no morir en el intento.

Más que gracias.

May the force be with you!

Este trabajo está dedicado a mi abuelo