



TESIS DE MAESTRÍA EN BIOLOGÍA. SUBÁREA: BIOLOGÍA CELULAR Y MOLECULAR.

PEDECIBA

**“SIGNIFICADO BIOLÓGICO DE PROTEÍNAS ARGONAUTA EN EUCARIOTAS
INFERIORES SIN EVIDENCIA DE INTERFERENCIA POR ARN”**



Lic. Juan Pablo Tosar Rovira
Orientador: Dr. Alfonso Cayota

Tribunal de Tesis:

Dr. Otto Pritsch
Dra. Mónica Marin
Dr. Marcelo Comini

Montevideo, 18 de noviembre de 2011

ÍNDICE

INTRODUCCIÓN	1
1. Del genoma basura a los ARNs pequeños reguladores	1
1.1 siARNs, microARNs, piARNs, tasiARNs, casiARNs, tasiARNs, natsiARNs, 21UpiARNs, 26GARNs... ¿fragmentos de tARNs?	2
1.2 La familia de proteínas Argonauta	3
1.3 Aspectos estructurales de las proteínas Argonauta	9
1.4 Duplicación, diversificación y evolución de proteínas Argonauta	13
1.5 El problema: las proteínas Argonauta en tripanosomátidos	14
OBJETIVOS	16
1. Hipótesis de trabajo	16
2. Objetivo general	16
3. Objetivos específicos	16
MATERIALES Y MÉTODOS	17
RESULTADOS EXPERIMENTALES	28
1. Estudios filogenéticos de las proteínas Argonautas permiten inferir un origen evolutivo común a las Argonautas de tripanosomátidos	28
2. Predicción de dominios en TcPIWI-tryp: detección del dominio PAZ previamente esquivo	32
3. Obtención de la secuencia codificante completa de la proteína TcPIWI-tryp en la cepa Dm28c de <i>Trypanosoma cruzi</i>	38
4. Amplificación de la secuencia codificante completa de TcPIWI-tryp y sus dominios, y clonado en vectores de expresión en <i>E.coli</i> y <i>T.cruzi</i>	39
4.1 Amplificación y clonado de TcPIWI-tryp en el plásmido pGEM-T Easy	39
4.2 Clonado de TcPIWI-tryp en el vector de expresión en <i>E.coli</i> pQE-80_TEV	41
4.3 Clonado de la secuencia de los dominios PAZ y MID/PIWI en el vector pQE-80_TEV	44
4.4 Clonado de la secuencia TcPIWI-tryp completa en el vector pTEX para sobre-expresión en <i>Trypanosoma cruzi</i>	46
5. Expresión y purificación de la proteína TcPIWI-tryp y del dominio MID/PIWI en bacterias	48
5.1 Expresión de la proteína TcPIWI-tryp en <i>E.coli</i>	48

5.2	<i>Expresión del dominio MID/PIWI en E.coli</i>	51
5.3	<i>Purificación de TcPIWI-tryp y del dominio MID/PIWI de la fracción insoluble</i>	51
6.	Obtención de anticuerpos policlonales anti-TcPIWI-tryp	53
6.1	<i>Inmunización de un conejo con la proteína TcPIWI-tryp y obtención de suero policlonal específico</i>	55
7.	Replegamiento <i>in vitro</i> de la proteína TcPIWI-tryp	55
7.1	<i>Replegamiento en columna</i>	56
7.2	<i>Replegamiento por dilución</i>	57
7.3	<i>Replegamiento por diálisis</i>	60
8.	Sobre-expresión de TcPIWI-tryp en <i>T.cruzi</i> e inmunoprecipitación	63
9.	Expresión y purificación en <i>E.coli</i> del dominio PAZ	64
9.1	<i>Purificación y estabilidad de la proteína purificada</i>	64
9.2	<i>Clivaje del 6x-HIS tag y purificación de la proteína clivada por IMAC y Gel-Filtración</i>	66
9.3	<i>Optimización de las condiciones de purificación del dominio PAZ</i>	69
9.4	<i>Caracterización de las proteínas PAZ-SH y PAZ-IAM por espectrometría de masas (MALDI-TOF y MALDI-TOF-TOF)</i>	72
10.	Ensayos de solubilidad, estabilidad y cristalogénesis de las proteínas PAZ-SH y PAZ-IAM	75
10.1	Estabilidad de las proteínas PAZ-SH y PAZ-IAM a temperatura ambiente	75
10.2	Ensayos de cristalogénesis de las proteínas PAZ-SH y PAZ-IAM	75
	DISCUSIÓN	81
	CONCLUSIÓN	92
	PERSPECTIVAS	93
	AGRADECIMIENTOS	97
	BIBLIOGRAFÍA	98
	APÉNDICE	103

RESUMEN

En la presente Tesis, para el acceso al título de Magister en Ciencias Biológicas (PEDECIBA), se puso a prueba la hipótesis – a la fecha del comienzo de este trabajo, globalmente aceptada – de que existen algunos protozoarios parásitos y patógenos para el hombre como *Trypanosoma cruzi* y *Leishmania major* que habrían perdido toda la maquinaria de regulación mediada por pequeños ARNs y proteínas de la familia Argonauta. Considerando la importancia de estas vías para el resto de los eucariotas, la distribución ubicua de las mismas, y el ciclo de vida de estos protozoarios (que implica una reprogramación del perfil de expresión génica, principalmente a nivel post-transcripcional) se creyó oportuno estudiar un grupo de proteínas presentes en los genomas de todos los tripanosomátidos, que poseen uno de los dominios característicos de las proteínas Argonauta (dominio MID/PIWI), pero que no han sido consideradas como tales por carecer de un dominio PAZ. Se demostró que dicho dominio sí existe, pese a ser de secuencia divergente. Se realizaron también los primeros estudios filogenéticos de las proteínas tipo-Argonauta de tripanosomátidos, evidenciando que las mismas poseen un origen evolutivo diferente al resto de las proteínas Argonauta eucariotas. En consecuencia, se propuso una nueva nomenclatura para las mismas, que ya comienza a ser adoptada en la literatura científica.

Para abordar el estudio de estas proteínas, se optó por la expresión recombinante en bacterias de la proteína Argonauta de *T. cruzi*. Como ha sucedido para el resto de las Argonauta eucariotas, la misma no se pudo obtener de forma soluble en *Escherichia coli*, pero permitió la obtención de suero policlonal inmunoreactivo. Pese a ser una proteína multi-dominio y de gran tamaño (120KDa), los ensayos de repliegamiento *in vitro* fueron parcialmente exitosos, aunque de muy bajo rendimiento. La expresión recombinante del dominio PAZ – que se creía ausente – permitió la purificación de más de 2 mg de proteína soluble y monomérica por litro de cultivo, lo que permitió focalizarse en la realización de estudios estructurales, actualmente en curso. Los resultados obtenidos permiten cuestionarse acerca de la importancia funcional de los residuos de cisteína presentes en la proteína, aunque toda evidencia al respecto es por el momento indirecta. La sobre-expresión de la proteína en *T. cruzi* y la obtención de anticuerpos específicos permitieron su inmunoprecipitación en extractos del parásito, resultando evidente la asociación de la proteína con fragmentos derivados de mitades de ARNs de transferencia, población previamente identificada por nuestro grupo y postulada como una nueva familia de ARNs pequeños reguladores. En conjunto, los resultados obtenidos permiten especular sobre una vía ancestral de regulación génica mediada por proteínas Argonauta y fragmentos de tARNs, presente aún en la familia de los tripanosomátidos. Por su simplicidad (una sola proteína Argonauta en su genoma, y la ausencia de mecanismos “clásicos” de pequeños ARNs) *T. cruzi* y *L. major* constituirían un excelente modelo para el estudio de estas vías que involucrarían fragmentos derivados de tARNs y proteínas Argonauta, y que parecerían existir también en eucariotas superiores, incluidos los seres humanos.

INTRODUCCIÓN

La primera década del siglo XXI estuvo marcada por la elaboración de numerosos conceptos y herramientas experimentales que indudablemente revolucionaron el campo de la biología molecular y celular. Por nombrar un hito conocido, la publicación del primer borrador del genoma humano en 2001 (Lander et al. 2001; Venter et al. 2001) y su “finalización” dos años más tarde (Collins et al. 2003) marcó un punto de inflexión y alteró definitivamente la dinámica de la investigación en biomedicina.

La secuenciación del genoma humano no estuvo exenta de sorpresas. Las primeras estimaciones situaron el porcentaje del genoma ocupado por exones en solo el 1.1%. El 99% de nuestro genoma no codificaría por tanto para proteínas, distribuyéndose entre intrones (25%) y regiones intergénicas (75%) (Venter et al. 2001). Frente a estas cifras, expresiones como “ADN basura” o “reliquia evolutiva” con las que frecuentemente se refería al ADN no codificante (Ohno 1972) debieron ser utilizadas con cautela. Más aún, pronto se descubrió que un porcentaje elevado de dicho “ADN basura” se transcribía activamente (Kapranov et al. 2002; Rinn et al. 2003) y que un porcentaje elevado del genoma no codificante se encontraba bajo los efectos de la selección natural purificadora (Lindblad-Toh et al. 2005; Waterston et al. 2002)

1. Del genoma basura a los ARNs pequeños reguladores

La visión tradicional en el campo de la biología molecular de la célula ha sido proteína-centrista hasta bien entrado el nuevo milenio (Groschans and Filipowicz 2008). Bajo tal perspectiva, el ADN es una molécula especializada en conservar fielmente la información necesaria para fabricar proteínas, mientras que el ARN no sería más que un intermediario entre ambos (ARN mensajero), o un conjunto de moléculas con funciones propias pero subordinadas al proceso de producción de proteínas (ARNs ribosomales, ARNs de transferencia). Si bien se conocían algunas excepciones tales como ARNs con capacidad catalítica (ribozimas), o algunos transcritos producidos a partir de promotores, las mismas eran consideradas como excepciones a la regla, pero el esquema general seguía siendo el mismo, basado en adaptaciones del “dogma central” de la biología molecular (ADN → ARN → Proteínas).

Sin embargo, la constatación de que la mayor parte del ADN transcrito no se traduce a proteínas movilizó el foco de atención hacia las moléculas de ARN; las causas en el origen o el desarrollo de numerosas patologías humanas poco comprendidas comenzaron a buscarse en este nuevo universo de transcritos no codificantes, que las células gastan considerable energía en producir.

Si no es ligada directamente a la producción de proteínas ¿Cuál es la función de estos transcritos? Lamentablemente, para muchos de ellos la función es aún hoy desconocida. En otros casos no solo se ha encontrado una función esencial para el mantenimiento de la homeostasis celular o la regulación de la expresión génica, sino que se ha posibilitado el desarrollo de nuevas tecnologías de vastas aplicaciones. En los últimos diez años, el término “*non-coding RNAs*” ha tenido un crecimiento explosivo en la literatura científica, siendo tapa de varias ediciones de las revistas de más amplia difusión.

Lo complejo, amplio y dinámico del tema exige establecer un punto de corte a los efectos de esta introducción. Por lo general, los ARNs no codificantes se clasifican por su tamaño, en ARNs pequeños

(de 20 a 40 nucleótidos) y ARNs no codificantes largos (de 70 nucleótidos en adelante, hasta tamaños de varios miles). Las funciones asociadas a estos últimos parecen ser tan variables como sus tamaños, participando en los mecanismos de *splicing*, en la producción de ribosomas, y un largo etcétera (Grosshans and Filipowicz 2008). Por el contrario, los ARNs pequeños no codificantes parecerían funcionar principalmente en mecanismos de silenciamiento génico. Su función principal sería el reconocimiento - mediante apareamiento de bases - de ARNs específicos, reclutando así la maquinaria de silenciamiento y/o degradación hacia los mismos. Si bien este tipo de silenciamiento a nivel post-transcripcional es el más estudiado, también permitirían el silenciamiento transcripcional induciendo la heterocromatización de regiones específicas del ADN.

Estos pequeños ARNs, hasta hace poco tiempo ignorados, constituyen hoy día una vasta familia de especies de diverso tamaño, origen y función, y son protagonistas de algunas de las vías más conservadas de regulación de la expresión génica.

2. siARNs, microARNs, piARNs, tasiARNs, casiARNs, natsiARNs, 21UpiARNs, 26GARNs... ¿fragmentos de tARNs?

La familia de ARNs reguladores pequeños no codificantes es una familia grande y en expansión, probablemente mucho mayor de lo que se supone. Si bien mantienen la integridad de los genomas eucariotas (al menos) y regulan qué genes se expresan y cuáles no - o a qué niveles - desde hace miles de millones de años, sabemos de su existencia desde hace poco más de diez.

En la página 806 del volumen 391 de la revista *Nature*, del 19 de febrero de 1998, puede encontrarse una comunicación que lleva por título "*Interferencia génica potente y específica mediada por ARN de doble hebra en *Caenorhabditis elegans**"¹ (Fire et al. 1998). Publicación que le valió a dos de sus autores, Andrew Fire y Craig C. Melo, el Premio Nobel en Fisiología o Medicina en el año 2006. Si bien existían ya algunos reportes previos relacionados, éste fue el primer trabajo en el que se evidenció la existencia de un nuevo tipo de mecanismo de regulación mediado por ARN de doble hebra (ARNdh), lo que dió origen a un nuevo campo de estudio y a una herramienta que revolucionó la biología experimental: la interferencia por ARN (*RNAi*).

El silenciamiento génico mediado por la expresión exógena de ARNs complementarios (anti-sentido) es un fenómeno conocido desde mediados de la década de los '80 (Izant and Weintraub 1984). La novedad del trabajo de Fire, Melo y colaboradores es que compararon el silenciamiento producido por un ARN sentido y antisentido por separado, con el silenciamiento producido por la inyección simultánea de ambos, o por el ARN dúplex resultante. Sorprendentemente, el efecto producido por el ARNdh fue mucho más potente que el producido por cualquiera de las dos hebras por separado (cerca de dos órdenes de magnitud, según el caso) demostrando que el silenciamiento no se debía a un fenómeno físico de interferencia por hibridación del ARN anti-sentido, sino que se estaba evidenciando la acción de algún tipo de maquinaria celular gatillada por la presencia de ARN dúplex. Más aún, los autores demostraron que el fenómeno de interferencia podía "transportarse" a tejidos distintos de aquellos en los que se había inyectado el ARNdh, y que podía incluso transmitirse a la progenie. Con el paso del tiempo se supo que este tipo de respuesta sistémica - frecuente en plantas - es excepcional en animales,

¹ "*Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans**"

no así el fenómeno “local” de interferencia por ARN, que es ubicuo en el dominio eucariota. Esto ha contribuido a hacer de *C. elegans* un modelo animal paradigmático.

Cerrando el artículo que los haría célebres, los autores afirmaban: “*Cualquiera sea el caso para el que se lo use, el mecanismo responsable de la interferencia por ARN probablemente exista por un motivo biológico. La interferencia por ARNdH podría ser utilizada por el organismo para el silenciamiento génico en condiciones fisiológicas*”² (Fire et al. 1998)

La puerta a la búsqueda de ARNdH endógenos quedaba abierta. Rápidamente se demostró que el silenciamiento por ARNdH no era exclusivo de *C. elegans*, ya que también funcionaba en *Drosophila melanogaster* (Kennerdell and Carthew 1998) y en *Trypanosoma brucei* (Ngo et al. 1998). En plantas ya se conocía un mecanismo de silenciamiento post-transcripcional mediado por apareamiento de secuencias complementarias y posterior clivaje de determinados mensajeros (Metzloff et al. 1997). Sin embargo, no se había propuesto que dicho mecanismo fuese estimulado por ARNdH. Pronto se descubrió que existían ARNs pequeños endógenos, complementarios a determinados mensajeros y de tamaños bien definidos, y que los mismos serían responsables del silenciamiento post-transcripcional en plantas (Hamilton and Baulcombe 1999). Para ese entonces, quedaba la duda si existiría alguna relación entre el mecanismo de silenciamiento propuesto en plantas (que no involucraba ARNdH, pero sí ARNs pequeños anti-sentido de aproximadamente 25 bases) y el mecanismo de silenciamiento gatillado por ARNdH y para entonces ya demostrado en varios organismos.

Después de un año de numerosos descubrimientos en donde se detectaron algunas de las proteínas involucradas en el fenómeno del ARNi, estudios realizados *in vitro* con un lisado de *Drosophila* permitieron demostrar que el ARNdH introducido era posteriormente procesado a ARNs pequeños de 21-23 nucleótidos, y que el ARNm blanco era clivado en la zona correspondiente al ARNdH y a intervalos discretos de 21-23 nucleótidos (Zamore et al. 2000). Simultáneamente, se describía la purificación de un complejo enzimático con actividad ribonucleasa, que co-sedimentaba con una población discreta de ARNs pequeños y era responsable del silenciamiento mediado por ARNdH. A dicho “complejo de silenciamiento inducido por ARN” se lo denominó RISC, por sus siglas en inglés (Hammond et al. 2000).

De este modo, se manifestaba la existencia de una maquinaria celular que - en un proceso dependiente de ATP y de la ribonucleasa III *Dicer* - sería capaz de procesar ARNdH a ARNs pequeños (siARNs), los cuales serían luego utilizados por un segundo complejo ribonucleoproteico (RISC) como guía para la degradación endonucleolítica de ARNm complementarios (Bernstein et al. 2001).

Pronto se demostró que la introducción exógena de ARNdH de 21-22 bases era suficiente para mediar el silenciamiento de mensajeros blanco, en la medida de que los mismos contuvieran extremos 3' protuyentes de dos ribonucleótidos (Elbashir et al. 2001a), lo que es una característica de los cortes realizados por ribonucleasas del tipo III. Más aún, el sitio de corte del mensajero blanco podía predecirse con precisión, en el enlace correspondiente entre los nucleótidos 10 y 11 contando desde el extremo 5' del siARN complementario, (Elbashir et al. 2001b) lo que es una característica del clivaje mediado por proteínas Argonauta, como se verá más adelante.

² “*Whatever their target, the mechanisms underlying RNA interference probably exist for a biological purpose. Genetic interference by dsRNA could be used by the organism for physiological gene silencing*”.

Respecto a la interferencia por ARN como mecanismo regulatorio, resultaba claro que el sistema permitiría una defensa celular contra la invasión de numerosos virus, debido a que durante su ciclo replicativo intracelular lo hacen frecuentemente a través de intermediarios de ARNdh. Sin embargo, la existencia de siARNs endógenos, codificados por la propia célula, resultaba más controversial. Las primeras evidencias al respecto se obtuvieron en plantas, con la descripción de pequeños ARNs ("*cis-acting RNAs*" o casiARNs) originados de transposones, elementos repetitivos y repetidos en tándem, y que provocan la formación de heterocromatina en el mismo loci que les diera origen (Xie et al. 2004). También se describieron en plantas otras poblaciones de pequeños ARNs interferentes endógenos tales como los tasiARNs ("*trans-acting RNAs*"), originados por la acción de una ARN polimerasa ARN dependiente (Vazquez et al. 2004; Yoshikawa et al. 2005) o los natsiARNs ("*natural antisense transcript-derived siRNAs*"), originados a partir de genes bidireccionales parcialmente solapantes (Borsani et al. 2005).

Aunque presentando mecanismos de biogénesis considerablemente distintos, que no requieren la acción de ARN polimerasas ARN dependientes, la expresión de siARNs endógenos también parecería ser frecuente en *Drosophila* y en mamíferos. En estos organismos, las fuentes de ARNdh endógenas son variadas, pudiendo originarse a partir de ARNs autocomplementarios, transcripción convergente entre dos genes adyacentes, "*read-through transcription*" entre dos copias de un mismo transposón orientadas de forma opuesta, transcripción bidireccional de un mismo gen, interacción en trans entre dos ARNs parcialmente complementarios, o a partir de la duplicación de un pseudogen en dos copias adyacentes e invertidas (Ghildiyal and Zamore 2009).

Si bien la importancia fisiológica de los siARNs reguladores codificados en el propio genoma (endo-siARNs) parecería ser mayor a lo inicialmente supuesto, se sigue considerando que la maquinaria evolucionó como un sistema de protección frente a ARNdh exógenos. Aparentemente, la función estrictamente regulatoria de la expresión génica recae en una segunda familia de pequeños ARNs, los llamados micro-ARNs o miARNs, codificados en el propio genoma y capaces de provocar el silenciamiento génico ya sea por degradación del ARN mensajero, inhibición de su traducción, estimulación de su degradación o secuestro a compartimentos intracelulares específicos, tales como los gránulos de GW o cuerpos – P (Hutvagner y Simard 2008).

Actualmente se estima que más de la mitad de los genes del genoma humano se encuentran sujetos a regulación por miARNs (Rossi 2004). Si bien es innegable la importancia de los mecanismos clásicos de control transcripcional de la expresión génica (ej: interacción de las ARN polimerasas con factores de transcripción), la importancia del control post-transcripcional por ARNs pequeños no puede ser subestimada.

En realidad, los miARNs fueron descubiertos varios años antes de que el trabajo de Fire y Mello atrajeran la atención hacia el mundo del ARNdh, y posteriormente, de los ARNs pequeños. El primer ejemplo de un miRNA actuando sobre un mensajero específico - por complementariedad de bases con una zona en su región 3' UTR - data del año 1993 (Lee et al. 1993; Wightman et al. 1993). Como sucede frecuentemente, este descubrimiento permaneció muchos años como un dato curioso aislado, hasta que numerosos miARNs comenzaron a ser identificados a partir del año 2001 (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001), y se descubrió que parte de la maquinaria molecular implicada (por ejemplo, las proteínas *Dicer*) era compartida entre las vías de miARNs y siARNs (Grishok et al. 2001; Hutvagner et al. 2001; Ketting et al. 2001).

Dado que los miARNs se procesan a partir de transcritos endógenos parcialmente autocomplementarios (pri-miARNs, conteniendo caperuza 5' y cola poli-A), las primeras etapas de su procesamiento y maduración ocurren dentro del núcleo. En animales, son procesados inicialmente a fragmentos más pequeños de 60-70 nucleótidos (pre-miARNs, con una característica estructura en forma de horquilla) por la ARNasa del grupo III *Drosha* (salvo en casos de procesamiento de miARNs a partir de intrones), exportados al citoplasma, y procesados nuevamente por una ARNasa del grupo III: *Dicer*. Fruto del procesamiento por esta clase de ARNasas, los miARNs maduros poseen una estructura de ARNdH con extremos 3' protuyentes de dos nucleótidos, extremos 3' OH (en plantas: 2'O-metilados) y 5' monofosfato. De tamaño similar o ligeramente superior a los siARNs (≈ 22 nt), la diferencia más obvia entre ambos es que el grado de complementariedad entre ambas hebras (miARN y miARN*) no suele ser perfecto.

Las vías de siARNs y miARNs convergen en el procesamiento de los ARNdH precursores por la enzima *Dicer* para dar origen a los ARNs pequeños maduros. El paso siguiente consiste en la transferencia de estos pequeños ARNs a los complejos ribonucleoproteicos efectores, que contienen siempre una proteína de la familia Argonauta, tal como se detallará más adelante. En complejo con estas proteínas, los siARNs y miARNs son nuevamente procesados, de modo tal que una sola de las dos hebras permanezca unida a la proteína Argonauta y actúe como guía para unir ARNs mensajeros - u otros ARNs - específicos, ya sea por hibridación perfecta dentro de la secuencia codificante (siARNs o miARNs de plantas) o por hibridación parcial en la región 3' UTR (miARNs de animales). En el primer caso, el silenciamiento se produce por clivaje endonucleolítico del mensajero, en la región reconocida por el ARN pequeño guía. En el segundo caso, los complejos Argonauta/miARN interfieren con el inicio o la elongación de la traducción del mensajero blanco, secuestran al mismo a cuerpos de procesamiento - excluyéndolo temporalmente de la maquinaria traduccional - y/o estimulan la degradación del mismo convocando enzimas especializadas.

Con el título "*Small RNAs just got bigger (...)*"³, una revisión del año 2006 anunciaba el descubrimiento de una nueva familia de ARNs pequeños de 26-31nt, los piARNs, o "ARNs asociados a Piwi" (Kim 2006). Los primeros piARNs fueron descubiertos en células de la línea germinal de *D. melanogaster* en 2001, e implicados en el silenciamiento de genes y de elementos transponibles (Aravin et al. 2001). Sin embargo, se creyó que el fenómeno observado era una variante del mecanismo general de ARNi mediado por la transcripción bidireccional de elementos repetitivos en el genoma, por lo que se llamó a los ARNs pequeños resultantes "rasiARNs" por "*repeat-associated siARNs*". Años más tarde se descubriría que la biogénesis de los rasiARNs es independiente de proteínas *Dicer*, que se originan a partir de transcritos antisentido, poseen modificaciones 2' O-metilo en sus extremos 3', y se asocian a proteínas de la subfamilia Piwi, un grupo particular de proteínas Argonauta de expresión confinada a la línea germinal y células estaminales (Vagin et al. 2006). Su función parecería estar relacionada al silenciamiento de elementos transponibles en la línea germinal, manteniendo la integridad del genoma en las etapas críticas del desarrollo. En conjunto con la reclasificación de los "rasiARNs" como "piARNs", también se incluyeron en esta categoría los "21U ARNs" descritos en *C. elegans*, aunque de tamaño considerablemente más corto (Wang and Reinke 2008).

Diez años después de descubierta la interferencia por ARNdH, los ARNs pequeños seguirían dando sorpresas. Para entonces, se podía hablar de tres grandes familias de ARNs pequeños (siARNs, miARNs y

³ "Los ARNs pequeños se han vuelto más grandes"

piARNs), todos ellos con funciones específicas y asociados a una maquinaria celular especializada, aunque con numerosos puntos de contacto (figura 1). Es en este panorama donde comienzan las primeras contribuciones de nuestro grupo en el tema en cuestión.

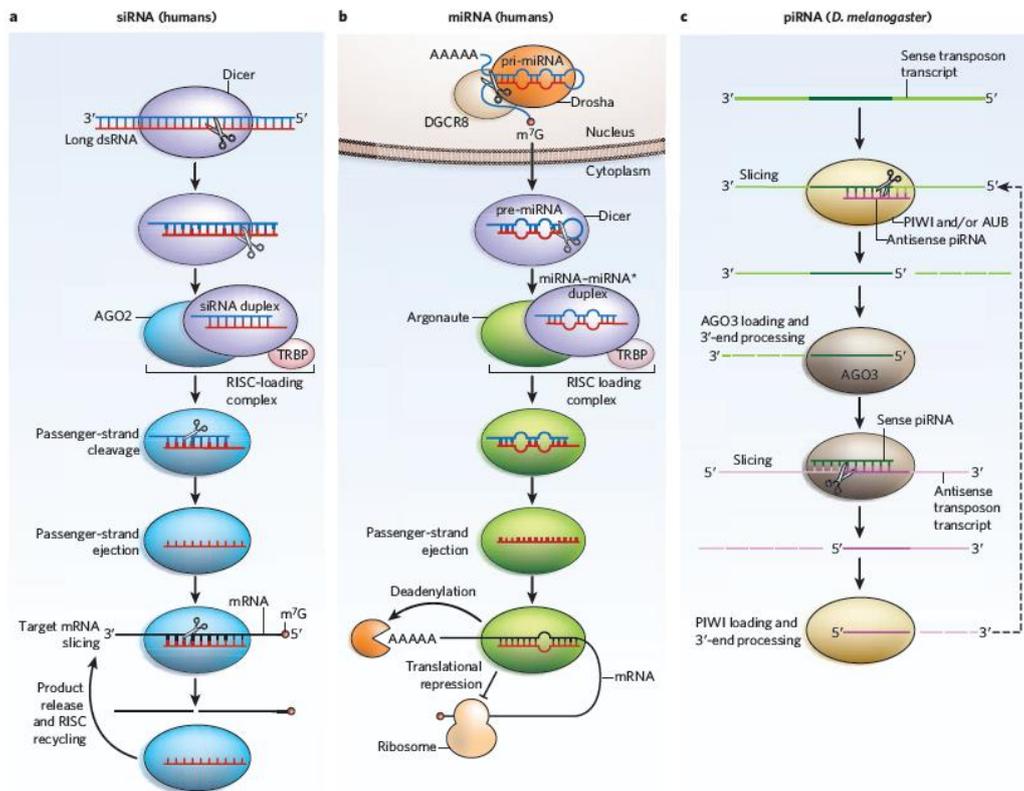


Figura 1: Vías “clásicas” de silenciamiento por ARNs pequeños. La vía del ARNi se muestra a la izquierda, con la proteína Dicer en violeta y la proteína Argonauta en celeste. En el centro se muestra el mecanismo de silenciamiento por miARNs, con la ribonucleasa III Drosha en naranja, la enzima Dicer en violeta, y la proteína Argonauta en verde, interaccionando con enzimas deadenilasas y con la maquinaria traduccional. A la derecha se muestra el mecanismo de “ping-pong” postulado para los piARNs, en el cual se alternan dos proteínas Argonauta: Piwi o Aubergine (en naranja) y la proteína Ago 3 (en negro). Los nombres de estas proteínas refieren al modelo de *D. melanogaster*. Imagen extraída de Jinek y Doudna (2009). RISC = “RNA-induced silencing complex” (Complejo de silenciamiento inducido por ARN)

Realizando un clonaje y secuenciado de pequeños ARNs en células de leucemia linfocitaria crónica, nuestro grupo describió diferencias en el perfil de expresión de algunos miARNs conocidos, al tiempo que detectaba nuevos genes de miARNs, algunos de los cuales aparentemente regularían supresores tumorales (Marton et al. 2008). Alentados por estos resultados, se procedió a realizar un secuenciado de pequeños ARNs en un organismo modelo: el protozooario parásito patógeno para el hombre *Trypanosoma cruzi*, considerado un organismo paradigmático en cuanto habría perdido todas sus proteínas Argonauta, así como el resto de la maquinaria de silenciamiento implicada (Hutvagner and Simard 2008)

¿Habría encontrado *T. cruzi* la manera de dispensar por completo de una vía de regulación por pequeños ARNs, cuando estas vías parecerían ser tan relevantes y conservadas en la evolución de los eucariotas, y presentes en organismos emparentados como *T. brucei*? Lo cierto es que el clonaje y secuenciado no permitió detectar una población que pudiese ser catalogada como de miARNs, al menos

no con las herramientas de análisis disponibles. Sin embargo, llamó la atención el alto porcentaje de pequeños ARNs que mapeaban en el genoma con genes de tARN (23%). Los porcentajes obtenidos parecían muy elevados para tratarse de simples fragmentos de degradación. Analizando los mismos, se constató que provenían principalmente de algunos genes de tARNs específicos (principalmente isoaceptores de ácido aspártico y glutámico), y de que no se distribuían al azar en la molécula, sino que mapeaban con la mitad 5' del tARN en cuestión, presentando un presunto "corte" a nivel del *loop* anticodón. Ensayos de *Northern blot* mostraban la expresión diferencial de las mitades 5' y 3' de numerosos tARNs, reflejando también los resultados del clonaje, en cuanto a la abundancia de mitades 5' de isoaceptores de ASP y GLU (figura 2). Más aún, estos fragmentos de tARNs presentaban una distribución granular en experimentos de hibridación *in-situ* con fluorescencia, y su expresión se veía incrementada en condiciones de stress, aunque no co-localizaban con los gránulos de stress previamente descritos.

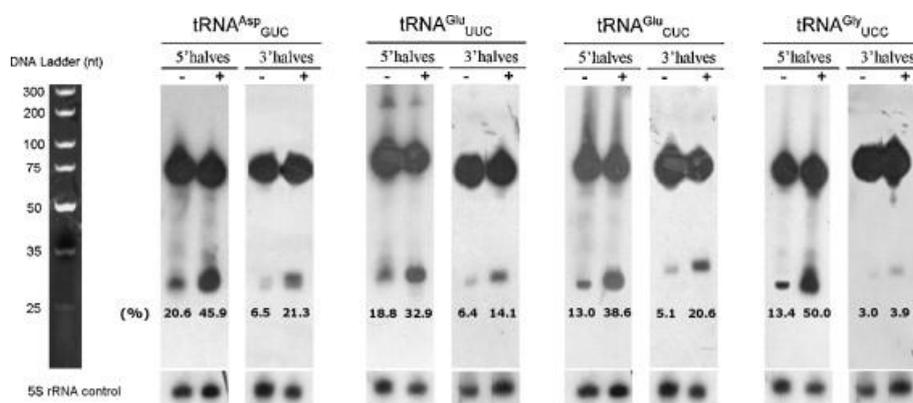


Figura 2: Análisis por Northern blot de las mitades 5' y 3' de tARNs en epimastigotes de *T. cruzi*, en condiciones óptimas de cultivo (-) o en stress nutricional (+) en medio TAU por 3 horas. Se marcan los respectivos tARNs maduros de aproximadamente 72 nucleótidos, y las mitades 5' y 3', utilizando sondas específicas para las mismas. Los porcentajes corresponden al porcentaje de la señal del tARN maduro correspondiente. Los valores se normalizaron contra el rARN 5S. Extraído de García-Silva et al. (2010).

¿Constituirían estos fragmentos de tARNs una nueva familia de pequeños ARNs reguladores, o serían simples fragmentos de degradación de los mismos? Esto último, en particular, era lo más difícil de demostrar. En condiciones de stress, podría tener sentido clivar algunos tARNs como forma de inhibir la traducción. Un mecanismo similar sucede en *E. coli* frente a la infección por el fago T4 (Amitsur et al. 1992). Sin embargo, este mecanismo sería demasiado poco específico para servir como un mecanismo efectivo de regulación, y los fragmentos de tARNs también podían ser detectados en condiciones de no-stress. Además, los ensayos de Northern blot permitían constatar que los niveles de los correspondientes tARNs maduros no clivados seguían siendo elevados en condiciones de stress, lo que se contrapone al argumento del clivaje de tARNs como llave maestra para inhibir la traducción.

La presencia de una vasta población de pequeños ARNs derivados de fragmentos de tARNs específicos clivados a nivel del anticodón resultó ser un fenómeno general y esparcido por todo el dominio eucariota. Resultados similares fueron reportados en el hongo filamentoso *Aspergillus fumigatus* (Jochl et al. 2008), el protozoario *Giardia lamblia* (Li et al. 2008), la levadura *Sacharomyces cerevisiae* (Thompson et al. 2008), la planta herbácea *Arabidopsis thaliana* (Hsieh et al. 2010; Thompson et al. 2008), e inclusive en líneas celulares murinas y humanas (Fu et al. 2009; Kawaji et al. 2008; Thompson et

al. 2008). A estos se suma - más cercano en el tiempo, fruto de numerosas idas y vueltas - nuestro reporte en *T. cruzi* (Garcia-Silva et al. 2010), y en diversas especies de plantas y hongos (Chen et al. 2011; Nunes et al. 2011; Wang et al. 2011). La reciente publicación del transcriptoma de pequeños ARNs no codificantes en *T. cruzi* corrobora los resultados obtenidos por nuestro grupo, aunque muestra una mayor representación de mitades 3' derivadas de tARNs (Franzen et al. 2011)

Al día de hoy, está claro que la existencia de fragmentos derivados de tARNs es un hecho, lo cual es corroborado por datos de secuenciado masivo (Cole et al. 2009). Si bien la cantidad de estudios disponibles no es suficiente, existen motivos para suponer cambios cuantitativos y cualitativos en las poblaciones de fragmentos de tARN durante la progresión tumoral. Un estudio reciente evidencia cambios en la abundancia relativa y el tipo de procesamiento de estos fragmentos entre células malignas de cáncer de colon primario y metastásico (Martens-Uzunova et al. 2011). Reanalizando datos depositados de secuenciado masivo de pequeños ARNs, también hemos constatado diferencias en la expresión de determinados fragmentos de tARNs entre células de origen tumoral y tejido normal.

En los últimos años, han aparecido las primeras evidencias experimentales de que fragmentos derivados de tARNs podrían tener un rol en la regulación de la expresión génica (Emara et al. 2010; Haussecker et al. 2010; Lee et al. 2009; Yamasaki et al. 2009). Más aún, se ha relacionado en *Tetrahymena thermophila* a una proteína Piwi (de la familia de las Argonautas) con este tipo de fragmentos derivados de tARNs (Couvillion et al. 2010), a los que se les ha dado el nombre de tRFs (por *tRNA-derived RNA fragments*) y clasificado como “una nueva clase de pequeños ARNs” (Lee et al. 2009). Estudios recientes aportan nuevas evidencias sobre la asociación de estos fragmentos derivados de tARNs y proteínas Argonauta (Burroughs et al. 2011; Haussecker et al. 2010)

3. La familia de proteínas Argonauta

Mediante estudios de mutagénesis en *Arabidopsis thaliana*, se aislaron una serie de mutantes que afectaban la arquitectura general de toda la planta. Aparentemente, la morfología inusual de estas plantas se asemejaba a la forma de un pequeño calamar, por lo que los investigadores denominaron a estos mutantes “argonauta”, por el cefalópodo *Argonauta argo* (Bohmert et al. 1998). Analizando el sitio de origen de la mutación, se definió y mapeó el locus AGO1, y se aisló un correspondiente cADN que daría origen a una proteína de 115KDa, con homología a otras proteínas putativas de plantas y animales. Aparecía así la primera descripción de la familia de proteínas Argonauta.

Las proteínas Argonauta son proteínas de entre 800-1500 aminoácidos, altamente básicas, con puntos isoeléctricos de 9.0-9.5. Se encuentran presentes en Archea, Bacteria, y en todo el dominio Eukarya, con contadas excepciones a las que se referirá más adelante. Su función principal consiste en unir ARNs pequeños reguladores, y - en conjunto con otras proteínas - silenciar la expresión de secuencias blanco parcial o totalmente complementarias. Al ser las proteínas que unen directamente a los ARNs pequeños, forman el corazón de los complejos ribonucleoproteicos de silenciamiento (llamados RISC en el caso del silenciamiento por siARNs y miRISC en el caso de silenciamiento por miARNs). Al día de la fecha, no se conocen ejemplos de organismos que posean vías de silenciamiento mediadas por pequeños ARNs y carezcan de proteínas Argonauta.

3.1 Aspectos estructurales de las proteínas Argonauta

Las proteínas Argonauta son proteínas globulares, compuestas por 4 dominios: un dominio N-terminal variable, el dominio PAZ, el dominio MID y el dominio PIWI. Originalmente, los dominios MID y PIWI fueron definidos como uno solo, y de hecho presentan una gran superficie de interacción, formando lo que podría definirse como el módulo C-terminal de la proteína, conectado con la mitad N-terminal por un bucle flexible (figura 3).

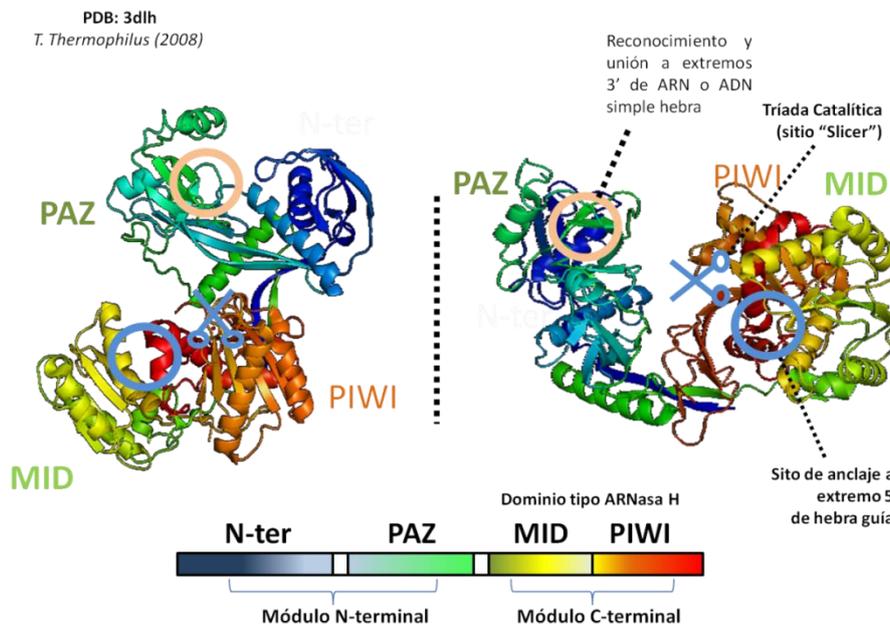


Figura 3: Estructura canónica de las proteínas Argonauta. La estructura corresponde a la proteína Argonauta de *T. thermophilus* (PDB: 3dlh). Se muestra el dominio N-terminal en azul, el dominio PAZ en verde, el dominio MID en verde/amarillo y el dominio PIWI en rojo. Se señalan los sitios de anclaje de los extremos 5' y 3' del pequeño ARN guía, y la tríada catalítica que confiere la actividad "slicer" (como una tijera). Las figuras fueron construidas a partir de los archivos pdb con el programa PyMol

Si bien la función de las proteínas Argonauta procariontas no es del todo conocida - aunque se especula que actuarían de modo similar (Makarova et al. 2009) - la mayor fuente de información estructural proviene de estas proteínas. Hasta el momento, se ha resuelto la estructura cristalográfica de cuatro proteínas Argonauta de procariontas hipertermófilos, provenientes de las Archea *Pyrococcus furiosus* (Song et al. 2004) y *Archeoglobolus fulgidus* (Ma et al. 2005; Parker et al. 2004, 2005), y de las bacterias *Aquifex aeolicus* (Rashid et al. 2007; Yuan et al. 2006; Yuan et al. 2005) y *Thermus thermophilus* (Wang et al. 2008a; Wang et al. 2009; Wang et al. 2008b).

Con varias estructuras resueltas de cada proteína - en forma libre, o unidas a pequeños ARNs - las cuatro Argonautas procariontas resueltas muestran un alto grado de conservación en los dominios MID/PIWI, en cuya interfase se encuentra el sitio que permite el anclaje del extremo 5' del ARN pequeño guía (que en Argonautas procariontas podría ser también ADN). Dicha región, denominada "bolsillo de unión a G1", se compone de residuos predominantemente básicos provenientes de los dominios MID y PIWI, y del carboxilato C-terminal de la proteína, que se entierra en esta región (prácticamente todas las proteínas Argonauta terminan con tres aminoácidos hidrofóbicos). Estos residuos permiten acomplejar un catión divalente, que participa en la unión de los fosfatos 1 y 3 del extremo 5' del ARN guía.

La Argonauta de *Archeoglobus fulgidus* se diferencia del resto en cuanto posee la mitad de tamaño, y codifica únicamente para los dominios MID y PIWI. Esto parecería ser relativamente común en procariontes, distinguiéndose familias de argonautas completas y bi-modulares, y argonautas más cortas que poseen únicamente el módulo MID/PIWI (Makarova et al. 2009). En muchos de estos casos, se suele expresar en el mismo policistrón una segunda proteína de unión al ARN que podría compensar la falta de los dominios N-ter y PAZ.

El dominio PIWI posee una estructura tipo ARNasa H. Ésta familia de endonucleasas es capaz de hidrolizar un enlace fosfodiéster interno a una molécula de ARN, cuando estas se encuentran formando híbridos ARN/ADN (por ejemplo, participan en la eliminación de los cebadores de ARN durante la duplicación del genoma). La actividad nucleasa de estas enzimas viene conferida por una tríada catalítica compuesta por dos residuos de ácido aspártico y uno de ácido glutámico (DDE), los que participan en la coordinación de dos cationes de Mg^{2+} (Nowotny et al. 2005). En Argonautas eucariotas, la tríada catalítica suele ser DDH, aunque la tercera posición también puede estar ocupada por ácido aspártico, glutámico, o lisina (Rivas et al. 2005; Tolia and Joshua-Tor 2007). En algunos casos, como el de la Argonauta-2 humana o la Argonauta-1 de *A.thaliana*, se ha demostrado que esta tríada catalítica confiere la actividad endonucleolítica “*slicer*” (clivaje de los ARNm blancos) característica de los complejos RISC (Liu et al. 2004).

Por muchos años desconocida, la función del dominio PAZ parecería ser la unión del extremo 3' de los ARNs pequeños. Su nombre proviene de la identificación de este dominio en las proteínas Piwi, Argonaute y Zwiille, que posteriormente se clasificarían como proteínas de la familia Argonauta. Estructuralmente, el dominio PAZ se compone de dos hélices α unidas a un OB-fold (*oligonucleotide/oligosaccharide binding fold*), un motivo estructural común presente en numerosas proteínas de unión a ácidos nucleicos de hebra simple. Este tipo de motivos presenta una estructura de “llave griega” de dos hojas β , cada una conteniendo tres hebras antiparalelas, una de las cuales pega un giro y es compartida por ambas hojas. Si bien este motivo se halla estructuralmente conservado, la conservación a nivel de secuencia entre diversos miembros de la familia es muy baja (Bochkarev and Bochkareva 2004; Theobald et al. 2003). Este hecho tiene implicancias en el desarrollo experimental de la presente tesis, que se discutirán más adelante.

El dominio PAZ también se encuentra en otro tipo de proteínas, distintas pero funcionalmente relacionadas: las ARNasas del grupo III “Dicer”. Se cree que el dominio PAZ permitiría - en estas proteínas - unir los extremos 3' protuyentes de los pre-miARNs procesados por Drosha (o por la propia Dicer en el caso de los siARNs), posicionando los dos dominios ARNasa III a una posición de 21-25 nucleótidos desde el extremo 3', actuando así como una regla molecular que definiría los largos de los pequeños ARNs (Macrae et al. 2006).

La serie de estudios del grupo de Dinshaw J. Patel sobre la Argonauta de *Thermus thermophilus* es una de las fuentes más ricas de información sobre cómo la estructura de las proteínas Argonauta se correlaciona con su función durante la unión y procesamiento de pequeños ARNs, así como de sus mensajeros blanco. La estructura cristalográfica de la Argonauta con un ADN⁴ pequeño guía muestra cómo este se une por su extremo 5' al bolsillo de unión en el dominio MID, y por su extremo 3' al

⁴ Este ADN guía de 21 nucleótidos emularía la hebra guía de los siARNs o miARNs unidos a Argonautas eucariotas. El problema consiste en que muchas Argonautas procariontes tienen mayor afinidad por el ADNsh que por el ARNsh.

dominio PAZ (Wang et al. 2008b). De este modo, la Argonauta actúa como un gancho que contiene al ARN guía por sus dos extremos.

En complejo con un ARN complementario, pero presentando un “*miss-match*” en las posiciones 10 y 11 (contando desde el extremo 5’ del ADN guía, como suele suceder en la interacción miARN-mARN), puede observarse cómo el ARN complementario hibrida con el extremo 5’ del ADN guía, pero “se pierde” más allá de la posición 10 (Wang et al. 2008a; Wang et al. 2009; figura 4A). En cambio, en caso de un ARN perfectamente complementario, la hibridación se extiende hasta la posición 19, ya que el dominio N-ter se interpone y provoca una “salida” del ARN blanco. Para que la hibridación sea posible más allá de la posición 10-11, debe ocurrir un cambio conformacional en la proteína Argonauta, que implica una rotación del módulo N-ter/PAZ respecto del módulo MID/PIWI, liberándose el extremo 3’ del ADN guía del dominio PAZ (Wang et al. 2009; figura 4B).

Los autores proponen un modelo en el cual el reconocimiento ARN guía/ARN mensajero ocurre inicialmente en las 7 primeras bases contando desde el extremo 5’ del ARN guía y excluyendo la primera, ya que esta se encuentra enterrada en el bolsillo de unión del dominio MID. Esto coincide con la región del ARN guía inicialmente más expuesta y rígida, y corresponde a una primera vuelta de hélice del híbrido. Esto se correlaciona con lo observado para la mayoría de los miARNs eucariotas, en cuanto el mayor determinante de la interacción miARN/ARNm recae en la posición “semilla”, compuesta por las 7-8 primeras bases del miARN (Brennecke et al. 2005). De haber complementariedad más allá de estas posiciones (caso de los siARNs y miARNs de plantas), la propia hibridación implicaría un segundo giro de hélice, acompañado por una rotación del módulo N-ter/PAZ respecto del módulo MID/PIWI, al que seguiría unido el ARN guía. De este modo, el enlace fosfodiéster entre los nucleótidos 10-11 del ARN blanco queda posicionado de forma adyacente a la tríada catalítica “*slicer*” del dominio PIWI. Más aún, para obtener la estructura de estos complejos fue necesario recurrir a mutantes en la tríada catalítica, o los ARN blancos serían subsecuentemente clivados en esta posición (Wang et al. 2009).

En caso de híbridos no perfectamente complementarios (caso de los miARNs de animales y sus blancos), el enlace fosfodiéster quedaría por fuera del sitio “*slicer*”, lo que explica por qué los miARNs no suelen actuar vía clivado de sus mensajeros blanco, a pesar de ser incorporados en proteínas Argonauta catalíticamente activas. En estos casos, el silenciamiento suele darse por otras vías, tales como la inhibición de la traducción. Al respecto, se ha propuesto que el dominio MID uniría por un sitio alostérico la estructura de caperuza 5’ de los ARNm blanco, brindando especificidad al reconocimiento y compitiendo con la maquinaria traduccional (Djuranovic et al. 2010). Esta hipótesis fue inicialmente respaldada por la resolución del dominio MID de la Argonauta QDE-2 de *Neurospora Crassa* (Boland et al. 2010), pero luego puesta en duda al resolver la estructura de los dominios MID/PIWI en conjunto (Boland et al. 2011), así como del dominio MID de la proteína Argonauta-2 humana (Frank et al. 2011).

Como queda en evidencia, la mayor fuente de información estructural proviene de proteínas Argonauta procariontas. Al día de la fecha, no se dispone de la estructura de ninguna Argonauta eucariota completa. No obstante, se ha resuelto la estructura de varios dominios, principalmente los dominios PAZ de las proteínas Piwi-like 1 humana (Tian et al. 2011; Zeng et al. 2011), Piwi-like 1 de ratón (Simon et al. 2011); Ago-2 humana (Ma et al. 2004), Ago-2 de *Drosophila* (Lingel et al. 2003, 2004; Song et al. 2003), y Ago-1 de *Drosophila* (Yan et al. 2003). A los dominios PAZ, se ha sumado la reciente resolución de los dominios MID de las proteínas Ago-2 humana (Frank et al. 2011; Frank et al. 2010) y QDE-2 del hongo *Neurospora*

crassa (Boland et al. 2011; Boland et al. 2010), y la primera estructura eucariota de los dominios MID/PIWI combinados (Boland et al. 2011).

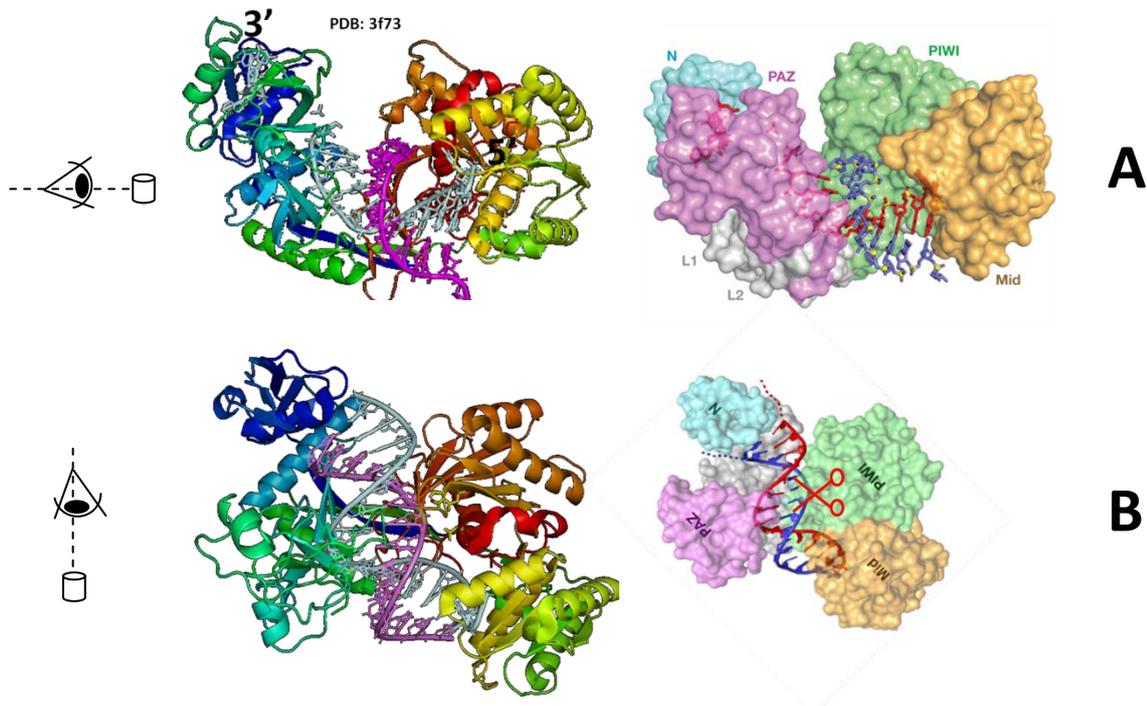


Figura 4: Estructura cristalográfica de la proteína Argonauta de *T. thermophilus* unida a un ADN guía de 21 nt (celeste) y a un ARN (violeta) presentando un “miss-match” en las posiciones 10-11 (A) o totalmente complementario (B). En el primer caso, el ADN guía está anclado por sus extremos 5’ y 3’ a los dominios MID y PAZ respectivamente. El ARN parcialmente complementario hibrida durante un giro de hélice, pero desaparece del mapa de densidad electrónica más allá de la posición del *miss-match*. En el caso B, el híbrido se extiende por dos giros de hélice, lo que es acompañado de un cambio conformacional de la proteína Argonauta, liberándose el dominio PAZ del extremo 3’ del ADN guía. Las figuras de la derecha fueron extraídas de Wang et al. (2008b y 2009). Las figuras de la izquierda fueron construídas con el programa PyMol en base a las estructuras pdb: 3f73 (A) y 3hxm (B).

En líneas generales, la estructura de estos dominios eucariotas avala la mayoría de las extrapolaciones que podrían realizarse a partir de los conocimientos obtenidos en Argonautas procariotas. Más allá de la estructura aislada de los dominios, lo que resulta realmente relevante es la disposición espacial de los dominios entre sí, y la dinámica de los mismos en complejo con ARNs pequeños y ARNs blanco. De obtenerse la estructura completa de una proteína Argonauta eucariota - ambiciosa empresa - el aporte al área de conocimiento sería profundo.

3.2 Duplicación, diversificación y evolución de proteínas Argonauta.

Las proteínas Argonauta están presentes tanto en eucariotas como en procariotas, y entre los primeros, pueden detectarse en todos los genomas disponibles con algunas notables excepciones: *S. cerevisiae*, *Plasmodium falciparum*, y los tripanosomátidos *Leishmania major* y *T. cruzi* (Hutvagner and Simard

2008). Considerando que la presente tesis se centra en el estudio de una proteína Argonauta de *T. cruzi*, se entenderá que el carácter excepcional de estos organismos es cuestionable.

A lo largo de su evolución, estas proteínas han estado sujetas a un intenso grado de duplicación génica, con la consecuente diversificación y especialización funcional. A modo de ejemplo, los seres humanos poseemos 8 proteínas Argonauta, 4 de cada sub-familia (ver más adelante). De ellas, la Argonauta-2 es la única con actividad “*slicer*” reportada, y por tanto esencial para la ARNi (Amitsur et al. 1992; Liu et al. 2004; Meister et al. 2004). La Argonauta-1 participaría también en el mecanismo nuclear de silenciamiento transcripcional por modificación de histonas y formación de heterocromatina (Janowski et al. 2006). *Arabidopsis thaliana* posee 10 proteínas Argonauta, todas de una misma sub-familia. No obstante, el record de proteínas Argonautas por genoma lo tiene *C. elegans*, con 26 representantes. Esto se correlaciona con el hecho de que plantas y nematodos poseen vías de amplificación de la interferencia, mediadas por enzimas ARN polimerasas ARN dependientes (ARNpARNd). En consecuencia, poseen proteínas Argonautas especializadas ya sea en la unión de siARNs primarios o secundarios (originados por amplificación mediada por ARNpARNd).

Si bien existe especialización funcional, muchas de las funciones que desempeñan distintas proteínas Argonauta se encuentran, al menos parcialmente, solapadas. Por ejemplo, los complejos de Argonauta 1-4 humanas unirían miARNs (Meister et al. 2004), aunque el tipo y modo de silenciamiento podría depender de la Argonauta en cuestión.

El primer estudio a gran escala sobre la evolución de las proteínas Argonauta y el resto de la maquinaria de silenciamiento fue realizado en 2006, utilizando las secuencias combinadas de los dominios PAZ y PIWI de numerosas proteínas Argonauta eucariotas (Cerutti and Casas-Mollano 2006). Dicho estudio mostró que todas las Argonauta eucariotas se segregaban en dos grupos filogenéticos presumiblemente parálogos, ya que muchos organismos - los animales, por ejemplo - presentaban miembros de ambos grupos. Uno de ellos comprendía a la proteína Argonauta-1 de *Arabidopsis* y a todas las proteínas caracterizadas como funcionales en los mecanismos de siARNs y miARNs, y fue denominado “subfamilia *Argonaute-like*”. El segundo grupo contenía a la proteína Piwi de *Drosophila* y a muchas de las proteínas demostradas en las vías de piARNs, y fue denominado “subfamilia *Piwi-like*”. Entre los resultados curiosos, el estudio dejó en evidencia que las plantas poseen únicamente proteínas “*Argonaute-like*”, mientras que los amebozoos, *Paramecium tetraurelia* y *T. thermophila* solo contienen proteínas “*Piwi-like*”.

Dado que ambas subfamilias se encuentran distribuidas por todo el dominio eucariota, encontrando proteínas “*Argonaute-like*” y “*Piwi-like*” en los unicelulares más basales, se ha propuesto - y es aceptado - que el último ancestro común de todos los eucariotas poseía al menos una proteína de cada subfamilia. Luego la evolución habría introducido numerosos eventos de duplicación génica (y en otros casos, deleciones linaje-específicas), con la consecuente diversificación y especialización funcional.

Del total de proteínas analizadas, hubo dos que no pudieron ser confidentemente asignadas a ningún grupo: las proteínas Argonauta de *G. lamblia* y la proteína Argonauta-1 de *T. brucei*. Esta última, fue la única de las proteínas de tripanosomátidos introducidas en el análisis, dado que muchas de las otras carecen de un dominio PAZ identificable.

3.3 El problema: las proteínas Argonauta en tripanosomátidos

Los tripanosomátidos (orden Kinetoplastidae) son un grupo de protozoarios parásitos flagelados, muchos de los cuales son patógenos para el hombre. Por sus implicancias biomédicas, los géneros más estudiados son *Trypanosoma* y *Leishmania*, y varias de sus especies cuentan con proyectos genoma completos o en estado avanzado. Más allá de su interés sanitario, estos microorganismos - que divergieron muy temprano del resto de los eucariotas - presentan características muy particulares a nivel de su expresión génica, lo que los ha hecho clásicos modelos de estudio. Su ciclo de vida parásita en diversos hospederos ha provocado su adaptación a ambientes hostiles altamente cambiantes, por lo que cuentan con sistemas eficientes para reprogramar la expresión de sus genes, metabolismo y morfología en función del ambiente (Siegel et al. 2011).

La transcripción en tripanosomátidos es policistronica, pero los genes transcritos en una misma unidad de transcripción (a veces, más de 100) no se encuentran - necesariamente - funcionalmente relacionados. En realidad, el énfasis en el control de la expresión génica parecería ser a nivel post-transcripcional, lo que hace de estos organismos modelos interesantes para el estudio de los mecanismos moleculares asociados a estas vías.

En su inmensa mayoría, los genes que codifican para proteínas de estos parásitos no contienen intrones; los transcritos policistronicos son procesados a ARNm maduros mediante un proceso acoplado de *trans-splicing* y poli-adenilación. Un mini-exón común (o *splice-leader*), capeado, se incorpora en la región 5' UTR de cada gen, y el corte provoca el procesamiento 3' y poliadenilación del gen ubicado corriente arriba. La maduración por *trans-splicing* provoca que los primeros nucleótidos de cada ARNm maduro sean comunes a todos los mensajeros. Otra particularidad que presentan estos organismos es la edición post-transcripcional del ARNm mediada por ARN, necesaria para la maduración de los mensajeros mitocondriales.

Curiosamente, las vías de silenciamiento post-transcripcional por pequeños ARNs y proteínas Argonauta - que tan importantes son para el resto de los eucariotas - parecerían estar presentes en algunos tripanosomátidos, y totalmente ausentes en otros. El silenciamiento génico por ARN interferencia fue tempranamente demostrado en *T. brucei* (Shi et al. 2000), y más recientemente, también en *T. congolense* y *L. braziliensis* (Lye et al. 2010). A pesar de varios esfuerzos al respecto, no ha sido posible silenciar genes por ARNi en el agente etiológico de la enfermedad de Chagas: *Trypanosoma cruzi* (DaRocha et al. 2004), *L. major* ni *L. infantum*. Dado que en ninguno de estos genomas se han detectado homólogos de proteínas Dicer ni de Argonautas - lo cual no es estrictamente cierto, como se discutirá a lo largo de esta tesis - se ha planteado que estos tripanosomátidos habrían perdido independiente toda la maquinaria molecular necesaria para la ARNi (Hutvagner and Simard 2008; Lye et al. 2010)

Trypanosoma brucei posee dos proteínas Argonauta en su genoma, una de las cuáles, TbAGO-1 (*Trypanosoma brucei argonaute-like protein 1*), es indispensable para la ARNi y su delección provoca un enlentecimiento en el crecimiento y defectos en la formación del huso mitótico (Durand-Dubief and Bastin 2003; Shi et al. 2004). La segunda proteína Argonauta de *T. brucei* carece de un dominio PAZ, y su delección no afecta el silenciamiento por ARNi ni provoca un fenotipo evidente. Dado que posee únicamente un dominio Piwi, los autores la denominaron TbPWI-1 (*Trypanosoma brucei piwi-like protein 1*).

Si bien los nombres de las proteínas TbAGO-1 y TbPWI-1 coinciden con los nombres de ambas subfamilias de proteínas Argonauta eucariotas, esto es una desafortunada casualidad, ya que la nomenclatura es previa a la definición filogenética de los grupos “Argonaute-like” y “Piwi-like” y obedece a la presencia/ausencia de un dominio PAZ (y por tanto, de un dominio Piwi aislado en TbPWI-1). Más aún, solamente TbAGO-1 fue incluida en análisis filogenéticos previos, y no pudo ser asignada a ninguno de ambos grupos (Cerutti and Casas-Mollano 2006; Farazi et al. 2008). En otros estudios filogenéticos, directamente no se incluyó ninguna Argonauta de tripanosomátidos (Murphy et al. 2008).

Mediante búsquedas simples por BLAST, pueden identificarse ortólogos de TbAGO-1 y TbPWI-1 en el resto de los tripanosomátidos. Mientras que proteínas ortólogas de TbAGO-1 pueden detectarse solo en algunos casos (*T. brucei*, *T. congolense*, *T. vivax*, *L. braziliensis*, todas conteniendo una típica arquitectura de dominios, con dominios PAZ y MID/PIWI fácilmente identificables), existe un ortólogo de TbPWI-1 en cada genoma en el que se buscó (*T. brucei*, *T. cruzi*, *T. congolense*, *T. vivax*, *L. major*, *L. infantum*, *L. braziliensis*), y en posición sinténica. Si bien todas estas proteínas presentan un dominio MID/PIWI conservado, parecerían carecer por completo del dominio PAZ.

En resumen, todos los tripanosomátidos analizados poseen al menos una proteína del tipo Argonauta, aunque las mismas carecerían del dominio PAZ (TbPWI-1 y sus presuntos ortólogos). Algunas especies poseen además un segundo gen Argonauta, con una arquitectura de dominios completa. Existe una correlación entre aquellas especies que poseen este segundo gen (TbAGO-1 y sus presuntos ortólogos) y aquellas especies en las que se ha demostrado o se presume un mecanismo de ARNi funcional, coincidente con el hecho de que TbAGO-1 – pero no TbPWI-1 – es esencial para la interferencia por ARN.

La ausencia de una función conocida para TbPWI-1 y sus presuntos ortólogos sumada a la ausencia de un dominio PAZ, han suscitado una falta de interés por esta familia de proteínas, lo que se ve claramente reflejado – por su ausencia – en la literatura científica. Sin embargo, llama la atención de que estas proteínas estén conservadas en todos los tripanosomátidos, y que para casos como *T. cruzi*, *L. major* o *L. infantum*, sean la única proteína “Argonauta” disponible, si bien hay autores que sostienen que estos protozoarios “habrían perdido todos sus genes Argonauta junto con el resto de la maquinaria de ARNi” (Hutvagner and Simard 2008).

Considerando que ahora sabemos que *T. cruzi* posee una población significativa de fragmentos derivados de tARNs específicos, y que dichos fragmentos hoy se aceptan como una nueva familia de ARNs pequeños reguladores, capaces de asociarse a proteínas Argonauta (Burroughs et al. 2011; Couvillion et al. 2010; Haussecker et al. 2010), resulta tentativo especular que las proteínas “Argonauta” de todos o algunos tripanosomátidos carentes de un dominio PAZ pudieran jugar un importante rol en la regulación de la expresión génica, y que dicho rol podría estar mediado por fragmentos de tARN, al menos en *T. cruzi*.

En estudios preliminares de nuestro grupo, hemos demostrado la expresión de la única proteína “Argonauta” de *T. cruzi* (que denominamos TcPIWI-tryp por razones que se discutirán en esta tesis) en distintos estadios de vida del parásito en cultivo (epimastigote, tripomastigote y amastigote), tanto a nivel del ARNm como de proteína.

En la presente tesis se abordará el estudio de esta proteína y del resto de las proteínas Argonauta de tripanosomátidos, ya sea aquellas que poseen o que carecen de un dominio PAZ. Comprender la

biología de proteínas Argonauta en organismos eucariotas unicelulares que divergieron muy temprano en la evolución podría permitir elucidar mecanismos ancestrales conservados en otros eucariotas, o bien adaptaciones específicas al ciclo de vida de estos parásitos. Además, *T. cruzi* se ofrece como un modelo privilegiado para el estudio de una presunta vía de regulación Argonauta-fragmentos de tARN (vía presente también en eucariotas superiores, y aparentemente implicada/afectada en algunos tipos de cáncer humano), ya que contaría con una única proteína Argonauta y con una única familia – identificada – de pequeños ARNs.

OBJETIVOS

1. HIPÓTESIS DE TRABAJO

Dado que *T. cruzi* posee una vasta población de pequeños ARNs no codificantes derivados de tARNs, nuestra hipótesis es que los mismos serían capaces de unirse a la proteína Argonauta de dicho organismo, a efectos de realizar una función regulatoria no aún conocida (distinta de la ARNi), pero tal vez conservada en otros organismos. En otras palabras, consideramos que la única proteína de *T. cruzi* que codifica para un dominio MID/PIWI es de hecho una proteína Argonauta, aunque carezca de un dominio PAZ, esencial para la función de este tipo de proteínas. Lo mismo sería aplicable para todo el resto de las proteínas Argonauta de tripanosomátidos, que carecen de un dominio PAZ. Más aún, dudamos que la ausencia de dicho dominio sea real, pues son conocidas las limitaciones para detectar dominios PAZ en proteínas de secuencia altamente divergente.

2. OBJETIVO GENERAL

Estudiar la única proteína Argonauta de *Trypanosoma cruzi* (TcPIWI-tryp; GenBank: ADI72731.1; TritrypDB: Tc00.1047053511367.240) y obtener evidencias a favor del rol de las proteínas Argonauta sin dominio PAZ en protozoarios patógenos en los cuales la maquinaria de la ARNi parece haberse perdido.

3. OBJETIVOS ESPECÍFICOS

- Estudio de las proteínas Argonauta de tripanosomátidos desde el punto de vista filogenético, procurando determinar cuáles de ellas pertenecen a las subfamilias “*Argonaute-like*” y “*Piwi-like*”.
- Estudio bioinformático de la proteína Argonauta de *Trypanosoma cruzi* (TcPIWI-tryp) y de todas las proteínas Argonauta de tripanosomátidos carentes de un dominio PAZ. Se busca comprender por qué estos organismos han conservado una proteína carente de un dominio tan relevante a la hora de unir ARNs pequeños.
- Obtención de la secuencia codificante completa de TcPIWI-tryp en la cepa Dm28c de *T. cruzi*, amplificación de la misma y clonado para la expresión recombinante en *Escherichia coli*. Purificación de la proteína en condiciones adecuadas – de ser posible – para la realización de estudios bioquímicos y estructurales.
- En vistas de lo ambicioso y arriesgado del objetivo anterior: amplificar, clonar, expresar en bacterias y purificar alguno o algunos de los dominios funcionalmente más importantes de la proteína, con miras a la realización de estudios bioquímicos y estructurales.

- Sobre-expresión de la proteína TcPIWI-tryp en el propio parásito, a fin de evaluar posibles fenotipos y facilitar la inmunoprecipitación y secuenciado de ARNs pequeños co-inmunoprecipitados.
- Obtención de suero policlonal anti-TcPIWI-tryp, de forma de poder desarrollar los ensayos de inmunoprecipitación antes mencionados.
- Evaluar si la fracción inmunoprecipitada contiene un enriquecimiento en fragmentos derivados de tARNs, y si existe enriquecimiento en fragmentos específicos.

MATERIALES Y MÉTODOS

Análisis filogenéticos

Las siguientes secuencias de aminoácidos correspondientes a proteínas Argonautas fueron extraídas de la base de datos de proteínas del *National Center for Biotechnology Information* (NCBI)⁵. En la medida de lo posible, secuencias curadas de la base RefSeq fueron utilizadas. Los números de acceso de las secuencias utilizadas se presentan a continuación:

Proteínas “Argonaute-like” y “WAGOs”

Alg-2 *C. elegans* NP_871992.1; Alg-1 *C. elegans* NP_510322.2; AGO *P.monodon* ABC68592.1; Ago3 *H.sapiens* Q9H9G7.2; Ago4 *H.sapiens* Q9HCK5.2; AGO4 *M.musculus* AAN75581.1; Ago1 *H.sapiens* Q9UL18.3; Ago2 *H.sapiens* NP_036286.2; AGO2 *M.musculus* AAN75580.1; Zwilli *A.thaliana* NP_199194.1; AGO1 *A.thaliana* NP_849784.1; AGO *S.pombe* CAA19275.1; AGO6 *A.thaliana* NP_180853.2; AGO4 *A.thaliana* NP_565633.1; AGO2 *A.thaliana* NP_174413.2; AGO_long *T.gondii* ABA70456.1; AGO2 *B.mori* NP_001036995.2; ppw-2 *C. elegans* NP_491535.1; ppw-1 *C. elegans* NP_740835.1; Sago-2 *C. elegans* NP_490758.1; Sago-1 *C. elegans* NP_504610.1

Proteínas “PIWI-like”

PIWI *D. melanogaster* AAF53043.1; PIWI *A.aegypti* EAT34046.1; Aubergine *D. melanogaster* CAA64320.1; Aubergine *B.mori* NP_001098066.2; prg-1 *C. elegans* NP_492121.1; prg-2 *C. elegans* NP_500994.1; PIWI *C.hemishaerica* ABY67112.1; PIWI1 *H.sapiens* (HIWI) NP_004755.2; PIWI1 *M.musculus* (MIWI) NP_067286.1; PIWI *G.gallus* ABR09543.1; PIWI *D. rerio* AF336369_1; PIWI4 *H.sapiens* Q7Z3Z4.2; PIWI *B.primigenus* BAG69146.1; PIWI2 *M.musculus* NP_067283.1; PIWI *S.mediterranea* ACC97187.1; PIWI *C.quinquefasciatus* EDS45390.1; PIWI *P.tetraurelia* CAI39076.1; PIWI *P.tetraurelia* CAI39075.1; PIWI *P.tetraurelia* CAI44468.1; Twi2p *T.thermophila* ABI17944.1; Twi7p *T.thermophila* ABP68416.1; Twi10p *T.thermophila* ABP68415.1; PIWI *G.lambli* XP_001707926.1

Argonautas procariotas

pAGO *S.fumaroxidans* ABK18163.1; pAGO *A.fulgidus* 2W42A; pAGO *P.furiosus* 1Z25A

Las secuencias de las proteínas Argonauta provenientes de tripanosomátidos fueron extraídas de la base de datos TritrypDB⁶ y se presentan los números de acceso correspondientes. Los nombres de las proteínas fueron asignados *a posteriori* en función de los resultados obtenidos

TbAGO-tryp *T. brucei* (TbAGO1) Tb927.10.10850; TcoAGO-tryp *T. congolense* TcIL3000.10.9150; TvAGO-tryp *T. vivax* TvY486_1010580; LbAGO-tryp *L. braziliensis* LbrM11_V2.0360; TbPIWI-tryp *T. brucei* (TbPIWI1) Tb927.10.2220; TcoPIWI-tryp *T. congolense* TcIL3000.10.1830; TvPIWI-tryp *T. vivax* TvY486_1002240; LmPIWI-tryp *L. major* LmjF21.0410; LiPIWI-tryp *L. infantum* LinJ21_V3.0470; LbPIWI-tryp *L. braziliensis* LbrM21_V2.0470. Por último, se incluyó la secuencia de la proteína TcPIWI-tryp de *T. cruzi* secuenciada por nuestro grupo en la cepa Dm28c, cuyo número de acceso en la base de datos de NCBI es: ADI72731.

Para la construcción de los árboles correspondientes a la secuencia de los dominios PIWI, se utilizaron los datos de anotación de las respectivas secuencias, o se analizaron mediante el programa CDD search⁷.

⁵ <http://www.ncbi.nlm.nih.gov/protein>

⁶ <http://tritrypdb.org/tritrypdb/>

Las secuencias se alinearon mediante el programa PRANKSTER, interface gráfica del programa PRANK⁸. El programa utiliza el modelo WAG de sustitución de aminoácidos, y se fijaron las penalizaciones por apertura y extensión de *gaps* en 1. Se definieron tres ciclos de iteración para la definición del árbol guía a partir de los alineamientos, y para realizar los alineamientos en función del árbol guía. Se incluyeron las opciones +F (*force*) y *post. Strict*. Otros programas de alineamiento múltiples utilizados (ClustalW, T-coffe) produjeron árboles de topología similar pero menor confianza estadística en sus nodos, por lo que no aparecen en la sección de resultados.

Con los alineamientos obtenidos se construyeron árboles por distintos métodos de inferencia filogenética. Para métodos bayesianos se utilizó el programa MrBayes 3.1⁹ y el modelo WAG de sustitución de aminoácidos. La resolución de la cadena de Markov por la técnica de Montecarlo (mcmc) se efectuó por 1.000.000 generaciones, tomando muestras cada 100 generaciones. Se utilizó la variante “*Metropolis-coupled Markov chain Monte Carlo*” (MCMCMC) utilizando cuatro cadenas de búsqueda o “*caminantes*”, siendo tres de ellas “*calientes*” y una “*fría*”. Habiendo alcanzado convergencia, se eliminaron las generaciones iniciales previo al estado estacionario.

Para la construcción del árbol filogenético realizado por el método de máxima verosimilitud, se utilizó el programa PhyML 3.0, disponible en línea¹⁰. Se utilizó el mismo alineamiento múltiple que en el caso anterior. Se utilizó el modelo de sustitución de aminoácidos WAG, el número de categorías discretas de tasas de sustitución se fijó en 4 y se dejó estimar a partir de los datos la proporción de sitios invariables y el parámetro de distribución gamma (obteniendo valores finales de 0,003714 y 3,6174 respectivamente). Los árboles iniciales se determinaron por el método BIONJ, y se mejoraron por NNI. No se adicionaron árboles al azar, se optimizaron los largos de las ramas y los parámetros del modelo de sustitución. Se utilizó aLRT (*SH-like*) en vez de bootstrap para asignar valores de precisión y exactitud para cada rama del árbol. Los árboles obtenidos se visualizaron con el programa MEGA 4.1

El mismo alineamiento múltiple fue introducido en el programa MEGA 4.1 para la construcción de árboles por el método de “*Neighbor-joining*”. La presencia de huecos en el alineamiento se trató mediante la opción “*pairwise deletion*”. En este caso el modelo de sustitución utilizado fue la corrección de Poisson, luego de numerosas pruebas de ensayo y error con otros métodos (que - vale aclarar - produjeron árboles de similar topología). Se incluyeron todas las sustituciones, con una tasa de variación homogénea entre linajes, y tasas uniformes entre sitios. Para establecer una medida de confiabilidad estadística de cada nodo, se realizaron 5000 pseudo-replicas por el método de “*bootstrap*”.

Predicción del dominio PAZ y modelado por homología

La secuencia completa de la proteína TcPIWI-tryp de *T. cruzi* en la cepa Dm28c fue introducida en la interface web del programa HHpred¹¹. Se seleccionó la base “pdb70” como base de datos de modelos de Markov ocultos (HMM) para realizar los alineamientos y la búsqueda de homólogos remotos. Se fijó un límite máximo de tres iteraciones de PSI-BLAST a efectos de la búsqueda de homólogos cercanos de TcPIWI-tryp para la construcción del modelo HMM correspondiente. Se habilitó la puntuación de la estructura secundaria predicha por PSIPRED, en el entendido que el peso asignado por el programa a dicha puntuación es bajo. Se eligió un modo de alineamiento local en vez de global y el resto de las opciones se dejaron por defecto.

Para el modelado por homología de los dominios PAZ-MID/PIWI de TcPIWI-tryp, se seleccionó el alineamiento de mayor puntaje realizado por HHpred (PfAGO; PDB: 1u04) y se introdujo automáticamente en la interface web del

⁷ <http://www.ncbi.nlm.nih.gov/cdd/>

⁸ <http://www.ebi.ac.uk/goldman-srv/prank/>

⁹ <http://mrbayes.csit.fsu.edu/>

¹⁰ <http://www.atgc-montpellier.fr/phyml/>

¹¹ <http://toolkit.tuebingen.mpg.de/hhpred>

programa Modeller. El modelo obtenido se visualizó y procesó gráficamente con el programa Pymol v.0.99¹², eliminando los aminoácidos no alineados y por consiguiente no modelados.

Cultivos de *T. cruzi*

Epimastigotes de la cepa Dm28c de *T. cruzi* fueron utilizados a lo largo de este trabajo. Los epimastigotes fueron crecidos a 28°C en medio LIT (*liver infusion tryptose*) suplementado con 10% de suero fetal bovino inactivado por calor. Los cultivos de parásitos fueron iniciados a partir de 1×10^6 células/mL y finalizados en fase logarítmica tardía a una densidad celular de 50×10^6 células/mL.

Extracción de ARN total y retrotranscripción del ARN a ADN copia

Un total de 1×10^7 células *T.* fueron centrifugados por 10min. A 3500g, resuspendidas en 1mL de reactivo del Trizol (Sigma), e incubadas por 5 minutos a temperatura ambiente de modo de lisar las células, desnaturalizar y solubilizar sus componentes -disociando los complejos nucleoproteicos e inactivando las ARNasas-. Posteriormente se procedió según descripción del fabricante, utilizando un protocolo adaptado del método de paso único desarrollado por Chomczynski & Sacchi (Chomczynski and Sacchi 1987, 2006). Brevemente, se realizó una extracción con 0.2 mL de cloroformo, se centrifugó, recuperó la fase acuosa, y se precipitó el ARN mediante el agregado de 0.5mL de isopropanol. El precipitado (en adelante, “*pellet*”) se lavó con etanol 75% y se volvió a centrifugar. Una vez seco, se resuspendió en agua libre de ARNasas a una concentración final de 1,77µg/mL. La integridad del ARN extraído fue evaluada por visualización en gel de Agarosa 1%/TBE.

Para la síntesis del ADN copia (ADNc), se tomaron 2µg de ARN total y se incubaron con 500ng de cebador oligo d(T)₂₀, 1µL de una solución 10mM de dNTPs y agua ultra pura en cantidad suficiente para 12µL. Se calentó la mezcla por 5 minutos a 65°C y se enfrió inmediatamente en hielo. De acuerdo a las instrucciones del fabricante (Invitrogen, Alemania), se adicionaron 4µL de tampón de síntesis de primera hebra y 2µL de una solución 0,1M de DTT. Se mezcló la reacción y se incubó por 2 minutos a 42°C de forma de permitir la hibridación del cebador oligo d(T) con las colas poliA de los ARN mensajeros maduros. Posteriormente, se añadieron 200U de la enzima retrotranscriptasa SuperScript II y se incubó a 42°C por 50 minutos, inactivando la reacción mediante calentamiento a 70°C por 15 minutos.

Obtención de la secuencia codificante completa de TcPIWI-tryp en la cepa Dm28c de *T. cruzi*

A partir del ADNc (correspondiente a la fracción de mensajeros maduros poliadenilados), se amplificaron por PCR diversas secuencias internas solapantes, específicas al gen de interés (tomando como referencia la secuencia depositada en las bases de datos). Las secuencias de todos los cebadores utilizados para esta actividad se presentan en la Tabla 1. Los productos de PCR se analizaron por electroforesis en gel de agarosa y se clonaron en el vector pGEM-T Easy (Promega), mediante clonaje “T”. Se transformó con dichos vectores recombinantes bacterias competentes de *Escherichia coli* (cepa XL1blue) y se secuenciaron tres clones para cada amplicón. Las secuencias obtenidas se ensamblaron posteriormente dado que eran parcialmente solapantes. De encontrarse diferencias en alguna posición entre secuencias correspondientes en distintos clones, se aumentó el número de clones secuenciados a cinco, de modo de poder discriminar entre posiciones polimórficas genuinas y artefactos de amplificación y/o secuenciación.

Para la obtención del ATG inicial y el extremo 5' codificante, se diseñó un cebador complementario al mini-exón (“SL”; en la región que es mantenida en los ARNm) y se amplificó por PCR el extremo 5', utilizando cebadores reversos internos (“R1”-“R4”). Para la amplificación del extremo 3', se utilizó un cebador directo interno a la secuencia codificante (“F9”), y un cebador reverso correspondiente a la región 3' no codificante (“R8”). Se realizó también una segunda PCR con un cebador correspondiente al gen que se encuentra río debajo de TcPIWI-tryp

¹² <http://www.pymol.org/>

(“R9”). Como cabría esperarse, el producto “F9-R9” no pudo amplificarse desde el ADNc ya que “R9” se encuentra río abajo del sitio de corte y poliadenilación, pero si desde el ADN genómico.

Tabla n°1: Cebadores de PCR utilizados. Se muestran las posiciones esperadas de inicio y final para cada producto de PCR (correspondientes a la secuencia anotada en CL Brener), y las obtenidas experimentalmente. N/D: no determinado. No Amp: no se obtuvo producto de PCR. Se colorean en verde las PCR destinadas a la obtención de los extremos 5’ y 3’. Todos los productos de PCR fueron amplificados del ADNc, salvo el F9-R9 que fue obtenido del ADNg y se colorea en celeste. En rojo se muestran las secuencias correspondientes a los sitios Bam HI (cebadores directos) y Hind III (cebadores reversos)

CEBADORES DIRECTOS		CEBADORES REVERSOS		PCR	ESPERADO	AMPLIFICADO
SL	CGCTATTATTGATACAGTTTCTG	R1	GG AAGCTT AGAGGTTTCGACGACACACCTCTT	SL-R1 (5')	?-424	37-424
F1	GGGGATCCATGTTTTGTTTTCCCTCATGATCT	R2	GG AAGCTT ACCTTCTGCGAACATTCACAAACG	SL-R2 (5')	?-726	37-726
F2	GGGGATCCCTCTCAACACGGCGAATAGTGGCGG	piR	CTTCCACAAGGAGGCATGATAAC	SL-R3 (5')	?-1325	37-1325
piF	GTGTTGTACGACGGTTCCTGCAG	R3	GG AAGCTT GGTCTCAAACACGCGAACACGGT	F1-R1 (5')	1 - 424	No Amp.
F3	GGGGATCCACGACAGCTTATTGTTGGGAGGGA	R4	CAAGACTTCATGCGGGGGTTCGTAG	F2-R2	208-726	208-726
F4	GTGGGGGAGACGGCACTGAAGC	R5	ACAGGCGTCCACGCCACAAC	pi1F-pi1R	679-1249	679-1249
F5	CGCCGTGCTGAAAGACCACAGC	R6	CCACAAACGTTTCATTAGAGGTGCAG	F3-R3	962-1325	962-1325
F6	GAAGTTCGCGGTTGGCACAG	R7	CCTCCAATTAACCCGGAAGAAAGAGCAGAGG	F3-R4	962-1913	962-1913
F7	CAATGTCGTTAAGGGAGCCGTCG	R8	CCCGACTTACGCTTATCAGCCATCG	F4-R4	1354-1913	1354-1913
F8	GGATGACTCCGTTGTTCCATCCTTCTAC	R9	CGCGAAATATAATCGCCGTGTCGCATG	F5-R5	1839-2628	1839-2628
F9	ACCTGCGACATGATCCCTGCACCTCTG			F6-R6	2575-3306	2575-3306
				F7-R7 (3')	2970-#433	N/D
				F8-R8 (3')	3021-#685	N/D
				F9-R7 (3')	3265-#433	No Amp.
				F9-R8 (3')	3265-#685	3265-#685
				F9-R9 (3')	3265-#1166	3265-#1166

Extracción de ADN genómico de *T. cruzi*

Se resuspendieron 750x10⁷ parásitos en 4mL de tampón A (0,2M Tris-Cl pH=8; 0,1M EDTA pH=8) y se agregaron 40µL de proteinasa K (10mg/mL) y 0,2mL de SDS 10%. Posteriormente, se realizó la extracción del ADN con solventes orgánicos agregando 0,8mL de NaCl 0,5M y 5mL de solución de extracción (1 parte de cloroformo/alcohol isoamílico 24:1; 1 parte de fenol y 2 partes de tampón A). Se homogenizó vigorosamente y se dejó toda la noche a temperatura ambiente. Se centrifugó a 825g por 10 minutos, se tomó la fase acuosa, y se repitió la extracción con solventes. Por último, se realizó una tercera extracción con igual volumen de cloroformo/isoamílico 24:1, de modo de extraer de la fase acuosa la fracción de fenol disuelta en agua. Después de centrifugar y tomar la fase acuosa, se precipitó la misma con 2 volúmenes de etanol absoluto, centrifugando a 13200g por 15 minutos. El precipitado se lavó con etanol 70% y se resuspendió en 4mL de tampón TE (10mM Tris-Cl pH=7,4; 1mM EDTA pH=8). Para eliminar el ARN de la muestra, se adicionaron 4µL de ARNasa A (10ng/mL), incubando toda la noche a temperatura ambiente. A efectos de eliminar la ARNasa, se repitió todo el protocolo de extracción con solventes descrito anteriormente. El ADN resultante, luego del lavado con etanol 70%, se resuspendió en 200µL de agua ultrapura (resistividad > 16Mohm) estéril y se almacenó a -20°C hasta su uso.

Amplificación por PCR de la secuencia codificante completa de TcPIWI-tryp

En volumen final de 50µL, se agregaron 5µL de tampón 10x + MgCl₂ (para concentración final de Mg²⁺ de 1,5mM); dATP, dGTP, dTTP y dCTP (Invitrogen) para concentración final de 0,25mM c/u; cebador INI-F y FIN-R para concentración final de 0,3µM cada uno (la secuencia de los cebadores se muestra en la sección correspondiente de resultados); 0,64ng de ADN genómico y 1,25U de “High fidelity PCR Enzyme Mix” (Fermentas, #K0192) compuesta por una mezcla de las enzimas Taq y Pfu polimerasas. Las reacciones de PCR se efectuaron en un termociclador “Gradient Palm Cycler” (Corbett Life Science) utilizando un programa compuesto por un paso de desnaturalización inicial de 3 minutos a 94°C y seguido por 10 ciclos de: 45 segundos a 94°C, 30 segundos a 57°C y 186 segundos a 68°C, seguidos de 25 ciclos en los cuales el tiempo de extensión se aumentó 10 segundos por ciclo. La secuencia de los cebadores INI-F y FIN-R se muestra a continuación, con los sitios de restricción para Bgl II y Sal I en verde y rojo respectivamente, y los codones ATG y TTA (reverso complementario de TAA) subrayados:

INI-F: GGAAGATCTATGCGAAAAACAGTTTTGGCCTTGC; **FIN-R:** TGCCTCGACTTAAAAAACCAACGTTTCATTAGAGGTG

Clonado de la secuencia TcPIWI-tryp en el plásmido pGEM-T Easy y transformación de *E. coli*

El producto de PCR de la secuencia de TcPIWI-tryp amplificada con los cebadores INI-F y FIN-R se corrió en un gel de agarosa 1% en TAE 1X (40mM Tris-Acetato, 1mM EDTA; pH=8) teñido con bromuro de etidio (1µg/mL) y analizado con luz ultravioleta. Se cortó del gel la banda de aproximadamente 3200pb y se purificó mediante el kit "QIAquick gel extraction" (QIAGEN). Para el agregado de adeninas salientes en los extremos 3' de los productos de PCR purificados, se agregaron 5µL del producto purificado (30ng/µL); 1 µL de tampón 10x de la enzima Taq polimerasa (Invitrogen); MgCl₂ para una concentración final de 1,5mM; dATP (Invitrogen) para una concentración final de 0,2mM; 5U de enzima Taq ADN polimerasa (Invitrogen) y agua ultrapura en cantidad suficiente para 10 µL. La reacción se incubó media hora a 70°C. Se tomaron 2 µL del producto de reacción y se agregaron 5µL del tampón "2x Rapid ligation buffer" (Promega Corp.); 50ng del plásmido-T abierto pGEM-T Easy (Promega Corp.); 3 Unidades Weiss de la enzima ADN ligasa del fago T4 (Promega Corp.) y agua ultrapura en cantidad suficiente para 10 µL. La ligación se incubó toda la noche a 4°C. Se transformaron 120 µL de bacterias de la cepa XL1Blue de *Escherichia coli* (que fueron previamente tratadas con CaCl₂ para volverlas competentes a la transformación según protocolo extraído de Seidman et al., 2001; eficiencia: 4x10⁵ ufc/µg) con 3µL del producto de ligación, incubando ambos 30 minutos en hielo, con un posterior *shock* térmico de 2 minutos a 42°C. Inmediatamente, se agregó medio LB (*Lysogeny broth*, Miller) hasta completar 2mL en tubos falcon de 15mL y se incubó por una hora a 37°C con agitación (220rpm) de modo que las bacterias puedan readaptarse. El cultivo se centrifugó, se resuspendió en un volumen de 200µL y se sembró en una placa de LB (25g/L) - Agar (15g/L) - Ampicilina (100µg/mL) - X-Gal (5-bromo-4-cloro-3-indolil-β-D-galactopiranosido; 80 µg/mL) - IPTG (Isopropil-β-D-1- tiogalactopiranosido; 0,5mM). La placa se incubó toda la noche a 37°C, observándose la presencia de 7 colonias de color blanco. La presencia del gen TcPIWI-tryp en todas ellas fue confirmada por PCR.

Purificación y secuenciado del plásmido recombinante pGEM/TcPIWI-tryp

Todas las preparaciones a pequeña escala de ADN plasmídico bacteriano ("*mini-preps*") se realizaron mediante un protocolo convencional de lisis alcalina con SDS y precipitación del ADN y las proteínas desnaturalizadas por neutralización con acetato de potasio en medio ácido. El ADN plasmídico se obtuvo del sobrenadante, se precipitó con etanol en presencia de sales y se resuspendió en agua ultrapura con ARNasa A, para eliminar el ARN.

La secuenciación del plásmido purificado a partir de colonias positivas para el inserto TcPIWI-tryp (colonias blancas resistentes a ampicilina, con presencia del inserto confirmada por PCR) se realizó en el servicio de secuenciación de ADN del IPMon (UBM), utilizando un secuenciador automático de 4 capilares, modelo ABI3130 (Applied Biosystems). Brevemente, se enviaron al servicio 5µL de una dilución a 100ng/µL del plásmido, estimada por visualización en gel de agarosa teñido con bromuro de etidio respecto a marcadores de tamaño de concentración conocida. Se utilizaron los siguientes cebadores para el secuenciado por el método de Sanger de regiones solapadas: INI-F1; F2; F4; R4; PAZR1; R5; F6; F8; piF; PIWI-F1; M13R. Nótese que todos los cebadores son complementarios a distintas regiones del inserto, a excepción del cebador M13R que es complementario a una región del vector pGEM-T Easy próxima al sitio de clonado. Las secuencias obtenidas fueron visualizadas con el programa Sequence Scanner¹³ y ensambladas con ayuda del programa Vector NTI v.10¹⁴. La secuencia completa del inserto TcPIWI-tryp clonado fue comparada contra la secuencia obtenida a partir del ARN mensajero, no detectándose diferencias.

Clonado de TcPIWI-tryp y sus dominios PAZ y MID/PIWI en vectores de expresión.

¹³ <https://products.appliedbiosystems.com/ab/en/US/adirect/ab?cmd=catNavigate2&catID=600583>

¹⁴ <http://www.invitrogen.com/site/us/en/home/LINNEA-Online-Guides/LINNEA-Communities/Vector-NTI-Community/vector-nti-software.html>

Para el clonado de la secuencia completa, se digirió el plásmido pGEM/TcPIWI-tryp purificado y secuenciado con 10U de las enzimas Bgl I (Fermentas), Bgl II (Fermentas) y Sal I (Roche) por 2 horas a 37°C en tampón "0" (fermentas) conteniendo 50mM Tris-Cl pH=7.5, 10mM MgCl₂, 100mM NaCl, 0,1mg/mL BSA (solución 1X). Según reporta el fabricante, las tres enzimas tienen su máximo de actividad en dicho tampón. La reacción se inactivó parcialmente calentando a 70°C durante 20 minutos y el producto de digestión se corrió en un gel de agarosa 1% teñido con bromuro de etidio en tampón TAE 1X. Se purificó la banda de 3200 pb utilizando el kit "GenElute Agarose Spin Columns" (Sigma) siguiendo las instrucciones del fabricante. El ADN purificado se precipitó ajustando la concentración de NaCl a 0,3M y agregando 2,5 volúmenes de etanol absoluto, incubando a temperatura ambiente por 2 horas y centrifugando a máxima velocidad por 15 minutos. El *pellet* resultante se lavó con etanol 70%, luego de lo cual se resuspendió en 30µL de agua ultrapura. En paralelo, se realizó una digestión del vector pQE-80_TEV con 10U de las enzimas BamHI (Fermentas) y Sal I (Roche) en tampón "SuRE/Cut Buffer B" (Roche) por 2 horas a 37°C, corriendo posteriormente en gel y purificando la banda de 4761 pb. La ligación entre cantidades equimolares de inserto y vector (100ng) se realizó por medio de la enzima T4 ADN ligasa (1 unidad Weiss; Fermentas), incubando en el tampón recomendado por el fabricante (40mM Tris-Cl, 10mM MgCl₂, 10mM DTT, 0.5mM ATP; pH=7,8 a 25°C) por 1 hora a temperatura ambiente, con posterior inactivación de la ligasa a 65°C por 10 minutos.

Para la transformación con el producto de ligación, se prepararon bacterias electro-competentes de *E. coli* (cepas XL1Blue y M15[pREP4]), siguiendo el protocolo descrito por Seidman (Seidman et al. 2001). Las mismas fueron divididas en alícuotas de 50µL a una densidad celular de aproximadamente 2×10^{11} /mL y almacenadas a -80°C con glicerol 10%. La eficiencia de transformación de las mismas se determinó utilizando el plásmido control pTZR cerrado, superando un valor de 1×10^7 ufc/µg de plásmido control.

El producto de ligación se diluyó diez veces en agua ultrapura, y se adicionó 1µL de la dilución fría en hielo a una alícuota de bacterias competentes. Luego de 5 minutos de incubación en hielo, se transfirió la mezcla a una cubeta de electroporación de 1mm de espacio entre placas (Sigma) y se colocó en un electroporador "Electro Cell Manipulator ECM 600" (BTX Harvard Apparatus), otorgando una onda eléctrica de caído exponencial a una diferencia de potencial de 2500V y una resistencia de 129Ω. Inmediatamente después del pulso eléctrico, se agregó 1mL de medio SOC (0,5% extracto de levadura, 2% triptona, 10mM NaCl, 2,5mM KCl, 10mM MgCl₂, 10mM MgSO₄, 20mM glucosa) y se transfirió a un tubo falcon de 15mL estéril, incubando a 37°C y 180 rpm por 1 hora previo a la centrifugación y sembrado en placa de LB-Agar conteniendo ampicilina.

La placa se dejó crecer toda la noche en estufa a 37°C, y las colonias resultantes fueron analizadas por "colony-PCR" para evidenciar la presencia del inserto TcPIWI-tryp. Las colonias positivas fueron crecidas en medio LB conteniendo ampicilina, y se purificó su ADN plasmídico por el método ya descrito. Los plásmidos purificados fueron analizados por restricción, digiriendo con la enzima Sal I, con la enzima Bgl II, o con las enzimas Sal I y Nde I. Se seleccionó una de las preparaciones de ADN plasmídico y se envió al servicio de secuenciador a efectos de verificar que la secuencia de TcPIWI-tryp se encontrase en marco con la secuencia codificante del vector. El plásmido recombinante obtenido se denominó pQE-80_TEV/TcPIWI-tryp.

El clonado de la secuencia codificante de TcPIWI-tryp en el vector pTEX se realizó del mismo modo, digiriendo al vector pGEM/TcPIWI-tryp con Bgl I, Bgl II y Sal I, y al vector pTEX con BamHI y Sal I, y ligando para obtener el plásmido pTEX/TcPIWI-tryp. Se realizó una segunda construcción digiriendo al vector pQE-80_TEV/TcPIWI-tryp con Eco RI y Sal I, al igual que el vector pTEX. Luego de la ligación, transformación, selección y recuperación del plásmido, se obtuvo una construcción destinada a la expresión en *T. cruzi* de la proteína TcPIWI-tryp conteniendo un tag N-terminal de 6 histidinas. Se denominó a dicha construcción pTEX_HisTEV/TcPIWI-tryp.

Para el clonado de las secuencias codificantes de los dominios PAZ y MID/PIWI se realizó un análisis bioinformático para elucidar las regiones a amplificar (ver resultados). Las mismas fueron amplificadas a partir del plásmido pGEM/TcPIWI-tryp con cebadores diseñados *ex profeso*, utilizando una mezcla de Taq y Pfu ADN polimerasas, y los mismos protocolos de reacción utilizados para la secuencia completa, pero disminuyendo los tiempos de

extensión. Los productos de PCR se precipitaron con etanol, se resuspendieron en agua y se digirieron con las enzimas Bgl II y Sal I. Los productos de digestión se corrieron en distintos geles, de los cuales se purificaron las bandas correspondientes y se ligaron al vector pQE-80_TEV digerido con BamHI y Sal I, del mismo modo a como fue descrito para el caso de la secuencia completa. Combinando los cebadores utilizados, se amplificaron distintas regiones de la secuencia codificante, dando lugar a las construcciones: pQE-80_TEV/Nter-PAZ; pQE-80_TEV/PAZ; pQE-80_TEV/PAZ-PIWI y pQE-80_TEV/PIWI.

Las secuencias de los cebadores utilizados para la amplificación por PCR de los dominios PAZ y MID/PIWI se muestran a continuación. En verde y rojo se muestran los sitios Bgl II y Sal I respectivamente. Como el extremo C-terminal del dominio PIWI coincide con el extremo C-terminal de la proteína, se utilizó el cebador **FIN-R** como cebador reverso para la amplificación del dominio MID/PIWI. La secuencia en subrayado corresponde a un codón "STOP", incluido para evitar el agregado de aminoácidos en el extremo C-terminal.

PAZF2: GGAAGATCTGGTGCCTTGCCTGAAGATTG; **PAZR2:** TGC**GTGAC**TTACTTCATGCGGGGTCGTAGCAG;
PIWI-F1: GGAAGATCTCCAACACGACGACTGGATGTCCTAC

Expresión de TcPIWI-tryp en *Escherichia coli*.

A partir de una colonia transformada, positiva y aislada en placa de LB-agar/ampicilina, se inocularon cultivos de 5mL de LB/ampicilina toda la noche a 37°C y 220 rpm. Pequeñas alícuotas se guardaron a -80°C con glicerol 20% para posteriores inoculaciones. Los pre-cultivos crecidos fueron diluidos en 1L de cultivo a razón de 1/100 y crecidos a 37°C y 220 rpm. en medio con ampicilina hasta alcanzar una densidad óptica a 600nm de 0,4-0,6. Dependiendo del ensayo, se indujo la expresión de la proteína recombinante con 1mM o 0,5mM IPTG y se mantuvo la inducción por 4 horas a 37°C, 8 horas a 30°C, o toda la noche a 25°C o 20°C. Para los ensayos de replegamiento, la inducción se realizó con 0,5mM IPTG y los cultivos inducidos se crecieron a 37°C por 4 horas.

Expresión de los dominios MID/PIWI y PAZ en *Escherichia coli*.

La expresión del dominio MID/PIWI se realizó igual que para el caso de la proteína completa. La expresión del dominio PAZ se realizó creciendo pre-cultivos en medio LB/ampicilina a 37°C toda la noche, y luego diluyendo a razón de 1/100 en medio 2xYT/ampicilina, y creciendo a 37°C y 220 rpm hasta una densidad óptica a 600nm de 1,0. La expresión de la proteína recombinante se indujo con 0,5mM IPTG, por 4 horas a 30°C o toda la noche a 20°C. Según la preparación, los volúmenes de medio de cultivo fueron de 1 o 3L, distribuidos en matraces de vidrio de 2L de capacidad, con 500mL de medio de cultivo por matraz.

Purificación de TcPIWI-tryp y del dominio MID/PIWI de la fracción insoluble.

Los cultivos fueron centrifugados para decantar el componente celular, lavados en tampón Tris-Cl 50mM pH=8; NaCl 150mM, centrifugados nuevamente, y las bacterias fueron resuspendidas y lisadas en Tris-Cl 50mM pH=8; NaCl 500mM, conteniendo lizosima de clara de huevo (Sigma) a 0,5mg/mL y Tritón X-100 0,1% v/v. Luego de 1 hora de incubación a temperatura ambiente, la lisis se completó (y se redujo la viscosidad del medio) mediante tratamiento con un equipo de ultrasonido Branson digital Sonifier por 1 minuto (a intervalos de 10s.), trabajando al 30% de su amplitud máxima. Se separó la fracción soluble de la insoluble por centrifugación a 10000g, a 4°C por 45 min. La fracción insoluble se resuspendió y lavó en Tris-Cl 50mM pH=8, NaCl 500mM, urea 2M, se centrifugó, y se resuspendió finalmente en Tris-Cl 50mM pH=8, NaCl 500mM, urea 8M y β-mercaptoetanol (BME) 10-20mM.

El extracto conteniendo a la proteína resolubilizada en urea 8M se purificó por cromatografía de afinidad de iones de níquel inmovilizados (en adelante, IMAC) al menos a las 24 horas de incubación en urea 8M, y luego de centrifugado a 10000g y/o filtrado por membranas de 0,45µm. El pH se mantuvo durante toda la purificación a un

valor de 8 en tampón Tris-Cl, NaCl 500mM, urea 8M. El extracto se introdujo a una concentración de imidazol de 20mM, y luego se lavó la columna con 10 volúmenes de tampón conteniendo imidazol 20mM, y otros 10 volúmenes de tampón conteniendo imidazol 50mM. La elución se realizó aumentando la concentración de imidazol a 500mM, y recogiendo fracciones de 1mL.

Replegamiento *in vitro* de la proteína TcPIWI-tryp y del dominio MID/PIWI.

El replegamiento en columna se realizó del mismo modo que la purificación por IMAC, pero se introdujo un paso de replegamiento luego del lavado con imidazol 50mM. Se probó agregar 10 volúmenes de columna bajando la concentración de urea a 0M y eluyendo posteriormente en imidazol 500mM sin urea, o bien incorporando un gradiente de urea 8M-0M (en 20 volúmenes de columna) previo a la elución, mediante el equipo de FPLC ÄKTAprime (Amersham, GE Healthcare). En algunos casos se agregó BME 10mM durante toda la purificación.

El replegamiento por dilución se realizó diluyendo la proteína purificada por IMAC en urea 8M en un volumen 50 veces mayor de tampón Tris-Cl (pH=6 o pH=8), NaCl 500mM, y conteniendo distintas combinaciones de agente reductor (GSH:GSSG 10mM:1mM; BME 10mM o DTT 5mM). La proteína fue agregada de a volúmenes de 50µL y a intervalos de 10s. entre agregados, manteniendo agitación moderada a 4°C. Finalizado el agregado de proteína, la disolución se dejó toda la noche a 4°C con agitación suave. Al día siguiente, se centrifugó para eliminar material insoluble y la proteína replegada se concentró por centrifugación (utilizando membranas de 30KDa de corte), previo a su análisis por cromatografía de exclusión molecular (en adelante, SEC).

El replegamiento por diálisis se realizó introduciendo 10mL de proteína purificada por IMAC en urea 8M en una membrana de diálisis benzoilata de 2KDa de paso (Sigma), y dializando contra un volumen de 200mL de tampón Tris-Cl pH=8, NaCl 500mM, BME 2mM (o DTT 1mM) toda la noche, con dos cambios de tampón al día siguiente, de tres horas cada uno. Se centrifugó para eliminar material insoluble y la proteína replegada se concentró por centrifugación, previo a su análisis por SEC.

El análisis por SEC se realizó utilizando el equipo de FPLC ÄKTApurifier (Amersham, GE Healthcare) y una columna Superdex 200 10/300 (Amersham, GE Healthcare), equilibrada en el mismo tampón en que se realizaron los ensayos de replegamiento. Las columnas de SEC y equipos de FPLC pertenecen a la Unidad de Proteínas Recombinantes del IPMon (UPR).

Como norma general, todo análisis por SDS-PAGE de la proteína fue realizado en condiciones reductoras y no reductoras, para evidenciar la formación de agregados covalentes.

Purificación del dominio PAZ por IMAC, clivaje del 6xHis-tag, purificación de la proteína clivada por IMAC reversa y purificación final por cromatografía de exclusión molecular (SEC).

La lisis de las bacterias transformadas e inducidas para la expresión del dominio PAZ se realizó de manera similar a lo descrito para la proteína completa y el dominio MID/PIWI, con la diferencia de que la proteína se recuperó de la fracción soluble luego de la centrifugación a 10000g. En muchos casos, las bacterias se centrifugaron inmediatamente después de los ensayos de expresión, se lavaron y resuspendieron en el tampón de lisis con lisozima, e inmediatamente se congelaron a -80°C. Días después, se descongeló y se dejó completar la lisis, se sonicó, separó la fracción soluble, y se procedió inmediatamente a la purificación por IMAC.

La purificación por IMAC se realizó manteniendo las mismas concentraciones de pH y de sal durante toda la purificación (Tris 50mM pH=8; NaCl 500mM), introduciendo el extracto con imidazol 20mM, lavando con 10 volúmenes de columna y una concentración de imidazol de 20mM, y luego otros 10 volúmenes con imidazol 60mM. La elución se realizó elevando la concentración de imidazol a 500mM y colectando fracciones de 0,5mL. Luego de probar distintos agentes reductores y a distintas concentraciones antes, durante y después de la purificación, en los protocolos finales se incubó el extracto con BME 4mM, se filtró por 0,22µm inmediatamente

antes de la inyección en la columna, se lavó con tampones libres de agente reductor, y se introdujo DTT 1mM inmediatamente después de eluída la proteína, para evitar su desestabilización, agregación y/o precipitación.

La proteína purificada se digirió agregando la proteasa del virus del mosaico del tabaco (en adelante: proteasa TEV, conteniendo un sitio 6xHis-tag) a razón de 1mg de proteasa por cada 40mg de proteína PAZ (estimada espectrofotométricamente), en Tris 50mM pH=8, NaCl 500mM, Imidazol 500mM, DTT 1mM, glicerol 5%. La digestión se realizó a 4°C dentro de una membrana de diálisis, dializando toda la noche contra el mismo tampón sin proteasa ni imidazol, y ajustando el volumen de modo tal que la concentración de imidazol en el equilibrio fuese 20mM. El producto de digestión proteolítica se separó de la proteasa, del péptido clivado, de los remanentes de proteína sin clivar y de proteínas de *E. coli* con afinidad intrínseca por el níquel mediante una segunda IMAC, recogiendo el dominio PAZ purificado y clivado en la fracción no retenida por la columna. Luego de lavar la columna con concentraciones crecientes de imidazol, se eluyó la proteasa TEV, la proteína no clivada y el péptido clivado con imidazol 500mM.

La proteína clivada se purificó hasta homogeneidad electroforética por SEC. Brevemente, 5mL de la preparación proteica se inyectaron en una columna Superdex 75 16/600 (Amersham, GE Healthcare), y se corrió en tampón Tris 20mM pH=8, NaCl 150mM, DTT 1mM, a un flujo de 1mg/mL mediante el equipo de FPLC ÄKTApurifier (Amersham, GE Healthcare). Las fracciones correspondientes al pico de elución de la proteína pura se juntaron y concentraron a 10mg/mL y se almacenaron a -80°C hasta su uso. Dado que la proteína fue purificada en condiciones reductoras, se la denominó PAZ-SH para enfatizar el estado de oxidoreducción de sus cisteínas.

Purificación de la proteína PAZ-IAM

El protocolo de obtención y purificación de la proteína PAZ con todas sus cisteínas bloqueadas (PAZ-IAM) fue el mismo que el descrito para la proteína no modificada, con la diferencia que el extracto soluble de proteínas de *E. coli* fue tratado con BME 4mM para reducir los tioles proteicos, y luego de 15 minutos, con 2-iodoacetamida (IAM) 15mM. La reacción se dejó proceder por una hora en hielo, luego de lo cual se introdujo el extracto en una columna de IMAC para su purificación. Los restantes pasos de purificación se realizaron igual que para la proteína no tratada. Se agregó DTT 1mM después de la elución de la primera IMAC, no porque fuera necesario, sino a causa de que es requerido para la acción enzimática de la proteasa TEV, así como a efectos comparativos. Del mismo modo, la purificación por SEC se realizó también en DTT 1mM, de modo de poder comparar los perfiles de elución de ambas proteínas (PAZ-SH y PAZ-IAM). El DTT fue posteriormente eliminado por diálisis, realizando varios cambios de solución tampón para asegurar su remoción completa.

Caracterización de la proteína PAZ-SH y PAZ-IAM por DLS y Espectrometría de masas

Las medidas de dispersión dinámica de luz (DLS) se realizaron en la Unidad de Biofísica de Proteínas del IPMon (UBP), utilizando un equipo Zetasizer Nano-S (Malvern Instruments, UK). Para realizar tales medidas, la concentración de proteína se ajustó a 0,5mg/mL y se centrifugó a 16000g para eliminar material particulado (principalmente partículas de polvo). Las medidas se realizaron a 4°C y a temperatura ambiente.

Ambas proteínas fueron analizadas por espectrometría de masas (mapeo peptídico por MALDI-TOF y análisis por MALDI-TOF-TOF de los péptidos conteniendo cisteínas) para corroborar la eficacia del tratamiento con iodoacetamida, además de obtener una medida que sirviese como control de calidad de la preparación y del proceso. Ambas proteínas fueron digeridas con tripsina siguiendo el protocolo utilizado por la Unidad de Bioquímica y Proteómica Analítica del IPMon (UByPA) y los espectros de masa MS y MS-MS fueron adquiridos por personal de dicha Unidad utilizando un equipo 4800 MALDI-TOF-TOF (Applied Biosystems). Para la identificación de la proteína y de los péptidos correspondientes a los valores de m/z experimentales se utilizó el programa MASCOT, admitiendo carbamidometilación en residuos de cisteína y oxidación de metioninas.

Estabilidad de las proteínas PAZ-SH y PAZ-IAM a temperatura ambiente

Además de estudios simples de estabilidad en el tiempo (medida de la concentración luego de centrifugar a 16000g a distintos intervalos de tiempo), se realizó un estudio de la estabilidad de ambas proteínas frente a la oxidación. Dicho estudio debió ser diseñado considerando que la proteína PAZ-SH debe ser mantenida en una concentración de DTT 1mM en todo momento, o rápidamente comienza a agregarse y precipita.

Brevemente, las proteínas congeladas a -80°C a una concentración de 10mg/mL fueron descongeladas en hielo, y diluidas a 0,25 mg/mL en Tris-Cl pH=8 20mM, NaCl 150mM, DTT 1mM (tampón preparado en el momento). Los espectros UV-visible en el rango 230-500nm fueron registrados mediante un espectrofotómetro UV-Visible Cary 50 Bio (Varian) y corregidos contra el espectro del blanco (correspondiente al mismo tampón, sin proteína). Se registraron espectros a temperatura ambiente de cada proteína a $t=0, 8, 12, 16, 20$ y 24 min (post dilución e incubación a temperatura ambiente). La velocidad de aumento de las absorbancias a 280nm y a 310nm (atribuibles al aumento en la concentración del DTT oxidado) fue comparada para cada proteína con la velocidad de aumento esperable por la oxidación espontánea del DTT en tales condiciones experimentales. Para determinar esta velocidad, se realizaron estudios cinéticos durante 20 minutos, midiendo la absorbancia a 280nm y a 310nm del tampón Tris-Cl pH=8 20mM, NaCl 150mM, DTT 1mM.

Ensayos de cristalogénesis de las proteínas PAZ-SH y PAZ-IAM

Para la proteína PAZ-SH se realizó un paneo robótico de 384 condiciones de cristalización, utilizando un equipo robótico HoneyBee 963 (Genomic Solutions, UK) y el kit Classics-E8 (QIAGEN). Dicho equipo utiliza el método de difusión de vapor por gota sentada, mezclando 0,3 μL de reservorio con 0,3 μL de proteína (a 10mg/mL en Tris 20mM pH=8, NaCl 150mM, DTT 1mM). Las placas fueron mantenidas a 20°C y observadas a distintos tiempos.

Los ensayos de optimización se realizaron utilizando placas de 24 pocillos, mediante el método de difusión de vapor por gota colgante. El volumen de reservorio se mantuvo en 1mL, y se mezcló 2 μL de reservorio con 2 μL de proteína. Se realizaron dos gotas por condición variando la concentración de DTT en la que se encontraba inicialmente la proteína (1mM y 5mM). Se testearon 6 condiciones, todas con tampón HEPES 100mM pH=7.5, pero variando la concentración del agente precipitante (Jeffamina D-600, Sigma): 6%, 8%, 10%, 12%, 14%, 16% v/v . En un ensayo de optimización independiente, se elevó la concentración de proteína de partida de 10mg/mL a 20mg/mL (en Tris 20mM pH=8, NaCl 150mM, DTT 1mM), y se utilizó un rango más amplio de concentraciones de Jeffamina: 6%, 10%, 14%, 18%, 22%, 26% v/v .

Para la proteína PAZ-IAM se realizó únicamente un paneo robótico de 384 condiciones, pero utilizando el kit JSCG Core Suite (QIAGEN). La proteína se utilizó a 8.5mg/mL en Tris 20mM pH=8, NaCl 150mM (sin agente reductor).

Todos estos ensayos se realizaron en la Unidad de Cristalografía de Proteínas del IPMon (PXF).

Obtención de suero policlonal anti-TcPIWI-tryp

La proteína TcPIWI-tryp, obtenida de la fracción insoluble de *E. coli* y purificada por IMAC en presencia de urea 8M, fue utilizada para inmunizar un conejo (New Zealand) por parte de personal calificado de la Unidad de Biomodelos Animales de Experimentación de la Facultad de Medicina (UdelaR). Se realizó una primera inmunización con adyuvante completo de Freund (50%), seguido de tres recuerdos con adyuvante incompleto de Freund los días +35, +77 y +110. En todos los casos, la concentración de proteína fue de 240 $\mu\text{g/mL}$. Se tomó una muestra de sangre a los 20 días del segundo recuerdo (obteniendo así el antisuero denominado Full2B) y se realizó una sangría final a los 23 días del tercero (obteniendo así el antisuero denominado Full3B). El título de anticuerpos inmunoreactivos fue evaluado por ELISA indirecto, sensibilizando placas MaxiSorp (Nunc) con la proteína recombinante TcPIWI-tryp (replegada, sin urea), el dominio PAZ o un extracto proteico de epimastigotes de *T. cruzi*.

Sobre-expresión de TcPIWI-tryp en *Trypanosoma cruzi*

Epimastigotes de *T. cruzi* (5×10^7) crecidos en fase exponencial fueron electroporados con 50µg el plásmido pTEX/TcPIWI-tryp , mediante dos pulsos de 450V y 500µF. Los parásitos electroporados se crecieron en medio LIT suplementado con suero fetal bovino 10% por 24hs, y luego seleccionados en neomicina (250 µg/mL por 24hs, y luego 500 µg/mL). Como control, se transfectaron parásitos con el vector pTEX cerrado, que fueron utilizados a efectos comparativos.

Inmunoprecipitación de TcPIWI-tryp

Esferas magnéticas (Dynabeads, Invitrogen) conteniendo anticuerpos anti-IgG de conejo fueron lavadas con PBS, incubadas con el suero policlonal reactivo a TcPIWI-tryp (suero Full 3B, en PBS/nonidet 0,5%), y vueltas a lavar dos veces en PBS/nonidet 0,5%. Posteriormente, las mismas se incubaron con extractos de epimastigotas de *T. cruzi* (sobre-expresantes de TcPIWI-tryp), lisados en condiciones no desnaturalizantes (Tris-Cl pH=8 20mM; NaCl 150mM; glicerol 10%; nonidet 0,5%, cocktail inhibidor de proteasas 1x), incubando toda la noche a 4°C en rueda giratoria. Las esferas aisladas fueron lavadas tres veces con PBS/nonidet 0,5%, dos veces con PBS, y finalmente aisladas magnéticamente para la extracción de proteínas y ARN por el método del Trizol.

Secuenciación de pequeños ARNs que co-inmunoprecipitan con TcPIWI-tryp

Las esferas magnéticas aisladas luego de la inmunoprecipitación de TcPIWI-tryp fueron incubadas con el reactivo de Trizol, realizando una extracción de ARN según descrito anteriormente. El ARN obtenido fue fraccionado en un gel de poliacrilamida desnaturalizante al 15%, seleccionando la población entre 18 y 40nt. Se ligaron adaptadores 5' y 3' conteniendo sitios Ban I, se amplificaron las secuencias por PCR utilizando cebadores complementarios a los adaptadores, se digirieron con la enzima Ban I, y se ligaron para dar origen a concatámeros. Los mismos se clonaron en el vector pGEM-T Easy, con el cual se transformaron bacterias de *E. coli*. Los clones positivos para la presencia de los concatámeros fueron enviados al servicio de secuenciado (Sanger) del IPMon.

RESULTADOS EXPERIMENTALES

1. Estudios filogenéticos de las proteínas Argonautas permiten inferir un origen evolutivo común a las Argonautas de tripanosomátidos.

Como se explicó en la introducción, la nomenclatura y clasificación de las proteínas Argonauta de tripanosomátidos en las subfamilias “*Argonaute-like*” o “*Piwi-like*” se ha realizado por criterios arbitrarios, ya que presentan un porcentaje de identidad demasiado bajo con el resto de las argonautas eucariotas como para una clasificación confiable. En la gran mayoría de los casos, no se dispone de evidencia funcional suficiente como para efectuar una clasificación por similitud de función (salvo para el caso de la proteína TbAGO1) y no existen – a nuestro saber – estudios filogenéticos destinados al estudio de las proteínas Argonautas de tripanosomátidos.

Para intentar zanjar esta cuestión, se decidió realizar estudios de filogenia que incluyeran numerosas proteínas Argonauta eucariotas de ambas subfamilias, y todas las proteínas Argonauta (según ORFs predichos) de tripanosomátidos que pudiesen ser adquiridas de las bases públicas de datos. También se decidió incluir algunos representantes de Argonautas procariotas, lo que hasta entonces tampoco había sido realizado. Desde un punto de vista estratégico, se decidió utilizar las secuencias de aminoácidos completas de las distintas proteínas para el análisis, así como también las secuencias del dominio PIWI en análisis independientes.

La figura 5 muestra el árbol obtenido mediante métodos bayesianos a partir de las secuencias completas de numerosas proteínas Argonauta. Para mayor claridad, los valores de probabilidad posterior de cada nodo (entre 0 y 1) fueron multiplicados por un factor de 100. Se reproduce lo reportado en la literatura, y aceptado hasta el momento, en cuanto a la formación de dos grupos parálogos que reúnen todas las proteínas eucariotas previamente identificadas como “*Argonaute-like*” y “*Piwi-like*”. La monofilia de ambos grupos es apoyada por un nodo de probabilidad 1, y en conjunto, también forman un grupo monofilético. Si se considera que en el análisis se introdujeron proteínas de organismos que abarcan todo el espectro de la escala biológica, desde procariotas, protozoarios, hongos filamentosos y levaduras, plantas, animales invertebrados y vertebrados, los resultados obtenidos son consistentes con las hipótesis realizadas por Cerutti y Casas-Mollano (2006) respecto al origen de las proteínas Argonauta eucariotas.

La mayor originalidad de nuestro análisis respecto a lo reportado en la literatura fue la inclusión de 11 secuencias de proteínas Argonauta pertenecientes a distintos tripanosomátidos de los géneros *Trypanosoma* y *Leishmania*. Contrario a lo que cabría esperarse, la proteína TbAGO1 no se agrupa con las proteínas “*Argonaute-like*” a pesar de ser esencial para el fenómeno de la interferencia por ARN, ni la proteína TbPIWI1 pertenecería al grupo de las “*Piwi-like*”. En cambio, todas las proteínas Argonauta de tripanosomátidos forman un grupo monofilético apoyado por un nodo de probabilidad 1. Consecuentemente, el nodo que agrupa a las subfamilias “*Argonaute-like*” y “*Piwi-like*” y que deja por fuera al grupo de los procariotas y de los tripanosomátidos, también tiene probabilidad de 1.

Varias lecturas complementarias pueden hacerse a raíz de esta observación. Tal vez la más inmediata se refiere al problema de nombrar a las proteínas argonautas de tripanosomátidos con los sufijos “AGO” y “PIWI”, ya que ninguna de ellas pertenecen filogenéticamente a una u otra subfamilia. Por el contrario, forman en conjunto un nuevo grupo al que hemos dado el nombre de AGO/PIWI-tryp, enfatizando con la terminación “tryp” que se trata de una nueva subfamilia específica de tripanosomátidos.

Dentro de la subfamilia AGO/PIWI-tryp, las secuencias - de manera análoga a lo que ocurre con el resto de las argonautas eucariotas - se subdividen a su vez en dos grupos monofiléticos con una alta confianza. Uno de los grupos contiene a la proteína TbAGO1 y proteínas homólogas en *T. congolense*, *T. vivax* y *L. braziliensis*. Para establecer una nomenclatura consistente con nuestros resultados filogenéticos pero que no altere demasiado a los nombres históricos ya arraigados, denominaremos a dicho sub-grupo como “AGO-tryp”. De este modo, la proteína TbAGO1 pasaría a llamarse TbAGO-tryp, y lo mismo para las demás. El otro grupo contiene a la proteína TbPIWI1, y presenta homólogos en todos los tripanosomátidos incluidos en el análisis. Utilizando el mismo criterio, llamaremos a dicho sub-grupo “PIWI-tryp”, y a las proteínas que lo componen TbPIWI-tryp (en vez de TbPIWI1), TcoPIWI-tryp (*T. congolense*), TvPIWI-tryp (*T. vivax*), TcPIWI-tryp (*T. cruzi*), LmPIWI-tryp (*L. major*), LiPIWI-tryp (*L. infantum*), LbPIWI-tryp (*L. braziliensis*), por nombrar sólo aquellas especies incluidas en el análisis.

La nomenclatura propuesta únicamente agrega la terminación “-tryp” a los nombres de todas las argonautas de tripanosomátidos (enfatizando que pertenecen a la subfamilia “AGO/PIWI-tryp” y no a las subfamilias “Argonaute-like” ni “Piwi-like”), y el sufijo “AGO” o “PIWI” según pertenezcan a los sub-grupos representados por TbAGO-tryp y TbPIWI-tryp respectivamente. Al primer grupo pertenecen proteínas con dominios PAZ y MID/PIWI identificables, y lo componen organismos con interferencia por ARN demostrada o probable. Al segundo grupo pertenecen proteínas a las que no es posible asignar un dominio PAZ, y se encuentran presentes en todos los tripanosomátidos analizados.

Como los dominios proteicos pueden comportarse como unidades evolutivas independientes, se realizaron también análisis filogenéticos utilizando únicamente secuencias curadas de dominios PIWI. Un árbol de dominios PIWI construido con el método de “Neighbour-joining” se muestra en la figura 6. Por simplicidad, solo se muestran los valores de *bootstrap* de los nodos que definen cada una de las subfamilias a las cuales nos hemos estado refiriendo. Como se ve, la topología del árbol es similar y comparable a la de aquellos árboles construidos en base a las secuencias proteicas completas como el de la figura 5. Por fuera del árbol y a efectos comparativos, se muestran entre paréntesis los valores de *bootstrap*, aLRT y probabilidad posterior bayesiana (x100) de los nodos correspondientes en los árboles construidos con las secuencias completas y los métodos de “Neighbour-joining”, “Maximum-likelihood” y bayesianos, respectivamente. En conjunto, es posible afirmar que la monofilia del grupo AGO/PIWI-tryp se encuentra bien respaldada, tanto a nivel de la secuencia completa como del dominio PIWI.

Llama poderosamente la atención que todos los tripanosomátidos introducidos en el análisis poseen miembros del sub-grupo “PIWI-tryp”, de función desconocida hasta el momento. Más aún, muchos autores no reconocen siquiera a estas proteínas como proteínas Argonauta, o las caracterizan como pseudogenes, dado que carecen de un dominio PAZ (Atayde et al. 2011; Hutvagner and Simard 2008; Lye et al. 2010). Por el contrario, los ortólogos de TbAGO-tryp, una proteína que sí desempeñaría funciones canónicas de silenciamiento mediado por ARNs pequeños, se encuentran solo en algunas especies, habiéndose perdido en numerosos casos y en especies de ambos géneros: *Trypanosoma* y *Leishmania*.



Figura 5: Análisis filogenético de las proteínas Argonata de tripanosomátidos. Las proteínas Argonatas procariotas incluídas se señalan con el nombre de “pAGOs”. La barra gris indica un sub-tipo de proteínas Argonata específicas de nematodos, que algunos autores han llamado “Wagos”. El árbol fue contruido con el programa MrBayes 3.1 utilizando una matriz de sustitución del tipo WAG. Los valores delante de cada nodo corresponden a la probabilidad posterior bayesiana multiplicada por un factor de 100.

Tal vez las preguntas más inquietantes que se desprenden de este primer capítulo de resultados sean las siguientes: ¿Por qué todos los tripanosomátidos han mantenido sus genes PIWI-tryp, mientras que muchos de ellos han perdido por completo sus genes AGO-tryp junto con el resto de la maquinaria del ARN interferente? ¿Qué función regulatoria cumplirán las proteínas PIWI-tryp, que justifique la presión evolutiva para mantener estas proteínas en vez de las AGO-tryp? Si las proteínas PIWI-tryp son filogenéticamente proteínas Argonata ¿Qué ARNs pequeños unen? Si las proteínas PIWI-tryp heredaron un dominio PAZ, esencial a la función de estas proteínas (pues las AGO-tryp lo tienen) ¿Será posible que lo hayan perdido? ¿O habrá el mismo divergido lo suficiente como para volverse indetectable frente a los algoritmos bioinformáticos clásicos de predicción de dominios?

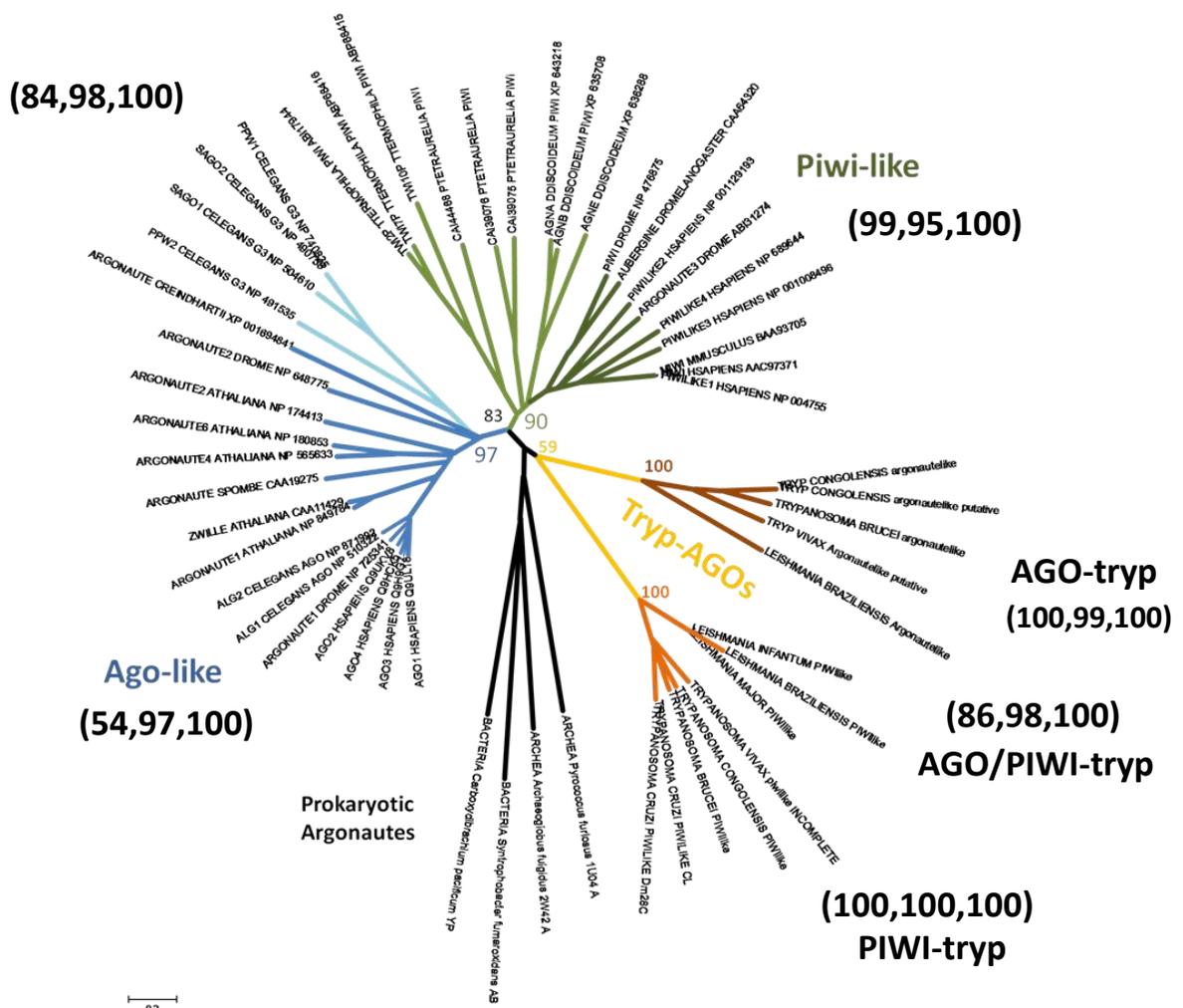


Figura 6: Árbol filogenético construido con las secuencias de los dominios PIWI de Argonautas eucariotas y procariotas. El árbol fue construido con el programa Mega 4.0, utilizando el método de “Neighbor-joining” y una corrección de Poisson para las distancias. Junto a los nodos que definen las distintas subfamilias y grupos, se muestran con el color correspondiente los valores de *bootstrap* (5000 pseudo-replicas). Por fuera del árbol, se muestran entre paréntesis los valores de *bootstrap*, aLRT y probabilidad posterior bayesiana (multiplicada por 100) de los respectivos nodos en los árboles construidos con las secuencias completas y utilizando los métodos de “Neighbor-joining”, “Maximum-likelihood”, y bayesianos (en ese orden).

2. Predicción de dominios en TcPIWI-tryp: detección del dominio PAZ previamente esquivo.

Como se explicó anteriormente, todas las proteínas Argonauta poseen una misma arquitectura de dominios, caracterizada por un dominio N-terminal variable, un dominio PAZ, y un dominio PIWI C-terminal (que suele ser subdividido en dominios MID y el dominio PIWI propiamente dicho). Inspeccionando el genoma de *T. cruzi* así como el de algunos otros tripanosomátidos, no se encuentra ningún marco abierto de lectura para el cual pueda predecirse dicha arquitectura de dominios. Únicamente TcPIWI-tryp destaca por poseer un dominio PIWI identificable por algoritmos clásicos basados en alineamientos de secuencias (NCBI CDD search¹⁵ y Pfam search¹⁶). Sin embargo, al carecer de un dominio PAZ evidente, no es posible catalogar a esta proteína como un miembro de la familia Argonauta. En la figura 7 se muestran capturas de pantalla al día 30 de abril de 2011 de ambos servidores, luego de introducir la secuencia de la proteína TcPIWI-tryp secuenciada por nuestro grupo¹⁷. Cabe señalar que la predicción de dominios no se ve afectada al introducir la secuencia de CL Brener ni ninguno de los ortólogos de esta proteína en los demás tripanosomátidos. El dominio PAZ tampoco aparece aún subiendo el E-valor (*expectation value*) a los valores más altos.

Numerosos autores de reconocida trayectoria en este campo ya habían destacado este caso paradigmático presente en algunos pocos eucariotas unicelulares, entre ellos *T. cruzi*. Gyorgy Hutvagner y Martin J. Simard, en una revisión muy citada de la revista Nature (2008) afirman que “*existen ejemplos de simplificación y pérdida de proteínas Argonauta en organismos taxonómicamente divergentes (...) S. cerevisiae así como algunos parásitos tales como T. cruzi y L. major habrían perdido sus genes Argonauta junto con toda la maquinaria de silenciamiento por ARN de manera independiente*”. Y en otro fragmento de la misma publicación: “*En algunos organismos como Trypanosoma cruzi y Leishmania major, las proteínas Argonautas habrían perdido su dominio PAZ y retenido únicamente el dominio PIWI. Sin embargo, estas proteínas serían o bien insensibles al ARN de doble hebra o no serían necesarias para la interferencia por ARN*” (Hutvagner and Simard 2008)

¹⁵ <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

¹⁶ <http://pfam.sanger.ac.uk/search>

¹⁷ La secuencia en cuestión corresponde a la traducción *in silico* de la secuencia codificante de la proteína TcPIWI-tryp, en la cepa Dm28c de *T. cruzi*. La metionina inicial corresponde al primer codón ATG desde el sitio de inserción del mini-exón, como se describirá en la sección 3 del capítulo de resultados. La numeración de residuos realizada en esta sección es la correspondiente a esta secuencia (GenBank: ADI72731.1).

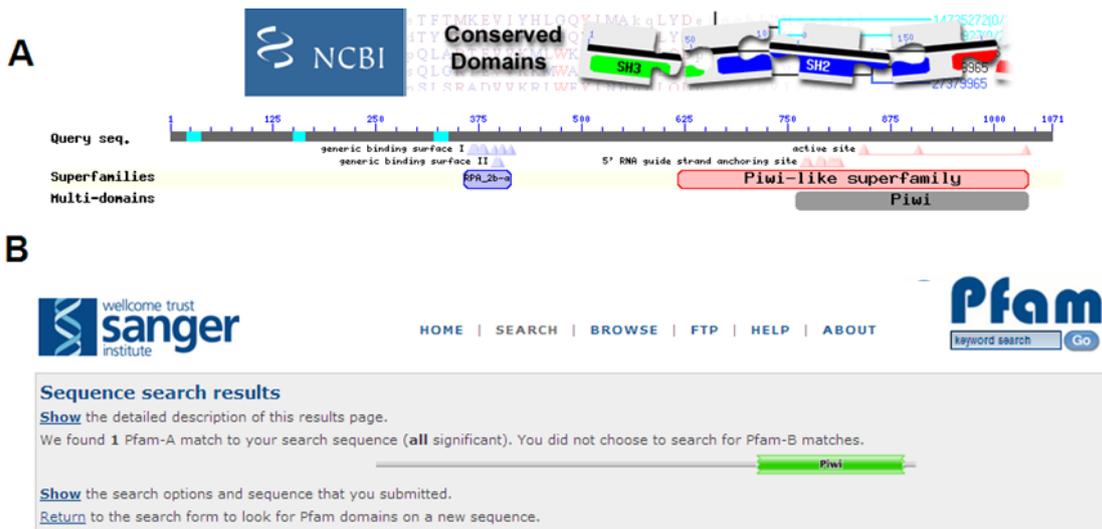


Figura 7: Captura de pantalla de los servidores NCBI Conserved domain (A) y Pfam (B) al día 30 de abril de 2011, luego de introducir la secuencia de TcPIWI-tryp (cepa Dm28c). En (A) se muestra el dominio PIWI en gris con un E-valor de $4,08 \times 10^{-8}$, los dominios MID/PIWI en rojo con un E-valor de $1,39 \times 10^{-13}$, y un dominio tipo “OB-fold” en azul con un E-valor de $9,68 \times 10^{-4}$. En (B) aparece únicamente el dominio PIWI con un E-valor de $1,2 \times 10^{-13}$

En la sección anterior se introdujo la interrogante sobre si la presunta pérdida del dominio PAZ en las proteínas “PIWI-tryp” sería real, o consecuencia de un problema técnico para detectar un dominio PAZ divergente. Sumado al argumento filogenético, existen otras razones para asentar esta duda. Cuando el equipo de la Dra. Leemor Joshua-Tor resolvió la primera estructura completa de una proteína Argonauta (proveniente de la Archea extremófila *P. furiosus*; pdb: 1U04) resultó claro que la misma poseía un dominio PAZ de estructura similar al dominio PAZ de la Argonauta 1 humana. Según los autores “A pesar de grandes semejanzas estructurales, la comparación de las secuencias primarias falló a la hora de detectar un dominio PAZ en PfAGO, mientras que la presencia y mapeo del dominio PIWI fue fácilmente detectado por búsquedas del tipo BLAST” (Song et al. 2004). Además, se sabe que los dominios OB-fold se encuentran conservados a nivel de estructura pero no de secuencia (Bochkarev and Bochkareva 2004; Murzin 1993; Theobald et al. 2003) y – como se dijo ya en la introducción – la mayor parte de un dominio PAZ corresponde a un OB-fold.

Por todo lo expuesto, se comprenderá que existían elementos para presumir la existencia de un dominio PAZ presente en las proteínas PIWI-tryp, aunque de secuencia lo suficientemente divergente como para no ser detectados. Más aún al descubrir que la proteína TcPIWI-tryp (así como el resto de sus ortólogos en tripanosomátidos) posee un dominio del tipo OB-fold entre los aminoácidos 350-425 aproximadamente. Si bien dicho dominio no es detectado confidientemente por Pfam, CDD search le asigna un E-valor de $9,68 \times 10^{-4}$. En la figura 7(A) se representa dicho dominio en azul, y aunque se le asigna el nombre “RPA2”, el mismo proviene de la proteína “Replication protein A, 2” y corresponde al dominio OB-fold de dicha proteína.

La presencia de dominios OB-fold en las proteínas “PIWI-tryp” se suma a una interrogante que se desprende de los resultados filogenéticos: si las proteínas PIWI-tryp y AGO-tryp se encuentran estrechamente emparentadas, y las proteínas AGO-tryp poseen dominios PAZ ¿Por qué habrían las proteínas PIWI-tryp perdido su dominio PAZ originariamente heredado? Por todo lo expuesto

anteriormente, parecería más probable que las proteínas PIWI-tryp sí hubiesen retenido su dominio PAZ, pero que la diversificación funcional de estas proteínas hubiese implicado una divergencia en la secuencia de dicho dominio que impidiese su detección por algoritmos convencionales.

Para abordar esta posibilidad, lo primero que se hizo fue recurrir a algoritmos bioinformáticos especializados en la detección de homólogos remotos, dado que nuestra hipótesis era la retención de un dominio PAZ conservado a nivel de estructura, pero no de secuencia. Dichos algoritmos deberían actuar a un nivel más allá del simple alineamiento de secuencias, incorporando información de estructura secundaria predicha, por poner un ejemplo.

La primera evidencia sólida provino del análisis de la secuencia de TcPIWI-tryp mediante el programa HHpred¹⁸. A diferencia de una búsqueda por BLASTp, basada en alineamientos de secuencias primarias de proteínas, o de PSI-BLAST, que incorpora perfiles de aminoácidos a la búsqueda, HHpred utiliza perfiles basados en modelos ocultos de Markov (HMM) que incorporan además de la frecuencia aminoacídica para cada posición y sus posiciones adyacentes, la frecuencia de inserciones y deleciones para cada sitio. Más aún, HHpred no solo construye perfiles HMM para la proteína de interés, sino que realiza alineamientos entre perfiles HMM-HMM de la proteína de interés y la base de datos, además de que incorpora información de estructura secundaria predicha para evaluar los alineamientos¹⁹ (Soding 2005).

Los resultados obtenidos con la proteína TcPIWI-tryp utilizando el programa HHpred se muestran en la figura 8. Los alineamientos HMM-HMM se realizaron únicamente contra proteínas anotadas en el “Protein Data Bank” (PDB), es decir, únicamente para aquellas proteínas de las cuales se dispone de información estructural. La barra horizontal superior representa la secuencia primaria de TcPIWI-tryp, y sobre ella se mapean las regiones que el programa fue capaz de alinear con mayor confianza con otras proteínas depositadas en el PDB. El código de cuatro letras dentro de cada proteína que logró ser alineada representa su código PDB, y entre paréntesis se muestra el puntaje que el programa asigna a cada alineamiento (cuánto mayor puntaje, menos probable que el alineamiento haya sido consecuencia del azar).

Como puede observarse, en la región entre los aminoácidos 325-425 aproximadamente, aparecen tres resultados representados por casilleros de color anaranjado. Los mismos corresponden a dominios OB-fold de distintas proteínas de las Archea: *M. mazei*, *S. solfataricus* y *T. acidophilum*. La región en cuestión es la misma en la cual CDD search es capaz de detectar un dominio OB-fold (figura 7, A), por lo cual se puede afirmar que TcPIWI-tryp posee dicho dominio en esa región.

¹⁸ <http://toolkit.tuebingen.mpg.de/hhpred>

¹⁹ About HHpred: http://toolkit.tuebingen.mpg.de/hhpred/help_ov

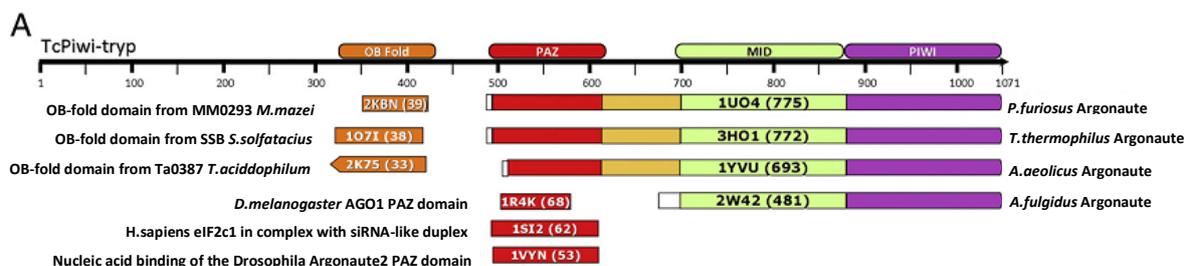


Figura 8: Análisis de la proteína TcPIWI-tryp por HHpred. La línea horizontal representa la secuencia primaria de la proteína y los casilleros de colores que se encuentran debajo corresponden a los resultados que el programa encuentra con mayor soporte estadístico. Para los alineamientos HMM-HMM, se utilizó como base de datos “pdb70” compuesta únicamente por proteínas depositadas en el PDB. Los casilleros de color anaranjado corresponden a los dominios OB-fold de proteínas alineadas en la región 325-425 de TcPIWI-tryp. Los tres casilleros superiores de la derecha corresponden a fragmentos PAZ-MID/PIWI de proteínas Argonaua procariotas, y por corresponder a varios dominios fueron coloreados con más de un color. Debajo de estos se representa la región alineada de la proteína Argonaua de *A. fulgidus*, que contiene únicamente los dominios MID/PIWI. Los tres casilleros rojos ubicados en la región medio inferior corresponden a los dominios PAZ de las proteínas Argonaua1 y Argonaua2 de *D. melanogaster*, y Argonaua2 humana. Dentro de cada casillero se muestra el “pdb id.” de cada proteína, y el puntaje asignado por el programa a cada alineamiento.

Los resultados de mayor puntaje se ubican en la región entre el aminoácido 500 y el extremo C-terminal de la proteína. Corresponden a las proteínas Argonaua de *P. furiosus*, *T. thermophilus* y *A. aeolicus*, 3 de las 4 argonauas procariotas de las cuales se dispone su estructura. Nótese que la región alineada es mayor a la región en la cual CDD search o Pfam mapeaban un dominio PIWI (del aminoácido 620 en adelante). Por otro lado, si se observa la región alineada de las proteínas Argonaua procariotas, se ve que las mismas abarcan las regiones correspondientes a los dominios PAZ y MID/PIWI. Esto quiere decir que el programa es capaz de alinear también la región correspondiente al dominio PAZ de dichas proteínas con la región comprendida entre los aminoácidos 500-620 de TcPIWI-tryp. Más aún, la cuarta Argonaua procariota de la cual se dispone su estructura también aparece en los resultados, debajo de las otras tres (“*A.fulgidus argonaute*”). Una característica de esta proteína es que es más corta que las demás, y posee únicamente el dominio MID/PIWI, no así el PAZ. Toda esta información permitiría mapear los dominios PAZ, MID y PIWI en TcPIWI-tryp.

No obstante lo anterior, los resultados más relevantes para nuestro análisis son los tres casilleros de color rojo que se ubican en la región comprendida entre los aminoácidos 500 y 610 de TcPIWI-tryp. Corresponden justamente a las tres estructuras de dominios PAZ eucariotas disponibles al momento del análisis²⁰. Los mismos provienen de las proteínas Argonaua1 y Argonaua2 de *D. melanogaster*, y de la proteína Argonaua2 (eIF2c) de *H. sapiens*. Es de destacar que no se dispone la estructura completa de dichas proteínas, sino únicamente de sus dominios PAZ, por lo que resulta claro que es el dominio PAZ el que se está alineando a la región correspondiente de TcPIWI-tryp.

Queda contestada por tanto la interrogante sobre la presunta pérdida del dominio PAZ en las proteínas “PIWI-tryp”. A la luz de los resultados obtenidos con HHpred, resulta claro que la proteína TcPIWI-tryp y demás proteínas “PIWI-tryp” poseen una arquitectura de dominios típica de cualquier proteína

²⁰ Si el análisis se repite al día de la fecha: 20 de setiembre de 2011, estos resultados se ven reforzados por la disponibilidad de un mayor número de dominios PAZ resueltos estructuralmente, así como el de algunos dominios MID y MID/PIWI eucariotas.

Argonauta, con un dominio N-terminal variable, un dominio PAZ central, y los dominios MID/PIWI en el extremo C-terminal. Además, el dominio N-terminal posee un OB-fold adicional al del dominio PAZ, lo cual no se observa en otras proteínas Argonauta y podría conferirle propiedades particulares a las proteínas “PIWI-tryp”.

Una ventaja adicional de la interfaz web del programa HHpred es su acoplado con el programa MODELLER, muy utilizado para la construcción de modelos estructurales de proteínas basados en homología con otras proteínas de estructura conocida. El programa permite elegir alguno de los alineamientos realizados por HHpred, y utilizar dicho alineamiento para modelar la estructura de la proteína propia. Utilizando el alineamiento realizado con la proteína Argonauta de *P. furiosus* (pdb id.: 1UO4), se modeló la estructura de los dominios PAZ y MID/PIWI de TcPIWI-tryp, desde el aminoácido 497 hasta el extremo C-terminal. El modelo obtenido se muestra en la figura 9. La raíz de la desviación cuadrática media (RMSD) de las posiciones de los átomos entre el modelo y la estructura que se utilizó de molde es de 2,8 angstroms, y el porcentaje de identidad de secuencia es tan sólo del 12%, y aunque la gran mayoría de los residuos se ubican en las regiones más favorables del “Ramachandran plot” algunos pocos (sin contar Glicina) se ubican en regiones no permitidas. Todo esto hace del modelo una herramienta útil pero de dudosa validez, sobre la cual cualquier interpretación debe ser hecha con sumo cuidado.

Según fue previamente reportado, la actividad endonucleolítica “slicer” de aquellas proteínas Argonauta capaces de clivar ARNs mensajeros depende de dos residuos de Ácido Aspártico (D) y una Histidina (H) presentes en el dominio PIWI (Rivas et al. 2005). Una inspección visual de los alineamientos de secuencias múltiples realizados (figura 9, A) también evidencian la existencia de una tríada compuesta por los aminoácidos D841, D908 y H1039 en el dominio PIWI de TcPIWI-tryp, y sus proteínas ortólogas. Asimismo, se observa como muchos de los residuos conservados en todas las proteínas Argonauta eucariotas (mostrados en rojo) no se encuentran conservados en las proteínas “AGO/PIWI-tryp”, lo cual es esperable dados los resultados filogenéticos. Mapeando los residuos D841, D908 y H1039 en el modelo, los mismos se ubican próximos en la estructura tridimensional, siendo esto consistente con la posibilidad de que los mismos conformen una tríada catalítica. Las distancias entre los carboxilatos de los residuos D841 y D908 y el anillo imidazol de H1039 son de 3,9 y 5,9 angstroms (0.39 y 0.59 nm) respectivamente (figura 9, B). No debe perderse de vista que la presencia de una tríada catalítica es condición necesaria pero no suficiente para la actividad slicer de las proteínas Argonauta (Amitsur et al. 1992; Hock and Meister 2008; Tolia and Joshua-Tor 2007).

Si bien el tercer componente del motivo “slicer” suele presentar un cierto grado de degeneración, pudiéndose encontrar tríadas catalíticas de la forma D-D-E/D/H/K (Tolia and Joshua-Tor 2007), es de destacar que no fue posible detectar una tríada catalítica en la subfamilia “AGO-tryp”, a pesar de que la misma está compuesta por proteínas que, como TbAGO1, participan activamente en los mecanismos de interferencia por ARN.

A

TbAGO-tryp	682	LYADGSDCDHTS--MSEG-----EP	(24)	-KGN-AVSAYLVLRGCASRGEL	(64)	FFLTGANCTLCHARSTLVVVS	(7)	GELOALVLMGFLVFN-----
LbAGO-tryp	669	LYED-SDAASSH--MSSTTL--TVRR	(24)	-NGKSGKGMVLYRGMATDAEV	(78)	FYLSAANCTLCHAANTYVLVQ	(8)	MDLQKTYNMGMVFNKPDALPLPLKCY-----
TcPIWI-tryp	836	LVVGVDACTTT--FSTGCVGILCA	(37)	--YD-GLDEVVVLQGDVSSSEA	(66)	FYLQNHEDNMSTARTVQYTH	(8)	SDVQQLSHVLSHVL--SQATKLEPMARCAHRLS--
TbPIWI-tryp	862	ELVGVDSCTPT--VSTGCVGILCT	(37)	--YD-GLDDVVVFLQGDVYSBM	(72)	FYLQNHEDGETSTARTVQYTH	(8)	SDVQQLSHVLAHAM--ATKATKLEPMARCAHRLA--
LbPIWI-tryp	949	LILGVDSCTNT--HSVGAIVGILST	(37)	--SG-RVDEVVVFQGDVVFSEL	(93)	FYLQTHESCMSTTRIVQYTH	(8)	ADVQHIANIMANVL--APQATKLEPMARCAHRLA--
Ce prg-1	476	MIVCYDLYDST--LKQKTVGACVST	(38)	SNDQTLPSRLILLYRDGAGDGOI	(70)	FYLVPQFVNOCTVTVSYNII	(8)	DKEQQAFKLGHLVYNWQIVRVPAPCOYAKLAFI
Hs HIWI	627	MIVCYDCEDMT--AGRRSIAGFVAS	(38)	C-NEYMPSRIIVYRDGVGDGQL	(59)	FFIVSQAVRS-SVSPETHYVI	(8)	DHLQRLTYKLGHLVYNWQIVRVPAPCOYAKLAFI
Mm MIWI	628	MIVCYDCEDTT--AGRRSIAGFVAS	(38)	C-NEYMPSRVIVYRDGVGDGQL	(59)	FFIVSQAVRS-SVSPETHYVI	(8)	DHLQRLTYKLGHLVYNWQIVRVPAPCOYAKLAFI
Dr PIWI	624	MIVCYDCEDTA--AGRRSIAGFVAS	(39)	Y-NNSLPSRIIVYRDGVGDGQL	(59)	FFIVSQAVRFQCVAPETHYVI	(8)	DHMORLTYKLGHLVYNWQIVRVPAPCOYAKLAFI
Dm PIWI	609	MTICFDIAKSTR--DRKRAYGALIAS	(39)	E-HRKLPSRIVFYRDGVSSCSL	(58)	FYLVSQVROGTVSPITSYNVL	(8)	EKLQKLTVMKCHLYYNWQIVRVPAPCOYAKLAFI
Aa PIWI	645	MTICFDVCDTN--DKSKSYGAMVAT	(37)	E-YNELPKRIFFYRDGVGEGQL	(59)	FYLVSQVROGTVSPITAYNVL	(8)	DHLQRLTYKLGHLVYNWQIVRVPAPCOYAKLAFI
Dm Auber	631	MTVCFDVCSPK--NKNYSYGAFAVAT	(39)	Q-HRSLPERLIFFRDGVGDGQL	(59)	FFIVSQAVRI-TVSPITSYNVI	(8)	DKLQMLSYAMTHMYNYSQIVRVPAPCOYAKLAFI
Pt PIWI	546	MIVCYDVEKTV--LGLDSCVGFNQ	(38)	Y-NKKLFDTHIIFRDGVCTSOI	(53)	YFLIAQQVNOCTATPTHTVVL	(8)	DLFWKFTYQCFNRNWCQPVKIPACVONAHTAAYR
Ce Alg-1	735	IFFGCDITEPPACDSRSPSIAAVVGS	(38)	N-RRFKARIVVYRDGVSEGOE	(63)	FYLCSHAGIQTSRSESHYVVL	(8)	DELOQTYQMGHTVVRCTRSVSIIPAPAYAHVAFR
Hs AGO2	592	HFLGADVTEPPACDGKQPSIAAVVGS	(38)	S-RRFKTRIFRYRDGVSEGOE	(63)	FYLCSHAGIQTSRSESHYVVL	(8)	DELOQTYQMGHTVVRCTRSVSIIPAPAYAHVAFR
Pm AGO	645	HFLGADVTEPPACDNKQPSIAAVVGS	(65)	S-RRFKNRILLYRDGVSEGOE	(63)	FYLCSHAGIQTSRSESHYVVL	(8)	DELOQTYQMGHTVVRCTRSVSIIPAPAYAHVAFR
At Zwilli	704	HIFGADVTEPPAGESSPSIAAVVGS	(52)	A-TGQKPLRIIFRYRDGVSEGOE	(65)	FYLCSHAGIQTSRSESHYVVL	(8)	DGLOSITNNLCYTVRCTRSVSIIPAPAYAHVAFR
At AGO1	757	HIFGADVTEPPHGEDSSPSIAAVVGS	(52)	S-TGKHLRIIFRYRDGVSEGOE	(65)	FYLCSHAGIQTSRSESHYVVL	(8)	DGLOSITNNLCYTVRCTRSVSIIPAPAYAHVAFR
Sp AGO	575	HILGGVYHFGVG--ATGVSIASIVAS	(38)	M-TKQQQRIIFRYRDGVSEGOE	(62)	FYLISHPSIQSVVPHVTVL	(8)	DQFQTLQCNLVVWARATSAVSLVPPVYAHVSNL
Bm AGO2	784	MIVGADVTEPPDQSNIPSIAAVVAS	(38)	S-QGILPKVVFVFRDGVSEGOE	(60)	FYLVSQALIKGTARPTRYHAW	(8)	NEVEHAAVYLCHLVARCMVAVSYPAPTYAHACLIR

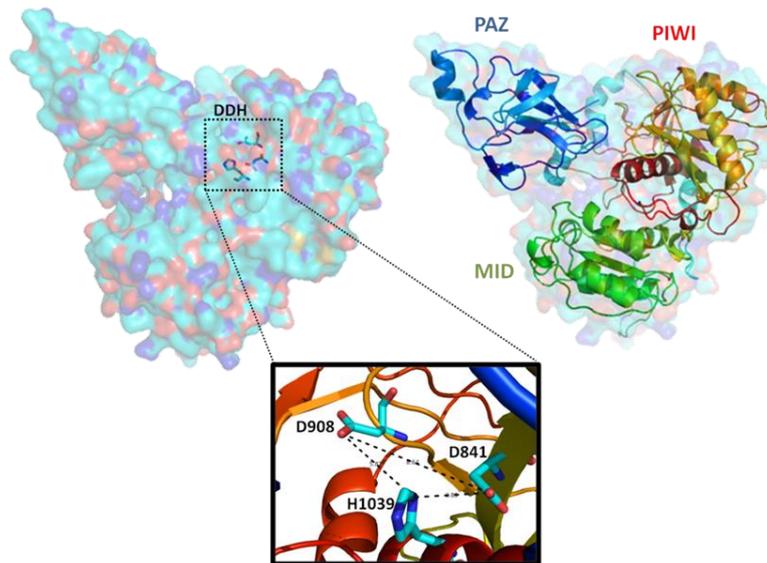
B

Figura 9: Modelado de TcPIWI-tryp y mapeo de una posible tríada catalítica DDH. (A) Alineamiento múltiple de secuencias de dominios PIWI de proteínas Argonauta representativas. Se muestran únicamente algunas regiones del alineamiento, mientras que las regiones intermedias menos conservadas se condensan y entre paréntesis se muestra el número de aminoácidos no representados. Aquellos aminoácidos que se encuentran conservados en todas las proteínas AGO/PIWI eucariotas se muestran en rojo, mientras que aquellos aminoácidos que parecen estar conservados únicamente en las proteínas “Argonaute-like” o “Piwi-like” se muestran en verde y azul respectivamente. En cuanto a las proteínas “AGO/PIWI-tryp”, se colorearon en base al criterio anterior según corresponda, y aquellas posiciones que parecerían ser características de las proteínas Argonauta de tripanosomátidos se muestran en negro. (B) Modelado de la estructura de los dominios PAZ-MID/PIWI de TcPIWI-tryp, utilizando los alineamientos generados por HHpred en el programa Modeller, y la proteína Argonauta de *P. furiosus* como molde (PDB id.: 1UO4). Se muestran en el recuadro los residuos D841, D908 y H1039, correspondientes a una presunta tríada catalítica en el modelo tridimensional. Las imágenes fueron generadas con el programa Pymol.

3. Obtención de la secuencia codificante completa de la proteína TcPIWI-tryp en la cepa Dm28c de *T. cruzi*

El genoma de *T. cruzi* que fue secuenciado corresponde a la cepa CL Brener, y es de hecho una mezcla de dos haplotipos diferentes (*Esmeraldo-like* y *non Esmeraldo-like*). Sin embargo, en nuestro laboratorio se cultiva la cepa Dm28c, dado que -a diferencia de CL Brener- la misma es susceptible a metaciclógenesis en cultivo. Además, el marco abierto de lectura anotado corresponde al marco abierto de lectura más inclusivo, pero ello no implica que los codones ATG inicial y “STOP” reales coincidan con los anotados. Dicho de otra forma, la secuencia proteica real puede ser distinta a la predicha. De cara a la expresión de la proteína recombinante, esto podría tener consecuencias negativas graves. Por todas estas razones, se entendió necesario obtener la secuencia codificante completa de nuestra proteína de interés en la cepa Dm28c, en la cual se realizarían todos los subsecuentes experimentos.

Esta actividad encontraba una pequeña dificultad técnica debido al tamaño predicho de la secuencia codificante (3315 pares de bases; número de acceso en tritrypDB: Tc00.1047053511367.240). Se diseñó por consiguiente una estrategia basado en RT-PCR con distintos juegos de cebadores que permitirían la amplificación de secuencias cortas solapadas entre sí. Cabe señalar que dichos cebadores se diseñaron sobre la secuencia anotada en la cepa CL Brener, por lo que podrían no ser efectivos para la amplificación de la secuencia en Dm28c.

Para la obtención de los extremos 5' y 3' codificantes, se siguió una estrategia diferente. Para la obtención de la secuencia 5' se aprovechó una particularidad que presenta la expresión génica en tripanosomátidos, que es el *trans-splicing*, de modo que se diseñó un cebador directo correspondiente al mini-exón común, o “*splice-leader*”. En todos los casos el sitio de inserción del mini-exón coincidió con la posición nucleotídica +37 en la secuencia referencia de CL Brener. Esto excluye que el ATG anotado como inicial sea efectivamente el que codifica para la metionina n^o1 de la proteína. Estos resultados son consistentes con lo predicho por métodos bioinformáticos, en donde se estableció que las primeras 78 bases del gen TcPIWI-tryp en CL Brener son no codificantes (datos no mostrados). La amplificación del extremo 3' se realizó mediante cebadores directos internos a la proteína, y cebadores reversos correspondientes al 3' UTR (amplificando desde el ADNc), o secuencias génicas río abajo del gen (desde el ADN genómico).

En total, se secuenció un fragmento de 3616 pares de bases (pb), con un marco abierto de lectura de 3210 pb (104 pb menos que el predicho en CL Brener), una región 5' UTR de 32 pb a partir del mini-exón, y una región 3' UTR de 371 pb, aunque su largo real no puede ser conocido dado que la región 3' se amplificó desde el ADNg, y solo parcialmente desde el ADNc. Dicha secuencia fue depositada en el GenBank y recibió el número de acceso: HM126462. El tamaño aproximado del ARNm maduro es consistente con lo observado previamente por Northern blot (García-Silva y Tosar et al. 2010).

Como ya se ha dicho, el sitio de inserción del mini-exón resultó ser interno al marco abierto de lectura predicho. Como consecuencia, el ATG inicial real se encuentra 78 pares de bases río abajo del ATG inicial predicho, dando lugar a una secuencia de 1070 aminoácidos, con un peso molecular de 118,2 KDa y un punto isoeléctrico teórico de 9,36.

Otras diferencias menores entre la secuencia de TcPIWI-tryp obtenida en Dm28c, y la predicha para CL Brener consisten en un acortamiento del tracto poli glutamina (poly-Q), pasando de un motivo compuesto por 12 glutaminas y 2 histidinas, a 6 glutaminas contiguas. Se detectaron también 54 sustituciones de nucleótidos, de las cuales 25 provocan cambios de aminoácidos. Por último, se detectaron dos variantes alélicas no sinónimas en las posiciones correspondientes a los nucleótidos 956 y 1059 (dos clones mostraban una determinada base en la posición en cuestión, mientras que los tres clones restantes mostraban una base diferente). Esto origina 4 posibles variantes en la secuencia de la proteína. Como se detallará, la proteína recombinante se obtuvo amplificando un único fragmento correspondiente al largo total del marco abierto de lectura, el cual fue posteriormente secuenciado en su totalidad. Los aminoácidos que quedaron codificados en las posiciones 956 y 1059 fueron Cys y Phe respectivamente. Cabe señalar que también en CL Brener las posiciones correspondientes están ocupadas por Cys y Phe.

4. Amplificación de la secuencia codificante completa de TcPIWI-tryp y sus dominios, y clonado en vectores de expresión en *E. coli* y *T. cruzi*

4.1 Amplificación y clonado de TcPIWI-tryp en el plásmido pGEM-T Easy

El largo del producto a amplificar (3,3 Kba.), así como la finalidad de esta actividad (la expresión de la proteína recombinante), hizo necesario la utilización de enzimas ADN polimerasas termoestables pero de alta fidelidad (con corrección de errores). El diseño de los cebadores se realizó de modo tal que el cebador directo hibridase desde el ATG inicial en adelante, mientras que el cebador reverso hibridase en la región 3' de la secuencia codificante, incluyendo el codón STOP, TAA. A los extremos 5' de ambos cebadores se les añadieron sitios de reconocimiento y corte para las enzimas Bgl II (directo) y Sal I (reverso), agregando 3 nucleótidos en los extremos para optimizar el corte de dichas enzimas. La elección de los sitios para las enzimas de restricción se discutirá más adelante.

La amplificación de la secuencia codificante completa se intentó realizar tanto a partir del ADN copia (cADN) como del ADN genómico, teniendo éxito únicamente en el segundo caso. Es probable que los bajos niveles de expresión del gen, sumada a la ineficiencia en la retrotranscripción de un transcrito tan largo, hiciesen dificultosa la amplificación desde el cADN. Si bien esto sería lo correcto a la hora de clonar una secuencia codificante para realizar ensayos de expresión proteica, es importante destacar que ya habíamos amplificado mediante PCRs cortas de productos solapantes la secuencia completa de TcPIWI-tryp a partir del ADN copia (sección 3). El secuenciado de estos productos y del producto completo amplificado del ADN genómico nos permite afirmar que ambas secuencias coinciden. No debe olvidarse que los genes que codifican para proteínas en *T. cruzi* no poseen, por lo general, intrones. El sitio de inserción del mini-exón - introducido por *trans-splicing* - fue debidamente estudiado en la sección anterior, por lo cual se conocían los codones ATG inicial y STOP reales.

La optimización de la reacción de PCR requirió considerable cantidad de ensayos, observando amplificación del tipo "todo o nada" según se ajustara la cantidad de ADN molde introducido, la temperatura de hibridación de los cebadores, o el protocolo de extensión utilizado (figura 10)

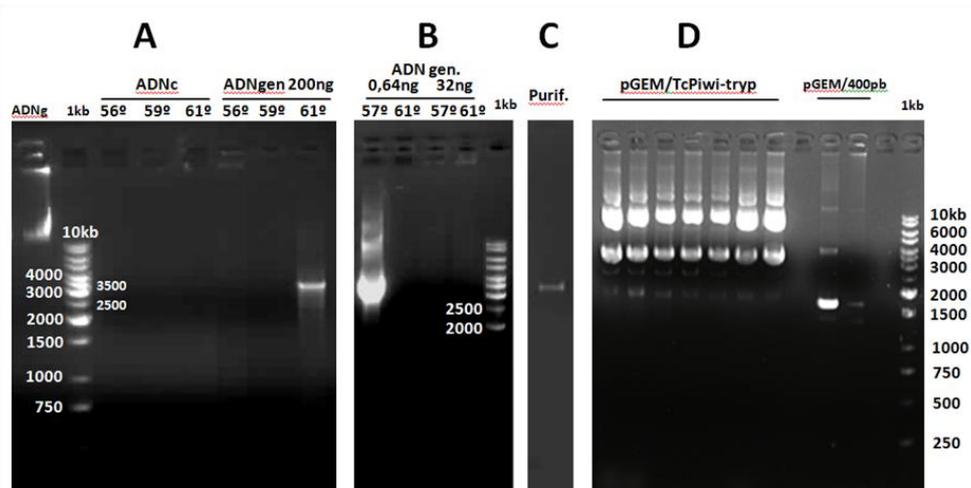


Figura 10: Amplificación por PCR y clonado en el plásmido pGEM-T Easy (Promega) de la secuencia codificante completa del gen TcPIWI-trypt. (A) Amplificación mediante Pfu/Taq polimerasas de TcPIWI-trypt a partir de 100ng de ADN copia y de 200ng de ADN genómico, utilizando distintas temperaturas de hibridación y un protocolo de amplificación de 35 ciclos. A la izquierda del marcador de tamaños (1kb) se corrieron 560ng (1 μ L) del ADN genómico extraído. (B) Amplificación mediante Pfu/Taq polimerasas de TcPIWI-trypt a partir de distintas cantidades de ADN genómico (0,64ng y 32ng) y a distintas temperaturas de hibridación, utilizando una temperatura de extensión de 68°C durante 180 segundos los primeros 10 ciclos, y aumentando los tiempos de extensión 10 segundos por ciclo durante los 25 ciclos restantes. La banda más intensa de aproximadamente 3200pb. que se observa a la izquierda fue cortada del gel, purificada, y vuelta a correr en (C). Mini-preparaciones de ADN plasmídico de bacterias transformadas con el plásmido pGEM-T Easy conteniendo el inserto TcPIWI-trypt amplificado, comparando con un vector pGEM-T Easy cerrado conteniendo un inserto de 400pb. (D). Todos las corridas electroforéticas se realizaron en geles de agarosa al 1% en tampón TAE 1x.

Previo al clonado de la secuencia en un vector de expresión, la misma se clonó en el plásmido de amplificación pGEM-T Easy, agregando adenosinas 3' terminales al producto de PCR, para facilitar la ligación "T". Se transformaron bacterias de la cepa XL1Blue de *E. coli*, y se seleccionaron en placa, obteniendo 7 colonias de color blanco, positivas para el inserto por PCR. Los plásmidos recuperados por "mini-prep" fueron analizados en gel de agarosa, junto con un plásmido pGEM-T Easy recombinante que poseía un inserto pequeño, de 400pb, a efectos comparativos (figura 10, D).

En la figura 10,D puede observarse un patrón característico al realizar corridas electroforéticas de plásmidos purificados. El tamaño esperado del plásmido recombinante pGEM/TcPIWI-trypt es de 6246 pb, mientras que el del plásmido pGEM con inserto control es de 3415 pb. Dicha diferencia de tamaños se manifiesta claramente en la corrida electroforética, pudiéndose observar como las bandas centrales (correspondientes a formas clivadas y linearizadas del plásmido) migran con un tamaño aparente similar al esperado.

No se detectaron diferencias entre la secuencia completa de TcPIWI-trypt clonada en pGEM-T Easy, y la secuencia codificante secuenciada por retrotranscripción a partir del ARN mensajero, por lo que se procedió al sub-clonado en vectores de expresión.

4.2 Clonado de TcPIWI-trypt en el vector de expresión en *E. coli* pQE-80_TEV

Una vez clonada la secuencia codificante en el plásmido pGEM-T Easy, se procedió a extraerla del mismo e insertarla en un plásmido adecuado para realizar ensayos de expresión proteica. El plásmido utilizado es una versión modificada del vector comercial pQE-80L (QIAGEN), a la cual se le introdujo una secuencia que codifica para el sitio de reconocimiento y corte de la proteasa del virus del mosaico del tabaco (proteasa TEV), inmediatamente después de la secuencia que codifica para el “6x-HIS tag” e inmediatamente antes del sitio BamHI donde se clona la secuencia de interés. De este modo, puede purificarse a la proteína por cromatografía de afinidad de iones inmovilizados (IMAC) gracias al tag N-terminal de histidinas, y luego remover esta secuencia exógena por tratamiento con la proteasa TEV. Un esquema del vector pQE-80_TEV utilizado se muestra en la figura 11.

El plásmido en cuestión posee un origen de replicación del tipo pMB1 (muy similar a ColE1), lo que lo hace un plásmido de bajo o moderado número de copias (15-20 copias por bacteria)²¹. Dado que expresa el gen de la beta-lactamasa (bla/AmpR), el vector confiere resistencia a ampicilina, que puede ser utilizada como marcador de selección. Una característica muy importante de este vector es que permite la expresión a altos niveles del represor lac i del operón lactosa (el nombre LacI⁹ hace referencia a la sobre-expresión del mismo debido a una mutación en la región -35 del promotor), por lo cual puede utilizarse para expresar proteínas en cepas bacterianas que no suelen ser utilizadas a tales fines, como XL1Blue.

El sitio BamHI se encuentra inmediatamente después de la secuencia que codifica para el sitio TEV (de hecho, se encuentran parcialmente solapados, figura 11B), por lo cual es un sitio preferencial para el clonado dado que minimiza el número de aminoácidos N-terminales que se añadirían a la proteína una vez clivada con la proteasa. Más aún, el sitio BamHI codifica para la secuencia Gly-Ser, dos aminoácidos que por su tamaño pequeño y propiedades no deberían afectar significativamente el plegamiento ni la actividad biológica de la proteína, aunque esto exigiría comprobación empírica. El vector posee también un sitio Sal I hacia 5' del terminador de la transcripción t₀, proveniente del fago λ.

Por todas estas razones se decidió utilizar un clonado direccional mediante BamHI y Sal I. El problema es que la propia secuencia codificante de TcPIWI-tryp contienen tres sitios BamHI, razón por la cual no puede ser cortada con tal enzima. Afortunadamente, no posee ningún sitio SalI ni Bgl II. Esto permite efectuar un rodeo al problema planteado pues BamHI y Bgl II generan extremos cohesivos compatibles. El único inconveniente es que luego del ligado se pierden ambos sitios de restricción, lo que impide el sub-clonado en otros vectores. Aún así, esto no representa un problema pues se dispone del plásmido pGEM/TcPIWI-tryp que – dado que no fue clonado por restricción – posee ambos sitios intactos.

Inicialmente, se intentó purificar el inserto TcPIWI-tryp digiriendo el plásmido pGEM/TcPIWI-tryp con las enzimas Bgl II y Sal I, como ya se ha dicho. Al realizar una electroforesis en gel de agarosa del producto de digestión, se comprobó la imposibilidad de purificar el mismo del gel, dado la similitud de tamaños entre el inserto y el vector (3222 y 3015 pb respectivamente; figura 12, A, izquierda).

²¹ <http://www.qiagen.com/plasmid/bacterialcultures.aspx>

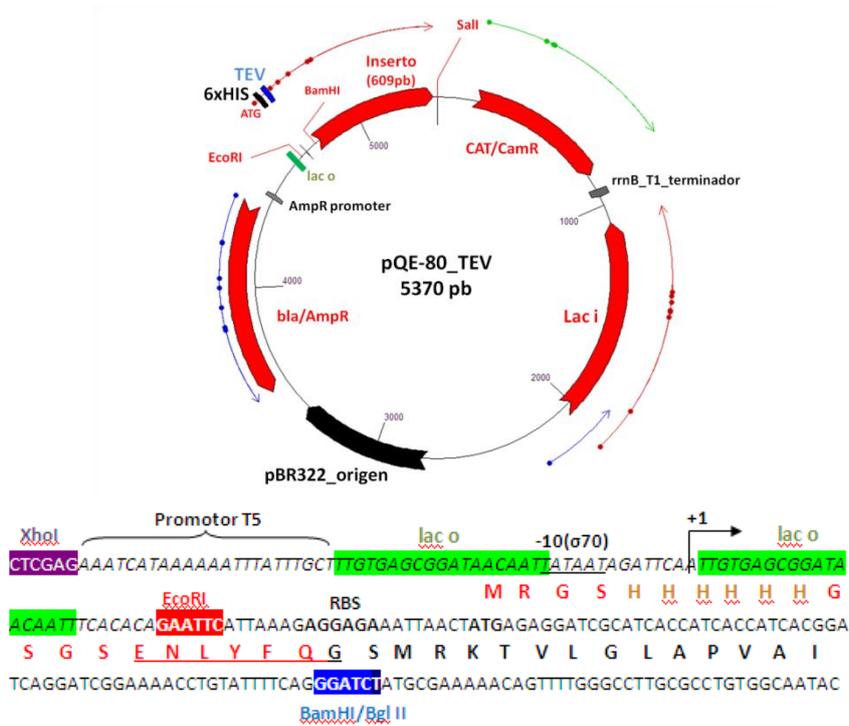


Figura 11: Mapa del vector de expresión en *E. coli* pQE-80_TEV (A) y detalle de los elementos regulatorios y funcionales de la región cercana al sitio BamHI dónde se clonó la proteína TcPIWI-tryp (B). El vector pQE-80_TEV de 5370 pb posee un inserto de 609 pb clonado en los sitios BamHI y Sall. Por fuera del vector se muestran marcos abiertos de lectura detectados automáticamente por el programa GENTle y la anotación de los mismos. ColE1: origen de replicación en *E. coli*; bla/AmpR: gen de la beta-lactamasa, que confiere resistencia a la ampicilina; lac o: operador lac; 6xHIS: tag N-terminal de 6 histidinas; TEV: sitio de reconocimiento y corte de la proteasa TEV; Lacⁱ: gen para la expresión constitutiva en *cis* del represor lac i; se señalan también las secuencias para la terminación de la transcripción t0 (del fago λ) y T1 (proveniente del fago T1). En la parte (B) se señalan también los sitios XhoI, EcoRI y el sitio formado luego del corte y ligado con BamHI y BglII, la región consenso a -10, que es reconocida por la subunidad σ70 de la ARN polimerasa de *E. coli*, el sitio de inicio de la transcripción (+1) a partir del promotor del fago T5, el sitio de unión del ribosoma (RBS) también conocido como secuencia de Shine-Dalgarno, el sitio de inicio de la traducción, la secuencia aminoacídica resultante antes (rojo + negro) y después (negro) del corte con la proteasa TEV, que reconoce la secuencia ENLYFQG, que se encuentra subrayada.

Luego de varios intentos en los que solamente se obtuvo colonias conteniendo el vector pGEM religado, se concluyó que el esfuerzo debería focalizarse en evitar por completo su inclusión en la reacción de ligación. Aparentemente, la contaminación con pequeñas cantidades del mismo sería suficiente para su aparición en las colonias resistentes a ampicilina. Se buscaron enzimas de restricción que cortasen al plásmido pGEM-T Easy sin afectar la secuencia del inserto TcPIWI-tryp. Afortunadamente, la enzima Bgl I (Bgl uno) satisface estos requisitos dado que corta al plásmido en cuatro sitios, originando fragmentos menores a 1500 pb., separables del inserto TcPIWI-tryp por electroforesis. Por consiguiente, se procedió a la purificación del mismo a partir del gel, luego de la triple digestión con Bgl I, Bgl II y Sal I (figura 12, A, centro izquierda).

En la figura 12 (centro-izquierda) puede observarse la banda de 3222 pb correspondiente a los 3210 pb del inserto más los sitios de restricción, y dos bandas inferiores correspondientes a los productos de mayor tamaño luego de la digestión del pGEM con Bgl I. Por encima de la banda de 3222 pb se observa una banda que migra a la altura de 4000 pb y de baja intensidad, que corresponde a los restos no digeridos del plásmido recombinante pGEM/TcPIWI-tryp en su forma super-enrollada (en la figura 10, D se observa como la forma super-enrollada del plásmido migra a dicho tamaño aparente). Si bien en

cantidades relativas se encuentra menos representado, se debió optimizar la triple digestión y la corrida electroforética para excluirlo de la purificación.

Una vez obtenido el inserto y el vector pQE-80_TEV purificado y digerido con enzimas compatibles, se procedió a realizar ensayos de ligación y transformación. Finalmente, se obtuvieron colonias positivas para el inserto TcPIWI-tryp y negativas para el plásmido pGEM (por PCR). Las mini-preparaciones de ADN plasmídico obtenidas de tales colonias se analizaron por restricción. El corte con Sal I generó una banda única de aproximadamente 8000 pb, coincidente con el tamaño esperado del vector pQE-80_TEV/TcPIWI-tryp (7983 pb; figura 12, A, derecha). Los plásmidos obtenidos fueron insensibles a Bgl II – como cabría esperarse dado que el sitio se pierde al ligar – y reprodujeron el patrón de bandas esperado luego del corte con Sal I + Nde I, que corta al vector pQE-80_TEV en un único sitio (datos no mostrados). El plásmido obtenido fue secuenciado a efectos de corroborar la presencia del ATG inicial de TcPIWI-tryp en marco con la secuencia codificante para el 6x-HIS *tag* y el sitio TEV, utilizando un cebador directo complementario al vector (figura 12, B).

4.2 Clonado de la secuencia de los dominios PAZ y MID/PIWI en el vector pQE-80_TEV

Como se describirá posteriormente en la discusión, a pesar de un gran interés manifiesto, únicamente existen dos ejemplos de argonautas eucariotas obtenidas en bacterias de forma recombinante, ambas expresadas como proteínas de fusión. De estos antecedentes se desprende que el objetivo de obtener la proteína TcPWI-tryp recombinante es altamente ambicioso, aunque las grandes disimilitudes de secuencia con el resto de las proteínas Argonauta eucariotas justifican la prueba. Adelantándonos a una predecible insolubilidad de la proteína recombinante completa, se realizaron otras construcciones a efectos de poder expresar algunos de los dominios más relevantes.

La amplificación por PCR de la secuencia que codifica para cada dominio no es una tarea trivial, pues exige conocer los sitios de inicio y fin de cada dominio por separado. La rotura de elementos de estructura secundaria, la inclusión de aminoácidos extra correspondientes a brazos conectores entre dominios o la incorporación innecesaria de aminoácidos hidrofóbicos normalmente enterrados en la estructura completa podrían acarrear grandes problemas.

Para resolver esta cuestión se utilizaron las herramientas bioinformáticas disponibles. La primera aproximación consistió en basarse en los resultados obtenidos con el programa HHpred, que ya fueron discutidos y analizados, y presentados en la figura 8. Según los alineamientos realizados, el dominio PAZ se ubica aproximadamente en la región comprendida entre los residuos 506-616, mientras que el dominio MID/PIWI se ubica desde el residuo 682 hasta el extremo C-terminal. El dominio MID/PIWI, que a diferencia del PAZ es detectado por CDD search, es mapeado en dicho programa desde el residuo 617 en adelante. El extremo C-terminal de la construcción correspondiente al dominio MID/PIWI quedaría definido por el extremo C-terminal de la proteína, pues el carboxilato libre terminal de las proteínas Argonauta es parte integral de los dominios MID/PIWI. Respecto al extremo N-terminal de dicha construcción, se decidió optar por la opción más abarcativa y posicionar el mismo en la prolina 617, que coincide con el extremo C-terminal de una α -hélice predicha por PSIPRED²².

²² <http://bioinf.cs.ucl.ac.uk/psipred/>

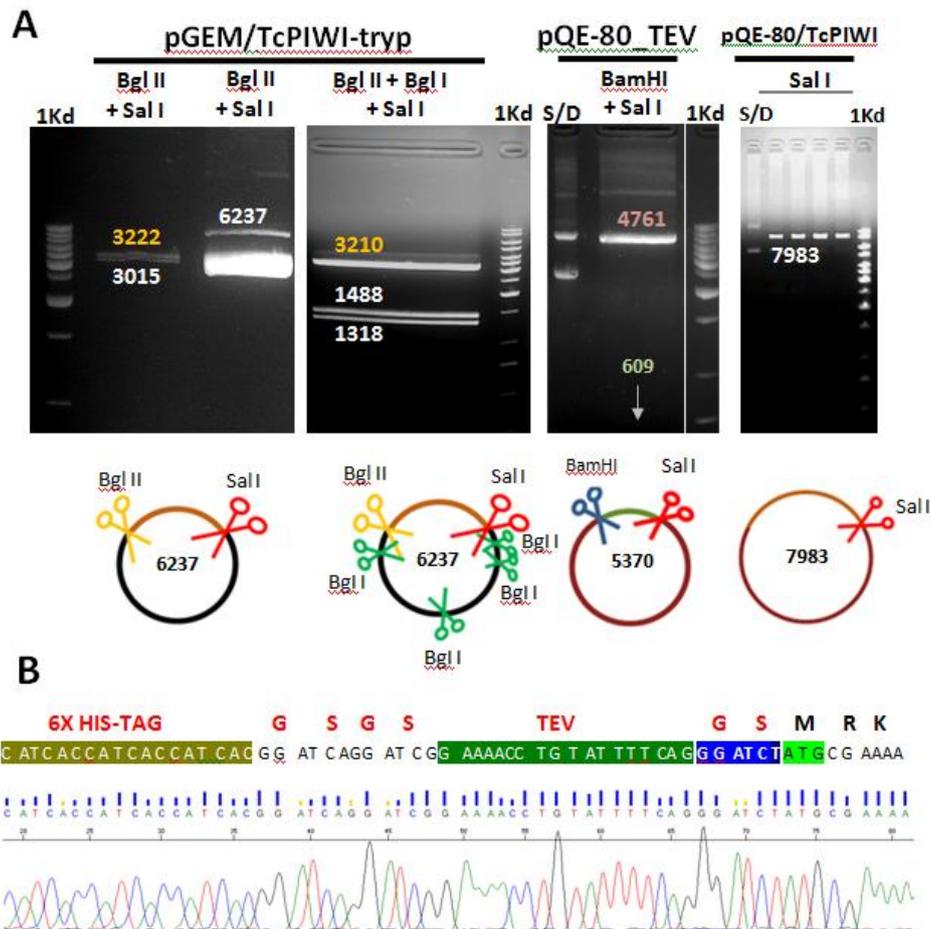


Figura 12: Clonado direccional por restricción de TcPIWI-tryp en el plásmido de expresión pQE-80_TEV. (A): El inserto TcPIWI-tryp fue liberado del plásmido pGEM/TcPIWI-tryp mediante digestión con las enzimas Bgl II y Sal I (izquierda), pero esto impide la purificación del inserto TcPIWI-tryp (de 3210 pb + 12 pb adicionales como sitios de restricción, más allá de que la integridad de los mismos se pierde luego de la digestión). Para facilitar la separación, se procedió a realizar una digestión triple con Bgl II, Bgl I y Sal I (centro-izquierda). El vector pQE-80_TEV, conteniendo un inserto de 609 pb, fue digerido con las enzimas BamHI y Sal I (centro-derecha). La ligación entre el inserto TcPIWI-tryp y el plásmido pQE-80_TEV digerido permitió la transformación y purificación del plásmido recombinante pQE-80_TEV/TcPIWI-tryp, de 7983 pb. Los plásmidos purificados se linealizaron con Sal I, obteniendo un fragmento del tamaño esperado (derecha). (B): Secuencia parcial del plásmido pQE-80_TEV/TcPIWI-tryp superpuesta a la secuencia esperada, con el sitio de ligación sombreado en azul-celeste y azul. La secuencia se obtuvo con un cebador directo complementario al vector

Para la definición de los límites del dominio PAZ, se procuró ampliar la información aportada en los alineamientos de HHpred con una predicción de estructura secundaria efectuada por PSIPRED. Sin embargo, se comprobó que PSIPRED (y otros algoritmos de predicción evaluados) falla por completo en la predicción de la estructura secundaria de dominios PAZ divergentes - tales como los de las Argonautas procariontas y de tripanosomátidos - asignando un exceso de α -hélices en la región del OB-fold, que es un dominio del tipo "todo beta" (figura 13, A y B).

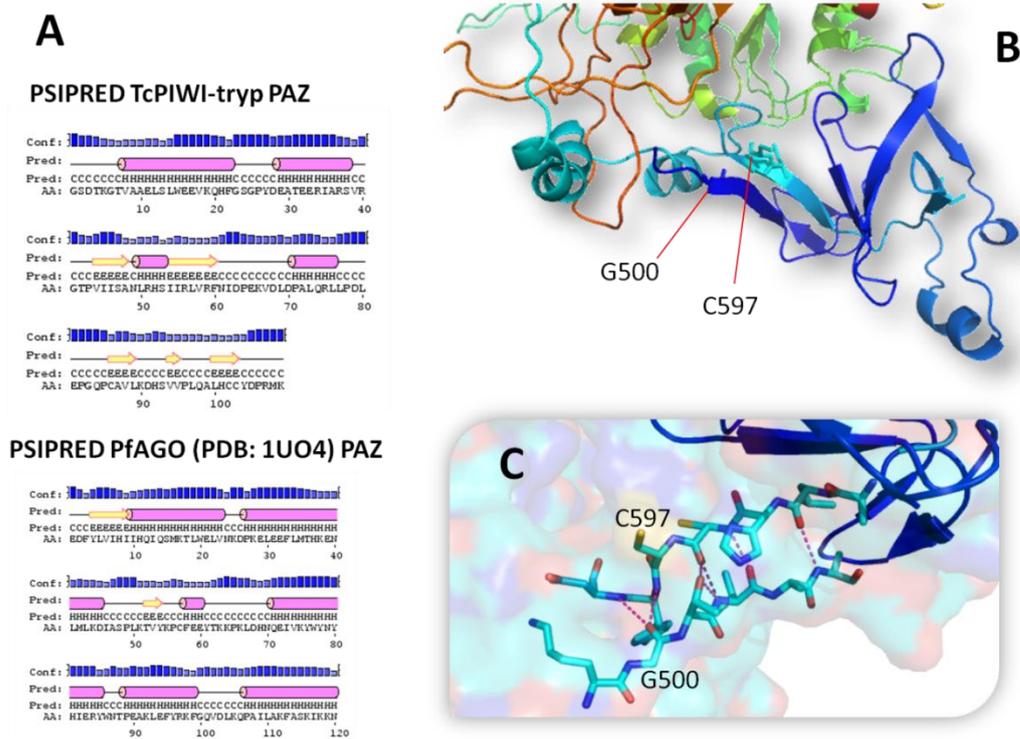


Figura 13: Predicción de los límites del dominio PAZ a efectos de la amplificación por PCR de su secuencia codificante; la región aproximada en donde se ubica el dominio PAZ fue predicha mediante HHpred. Se realizó una predicción de la estructura secundaria de dicha región mediante PSIPRED, a efectos de evitar suprimir elementos de estructura secundaria (A, recuadro superior). La idoneidad de este método fue analizada al evaluar la predicción por PSIPRED del dominio PAZ de la proteína Argonauta de *Pyrococcus furiosus* (A, recuadro inferior). Modelado por homología de la proteína TcPIWI-trypan, realizado con el programa MODELLER a partir de los alineamientos de HHpred, y utilizando la estructura de PfAGO como molde, detallando el dominio PAZ en azul (B). Un detalle de los residuos involucrados en una posible hoja β anti-paralela que presumiblemente mantendría adyacentes los extremos N-terminal y C-terminal del dominio PAZ se muestra en (C), señalando las interacciones polares en violeta, entre ellas los enlaces por puentes de hidrógeno entre los grupos amino y carboxilo que componen los enlaces amida de las hebras adyacentes, tal como se espera de una lámina β anti-paralela.

Si bien ya se aclaró que aún frente a las limitaciones que presenta cualquier modelo estructural por homología, éste es particularmente poco robusto, se decidió recurrir al modelado de la proteína TcPIWI-trypan como un elemento más de decisión. El modelo comienza a ser estructurado a partir del residuo Gly 500, dónde comienza una hebra β . En la figura 13-C, se muestra un detalle del dominio PAZ en el modelo, coloreado en azul y cian. Se observa una segunda hebra β adyacente a la misma, compuesta por los aminoácidos que se encuentran en la región C-terminal del dominio PAZ. En la figura 13-C se muestra un detalle de ambas hebras, señalando los contactos polares entre átomos de la proteína en violeta. Como puede apreciarse, los mismos satisfacen los enlaces por puente de hidrógeno que serían esperables en un par de hebras β que formasen una pequeña hoja β anti-paralela. Por consiguiente, se entendió favorable la inclusión de las secuencias correspondientes a cada hebra, como un posible determinante estructural del dominio. Los límites del dominio PAZ entendidos como mínimo quedarían compuestos por los residuos Gly-500 y Cys-597, diseñando un juego de cebadores para amplificar por PCR la región correspondiente. Dado lo especulativo de esta aproximación, se diseñó un segundo juego de cebadores de PCR (PAZF2, PAZR2) para amplificar una región más abarcativa, comprendida entre los

residuos D-497 y K-606, procurando que ambos extremos N-terminal y C-terminal tuviesen un carácter hidrófilo.

La figura 14 muestra la amplificación de las secuencias correspondientes a los distintos dominios y combinaciones de dominios. Como puede observarse, todos los fragmentos deseados pudieron amplificarse con alto rendimiento, pero únicamente se tuvo éxito con uno de los dos juegos de cebadores para el dominio PAZ. Finalmente, la lista de construcciones realizadas concernientes a la expresión de la proteína TcPIWI-tryp y sus dominios en *E. coli* quedó conformada por: pQE-80_TEV/TcPIWI-tryp (proteína completa); pQE-80_TEV/N-ter_PAZ (dominios N-terminal y PAZ; desde la metionina inicial al residuo 506); pQE-80_TEV/PAZ (dominio PAZ; desde el residuo 497 al 506); pQE-80_TEV/PAZ-PIWI (dominios PAZ y MID/PIWI, desde el residuo 497 hasta el 1071) y pQE-80_TEV/PIWI (dominios MID/PIWI, desde el residuo 617 hasta el 1071).

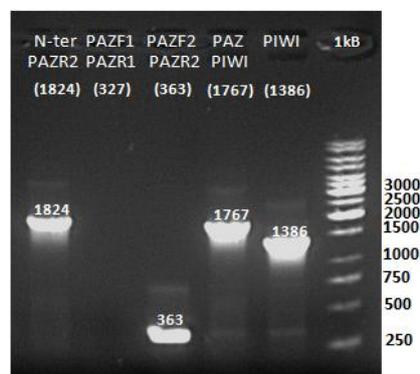


Figura 14: Amplificación por PCR (Pfu/Taq ADN polimerasas) de distintos fragmentos de TcPIWI-tryp destinados a la expresión de distintas combinaciones de dominios de la proteína. Los fragmentos fueron amplificados utilizando el plásmido pGEM/TcPIWI-tryp como molde. Los tamaños esperados para cada amplicón se muestran entre paréntesis, y sobre la banda correspondiente en caso de observarse producto de amplificación. La corrida electroforética se realizó en gel de agarosa 1% en tampón TAE 1X.

De las cinco construcciones realizadas, se trabajó con tres de ellas a nivel de la expresión de proteínas, como se detallará más adelante. Las proteínas expresadas fueron la proteína completa, el dominio PAZ, y el dominio MID/PIWI.

4.3 Clonado de la secuencia TcPIWI-tryp completa en el vector pTEX para sobre-expresión en *Trypanosoma cruzi*

Para la expresión de TcPIWI-tryp en *T. cruzi*, se procedió a sub-clonar su secuencia codificante completa en el vector de expresión pTEX (Kelly et al. 1992; figura 15, A), que confiere resistencia a aminoglucósidos como la neomicina o la geneticina (G418) para su selección en parásitos. El vector es del tipo "shuttle", lo que implica que es replicativo también en *E. coli*, confiriendo resistencia a ampicilina en este caso.

La secuencia codificante se liberó por restricción del plásmido pGEM/TcPIWI-tryp y se ligó al vector pTEX previamente digerido con las enzimas BamHI y Sall. Con el producto de ligación se transformaron

bacterias competentes de la cepa XL1Blue de *E. coli*. Se aisló uno de los clones positivos para la presencia del inserto (figura 15, B) y se secuenció parte del plásmido correspondiente a efectos de verificar que la secuencia coincidiese con la esperada. El mismo se amplificó en bacterias y se recuperó por “*midi-prep*”, con posterior análisis por restricción (figura 15, C). Con dichos plásmidos se transfectaron por electroporación epimastigotes de la cepa Dm28c de *T. cruzi*, que fueron seleccionados por incubación en concentraciones crecientes de neomicina.

Además de la sobre-expresión de TcPIWI-tryp en *T. cruzi*, se creyó oportuno realizar una segunda construcción para la sobre-expresión en el parásito de una versión de la proteína que contuviese un *tag* N-terminal de 6 histidinas. Esto podría facilitar la detección, inmunoprecipitación y eventual purificación de la proteína del parásito utilizando anticuerpos monoclonales anti 6x-HIS *tag* en lugar de – o en combinación con – el suero policlonal anti-TcPIWI-tryp.

Para lograr lo anterior, se retiró todo el *cassette* de expresión del vector pQE-80_TEV/TcPIWI-tryp digiriendo al mismo con las enzimas EcoRI y Sal I (figura 15, D). Como se observa en la figura 11, el sitio EcoRI se encuentra río arriba del ATG inicial codificado por el vector, y por consiguiente, de la secuencia del *6x-HIS tag*, del sitio TEV, y de la secuencia de TcPIWI-tryp previamente clonada en marco. Las dos bandas que se observan en la figura 15, D corresponden al vector pQE-80_TEV digerido (4668 pb; 4761 menos la secuencia comprendida entre el sitio EcoRI y el sitio BamHI) y al inserto TcPIWI-tryp más la secuencia comprendida entre los sitios EcoRI y BamHI del vector pQE-80_TEV (3303 pb). Dicha banda fue purificada del gel, y ligada con la banda purificada de 5667 pb correspondiente al vector pTEX digerido con EcoRI y Sal I (figura 15, E). Luego de la ligación y transformación, las colonias obtenidas en placa con antibiótico fueron analizadas por *colony-PCR* (figura 15, F).

A modo de resumen de esta sección, se construyó el plásmido pTEX/TcPIWI-tryp para la expresión de la proteína en *T. cruzi*, y el plásmido pTEX_HisTEV/TcPIWI-tryp para la expresión en el parásito de la proteína fusionada a un *tag* de histidinas en su extremo N-terminal. De las dos, únicamente la primera se utilizó efectivamente para transfectar *T. cruzi*, aguardando la segunda construcción posteriores ensayos.

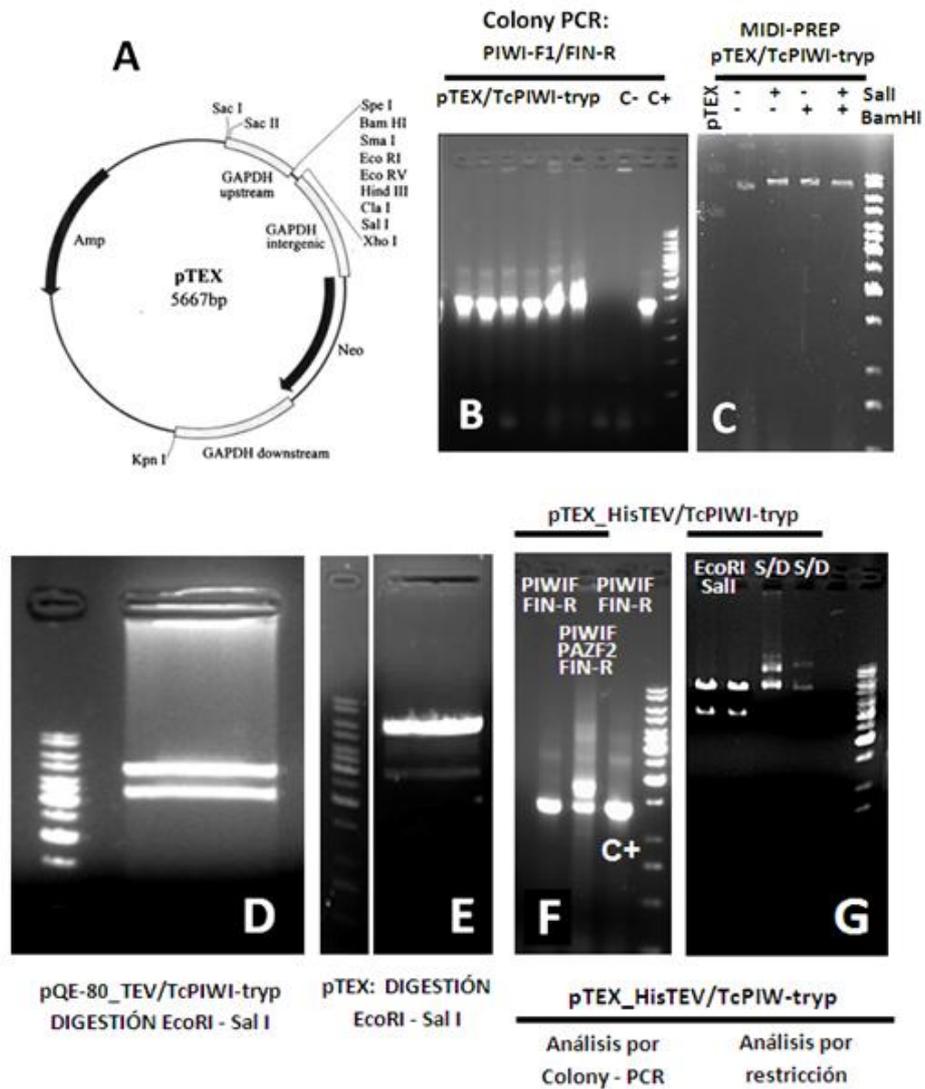


Figura 15: Clonación de TcPIWI-tryp en el vector pTEX para la expresión en *T. cruzi* de la proteína con y sin un tag de histidinas N-terminales. El mapa del vector pTEX (A) fue extraído de Martínez-Clavillo et al. (1997). B) Verificación de la presencia del inserto en el plásmido recombinante pTEX/TcPIWI-tryp por “colony-PCR”, trasapando colonias bacterianas resistentes directamente a la reacción de PCR. Se utilizaron los cebadores PIWI-F1 y FIN-R, los cuales amplifican una región de 1386 pb. Como control negativo (C-) se picaron colonias transformadas con el vector pTEX cerrado, mientras que el control positivo (C+) corresponde a colonias transformadas con el plásmido pGEM/TcPIWI-tryp. C) Análisis por restricción del vector pTEX/TcPIWI-tryp mediante las enzimas Sal I, Bam HI, o ambas. En paralelo, se realizó la construcción pTEX_HisTEV/TcPIWI-tryp digiriendo el plásmido pQE-80_TEV/TcPIWI-tryp (D) y pTEX (E) con EcoRI y Sal I. Las bacterias transformadas fueron analizadas por colony-PCR (F) y los plásmidos purificados de clones positivos fueron analizados digiriendo con las enzimas EcoRI y Sal I. S/D: sin digerir. Los marcadores de tamaño son los mismos utilizados en figuras anteriores.

5 EXPRESIÓN Y PURIFICACIÓN DE LA PROTEÍNA TcPIWI-TRYP Y DEL DOMINIO MID/PIWI EN BACTERIAS

5.2 Expresión de la proteína TcPIWI-tryp en *E. coli*

Los primeros ensayos de expresión se realizaron en cultivos de pequeña escala a efectos de estudiar la cinética de inducción y establecer las condiciones de expresión más adecuadas en términos de temperatura, medio de cultivo, concentración de inductor y tiempo de inducción. El principal objetivo en esta etapa fue determinar en qué condiciones – si alguna – se obtenían mayores cantidades de proteína en la fracción soluble.

Se probaron diferentes temperaturas de expresión a concentraciones constantes de inductor de 0,5mM: 37°C, 30°C, 25°C y 20°C (figura 16). La expresión se realizó por tiempos variables que fueron ajustados empíricamente. En la parte superior de la figura 16, a la derecha (30°C; 3hs.), se muestra un SDS-PAGE (10%) dónde se distinguen dos situaciones, separadas por un marcador de peso molecular. Los primeros cuatro carriles del gel corresponden a bacterias transformadas con el plásmido pQE-80_TEV, mientras que los últimos cuatro corresponden a bacterias transformadas con pQE-80_TEV/TcPIWI-tryp, y que deberían expresar la proteína de interés de 120KDa. Los primeros dos carriles de cada caso corresponden a lisados totales (T) de las bacterias a tiempo final (3 horas post-inducción), luego de haber sido (+) o no (-) incubadas con 0,5mM IPTG. En el caso del vector sin inserto ambos carriles lucen exactamente iguales, pero para el caso de interés se observa en el carril inducido una intensa banda a una altura acorde a 120KDa.

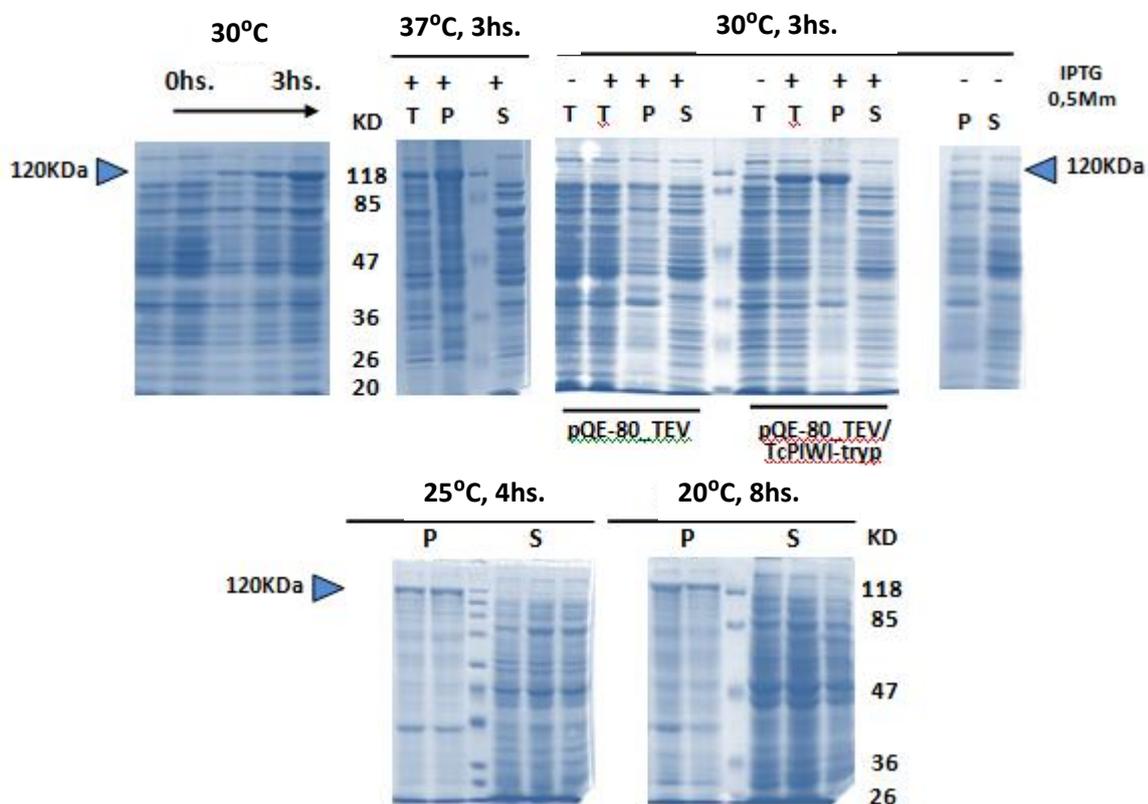


Figura 16: Expresión de TcPIWI-tryp como cuerpos de inclusión en todas las condiciones de temperatura ensayadas (37°C; 30°C; 25°C y 20°C). Se evaluaron distintas condiciones a efectos de establecer si la proteína se recuperaba en la fracción soluble (S) o en la fracción insoluble (P, de *pellet*). Por T se designa extracto total (sin separar soluble de insoluble). Todas las imágenes corresponden a ensayos de SDS-PAGE en geles de poliacrilamida al 10%. En todos los casos se utilizó el marcador de pesos moleculares #SM0441 (Fermentas) cuyos pesos moleculares se indican, salvo en la figura inferior izquierda que se utilizó el marcador S8445 (Sigma) cuyas bandas corresponden a los siguientes tamaños, en KDa: 205 (migra en lo más alto del gel), 116, 97, 84, 66, 55, 45, 36, 29, 24, y otras bandas que no se resuelven: 20.1, 14.2 y 6.5 KDa.

Una observación que resulta interesante es que – con mucho menor intensidad – algo se observa también en el carril no inducido. Puede descartarse que corresponda a una proteína de *E. coli* de similar tamaño pues para el caso del vector control no se observa una banda de intensidad siquiera comparable, aún en el caso inducido. Esto es una evidencia de lo que en inglés se conoce como “*leaking expression*”, es decir, de un cierto grado de expresión basal en ausencia de inductor, que puede traer serias dificultades cuando la proteína en cuestión resulta tóxica para *E. coli*

Mediante centrifugación a velocidades moderadamente altas (10.000g), se aislaron y purificaron los cuerpos de inclusión, que luego fueron re solubilizados en presencia de urea 8M. Por la forma cómo se realiza esta preparación, resulta más correcto hablar de fracción insoluble (P, de *pellet*) que de cuerpos de inclusión, ya que no se realizó una observación de las bacterias al microscopio electrónico para constatar su presencia, ni se utilizó una preparación específica para el aislamiento de los mismos. Por añadidura, al sobrenadante se lo designa fracción soluble (S). Como puede observarse, en todos los casos en que se observa proteína en la fracción inducida total (T, +), la misma aparece exclusivamente en la fracción insoluble. La escasa presencia de proteínas bacterianas en dicha fracción y el enriquecimiento en algunas proteínas específicas, permite descartar problemas técnicos de lisis incompleta. Más aún, se realizó el experimento de separar fracción soluble e insoluble en extractos no inducidos, crecidos a 30°C (figura 16, superior-derecha). También en este caso se observa la banda del tamaño correspondiente únicamente en la fracción P. Siendo ésta una expresión “*leaky*” producida en condiciones basales en las que la transcripción se encuentra reprimida, se espera que las tasas de síntesis de mRNA y proteína sean mínimas. La observación de la proteína en la fracción P resulta entonces un muy mal indicador.

Se intentó transformar bacterias electro-competentes de las cepas M15[pREP4] y Rosseta-gami(DE3)pLysS con el plásmido pQE-80_TEV/TcPIWI-tryp purificado, pero no se obtuvieron colonias. La elección de esta última cepa radica en que contiene genes de tARN adicionales, al tiempo que permite la formación de puentes disulfuro (ver más adelante). La cepa también codifica para la enzima T7 ARN polimerasa inducible por IPTG, y para su inhibidor codificado en el plásmido pLysS, pero todo este sistema resulta irrelevante en este caso, ya que la expresión se realiza desde un promotor T5 (reconocible por la ARN polimerasa de *E. coli*) y no T7. De todos modos, no se obtuvieron bacterias transformantes y los buenos resultados que se obtuvieron de forma simultánea con los ensayos de replagamiento *in vitro* (ver sección 7) desviaron los esfuerzos en esta última dirección. También es importante destacar que habiendo tantos ensayos infructuosos para la obtención de proteínas Argonauta eucariotas solubles en *E. coli* (Salvatore et al. 2010), no es posible predecir cuántas cepas y condiciones deberían ser ensayadas hasta obtener los primeros resultados positivos. Por el contrario, el re-plegamiento *in vitro* de proteínas Argonauta es una posibilidad no explorada en la literatura

científica²³, y los buenos resultados obtenidos en los primeros ensayos coadyuvaron en redirigir en tal sentido los experimentos.

5.3 Expresión del dominio MID/PIWI en *E. coli*

Resultados similares se obtuvieron al transformar bacterias de la cepa XL1Blue con el plásmido recombinante pQE-80_TEV/PIWI, que permite la expresión de la secuencia codificante para el dominio MID/PIWI. Los ensayos de expresión y distribución entre extracto soluble/insoluble no se muestran, pero – al igual que la proteína completa – únicamente se obtuvo proteína recombinante en la fracción insoluble, en todas las condiciones ensayadas.

5.4 Purificación de TcPIWI-tryp y del dominio MID/PIWI de la fracción insoluble

La expresión de TcPIWI-tryp y del dominio MID/PIWI como agregados insolubles es sin duda un resultado negativo de caras a la realización de ensayos estructurales y bioquímicos. No obstante, la expresión de proteínas recombinantes como cuerpos de inclusión también puede ser vista como un primer paso de purificación, dado que los mismos suelen ser bastante puros para la proteína de interés, como puede apreciarse en los extractos insolubles de la figura 16. Dado que existen aplicaciones para las cuales no se requiere disponer de la proteína de interés en su conformación nativa, tales como la inmunización y obtención de suero policlonal específico (salvo que se requiera reconocer epítopes conformacionales), la observación de la proteína en la fracción insoluble no implica que todo esté perdido. Además, como ya se ha adelantado, existe la posibilidad de resolubilizar, purificar y replegar la proteína *in vitro*, pudiéndose - en el mejor de los casos - recuperar el estado nativo y la actividad biológica.

Los cuerpos de inclusión (extracto insoluble, siendo estrictos) conteniendo una u otra proteína fueron obtenidos por centrifugación del lisado bacteriano a 10.000g y 4°C, como ya se ha explicado. Dependiendo de la aplicación, se realizaron lavados en tampón de lisis de modo de eliminar compuestos solubles contaminantes, que pudiesen haber sedimentado por oclusión o adsorción al precipitado. Posteriormente, se realizaron uno o dos lavados en idéntico tampón, pero con el agregado de urea en concentración 2M. Tal concentración de urea no debería ser suficiente para resolubilizar las proteínas agregadas, pero podría facilitar la remoción de algunos contaminantes. Por último, se incubaron los cuerpos de inclusión con urea 8M a 4°C toda la noche, luego de una leve disgregación mecánica del precipitado.

Una ventaja del 6x-HIS *tag* es que permite la purificación de la proteína aún en condiciones no nativas, como es este caso. Los cuerpos de inclusión resolubilizados en urea 8M se centrifugaron y filtraron para descartar elementos remanentes en suspensión y se introdujeron en columnas comerciales para la purificación por afinidad de proteínas conteniendo varias histidinas consecutivas (HisTrap HP, GE Healthcare), realizando lo que se conoce como cromatografía de afinidad de iones inmovilizados (en adelante, IMAC, por sus siglas in inglés).

²³ Al día 15 de junio de 2011, una búsqueda en Pubmed con los términos “Argonaute” y “refolding” (y similares) no arroja ningún resultado, ni tampoco figuran proteínas argonauta en la base de datos REFOLD.

La figura 17 muestra la purificación por IMAC de la proteína TcPIWI-tryp y del dominio MID/PIWI. Ambas proteínas fueron purificadas a partir de los cuerpos de inclusión correspondientes a 1 litro de cultivo inducido con IPTG 0,5mM. Dado que se decidió purificar las proteínas a partir de los cuerpos de inclusión, se realizó la expresión a 37°C para maximizar la producción de proteína insoluble. La lisis bacteriana se realizó en tampón Tris-Cl 50mM pH=8 y NaCl 0,5M, con el agregado de Tritón X-100 0,1% ν , EDTA 5mM y lisozima 1mg/mL, y posterior sonicado del extracto.

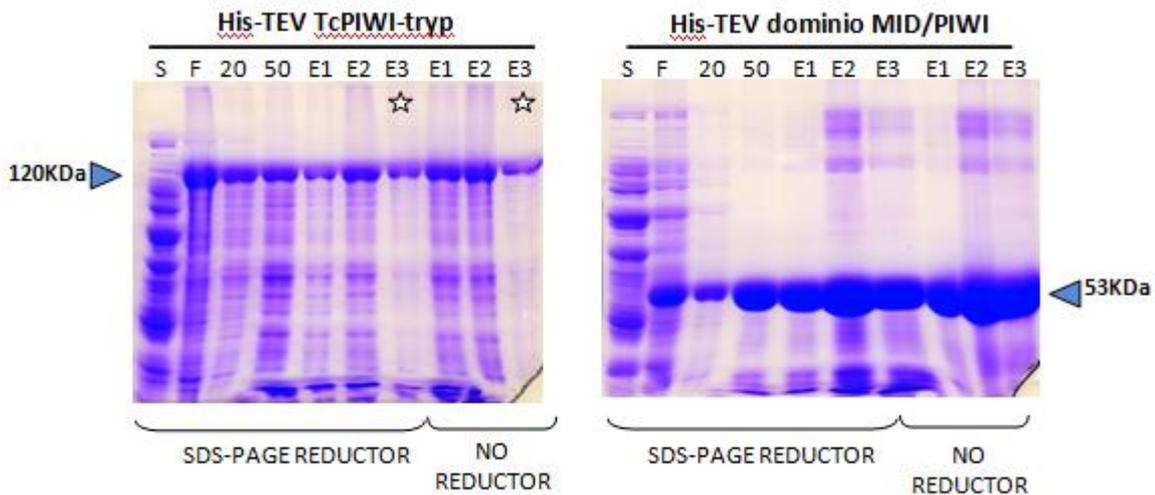


Figura 17: Purificación por cromatografía de afinidad de iones metálicos inmovilizados (IMAC) de la proteína TcPIWI-tryp (izquierda) y del dominio MID/PIWI (derecha). La elución se realizó utilizando una concentración constante de imidazol de 500mM, recogiendo tres fracciones de 1mL cada una (E1, E2 y E3). Todas las fracciones recogidas, junto con el extracto soluble (S), fueron analizadas por SDS-PAGE en geles al 8%. Las fracciones conteniendo la proteína purificada fueron también analizadas en condiciones no reductoras, utilizando un tampón de carga (*loading buffer*) sin el agregado de agente reductor. Las estrellas indican que la fracción E3 fue utilizada para la obtención de suero policlonal anti-TcPIWI-tryp.

Previo a la IMAC, se agregó una baja concentración de imidazol a las muestras (20mM). Esto permite minimizar el pegado inespecífico de proteínas con pocas histidinas a la columna, ya que la presencia de imidazol compite con las mismas, mejorando la pureza del producto final. Se lavó la columna con concentraciones crecientes de imidazol (20mM y 50mM), de modo de retener únicamente aquellas proteínas que presentasen una afinidad elevada por el Ni^{2+} .

La elución de las proteínas recombinantes se realizó elevando bruscamente la concentración de imidazol a 500mM (elución “en escalón”). Se recogieron fracciones de 1mL cada una, no obteniendo cantidades significativas de proteína más allá de cuatro volúmenes de columna (según absorbancia a 280 nm). Las fracciones purificadas y los lavados se analizaron por SDS-PAGE como se muestra en la figura 17. Una característica llamativa es la observación de grandes cantidades de ambas proteínas en la fracción no unida a la columna (F) y en ambos lavados, sobre todo para la proteína completa. Dado que se corrió un mismo volumen en cada carril, pero los lavados provienen de un volumen de 10mL mientras que las fracciones E1, E2 y E3 tienen 1mL cada una, queda claro que la mayor parte de la proteína recombinante producida se está perdiendo en los lavados. Podría pensarse en disminuir la fuerza de los mismos a efectos de maximizar el rendimiento, pero esto implicaría un compromiso en términos de pureza. De las

fracciones eluidas, la fracción E3 es la única relativamente pura, y la misma fue utilizada para inmunizar conejos de modo de obtener anticuerpos específicos (ver sección 6).

La purificación del dominio MID/PIWI se muestra en el gel adyacente. Comparativamente, dado que las condiciones de expresión y purificación son las mismas, puede decirse que el dominio MID/PIWI se expresa en mayores cantidades que la proteína completa (al menos en estas condiciones), lo que es esperable ya que se sabe que *E. coli* no produce eficientemente proteínas mayores a 100 KDa. Además, el costo metabólico de producir una proteína de 52,9 KDa (dominio MID/PIWI) es muy inferior al de producir una proteína de 120 KDa (proteína completa). Por último, es posible que la proteína TcPIWI-tryp resultase ser levemente tóxica para *E. coli*, lo que también implicaría menores niveles de producción a igual tiempo de expresión. Si bien TcPIWI-tryp es una proteína Argonauta, y por tanto, una proteína de potencial unión a ácidos nucleicos, una toxicidad “biológica” exigiría un correcto plegamiento de la proteína, aunque en algunos casos *E. coli* es capaz de suprimir la asistencia al plegamiento de una determinada proteína, cuando la misma le resulta tóxica, “derivándola” entonces a los cuerpos de inclusión.

6 OBTENCIÓN DE ANTICUERPOS POLICLONALES anti-TcPIWI-tryp

6.2 Inmunización de un conejo con la proteína TcPIWI-tryp y obtención de suero policlonal específico.

La fracción E3 de la proteína TcPIWI-tryp purificada en presencia de urea 8M (figura 17, izquierda) fue utilizada para la inmunización de un conejo por parte de personal calificado de la Facultad de Medicina (Universidad de la República). Se realizó una primera inmunización con adyuvante completo de Freund (50%), seguido de tres recuerdos con adyuvante incompleto de Freund los días +35, +77 y +110. Se tomó una muestra de sangre a los 20 días del segundo recuerdo (antisuero Full2B) y se realizó una sangría final a los 23 días del tercero (antisuero Full3B).

El título de los antisueros obtenidos fue determinado por la técnica de ELISA (“enzyme-linked immunosorbent assay”), como se muestra en la figura 18. Se utilizó un diseño convencional de ELISA indirecto para la determinación de anticuerpos, sensibilizando durante toda la noche a 4°C placas MaxiSorp (Nunc) con la proteína recombinante TcPIWI-tryp, el dominio PAZ o un extracto proteico de epimastigotes de *T. cruzi*. En el caso de la sensibilización con la proteína TcPIWI-tryp, no se utilizó la misma preparación con que se inmunizaron los conejos, dado que la misma se encontraba en presencia de urea 8M. En cambio, se sensibilizó con la proteína replegada *in vitro* en ausencia de agentes reductores, para que éstos no interfiriesen con el análisis. El replegamiento *in vitro* de TcPIWI-tryp se describirá más adelante, pero debe considerarse que la proteína se encuentra en la forma de agregados solubles en la preparación utilizada. La preparación del dominio PAZ recombinante también se describirá más adelante, pero no se recurrió en este caso al replegamiento *in vitro* pues la proteína se expresa de forma soluble.

Se compararon tres antisueros distintos: las dos muestras obtenidas luego del segundo y tercer recuerdo (o “booster”) con la proteína TcPIWI-tryp completa (Full2B y Full3B, respectivamente) y un antisuero obtenido previo al comienzo de esta tesis, que debería reconocer epítopes lineales comprendidos en una pequeña región ubicada en la zona N-terminal de TcPIWI-tryp (Nter; aminoácidos

287-392). Se ensayaron diluciones seriadas de los antisueros, diluidos en PBS en razón 1/2000 a 1/64000. Como control, se utilizó suero pre-inmune del mismo conejo, diluido 1/2000 y 1/500. Como control positivo del ensayo, se sensibilizaron los pocillos G4 y G5 con el propio antisuero Full3B, en una dilución 1/200 y 1/2000, respectivamente. De este modo, se asegura que quedarán inmunoglobulinas de conejo unidas a la placa, que deberían ser reconocidas por el anticuerpo secundario (un anticuerpo purificado específico de las colas Fc de las inmunoglobulinas IgG de conejo, conjugado a la enzima peroxidasa de rábano, HRP). También se incluyeron en los pocillos G1, G2, G3 los controles negativos de rutina: pocillos sensibilizados con PBS en lugar de proteína (NS) y pocillos sensibilizados con proteína pero con el agregado de PBS en lugar del antisuero (NA) o del anticuerpo secundario (NC).

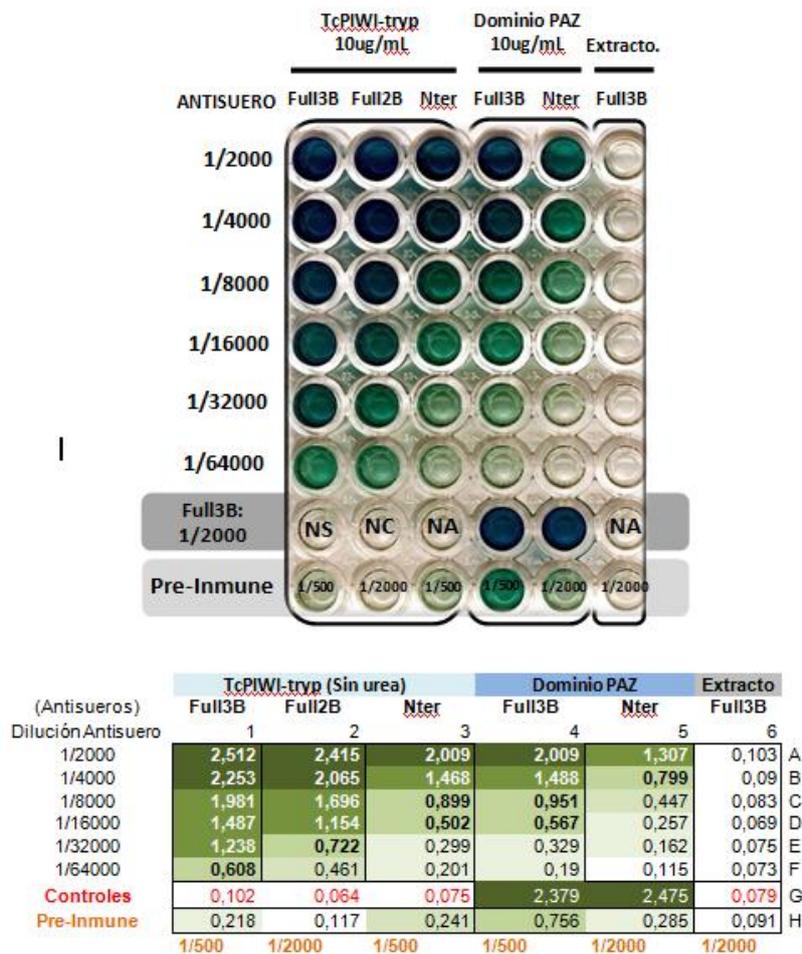


Figura 18: Determinación por ELISA del título de anticuerpos policlonales anti-TcPIWI-tryp. Se analizaron sueros inmunes extraídos de conejos inmunizados con la proteína completa (Full2B y Full3B, tomados a las 3 semanas del segundo y tercer recuerdo, respectivamente) y con un pequeño péptido del dominio N-terminal (Nter). La placa fue sensibilizada con la proteína recombinante completa replegada *in vitro* (TcPIWI-tryp [sin urea]), con el dominio PAZ soluble, o con un extracto proteico de *T. cruzi*. La detección se realizó mediante anticuerpos monoclonales anti-IgG de conejo unidos a la enzima peroxidasa de rábano (HRP) y se midió la absorbancia a 405nm correspondiente al producto de oxidación del ABTS. NS: no sensibilizado. NC: sin anticuerpos secundario. NA: sin antisuero. El control positivo corresponde a la sensibilización con el propio antisuero diluido 1/200 (G4) y 1/2000 (G5). La fila H fue sensibilizada igual que las filas A-F, pero se incubó suero pre-inmune

En la columna A5-F5 se incubó el antisuero Nter con pocillos sensibilizados con el dominio PAZ. Dado que el antisuero Nter se obtuvo inmunizando conejos de la misma raza (New Zealand) con un péptido

correspondiente al dominio N-terminal que no se solapa con el dominio PAZ (regiones 287-392 y 497-606 respectivamente), dicho antisuero podría ser categorizado como un suero inmune negativo para el dominio PAZ, y utilizado como referencia para estimar los títulos de los sueros positivos. Sin embargo, llama la atención los valores elevados de absorbancia en dicha columna. Existe para esto una explicación muy simple: el péptido utilizado para inmunizar con la región 287-392 fue expresado en *E. coli* a partir del vector de expresión pQE-30L, que – al igual que el vector pQE80_TEV – codifica para la secuencia N-terminal MRGSHHHHHGS. Dicha secuencia se encuentra también en el extremo N-terminal del dominio PAZ con el que se sensibilizó la placa, por lo que deja de ser válido que el antisuero Nter sea negativo para dicho dominio. Además, las mismas proteínas de *E. coli* contaminantes del dominio PAZ podrían estar presentes en el péptido y la proteína completa utilizadas para inmunizar.

Más correcto sería utilizar el suero-pre inmune como suero negativo, aunque no se trabajó con las mismas diluciones que los sueros reactivos. Como no se disponen de suficientes replicados para realizar aproximaciones estadísticas, y como solo se requieren medidas cualitativas comparativas, se puede tomar el promedio de las dos diluciones 1/500 del suero pre-inmune sometido a la proteína TcPIWI-tryp (0,2295). Conste que se está siendo conservador al utilizar una dilución del suero pre-inmune cuatro veces más concentrada que la menor dilución de los antisueros. Si se corrigen por este valor todas las medidas de absorbancia y se establece un punto de corte arbitrario de ABS = 0,5 (cinco veces superior al control NS), los títulos de los antisueros Full3B, Full2B y Nter para el reconocimiento de la proteína TcPIWI-tryp serían de manera respectiva y aproximada: 60.000, 32.000 y 12.500 (expresados como el inverso de la mayor dilución con absorbancia mayor a 0,5). Del mismo modo, los títulos de los antisueros Full3B y Nter para el reconocimiento del dominio PAZ serían: 12.500 y 5.000 respectivamente. Estos resultados son acordes a lo esperable, observándose una duplicación del título luego del tercer recuerdo, y un título mucho mayor al del antisuero Nter, ya que los sueros Full3B y Full2B son más recientes en el tiempo y – sobre todo – pueden reconocer un número mucho mayor de epítopes en la proteína TcPIWI-tryp.

Llama la atención los altos valores de absorbancia en el caso del suero pre-inmune sometido a la proteína PAZ (H4 y H5). También es importante destacar la columna A6-F6 que fue sensibilizada con un extracto proteico de *T. cruzi*. Para ninguna dilución de Full3B se observaron valores de absorbancia por encima de los controles negativos, lo que no implica que la proteína TcPIWI-tryp no se exprese. De ensayos previos de detección de la proteína por Western blot, se sabe que los niveles de expresión endógena de la proteína son bajos. Por consiguiente, es muy poco probable que niveles apreciables de la misma queden adheridos a la placa al sensibilizar con un extracto, dado que las proteínas mayoritarias de *T. cruzi* competirán por los sitios de unión en la placa. En cualquier caso, la prueba de detección de la proteína endógena en el extracto debió haberse realizado mediante un ELISA del tipo “sándwich”, prueba que finalmente no se realizó dado que pudo demostrarse detección de la proteína en extracto por Western blot.

7 REPLEGAMIENTO IN VITRO DE LA PROTEÍNA TcPIWI-tryp

Desde la purificación de cuerpos de inclusión conteniendo agregados insolubles de la proteína de interés hasta la obtención de la proteína en su forma nativa, soluble y monomérica (u oligomérica, cuando dichos oligómeros representen alguno de los estados “biológicamente activos” de la proteína), suele

seguirse una plétora de estrategias experimentales que esquemáticamente apuntan a disgregar los agregados proteicos mediante la adición de agentes caotrópicos, y la posterior remoción súbita o gradual de los mismos. El objetivo es eliminar el agente caotrópico en condiciones tales que sea vea favorecida cinéticamente la formación de interacciones intramoleculares que conduzcan al estado nativo, y se vea desfavorecida la formación de interacciones intermoleculares que conduzcan a estados agregados amorfos de la proteína.

7.2 Replegamiento en columna

La primera de las estrategias que se utilizaron fue el uso de la propia columna de IMAC para efectuar el replegamiento. En líneas generales, el procedimiento realizado consistió en introducir a la columna el extracto conteniendo a la proteína TcPIWI-tryp en urea 8M, de modo que ésta quede inmovilizada por su extremo N-terminal, en virtud del 6x-HIS *tag*. Posteriormente se realizaron una serie de lavados idénticos a los utilizados a efectos de la purificación de la proteína en presencia de urea, pero incluyendo un paso de replegamiento previo a la elución, donde la misma fue eliminada.

Una ventaja del replegamiento por IMAC es que las proteínas se mantienen “aisladas” unas de otras, dado que cada una se encuentra inmovilizada por su extremo N-terminal (Eiberle and Jungbauer 2010). Si bien las probabilidades de establecer interacciones intermoleculares son más bajas que en el caso del replegamiento libre en solución, también se pierden “grados de libertad” conformacionales, lo cual puede afectar la “ruta” del plegamiento tanto positiva como negativamente, por lo que la última palabra la tiene siempre la evidencia empírica.

En ensayos iniciales se quitó la urea de forma abrupta. Las fracciones eluidas al aumentar la concentración de imidazol se analizaron por SDS-PAGE convencional, identificando una proteína mayoritaria cuyo tamaño coincide con el tamaño esperado de TcPIWI-tryp de 120KDa. Sin embargo, al analizar la misma fracción en condiciones de SDS-PAGE no reductoras, desaparece la proteína de su ubicación en el gel separador, apareciendo dos poblaciones conteniendo proteína en el pocillo y en la interface entre el gel de *stacking* y el gel separador.

Debe notarse que en la figura 17 se muestran ejemplos de fracciones purificadas por IMAC en urea 8M, y analizadas por SDS-PAGE en condiciones reductoras y no reductoras, no observándose diferencias. En este caso, se eluye proteína soluble en ausencia de urea, pero la misma se encuentra formando agregados. Más aún, estos agregados son insensibles a la acción de detergentes iónicos como el SDS y la acción del calor, salvo que se agreguen también agentes reductores en altas concentraciones. La conclusión de estas observaciones es que tales agregados poseen enlaces covalentes intermoleculares (presumiblemente puentes disulfuro intermoleculares, al ser sensibles a agentes reductores como el β -mercaptoetanol). Se podría afirmar que el protocolo de replegamiento en columna utilizado provoca la elución de TcPIWI-tryp soluble en ausencia de urea, pero en la forma de agregados covalentes.

En función de lo anterior, se repitió el procedimiento manteniendo una concentración estable de β -mercaptoetanol (BME) de 1mM durante todo el ensayo. Sin embargo, al eluir la proteína con 500mM de imidazol, no se obtuvo ninguna fracción con absorbancia a 280nm por encima del tampón de elución. Se elevó la concentración de imidazol a 1000mM en el mismo y en subsecuentes ensayos, pero tampoco se logró eluir cantidades apreciables de proteína. Sospechando que el replegamiento pudiese no haber sido efectivo y que la proteína se hubiese agregado dentro de la columna, se volvió a equilibrar la

columna con urea 8M y eluir en presencia de urea, recuperándose grandes cantidades de proteína. Lo mismo sucedió al realizar el replegamiento mediante un gradiente suave de urea, descendiendo linealmente la concentración de urea de 8M a 0M en 20 volúmenes de columnas: no se eluyó proteína, sino hasta volver a elevar la concentración de urea a 8M.

7.3 Replegamiento por dilución

Una de las formas más simples de minimizar la formación de interacciones intermoleculares no deseadas durante el replegamiento de proteínas es realizar el replegamiento en disoluciones muy diluidas. Agregando gota a gota la solución conteniendo la proteína purificada de cuerpos de inclusión en urea 8M, en un recipiente conteniendo un volumen mucho mayor de tampón sin urea, con agitación suave, es posible replegar numerosas proteínas.

Cuando la primera gota conteniendo la proteína de interés cae en el tampón de replegamiento en agitación, la proteína pasa de un entorno conteniendo urea 8M, a un entorno casi sin urea. Además, la agitación produce que la concentración local de proteína sea muy baja, ocurriendo así un replegamiento en solución en condiciones de alta dilución.

Existen determinados parámetros que suelen ser críticos, entre ellos: la intensidad de la agitación, el ritmo con el que se agrega la proteína (volumen por unidad de tiempo), la concentración de proteína agregada y la concentración final de proteína obtenida (Eiberle and Jungbauer 2010). Siendo los estadios de replegamiento intermedios (no nativos) aquellos sensibles a la agregación, debe darse el tiempo suficiente para que la proteína agregada se haya plegado correctamente, antes de agregar nueva proteína desnaturalizada. De lo contrario, se perdería la condición de “alta dilución”. La concentración final de proteína también suele mantenerse baja para que la técnica sea efectiva, por lo general por debajo de 0,1mg/mL. En este caso se partió de 2mL de proteína a una concentración de 0.5mg/mL, replegando en un volumen total de 48mL, por lo que la concentración final de proteína fue de 0.02mg/mL y la concentración final de urea 0.32M (figura 19)

Mención aparte merece la temperatura, en ésta y otras técnicas de replegamiento. Las proteínas suelen haber evolucionado para presentar – en su forma nativa – máximas estabilidades termodinámicas a temperaturas cercanas a la temperatura fisiológica. Sin embargo, no debe perderse de vista que si bien el objetivo es obtener proteína en su conformación nativa, esta reacción compite con la formación de agregados. Ésta última es una reacción bi o multi-molecular, por lo cual presenta una dependencia marcada con la temperatura; cuanto menor sea la temperatura (Chi et al. 2003). Además, los fenómenos de agregación suelen ser dominados por fuerzas hidrofóbicas, y el efecto hidrofóbico es menor a temperaturas bajas (Xie and Wetlaufer 1996). Por estas razones, los ensayos de replegamiento por dilución se realizaron en cámara fría, a una temperatura de 6-7°C.

Una última consideración refiere al uso de agentes reductores. De los resultados obtenidos en los ensayos de replegamiento por IMAC, resulta claro que el agregado o no de agentes reductores puede tener influencia en la eficacia del proceso. El caso de la IMAC es más complejo en ese sentido, pues en todo momento se está en presencia de cationes divalentes (Ni^{2+}) que pueden catalizar reacciones de oxidación/reducción de proteínas. Para el caso del replegamiento por dilución, se estudió la influencia de diversos agentes oxido-reductores en el proceso de replegamiento de la proteína TcPIWI-tryp y el

dominio MID/PIWI, partiendo de la misma concentración de ambas proteínas en urea 8M. En todos los casos mostrados en la figura 19, se trabajó a distintos valores de pH (6 y 8), en alta concentración salina (NaCl 0,5M) y en presencia del par redox glutatión reducido / glutatión oxidado en relación 10:1 (GSH 10mM, GSSG 1mM).²⁴

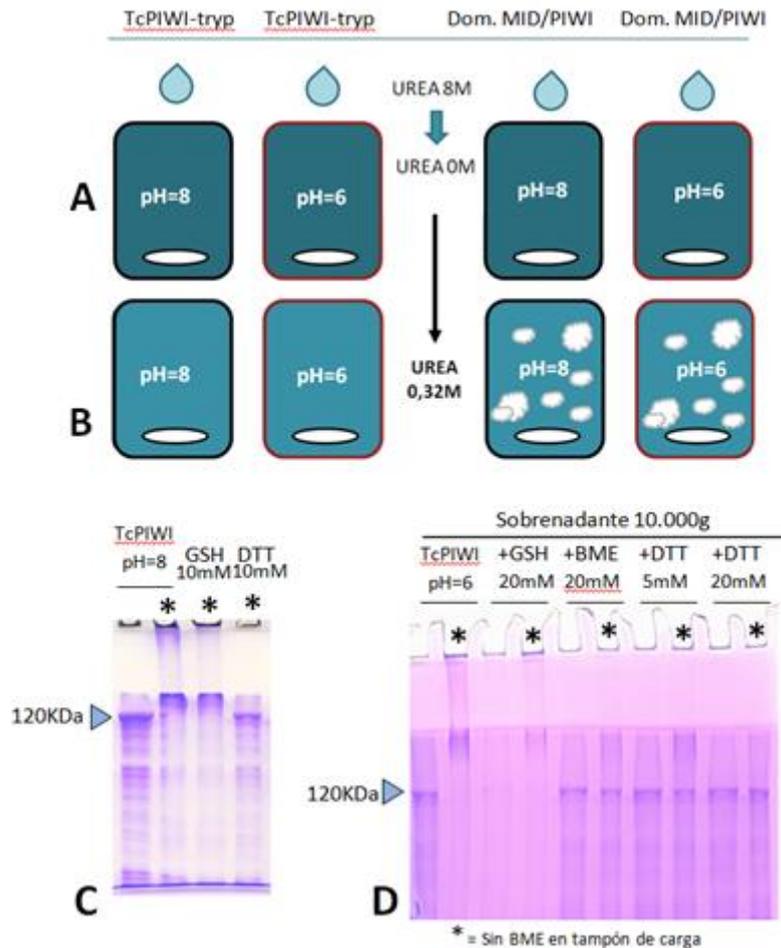


Figura 19: Replegamiento por dilución de la proteína TcPIWI-tryp y del dominio MID/PIWI (A y B). Distintas alícuotas de las proteínas fueron tratadas con diferentes agentes reductores, centrifugadas, y los sobrenadantes analizados por SDS-PAGE (C y D), con y sin (*) el agregado de BME 4% en el tampón de carga.

Ambas proteínas fueron agregadas gota a gota, a un flujo de aproximadamente 0,2mL/min. Finalizado el agregado, se mantuvieron las condiciones de agitación y temperatura, y se dejó la reacción proseguir toda la noche. Al día siguiente, la característica más llamativa fue la observación de grandes precipitados en el caso del dominio MID/PIWI, en cualquiera de los dos valores de pH de trabajo. Por el contrario, las soluciones de replegamiento conteniendo a la proteína completa no mostraban signos de turbidez (figura 19, B). Ambas proteínas se centrifugaron a 10.000g, observándose un gran *pellet* en el caso del dominio MID/PIWI, no así para la proteína completa. Una vez concentradas hasta un volumen final de 2mL, se tomaron alícuotas de ambas proteínas, que fueron analizadas por SDS-PAGE. Los gels correspondientes al dominio MID/PIWI no se muestran, pues no contenían cantidades detectables de

²⁴ El par GSH/GSSG puede no haber tenido el efecto buscado a pH=8, ya que a dicho pH el glutatión reducido puede oxidarse rápidamente de forma espontánea.

proteínas. Por el contrario, puede detectarse la proteína TcPIWI-tryp de 120KDa, ya sea a partir del replegamiento realizado a pH=8 (figura 19, C), como a pH=6 (figura 19, D).

Al igual que en el caso del replegamiento por IMAC, al realizar la corrida electroforética en condiciones no reductoras (es decir, sin el agregado de reductor más allá del contenido de por sí en la muestra; indicado en las figuras con un asterisco), la proteína deja de migrar hasta la posición esperada. Para el caso de la proteína replegada a pH=8, se agregó a alícuotas del mismo volumen glutatión reducido (10mM) o DTT (10mM) y se analizaron en condiciones “no reductoras”²⁵, recuperando el mismo perfil de migración. De este modo, queda claro que la proteína replegada se encuentra en forma de agregados covalentes solubles, que pueden ser disgregados en condiciones reductoras moderadas.

Lo mismo se realizó para el caso de la proteína replegada a pH=6, con la diferencia de que se tomó la precaución de centrifugar a 10.000g luego de adicionar los distintos agentes reductores, y analizar únicamente el sobrenadante. Como se muestra en la figura, se corrieron cinco alícuotas, ya sea sin reductor (carriles 1 y 2), con la adición de GSH 20mM (3 y 4), BME 20mM (5 y 6), DTT 5mM (7 y 8) y DTT 20mM (9 y 10). Los carriles impares corresponden a SDS-PAGE convencionales, mientras que los carriles pares – indicados con asteriscos – corresponden a SDS-PAGE en condiciones “no reductoras”. Llama la atención que para el caso del glutatión (GSH 20mM) no se observe casi proteína en ninguna de las dos condiciones. Esto quiere decir que los agregados solubles de proteína replegada precipitaron luego del agregado de GSH 20mM (se tomó la precaución de correr los mismos volúmenes en todos los carriles). Por el contrario, la adición de BME 20mM, DTT 5mM o DTT 20mM no afectó significativamente la solubilidad de los agregados, pero sí permitió romper los puentes disulfuro intermoleculares (presumiblemente). Algo que debe considerarse es que si bien las disoluciones de glutatión se realizaron a partir de la especie reducida (GSH), el estado redox del glutatión en las condiciones de trabajo no fue estudiado, y es probable que un porcentaje importante de este se hubiese oxidado espontáneamente en las condiciones de trabajo.

Considerando que la proteína de partida – en urea 8M – no se encontraba formando agregados covalentes, como queda claro de la figura 17, se arribó a la conclusión de que la presencia de glutatión oxidado, aún siendo éste minoritario frente al glutatión reducido, era suficiente para la formación de los mismos. Algo similar podría haber sucedido en los ensayos de replegamiento por IMAC, dónde la presencia de cationes divalentes pudiese haber catalizado la agregación oxidativa de la proteína. Por consiguiente, se decidió repetir el ensayo para la proteína TcPIWI-tryp, en condiciones netamente reductoras.

Se sustituyó el par GSH/GSSG por BME 10mM o por DTT 5mM en dos ensayos independientes, realizados en tampón Tris-Cl 50mM pH=8; NaCl 0,5M y el reductor correspondiente, obteniendo en ambos casos proteína soluble, y evitando la formación de agregados covalentes.

Lo anterior no implica de ningún modo que el replegamiento haya conducido a la obtención de proteína en su forma nativa y monomérica. Se corrieron geles nativos (sin SDS ni calor) a pH=8 y a pH=7,4 (siendo el punto isoeléctrico de la proteína 9,41), permaneciendo la proteína en la parte superior del gel, apenas por debajo del pocillo. Asimismo, se realizaron medidas de dispersión dinámica de luz (DLS, Malvern Instruments) que mostraron que la población proteica poseía un radio hidrodinámico equivalente a

²⁵ Vale aclarar que por “no reductoras” se entiende sin el agregado de β -mercaptoetanol 4% (aproximadamente 500mM) al tampón de carga. No quiere decir que las muestras no contengan de por sí agentes reductores.

una proteína globular de más de 700 KDa, indicando la presencia de agregados. Como se discutirá posteriormente, existen motivos para pensar que dichos agregados se deben – al menos en parte – a la presencia de ácidos nucleicos en la preparación.

7.4 Replegamiento por diálisis

Se probó una tercera estrategia de replegamiento *in vitro*, que consiste en la eliminación de la urea mediante diálisis. Este procedimiento es esencialmente distinto del anterior, en cuanto la concentración de proteína permanece relativamente constante durante todo el ensayo (disminuyendo un poco debido a la entrada de agua en los primeros minutos), y la concentración de urea disminuye lenta y gradualmente.

De las experiencias anteriores, se mantuvo siempre una concentración constante de BME de 10mM a ambos lados de la membrana (DTT 5mM en pruebas subsecuentes), para evitar la formación de agregados covalentes. La disminución en la concentración de urea es progresiva hasta alcanzar el equilibrio, luego de lo cual se detiene. Por esta razón, se dejó la proteína en diálisis toda la noche en cámara fría (6-7°C), y se hicieron al otro día tres cambios de solución de replegamiento. El objetivo de esto es eliminar completamente la urea, y también “reciclar” el BME, pues éste puede comenzar a oxidarse espontáneamente por reacción con el oxígeno disuelto y afectar el estado redox de la proteína.

Luego del tercer cambio en la solución de replegamiento, la proteína se centrifugó para eliminar todo aquello que hubiese precipitado y se concentró hasta 0,5mL (figura 20, A). Alícuotas del igual volumen fueron analizadas por SDS-PAGE antes (T) y después de centrifugar a 10.000g (P: *pellet*; S: sobrenadante). Si bien puede detectarse a la proteína TcPIWI-tryp en el precipitado, una fracción significativa permanece soluble. Más aún, la misma no se ve afectada por la presencia o ausencia de BME en el tampón de carga, lo que implica que la concentración de 10mM utilizada a lo largo del ensayo fue suficiente para prevenir la formación de puentes disulfuro intermoleculares.

Una vez concentrada, la proteína se inyectó en una columna de cromatografía de exclusión molecular (en adelante, gel filtración o SEC, de sus siglas en inglés). Se utilizó una columna Superdex S200 10/300 (Amersham, GE Healthcare) capaz de separar proteínas en el rango entre 10 y 600 KDa, operando a un flujo de 0,5mL/min y equilibrada en el mismo tampón utilizado para el replegamiento (Tris-Cl 50mM pH=8; NaCl 0,5M; BME 10mM). Según la calibración propia de la columna, se espera que una proteína globular de 120KDa eluya con un volumen de retención de 12-13mL. En efecto, se observó un pico claramente definido en 12,89 mL (figura 20, B), que al ser analizado por SDS-PAGE y teñido con plata, denota la presencia de una proteína que migra justo por encima de la banda del marcador de peso molecular de 118 KDa (figura 20, C). Esto quiere decir que la proteína TcPIWI-tryp pudo ser replegada *in vitro*, hasta la obtención de proteína soluble y monomérica.

Desde luego, aparece aquí un evidente problema de rendimiento. La intensidad del pico en cuestión (Abs 280nm) es de 0.006 unidades de absorbancia, que para una proteína de 120KDa con un coeficiente de extinción de $89840 \text{ M}^{-1} \text{ cm}^{-1}$ en condiciones reductoras²⁶, y con un ancho de pico de aproximadamente 1mL, corresponde a no más de 8µg. El resto de la proteína seguramente se encuentra

²⁶ Estimado mediante el algoritmo ExPASy Prot param: <http://ca.expasy.org/cgi-bin/protparam>

en forma de agregados solubles, que se evidencian por la presencia de un pico de alta intensidad (relativa) en el volumen muerto de la columna (V_m).

Resulta extraño que al analizar las fracciones correspondientes al volumen muerto por SDS-PAGE convencional, no se hubiese detectado una banda a 120KDa (ni a otros tamaños), aún mediante tinción con plata. Esto sería lo esperable si la absorbancia a 280nm en el volumen muerto correspondiese a agregados solubles de TcPIWI-tryp. Frente a la ausencia aparente de proteína, se juntaron las fracciones correspondientes al volumen muerto, se concentraron mediante el sistema Ultrafree-15 (NMWCO 30KDa) y se analizaron por SDS-PAGE, pero utilizando una tinción con bromuro de etidio en lugar de una tinción convencional con Azul de Coomasie o plata. Al revelar bajo luz ultravioleta, se observó en la fracción concentrada – no así en el líquido no retenido – una banda intensa en la región superior del gel separador, acompañada de lo que se denomina un “lagrimeo” o “smear”. Si bien algunas proteínas se tiñen parcialmente con bromuro de etidio, no se detectan bandas en la altura correspondiente en el gel teñido con plata (figura 20, C), aunque debe reconocerse que la muestra teñida con bromuro de etidio fue previamente concentrada.

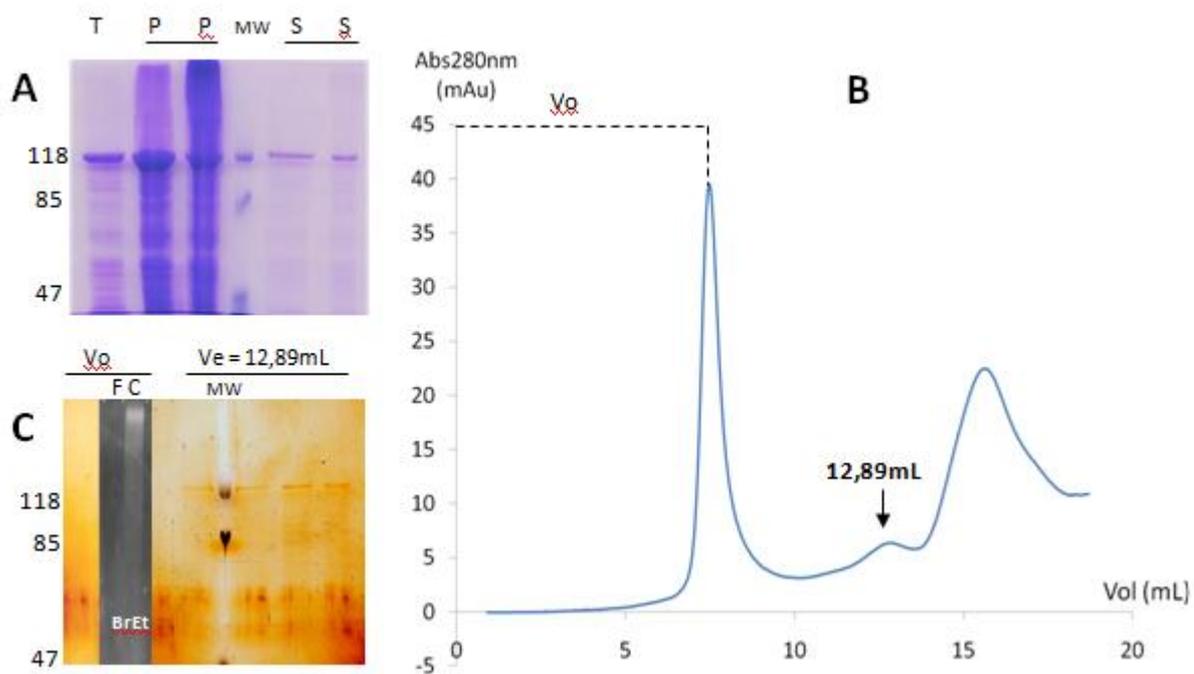


Figura 20: Replegamiento por diálisis de la proteína TcPIWI-tryp. A: Análisis por SDS-PAGE de la proteína TcPIWI-tryp antes (T) y después de centrifugar a 10.000g (P: pellet; S: sobrenadante). Los asteriscos indican SDS-PAGE corridos sin el agregado de BME 4% al tampón de carga. B) La proteína soluble replegada fue concentrada a 0,5mL y analizada por cromatografía de exclusión molecular en una columna Superdex S200 10/300 (Amersham, GE Healthcare). Las fracciones correspondientes al volumen de retención = 12,89mL fueron analizadas por SDS-PAGE (8%) y tinción con plata (C). Las fracciones correspondientes al volumen muerto o de exclusión (V_0) fueron analizadas en el mismo gel, o concentradas y analizadas por SDS-PAGE mediante tinción con bromuro de etidio (BrEt).

La presencia de ácidos nucleicos en la preparación de la proteína TcPIWI-tryp luego del replegamiento por diálisis también se manifiesta en el espectro de absorción en el ultravioleta de la proteína replegada.

El espectro muestra un incremento lineal en la absorbancia a medida que disminuye la longitud de onda, característico de fenómenos de dispersión por agregados, pero con un pico definido y pronunciado a 260nm (figura 21, A).

La proteína TcPIWI-tryp replegada en presencia de BME 1mM fue también analizada por DLS, previo a realizar la gel filtración (figura 21, B). Es importante señalar que la proteína fue centrifugada a 16.000g inmediatamente antes de las medidas, para descartar la presencia de material particulado. La muestra presenta un comportamiento por DLS con dos poblaciones claramente distinguibles entre sí. La de menor intensidad es mayoritaria (93%) en términos de porcentaje del volumen total ocupado, y posee un radio hidrodinámico de 10,67nm, con un ancho de pico de 1nm. Este radio hidrodinámico equivale al de una proteína perfectamente globular de 776KDa (de ahí se desprende que sea excluido de una columna empaquetada con Superdex 200). Desde luego, este valor no quiere decir nada por sí mismo, pero da una cierta idea del tipo de agregados solubles con los que se trabaja. Aún en un estado máximo de compactación, cada partícula de 10,67nm de radio no podría corresponder a más de 6 o 7 moléculas de TcPIWI-tryp. Es improbable que agregados totalmente inespecíficos de proteína mal plegada den lugar a partículas de tamaño discreto, conteniendo un número relativamente bajo de moléculas proteicas por partícula. Combinando esta información con las observaciones anteriores, resulta posible suponer la presencia de complejos entre una o más moléculas de TcPIWI-tryp y ácidos nucleicos bacterianos.

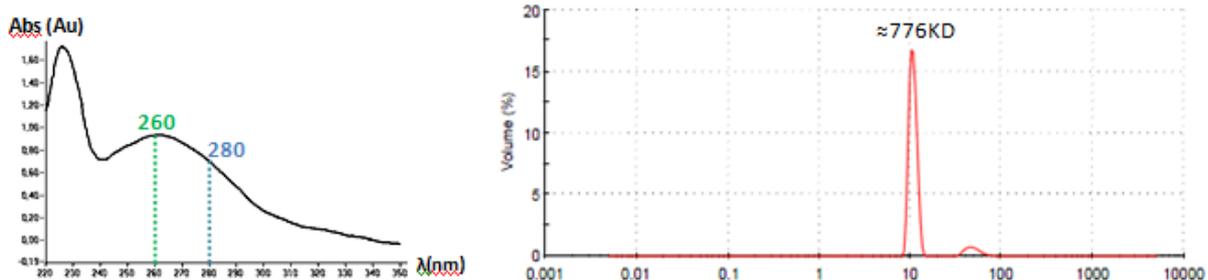


Figura 21: Co-purificación de ácidos nucleicos bacterianos con la proteína TcPIWI-tryp replegada. A: espectro de absorción en el ultravioleta cercano de la proteína TcPIWI-tryp replegada. B: análisis por DLS de la proteína replegada luego de centrifugar a 16.000g. La gráfica indica el porcentaje del volumen total ocupado en función del radio hidrodinámico de las partículas en nanómetros. La escala en el eje de las abscisas no es lineal.

Para corroborar esta hipótesis se incubó la proteína TcPIWI-tryp con 2 unidades de ADNasa I (libre de ARNasas, certificada), 32 μ g de ARNasa A (3 unidades Kunitz), o una mezcla de ambas. Luego de incubar toda la noche a 4°C, las alícuotas se centrifugaron a 16.000g, observándose la presencia de un precipitado. Los sobrenadantes se volvieron a centrifugar y se analizaron por DLS. En ambos casos, pero sobre todo para el tratamiento con ADNasa I, el radio hidrodinámico de la población disminuyó hasta casi la mitad. La doble incubación con ADNasa y ARNasa no incrementó esta disminución por debajo del valor obtenido con la ADNasa sola. La adición de ADNasa I al tampón de lisis bacteriana es a la luz de estos resultados una prueba pertinente a ensayar.

8 SOBRE-EXPRESIÓN DE TcPIWI-tryp EN *T. cruzi* E INMUNOPRECIPITACIÓN

Epimastigotes de la cepa Dm28c de *T. cruzi* fueron transfectados por electroporación, utilizando el plásmido recombinante pTEX/TcPIWI-tryp. Como control, se transfectaron también parásitos con el vector pTEX cerrado, es decir, sin inserto. Ambos grupos fueron cultivados utilizando concentraciones crecientes de neomicina a efectos de seleccionar aquellos parásitos que hubiesen incorporado los respectivos plásmidos.

Luego del mes post-transfección, los sobre-expresantes de TcPIWI-tryp nunca recuperaron la morfología característica de los epimastigotes, permaneciendo redondeados y en algunos casos, presentando una extensión apical.

La inmunoprecipitación de TcPIWI-tryp se realizó tanto en parásitos sobre-expresantes como en aquellos no transfectados, utilizando esferas magnéticas (Dynabeads, Invitrogen) unidas a anticuerpos anti-IgG de conejo. Las esferas magnéticas se incubaron con el suero policlonal obtenido según se describió en secciones anteriores, y posteriormente con extractos de epimastigotes de *T. cruzi* lisados en condiciones no desnaturizantes. La figura 22 (izquierda) muestra un análisis por SDS-PAGE de las proteínas inmunoprecipitadas por este medio, corridas en condiciones reductoras (+) y no reductoras (-). Por “condiciones no reductoras” se hace referencia únicamente a las condiciones de corrida electroforética, ya que se incluyó BME en concentraciones bajas a la solución de extracción de las proteínas inmunoprecipitadas (lo que explica por qué se observa la banda característica de las cadenas pesadas de las inmunoglobulinas a 55KDa en ambos casos).

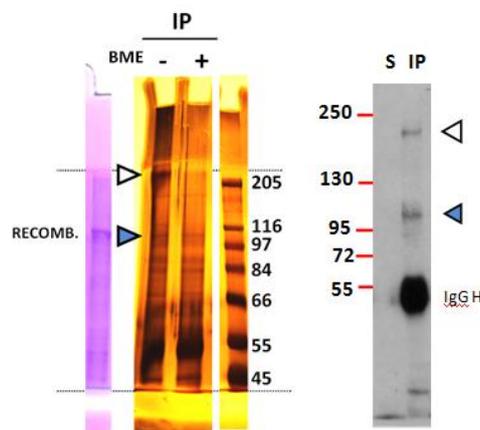


Figura 22: Inmunoprecipitación de TcPIWI-tryp utilizando un extracto no desnaturizante de epimastigotes de *T. cruzi* no transfectados (izquierda) y sobre-expresantes de la proteína TcPIWI-tryp (derecha). A la izquierda se muestra un SDS-PAGE (8%) teñido con plata, donde se analizó la fracción inmunoprecipitada. Los signos +/- indican si el SDS-PAGE se realizó en condiciones reductoras (+) o no reductoras (-). A la izquierda se muestra un gel corrido en paralelo y teñido con azul de coomasie, donde se corrió la proteína TcPIWI-tryp recombinante (RECOMB.) a efectos de facilitar la detección de la banda inmunoprecipitada (indicada con una flecha azul). La flecha blanca indica una banda de más de 205kDa, que solo aparece en condiciones no reductoras. La figura de la derecha muestra un análisis por Western blot de la fracción inmunoprecipitada (IP) y no unida a las esferas magnéticas (S), a partir de un extracto de parásitos sobre-expresantes de TcPIWI-tryp.

Entre las proteínas que aparecen en la fracción inmunoprecipitada del gel teñido con plata de la figura 22 se observa una banda que corre apenas por debajo del marcador de 116 pb (que presenta una diferencia de carga evidente, que puede afectar ligeramente su movilidad electroforética). De modo

comparativo, se corrió en paralelo a la proteína recombinante (en geles separados, para evitar contaminación de pocillos), que se muestra luego del teñido con azul de coomasie.

Entre las otras bandas presentes, llama la atención una que migra justo por debajo del límite entre el gel de *stacking* y el gel separador, que solamente se observa en condiciones no reductoras. Si bien el peso molecular asociado a dicha banda no puede estimarse ya que se encuentra por encima del marcador más alto, resulta llamativo que al realizar un Western blot de la fracción inmunoprecipitada también se observa, junto con la banda esperada a 118 KDa, una segunda banda inmunoreactiva que migra por debajo de los 250 KDa. Por su migración e inmuno-reactividad, dicha banda bien podría corresponder a un dímero de TcPIWI-tryp, aunque hasta el momento no se dispone de suficiente información como para aseverarlo. Tampoco se realizó - hasta el momento - un Western blot de la fracción inmunoprecipitada realizado en condiciones reductoras y no reductoras de SDS-PAGE, a efectos de determinar si la presencia de la banda de ≈ 240 KDa aumenta en proporción al impedir la rotura de puentes disulfuro intermoleculares.

De la proteína inmunoprecipitada a partir de parásitos sobre-expresantes, se lograron obtener cantidades suficientes de proteína como para realizar una extracción de ARN y analizar por electroforesis el ARN co-inmunoprecipitado con la misma. Dicho ARN se fraccionó en gel, clonando la fracción de menos de 35 nucleótidos según el protocolo convencional para el clonado de ARNs pequeños utilizado en nuestro laboratorio (Garcia-Silva et al. 2010) y obteniendo un total de 87 secuencias, distribuidas del siguiente modo: tARNs (55,6%); rARNs (18,5%), mARNs/ESTs (7,4%), sin asignar (18,5%).

Del 55,6% de secuencias de menos de 35 nucleótidos obtenidas, correspondientes a fragmentos de tARNs, el 53% de las mismas correspondieron a fragmentos 5' de isoaceptores de ácido aspártico, seguido por isoaceptores de ácido glutámico (13%), valina (13%), prolina (7%) y glutamina (7%).

Que la proteína TcPIWI-tryp co-inmunoprecipite con ARNs de pequeño tamaño, de los cuales más de la mitad corresponden a fragmentos de tARNs es desde luego un resultado sumamente relevante. Justamente por eso, requiere ser tratado con cuidado, reproducido y estudiado con otro método diferente. En trabajo de otros miembros de nuestro grupo, se realizaron sucesivas inmunoprecipitaciones, hasta lograr ARN en suficiente calidad y cantidad para realizar un estudio por secuenciado masivo (Illumina). El secuenciado se obtuvo y los datos se encuentran actualmente en procesamiento.

9 EXPRESIÓN Y PURIFICACIÓN EN *E. coli* DEL DOMINIO PAZ

9.1 Purificación y estabilidad de la proteína purificada

El dominio PAZ se expresó también en la cepa XL1Blue de *E. coli*, a partir del vector pQE-80_TEV/PAZ. A diferencia de lo ocurrido para la proteína completa o para el dominio MID/PIWI, esta vez se obtuvo una expresión mayoritaria en la fracción soluble (figura 23, A, recuadro interno).

Lo primero que se intentó fue expresar la proteína a partir de 1L de cultivo en medio LB, y realizar un primer paso de purificación por IMAC. Se eluyó proteína que pudo ser cuantificada por su absorbancia a

280nm (alrededor de 5mg), aunque es de suponer que la misma no se encontrase en estado puro. La proteína eluída se almacenó toda la noche a 4°C, y se retiró al día siguiente para su análisis por SDS-PAGE. Al retirarla, resultó claro que toda la proteína había precipitado durante su almacenamiento.

Se repitió entonces la expresión, pero modificando el paso de purificación. En vez de recurrir al uso de un escalón abrupto de imidazol (20mM-500mM), se eluyó por medio un gradiente desde 10mM a 500mM en 20 volúmenes de columna, utilizando un equipo de FPLC (figura 23, A). El cromatograma muestra la presencia de un primer pico bimodal cuándo el eluyente contiene un 21% de tampón de elución (*buffer B*, conteniendo imidazol 500mM), lo que corresponde a una concentración de imidazol de 110mM. Sin lograr una buena resolución del pico anterior, aparece luego el pico principal al 38% de *buffer B* (190-200mM imidazol). Dicho pico corresponde al dominio PAZ, como queda claro luego de su análisis por SDS-PAGE (figura 23, B). Integrando el pico y considerando el coeficiente de extinción a 280nm del dominio PAZ, se puede estimar una elución de casi 3mg de proteína.

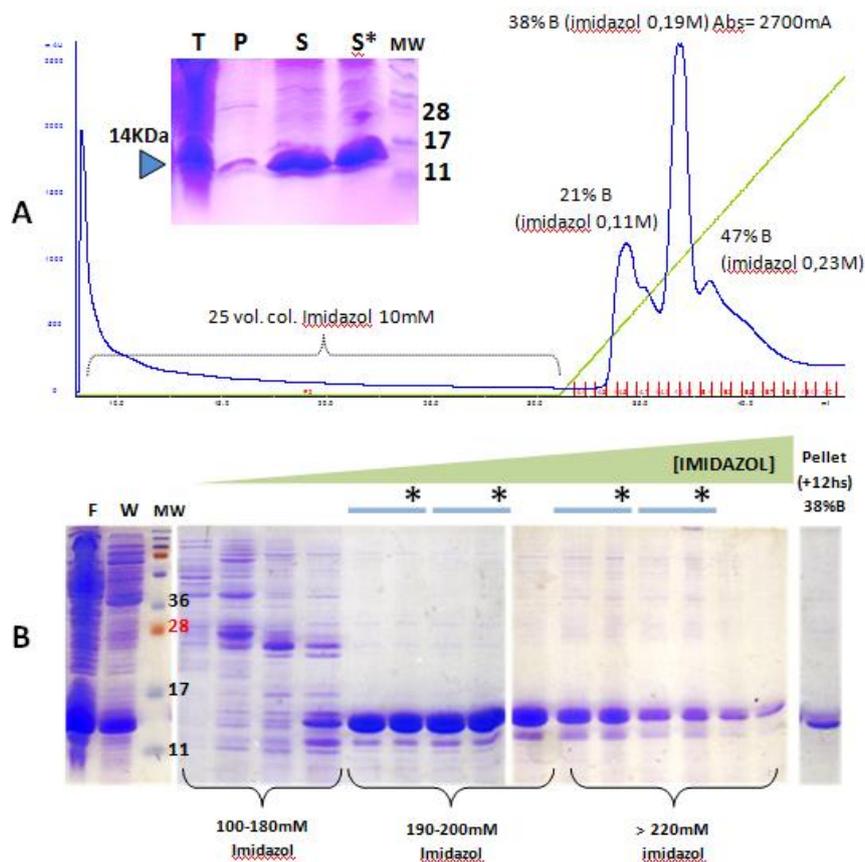


Figura 23: Expresión en *E. coli* y purificación del dominio PAZ. A (gel interno): en las condiciones de expresión utilizadas, la mayor parte de la proteína se recupera de la fracción soluble (S). La proteína no se encuentra formando agregados covalentes solubles, como puede evidenciarse al correr el gel en condiciones no reductoras (S*). La fracción insoluble y el extracto total se denotan P y T respectivamente. A (cromatograma): purificación por IMAC del dominio PAZ soluble, mediante un gradiente de elución de imidazol, desde 10mM hasta 500mM (verde). La curva azul corresponde a la absorbancia corregida a 280nm a la salida de la columna. B: Purificación del dominio PAZ por IMAC. F: fracción no unida a la columna. W: primer lavado de 5 volúmenes de columna con el tampón de equilibrado conteniendo imidazol 10mM. Las fracciones eluídas en gradiente de imidazol fueron analizadas por SDS-PAGE (15%). Los asteriscos indican que se corrió la misma fracción que en el carril de su

izquierda, pero en condiciones no reductoras. A las 12 horas de almacenamiento a 4°C, la mayor parte del pico eluído a 190-200mM imidazol había precipitado. Dicho precipitado se resuspendió en tampón de carga y se analizó en el último carril.

Curiosamente, aparece un segundo pico, de menor intensidad que el anterior, a un porcentaje mayor de *buffer* B (230mM imidazol) que luego decae con una pendiente suave. Analizando dichas fracciones por SDS-PAGE, resulta claro que también corresponden al dominio PAZ. Por tanto, habría dos poblaciones discretas que se caracterizan por una diferencia de afinidad por el Ni²⁺ presente en la columna, y eluyen por consiguiente a distintas concentraciones de imidazol.

Al igual que antes, la proteína purificada se almacenó a 4°C, y menos de una hora después se observó ya la presencia de precipitado. Se la centrifugó a máxima velocidad, y se dejó el sobrenadante por otras 12 horas, volviéndose a observar precipitado. El mismo se sedimentó por centrifugación, se resuspendió en tampón de carga, y se analizó por SDS-PAGE, constatando que el precipitado contenía mayoritariamente a la proteína PAZ. La conclusión de estos estudios fue que si bien la proteína se expresaba de forma soluble, era inestable al almacenamiento luego de su purificación.

Por un tiempo se intentó revertir esta inestabilidad de diversas formas, sin conseguirlo. Pasado unos meses, se comenzó a observar con el replegamiento de la proteína completa los problemas asociados a la formación de puentes disulfuro intermoleculares de los que ya se ha hablado. Se hipotetizó entonces que estas observaciones pudiesen ser aplicables también a los problemas de inestabilidad asociados al dominio PAZ, ya que la oxidación de las cisteínas podría catalizar la formación de agregados. Es así que en ensayos subsiguientes se comenzó a agregar DTT 1mM a todas las fracciones conteniendo la proteína de interés, inmediatamente después de su elución por IMAC. Sorprendentemente, el agregado de DTT 1mM fue suficiente para detener los procesos de agregación durante el almacenamiento, obteniendo de esta forma proteína soluble, pura y estable.

9.2 Clivaje del 6x-HIS tag y purificación de la proteína clivada por IMAC y Gel-Filtración.

Habiendo logrado estabilizar al dominio PAZ purificado, el paso siguiente fue la remoción de toda la secuencia N-terminal ajena a dicho dominio, de forma que ésta no interfiriese con la actividad o la estructura propia del mismo. También se buscó definir un protocolo de purificación adecuado, tal que permitiese obtener varios miligramos de proteína con “pureza electroforética” (banda única en un SDS-PAGE teñido con azul de coomasie), con un estado de oligomerización único y definido y correspondiente a una población “monodispersa” por DLS. Todos estos son requisitos de cara a la realización de ensayos de cristalogénesis a mediana escala.

En esta etapa se introdujo una primera modificación al protocolo de purificación, cultivando las bacterias en el medio 2xYT en vez de en el medio LB, lo cual permitió obtener mayor rendimiento de masa celular.

Por simplicidad, se volvió a realizar la purificación por IMAC en escalón de imidazol en vez de en gradiente (figura 24, izquierda; “IMAC 1”). Para que esto no implicase una concesión en términos de pureza, se pegó la proteína en presencia de imidazol 20mM y la columna se lavó con esta disolución por 10 volúmenes (W20), seguidos de otros 10 volúmenes con imidazol 70mM (W70). Puede observarse que

la fracción correspondiente a la proteína eluida (E) es igual o aún más pura que en el caso anterior. Creyendo que sería más eficiente que el agregado de DTT luego de la purificación, se incluyó DTT 1mM en todos los tampones de IMAC, lo que traería consecuencias que se discutirán más adelante.

Se estimó la concentración de proteína eluida espectrofotométricamente, y se agregó la proteasa del virus del mosaico del tabaco (proteasa TEV) a razón de 1mg de proteasa cada 40mg de proteína recombinante pura. La digestión se llevó a cabo toda la noche a 4°C en diálisis, de modo que la concentración final de imidazol en el equilibrio fuese de 20mM (para poder aplicar una segunda IMAC).

La proteasa TEV utilizada corresponde a una proteína recombinante que posee un 6x-HIS tag en su extremo N-terminal pero – a diferencia del dominio PAZ recombinante – no posee un sitio de reconocimiento y clivaje para sí misma. Esto permite purificar la proteína de interés clivada de una forma sencilla, tanto de la proteasa y otras proteínas bacterianas como de aquellos remanentes que por alguna razón no hubiesen sido clivados. La purificación consiste en una segunda IMAC (“IMAC 2”, figura 24), con la diferencia de que en esta oportunidad la proteína clivada y purificada se obtiene de la fracción no unida, ya que habría perdido el motivo rico en histidinas que permitía su unión a la columna.

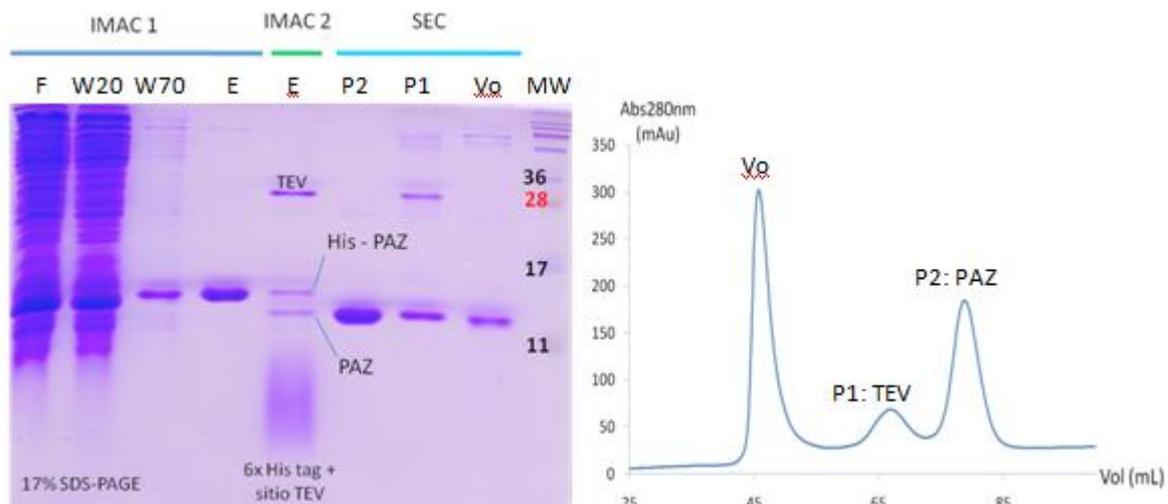


Figura 24: Purificación del dominio PAZ por IMAC, clivaje del 6x His-tag y purificación adicional por cromatografía de exclusión molecular (gel filtración). Los primeros 4 carriles del gel de la izquierda corresponden a las distintas fracciones obtenidas a partir de una IMAC convencional, eluyendo con un escalón de imidazol (F: no unido; W20: lavado por 10 vol. col. con imidazol 20mM; W70: lavado por 10 vol. col. con imidazol 70mM; E: eluido con imidazol 500mM). Toda la purificación se realizó en presencia de DTT 1mM, en todos los tampones. La proteína eluida se incubó toda la noche a 4°C en diálisis con la proteasa TEV, y la misma se purificó por medio de una segunda IMAC, obteniendo la proteína clivada en la fracción no unida. El último paso de purificación consistió en una gel filtración (SEC) en columna S75 16/600 (Amersham, GE Healthcare) (derecha). Los picos correspondientes al volumen de exclusión (Vo; 47.3mL), P1 (Ve=66,03mL) y P2 (Ve=78,32mL) fueron analizados en el mismo gel de la izquierda.

Luego de haber pasado el producto de la IMAC 1 por la IMAC 2 y haber recogido la fracción no unida, se procedió a eluir todas las proteínas retenidas en la columna volviendo a subir la concentración de imidazol a 500mM. Dichas fracciones se analizaron por SDS-PAGE y aparecen en el gel de la figura 24 bajo el título “TEV” (carril n° 5). Se puede observar una proteína que migra entre los marcadores de 28 y

36 KDa, no identificable en el extracto bacteriano, que corresponde a la proteasa TEV (cuyo peso molecular es de 27 KDa, más el péptido N-terminal que contiene el 6x-HIS tag). Por debajo de la proteasa se observan dos bandas, cuyos tamaños coinciden con los de la proteína sin clivar (His – PAZ) y la proteína clivada (PAZ). Si bien la diferencia de tamaños entre ambas es de solo 2KDa, la utilización de geles de poliacrilamida al 17% permite una muy buena resolución en este rango de tamaños. Cabe preguntarse por qué se recupera proteína clivada, puesto que sería de esperar que la misma no tuviera la capacidad de unirse a la columna y resistir a los lavados con imidazol 70mM. Lo más probable es que la misma no estuviese clivada al momento del pegado a la columna, sino que fuese un producto de digestión post-purificación. También llama la atención una “sombra” que se ubica en la región inferior del gel, que probablemente corresponda al péptido N-terminal clivado.

El paso siguiente de purificación consistió en una cromatografía de exclusión molecular (SEC) utilizando para ello una columna Superdex 75 16/600 (Amersham, GE Healthcare) y el equipo de FPLC ÄKTApurifier (Amersham, GE Healthcare). Además de posibilitar la purificación propiamente dicha y de permitir determinar el estado de oligomerización en el cual se encuentra la proteína, también se utilizó la SEC para reducir la concentración de sales en la que se encuentra la proteína, procurando dejarla en condiciones óptimas para la realización de ensayos de cristalogénesis por difusión de vapor. Se utilizó por tanto un tampón de corrida compuesto por Tris 20mM pH=8; NaCl 150mM y DTT 1mM, obteniendo el cromatograma que se muestra en la figura 24 (derecha).

El cromatograma obtenido muestra tres picos bien resueltos entre sí, correspondiendo el primero de ellos al volumen muerto de la columna (V_0 ; 47,3mL), y denominando P1 y P2 a los restantes, presentando volúmenes de elución de 66,03 y 78,32 mL respectivamente. Utilizando la calibración de la columna que suministra el fabricante, pueden estimarse los pesos moleculares de las proteínas correspondientes a los picos P1 y P2 en 26,9 KDa y 13,1 KDa respectivamente. Admitiendo ligeras variaciones (dependientes también de la estructura proteica), estos tamaños son coincidentes con el de la proteasa TEV y la proteína PAZ clivada (PM teórico: 12,1KDa). Los picos V_0 , P1 y P2 fueron también analizados por SDS-PAGE, como se muestra en el gel de la figura 24. En efecto, el pico P2 corresponde al dominio PAZ clivado y puro. Como se había predicho, en el pico P1 se detecta la proteasa TEV, aunque aparece también una cantidad apreciable de proteína PAZ clivada, arrastrada desde el pico P2 y/o el volumen muerto. En el volumen muerto la proteína mayoritaria es la proteína PAZ clivada, lo que implica que durante la purificación, una fracción considerable de la misma sufre agregación. Como se verá más adelante, la proporción de proteína que aparece en el volumen muerto guarda una estrecha relación con las condiciones redox utilizadas durante la purificación.

En resumen, el dominio PAZ recombinante pudo expresarse de forma soluble, estabilizarse por DTT, digerirse con la proteasa TEV, y purificarse hasta pureza electroforética en su forma monomérica. También se logró reducir la concentración de sales sin pérdida de estabilidad. El problema a este nivel fue una cuestión de rendimientos. El pico P2 aislado y concentrado corresponde a aproximadamente 1.4mg de proteína, pero el volumen de medio de cultivo de partida fue de 3L, por lo que el rendimiento final sería de 0,45mg de proteína pura por litro de cultivo. Desde luego, parte de la proteína expresada se pierde en la forma de agregados solubles que eluyen en la gel filtración en el volumen muerto, pero la mayor parte de la proteína parece perderse en la IMAC 1, ya sea en los lavados, o por agregación dentro de la columna, evidente a simple vista.

9.3 Optimización de las condiciones de purificación del dominio PAZ

Del cromatograma mostrado en la figura 24 (que es el mismo mostrado en la figura 25, A, curva roja) llama la atención que la mayor parte de la proteína inyectada en la columna se encuentra en forma de agregados que eluyen en el volumen muerto. La formación de tales agregados disminuye el rendimiento final obtenido en más de un 50%. De los ensayos de replegamiento realizados para la proteína completa, se constató que los agregados covalentes se formaban una vez que la proteína resolubilizada en urea 8M se ponía en contacto con la columna de IMAC. Más aún, el agregado o no de agente reductor durante la IMAC implicaba la elución de agregados solubles o la no elución de proteína. Todo esto condujo a pensar que la proteína TcPIWI-tryp es sensible a la oxidación catalizada por los iones Ni^{2+} , y que la presencia de agentes reductores en contacto con la proteína y con el Ni^{2+} podría tener el efecto opuesto al deseado. Se especuló que fenómenos similares podrían estar sucediendo en el caso del dominio PAZ, y que la inclusión de DTT 1mM en las soluciones de pegado, lavado y elución podría no haber sido una buena idea.

Para corroborar esta hipótesis, se alteró el protocolo de purificación, evitando poner en contacto agentes reductores con la columna HisTrap. De este modo, la IMAC1 se realizó en condiciones no reductoras, pero teniendo la precaución de agregar DTT 1mM inmediatamente después de la elución (de no hacerlo, la proteína es inestable y precipitaría, como se constató numerosas veces). También se alteraron ligeramente las condiciones de expresión, induciendo toda la noche a 20°C en vez de 4hs a 30°C, obteniendo un mayor rendimiento en el peso del cultivo (4 gr. / litro). La concentración de IPTG (0,5mM), el momento de inducción (Abs 600nm=1.0), el volumen de cultivo de partida (3L) y demás parámetros permanecieron constantes entre ambos ensayos. La mayor diferencia se vio luego de analizar por gel filtración la proteína purificada (figura 25, A, curva verde). Si bien las cantidades totales de proteína inyectadas a la columna son similares, el perfil de elución se invierte, disminuyendo a menos de la mitad la intensidad del pico en el volumen muerto, y duplicándose la intensidad del pico a 78-79mL, donde eluye la proteína PAZ clivada y monomérica. El pico eluido corresponde a 2,5mg de proteína pura, por lo que el rendimiento de proteína obtenida de forma pura pasó de 0.45 a 0.82mg por litro de cultivo.

La proteína obtenida se analizó por dispersión dinámica de luz (DLS) a 4°C, obteniendo la distribución de tamaños que se muestra en la figura 25, B. De dicho resultado se desprende que la solución conteniendo a la proteína purificada se comporta como una población monodispersa compuesta por partículas de 1.69 nm de radio (con un ancho de pico de 0.26nm). Dicho tamaño correspondería al de una proteína perfectamente globular de 15KDa, que “coincide” con los 12,1KDa del dominio PAZ. También se observa una segunda población más heterogénea y de mucho mayor tamaño (5nm de radio, con un ancho de pico de 4 nm) que podría corresponder a la presencia de agregados, pero en términos relativos los mismos representan una fracción muy minoritaria.

Utilizando un sistema de concentración por centrifugación con membranas de 3KDa de “cut-off” (Vivaspin 20, Sartorius) la proteína se concentró hasta 10mg/mL y se almacenó a -80°C. La proteína resistió adecuadamente los ensayos de congelación/descongelación que se realizaron, no observándose pérdidas cuantiosas de material luego de descongelar, centrifugar y cuantificar el sobrenadante. Cabe señalar que este lote de proteína, concentrado a 10mg/mL y almacenado a -80°C, fue el que se utilizó para realizar el *screening* de condiciones de cristalización, tal como se describirá más adelante.

Un tercer lote de proteína se preparó de manera similar, pero introduciendo una nueva modificación en la primera IMAC. Si bien las modificaciones ya descritas habían permitido elevar los rendimientos del

proceso, la decisión de no incluir agente reductor hasta haber eluido la proteína de la primera IMAC implicó una pérdida considerable de material que precipitó dentro de la columna HisTrap. Esto se evidencia por la coloración de la columna después de la elución. En esta oportunidad, se decidió probar un procedimiento intermedio: el extracto bacteriano soluble se trató con BME 4mM, de modo de impedir la formación de agregados covalentes antes y durante el pegado de las proteínas a la columna. Pero a diferencia del primer caso, no se agregó agente reductor a las soluciones de lavado ni de elución, de modo que el BME fue rápidamente eliminado de la columna. Inmediatamente después de la elución, se agregó DTT 1mM y se procedió a realizar el clivaje del *tag* de histidinas, la segunda IMAC y la gel filtración (figura 25, A, curva azul) en las condiciones ya estipuladas.

Dado que el extracto bacteriano proveniente de 3L de cultivo se dividió asimétricamente en dos fracciones (a efectos de tratar una parte con iodoacetamida, como se explicará a continuación), la cantidad de proteína inyectada en la columna de gel filtración corresponde en realidad a 1,25L de cultivo. Si bien el volumen efectivo de cultivo fue menos de la mitad, la intensidad del pico a 78-79mL en la gel filtración fue considerablemente mayor a la del caso anterior, mientras que la intensidad del pico correspondiente al volumen muerto permaneció inalterada. Por tanto, no sólo se logró elevar los rendimientos obtenidos luego de la purificación, sino que se mejoró aún más la relación entre la proteína soluble en su forma monomérica y la proteína agregada. En esta preparación se obtuvieron 3,3mg de proteína presentando “pureza electroforética” (figura 25, D, cuarto carril), lo que corresponde a un rendimiento de 2,64 mg de proteína pura por litro de cultivo, muy superior a los valores iniciales. Si bien este rendimiento está lejos de ser elevado, sí es aceptable para una producción eficiente de proteína recombinante destinada a la realización de estudios estructurales, biofísicos y bioquímicos.

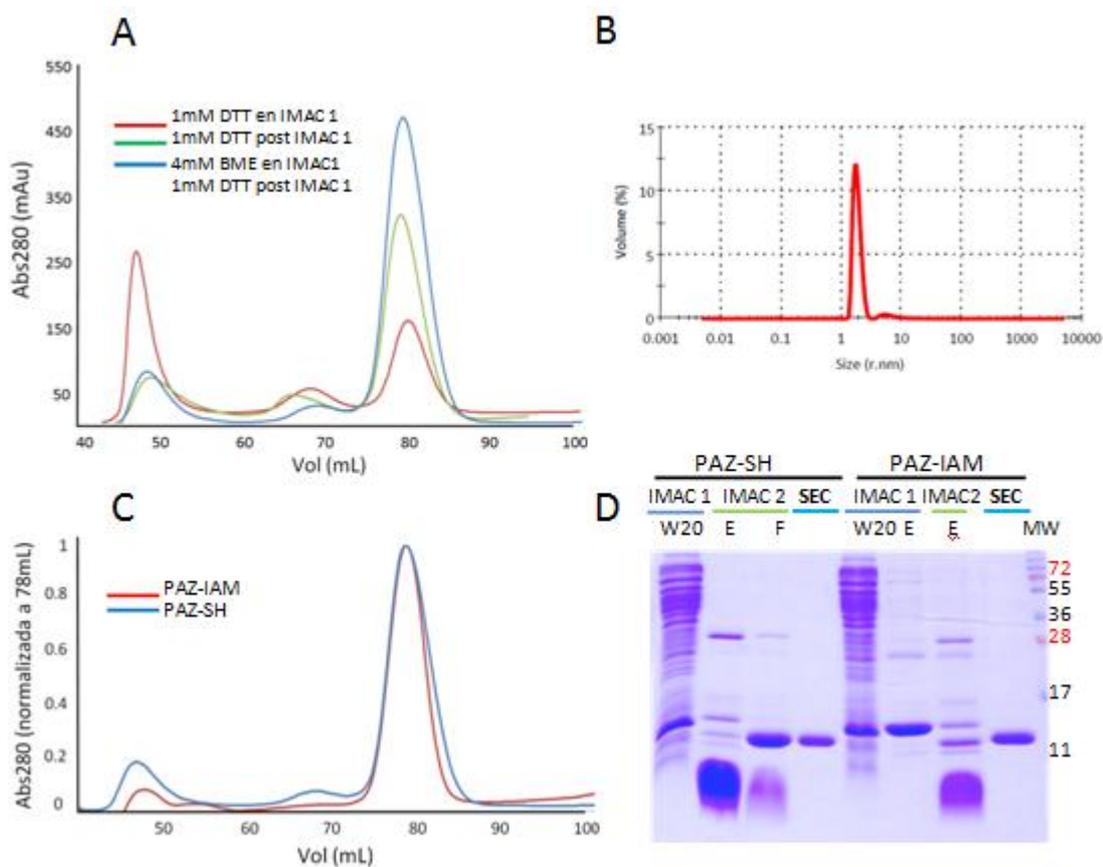


Figura 25: Optimización de las condiciones de purificación del dominio PAZ recombinante. A) Cromatogramas obtenidos luego de la gel filtración (S75 16/600) de tres lotes de proteína purificados de forma distinta a partir de 3L de cultivo 2xYT (1,5L para la curva azul. Rojo: adición de DTT 1mM en las soluciones de pegado, lavado y elución; verde: no adición de agente reductor sino hasta después de haber eluido la proteína; azul: adición de BME 4mM al extracto, y luego DTT 1mM después de eluir la proteína). B) Ensayo de dispersión dinámica de luz (DLS) de la proteína PAZ purificada (el lote utilizado corresponde a la curva verde en (A)). C) comparación de los cromatogramas obtenidos luego de la gel filtración de la proteína PAZ purificada en condiciones reductoras (PAZ-SH; azul), o luego de haber sido tratada con agente reductor y iodoacetamida (PAZ-IAM; rojo). Los valores de absorbancia se normalizaron por el valor correspondiente a 78mL. D: Análisis por SDS-PAGE (17%) de distintas fracciones correspondientes a la purificación de la proteína PAZ-SH y de la proteína PAZ tratada con IAM previo a la IMAC 1.

Otra variante que se decidió probar fue la alquilación de las cisteínas reactivas de la proteína. Como ya se explicó, existen razones para pensar que la oxidación del dominio PAZ catalizada por cationes metálicos es responsable de una pérdida importante de proteína, que se agrega y precipita dentro de la columna de IMAC. Si la formación de agregados se ve gatillada o estabilizada por la oxidación de la proteína y la formación de puentes disulfuro intermoleculares (como sucede con la proteína TcPIWI-tryp completa), el bloqueo de sus cisteínas debería ser una estrategia eficiente para impedirlo. Teniendo esto en mente, se tomó una alícuota del extracto bacteriano soluble, se trató con BME 4mM, y se agregó 2-iodoacetamida (IAM) a una concentración final de 15mM, incubando por media hora. La iodoacetamida es capaz de reaccionar con determinados nucleófilos como los grupos sulfhidrilos presentes en los residuos de cisteínas o en tioles de bajo peso molecular, produciendo un derivado S-carbamidometil, con salida del átomo de iodo. Para que la reacción sea posible, las cisteínas reactivas deben encontrarse en su forma reducida (-SH), razón por la cual se agregó BME previo al agregado de IAM.

A diferencia de la proteína no tratada, la proteína tratada con IAM (en adelante, PAZ-IAM) resultó ser estable en ausencia de DTT, como cabría esperarse. La misma se purificó de igual modo, pero se almacenó en ausencia de agente reductor.

A modo comparativo, la figura 25, C muestra los cromatogramas obtenidos para el dominio PAZ recombinante en condiciones reductoras (PAZ-SH)²⁷ y para la proteína PAZ-IAM, normalizados al valor respectivo de absorbancia a 280nm para el pico de elución. La intensidad del pico en el volumen muerto es – relativo al pico de elución – más chico en el caso de la proteína PAZ-IAM. Esta observación coincide con lo esperado, en cuanto el bloqueo de las cisteínas debería evitar (al menos en parte) la formación de agregados.

9.4 Caracterización de las proteínas PAZ-SH y PAZ-IAM por espectrometría de masas (MALDI-TOF y MALDI-TOF-TOF)

Se tomaron alícuotas de las proteínas PAZ-SH y PAZ-IAM, se diluyeron a 15pmol/ μ L, y se enviaron al servicio de espectrometría de masas del IPMon, solicitando un análisis de la proteína por mapeo peptídico con tripsina y medida por MALDI-TOF (MS), incluyendo un análisis de fragmentación peptídica y medida por MALDI-TOF-TOF (MS/MS) e incorporando la posibilidad de encontrar péptidos con

²⁷ De aquí en adelante, se referirá a la proteína PAZ purificada y presente en un entorno reductor como PAZ-SH, para diferenciarla de la proteína tratada con iodoacetamida, a la que se referirá como PAZ-IAM

cisteínas carbamidometiladas, lo que debería producir un incremento de masa de 57 daltons/z²⁸. También se habilitó la posibilidad de detectar metioninas oxidadas (+16 Da).

Para ambas proteínas se obtuvo una cobertura del 85% de la secuencia. Las regiones sin cobertura corresponden a regiones con menos de cuatro aminoácidos entre residuos de lisina o arginina (sitio de corte de la tripsina), y por tanto, a péptidos con una relación m/z muy baja, que no suelen ser detectables. En ambos casos, los únicos resultados significativos ($p < 0.05$) encontrados por el programa Mascot son distintas versiones depositadas de la proteína Argonauta de *T. cruzi*.

Este resultado corrobora la identidad de la proteína purificada y la ausencia de proteínas contaminantes en proporciones detectables por espectrometría de masas. Como cabría esperarse, la lista es encabezada por la proteína gi|298264515 (GenBank: ADI72731.1) que corresponde a la secuencia codificante (y traducción *in silico*) de la proteína TcPIWI-tryp, secuenciada por nuestro grupo en la cepa Dm28c durante el desarrollo de esta tesis (Sección 3 de resultados).

Dada la alta cobertura obtenida, la presencia de modificaciones post-traduccionales puede descartarse en la proteína PAZ-SH, salvo en aquellos aminoácidos que no pudieron ser mapeados debido a que pertenecerían a péptidos demasiado pequeños como para ser detectados. Para estar seguro de esto, se tomó la precaución de verificar que las señales de mayor intensidad del espectro correspondiesen a péptidos mapeados y que las diferencias de tamaño entre señales consecutivas o cercanas no coincidiesen con los valores esperados para una fosforilación (+79.9799Da), metilación (+14.0266Da), oxidación (+15.9994Da) o carbamilación N-terminal (+43.0247Da). Por consiguiente, salvando los 16 aminoácidos que no pudieron ser mapeados, puede afirmarse que la proteína PAZ purificada en condiciones reductoras (PAZ-SH) es homogénea en su composición, y no posee ninguna de las modificaciones post-traduccionales más frecuentes.

Respecto a la proteína tratada con iodoacetamida, puede decirse que el tratamiento fue efectivo y permitió el bloqueo de las tres cisteínas presentes, en todas las moléculas que componen la población (figuras 26 y 27). La figura 26 muestra una ampliación de los espectros MS en la región correspondiente a los dos péptidos que contienen cisteínas: LLPDLEPGQPCAVLK (A) y DHSVVPLQALHCCYDPR (B), con valores de m/z calculados de: 1592.8 y 1952.8 Da respectivamente²⁹. Ambos péptidos se detectan en la proteína PAZ-SH, pero están ausentes en la proteína PAZ-IAM. Por el contrario, se detecta únicamente en la proteína PAZ-IAM un péptido de 1649.9075 (A) y 2066.9604 (B). Las diferencias entre los valores de m/z para dichas señales diferenciales, y los valores de m/z esperados para los péptidos sin modificar son de +57.0584 Da y +114.0141 Da respectivamente. Estos valores son consistentes con la presencia de un grupo carbamidometil (+57 Da) en el péptido que contiene una sola cisteína, y dos grupos carbamidometil (+114 Da) en el péptido que contiene dos.

De la figura 26, A, llama la atención que no es una la señal de la proteína PAZ-SH que desaparece en la proteína PAZ-IAM, sino tres (1592.8491 Da; 1614.8390 Da; 1636.8191 Da). Asimismo, no es una la señal que aparece a +57 Da en PAZ-IAM, sino tres (1649.9075 Da; 1671.8848 Da, la tercera no se distingue casi del ruido de fondo). Llamativamente, las diferencias en los valores de m/z entre señales consecutivas son en todos los casos de 21.98 Da. Este es un fenómeno bien conocido en espectrometría de masas, y

²⁸ Dado que la mayoría de los iones encontrados por MALDI-TOF poseen un valor de $z=1$, se suelen expresar los valores de m/z en Daltons (Da), y así se hará de este punto en adelante. No obstante, no debe perderse de vista que no se están midiendo masas moleculares sino relaciones masa/carga.

²⁹ Se presentan únicamente los valores de m/z monoisotópicos.

se debe a la sustitución de un protón (H^+ ; PA=1.0073 Da) por un catión sodio (Na^+ ; PA = 22.9898 Da), originándose un aducto de sodio de igual carga pero con un incremento de masa de 21.98 Da.

Respecto al péptido DHSVVPLQALHCCYDPR, que contiene dos cisteínas consecutivas, resulta claro que ambas se encuentran modificadas en la proteína PAZ-IAM (figura 26, B). En este caso, habría que descartar la coexistencia de moléculas sin modificar, o con una sola de las dos cisteínas modificadas. Respecto a lo primero, es clara la desaparición de la señal de 1952.8859 Da, correspondiente al valor de m/z del péptido nativo. Una pequeñísima señal que se detecta en la misma región no parecería corresponder a dicho péptido, ya que está casi 2 Da por debajo (1950.9010 Da), y se encuentra también en la proteína sin modificar. Dicha señal podría corresponder a remanentes del péptido en forma de disulfuro, ya que implicaría la pérdida de dos protones. En cualquier caso, la intensidad de dicha señal es apenas apreciable. Tampoco se detecta proteína con una sola cisteína modificada, lo que correspondería a un péptido de $1952.8 + 57 = 2009.8$ Da. De nuevo, aparece una señal apenas detectable a 2009.947 Da que se podría considerar como el péptido buscado, pero dicha señal se encuentra también en la proteína sin modificar con una intensidad comparable (flechas negras verticales).

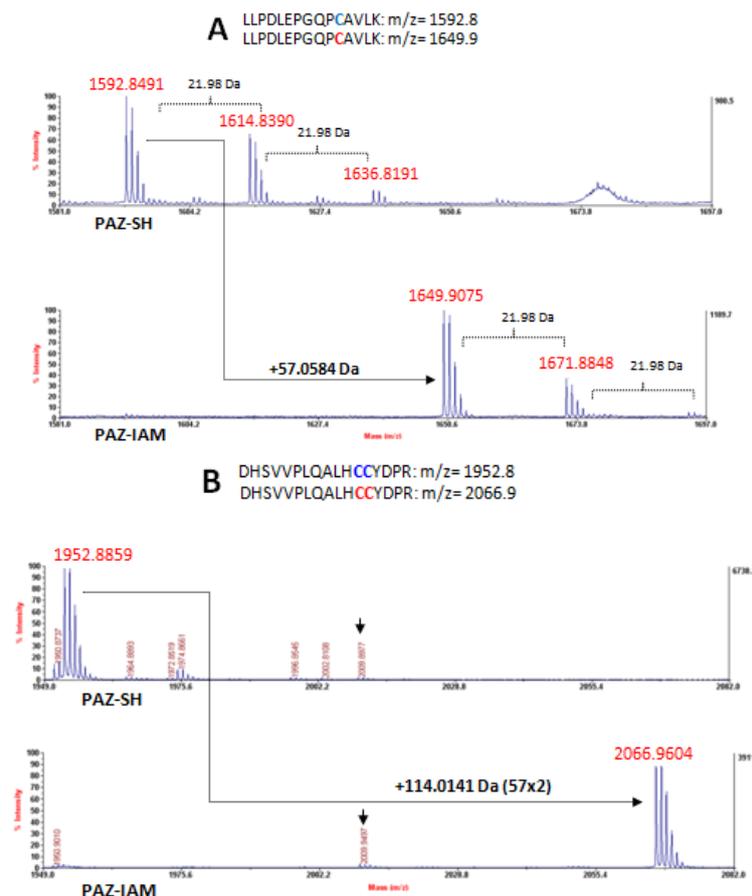


Figura 26: Espectros M/S obtenidos por MALDI-TOF en la región correspondiente a los péptidos con cisteínas LLPDLEPGQPCAVLK (A) y DHSVVPLQALHCCYDPR (B), para las proteínas PAZ-SH y PAZ-IAM digeridas con tripsina. En la parte superior de las figuras se muestran los valores de m/z esperados para los péptidos con y sin cisteínas modificadas (rojo y azul respectivamente)

La confirmación de que los péptidos que presentan un desfase de +57 o +114 Da en la proteína PAZ-IAM corresponden a péptidos con cisteínas carbamidometiladas, se realizó por espectrometría de masas en tándem (MS/MS) de los péptidos presuntos. En la figura 27 se muestran dos regiones del espectro MS/MS del péptido DHSVVPLQALHCCYDPR. La señal de 653.2272 Da (A) corresponde al péptido CYDPR fragmentado en y^{30} (y_6), mientras que la señal de 756.1828 Da (B) corresponde al péptido CCYDPR fragmentado en y (y_7). Ninguna de las dos señales aparece en el caso de la proteína PAZ-IAM. Por el contrario, surgen nuevas señales en 710.1819 Da (A) y 870.1854 Da (B). La primera de ellas corresponde al fragmento $y_6 + 56.95$ Da, mientras que la segunda corresponde a la masa del fragmento $y_7 + 114.00$ Da, quedando claro que las modificaciones se encuentran ubicadas sobre las cisteínas (también podrían mostrarse los fragmentos de la serie “b”, para evidenciar que las señales son iguales hasta b_{11} , y poseen un incremental de +114 Da a partir de b_{13}). La señal de 876.34-876.35 Da que se evidencia tanto en la proteína PAZ-SH como en la proteína PAZ-IAM corresponde al fragmento b_8 : DHSVVPLQ (m/z teórico: 876.458 Da), que al no contener cisteínas, se puede detectar en ambos casos.

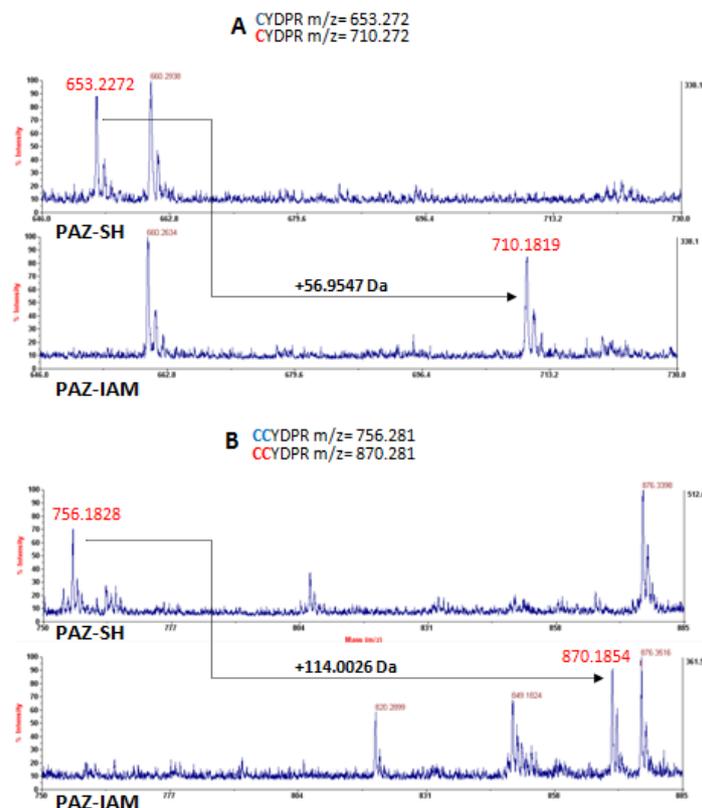


Figura 27: Espectros MS/MS correspondientes a la fragmentación del péptido DHSVVPLQALHCCYDPR de las proteínas PAZ-SH y PAZ-IAM. En la parte superior se muestra el valor de m/z esperado para cada fragmento de acuerdo a la fragmentación más probable (en y').

³⁰ Fragmentación más frecuente, correspondiente a la rotura del enlace peptídico entre el carbono y el nitrógeno que componen el enlace amida.

Los espectros MS/MS del péptido LLPDLEPGQPCAVLK no se muestran por cuestión de espacio, pero también permiten corroborar el corrimiento de +57 Da en las señales en “b” y en “y” a partir de la cisteína.

10 ENSAYOS DE SOLUBILIDAD, ESTABILIDAD Y CRISTALOGÉNESIS DE LAS PROTEÍNAS PAZ-SH Y PAZ-IAM.

10.1 Estabilidad de las proteínas PAZ-SH y PAZ-IAM a temperatura ambiente

Por lo general, las probabilidades de cristalizar una proteína aumentan cuando la misma se encuentra presente en aquellas condiciones en las que es más soluble, previo al agregado del agente precipitante. Al estabilizar a las proteínas en solución, las interacciones proteína-proteína pueden moderarse en el sentido de favorecer la nucleación y crecimiento de cristales, minimizando las probabilidades de formación de agregados amorfos (Izaak et al. 2006)

El caso de la proteína PAZ-SH es un caso muy particular, pues si bien la proteína pudo ser estabilizada por medio de la inclusión de agentes reductores y purificada en su forma monomérica por gel filtración, todos los intentos por eliminar o reducir la concentración del agente reductor post-purificación fallaron, conduciendo a una rápida precipitación. Queda claro entonces que la proteína es intrínsecamente sensible a la oxidación, y requiere de concentraciones moderadas o altas de DTT para mantenerse en solución. Considerando que un ensayo de cristalización requiere de varios días de incubación en los que la proteína no puede perder homogeneidad estructural, resulta pertinente saber si la misma es capaz de mantenerse en su estado reducido por períodos largos de tiempo.

Se procuró trabajar siempre a pH ligeramente alcalino, dado que el punto isoeléctrico de la proteína es de 5,7, y a una concentración moderada de sales (NaCl 150mM). Todas las etapas de purificación se realizaron a 4°C, incluyendo la digestión con la proteasa TEV, que se realiza en diálisis toda la noche. A 4°C, la proteína parece ser bastante estable, y la formación de agregados observada al principio se minimizó enormemente optimizando las condiciones redox durante la purificación (figuras 23-25). Una vez purificada, la proteína pudo ser concentrada a 10mg/mL – e incluso a 20mg/mL ($\approx 1,66$ mM) – sin mayores inconvenientes, pero trabajando siempre a 4°C.

También se realizó un *screening* de solubilidad, a efectos de estudiar la concentración óptima de proteína para los ensayos de cristalogénesis. Para esto, se realizaron distintas pruebas mezclando a la proteína PAZ-SH a distintas concentraciones (7 mg/mL y 14 mg/mL), con concentraciones crecientes de diversos agentes precipitantes. Observando la proporción de gotas conteniendo precipitado en cada caso, luego de un día de concentración por evaporación, se estableció que la concentración óptima de la proteína para los ensayos de cristalogénesis debiera ser cercana a 10mg/mL.

Hasta aquí, todos los observables que permitieron extraer conclusiones respecto a la estabilidad y solubilidad de la proteína en las condiciones de trabajo se relacionan a la formación o no formación de agregados (solubles e insolubles, covalentes y no covalentes). Sin embargo, la preparación proteica también puede volverse heterogénea por mecanismos que no necesariamente implican agregación. Sabiendo que la proteína es propensa a oxidarse, se torna necesario saber si la misma es capaz de

mantenerse en el tiempo en un estado redox determinado (en este caso, en su estado reducido). Para responder esta cuestión, se diseñó un ensayo indirecto basado en espectrofotometría de absorción que permite sacar provecho del hecho de que la proteína PAZ-SH debe ser mantenida siempre en presencia de DTT, lo que interfiere con métodos analíticos clásicos como la reacción de Ellman (Ellman 1959)

La figura 28 muestra espectros de absorción de las proteínas PAZ-SH y PAZ-IAM tomados a distintos tiempos, en la región correspondiente a la absorbancia de los residuos aromáticos. Se partió de disoluciones *stock* (10mg/mL) de ambas proteínas almacenadas a -80°C , las cuales se centrifugaron a máxima velocidad para eliminar todo el material insoluble. Ambas proteínas se diluyeron en tampón Tris 20mM pH=8, NaCl 150mM, DTT 1mM (utilizando una solución fresca de DTT) a una concentración aproximada de 0,2mg/mL (0.14 Au), registrando a distintos tiempos el espectro de absorción con corrección de línea de base (dado que las formas oxidadas y reducidas del DTT absorben luz en el rango de longitudes de onda de trabajo).

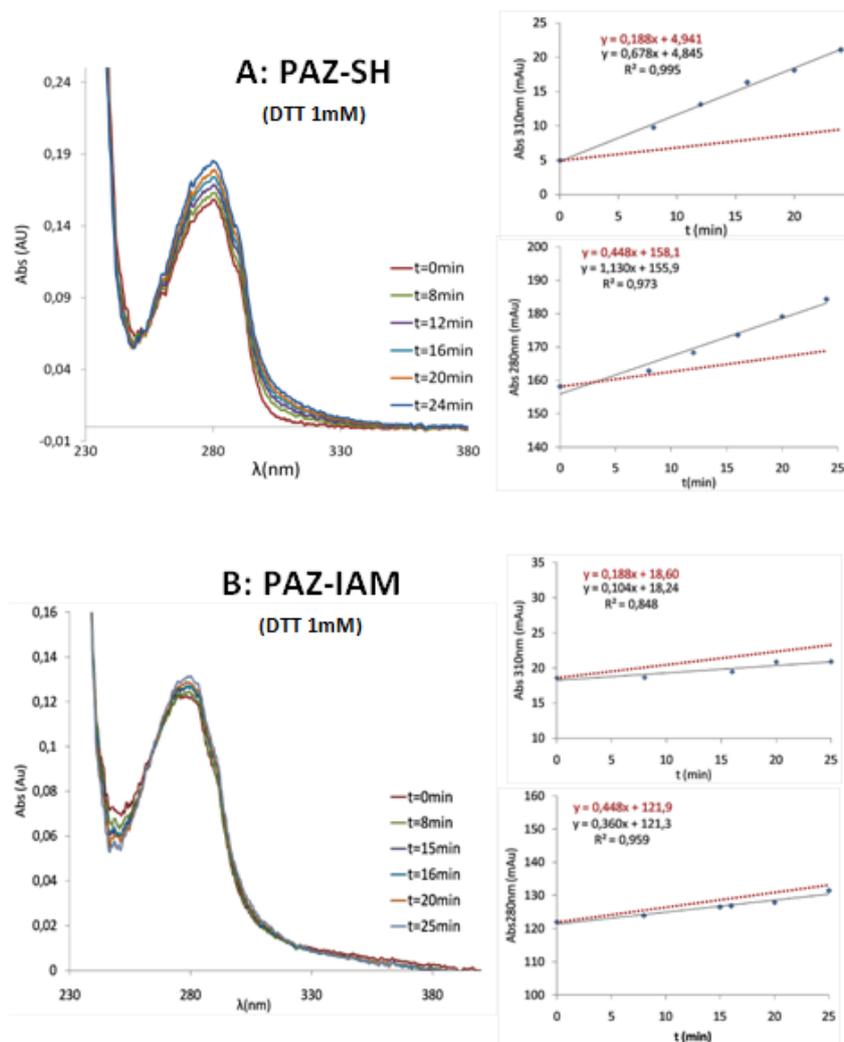


Figura 28: Estabilidad a temperatura ambiente de la proteína PAZ-SH (A) y PAZ-IAM (B). A efectos comparativos, se partió de una disolución a 10mg/mL de cada proteína almacenada a -80°C , centrifugada a 4°C e inmediatamente diluída en una solución tampón conteniendo Tris 20mM pH=8; NaCl 150mM; DTT 1mM, a temperatura ambiente. Los espectros de absorción en la región 230-400nm se registraron a distintos tiempos post-dilución, corrigiendo los espectros respecto al espectro de absorción del tampón sólo. Las gráficas de la derecha muestran los valores de absorbancia (en mAu) en función del tiempo, ya sea a

310nm (superior) o 280nm (inferior). En cada caso, se superpuso la pendiente correspondiente a la oxidación espontánea del DTT (línea roja punteada), obtenida en un ensayo independiente.

Lo primero que llama la atención de la figura 28 es un corrimiento del espectro hacia valores más altos de absorbancia a medida que aumenta el tiempo entre medidas, lo que se observa en el caso de la proteína PAZ-IAM, pero es mucho más acentuado para la proteína PAZ-SH. Las mayores variaciones se observan en el entorno de los 280nm, y en el rango entre los 300-330nm. En realidad, este comportamiento es esperado si se considera que la solución contiene DTT 1mM, y que los espectros se registran en un espectrofotómetro de haz único con corrección de línea de base, es decir, sustrayendo el espectro de absorción del DTT 1mM a $t=-1$ min (dado que los espectros se registran en un tiempo aproximado de 1 minuto). Si bien la forma reducida del DTT (DTT_{red}) absorbe poco por encima de 270nm, éste se oxida espontáneamente por reacción con el oxígeno disuelto, dando origen a la forma oxidada del DTT (DTT_{ox}) que presenta un máximo de absorción a 283nm. El ancho del pico de absorción del DTT_{ox} es considerablemente mayor al de los aminoácidos aromáticos pudiéndose detectar la presencia del DTT_{ox} por encima de los 310nm, dónde las proteínas casi no absorben (Iyer and Klee 1973)

En ausencia de cationes metálicos oxidantes como el Fe^{3+} o el Ni^{2+} , el DTT presenta una cinética de oxidación espontánea de orden 0, es decir, que la concentración de DTT_{ox} aumenta de forma lineal con el tiempo y de manera independiente de la concentración de DTT_{red} inicial (Getz et al. 1999). En base a lo anterior, se estudió la cinética de oxidación espontánea del DTT en las mismas condiciones de trabajo en que se realizaron los espectros de la figura 28, a dos longitudes de onda distintas: 280nm y 310nm. En tales condiciones (18°C, soluciones sin desoxigenar, $[\text{DTT}]=1\text{mM}$; $\text{pH}=8$), la oxidación espontánea del DTT provoca un aumento de absorbancia de 0.448mAu/min y de 0.188 mAu/min a 280nm y 310nm respectivamente (figura 28, derecha, líneas rojas punteadas). Considerando los valores de absorptividad molar reportados a pH 8 para el DTT_{ox} de $250 \text{ M}^{-1} \text{ cm}^{-1}$ a 280nm, y de $110 \text{ M}^{-1} \text{ cm}^{-1}$ a 310 nm (Iyer and Klee 1973), se desprende que la velocidad de reacción es de 1,79 $\mu\text{M}/\text{min}$ o de 1,71 $\mu\text{M}/\text{min}$ según se calcule a 280nm o 310nm; tomando 1,75 $\mu\text{M}/\text{min}$ como valor promedio. Nótese que no se agregaron quelantes de metales de transición como el DTPA, por lo que no puede descartarse que la oxidación del DTT se vea catalizadas por trazas de metales presentes. De todos modos, lo que importa a los efectos de este estudio es que las condiciones de trabajo con y sin proteína son idénticas.

Las gráficas a la derecha de los espectros de la figura 28 muestran la variación en la absorbancia a 310nm (superior) y 280nm (inferior) en función del tiempo, para la proteína PAZ-SH (A) y PAZ-IAM (B). Superpuesta, se muestra la curva correspondiente a la oxidación espontánea del DTT en las condiciones de trabajo (línea roja punteada). En el caso de la proteína PAZ-IAM, la variación temporal registrada en la absorbancia a 310nm y 280nm se ajusta a la variación esperada debido a la oxidación espontánea del DTT. Por consiguiente, se puede decir que la proteína no interfiere con esta reacción, lo que es consistente con la ausencia de cisteínas redox-activas a raíz del bloqueo con IAM, confirmado por espectrometría de masas. Por el contrario, la presencia de la proteína PAZ-SH – a una concentración similar – provoca un aumento considerable en la pendiente temporal a ambas longitudes de onda. Este aumento no se puede explicar por “*light scattering*” (dispersión de luz por agregación) porque desaparece por encima de 350nm, que coincide con la longitud de onda a la cual el DTT_{ox} deja de absorber luz. El aumento en la pendiente temporal a ambas longitudes de onda parecería reflejar un aumento en la velocidad de oxidación del DTT, hipótesis que se ve fortalecida dada que las velocidades

calculadas a ambas longitudes de onda son comparables (4.52 $\mu\text{M}/\text{min}$ y 6.16 $\mu\text{M}/\text{min}$ a 280nm y 310nm respectivamente; tomando 5.34 $\mu\text{M}/\text{min}$ como valor promedio).

Lo que estos resultados indican es que la proteína PAZ-SH es capaz de catalizar la oxidación del DTT, aumentando la velocidad de la reacción por un factor de 3, y que este efecto se anula al bloquear las tres cisteínas de la proteína con IAM. Más aún, no se observa una cinética de saturación en un tiempo de 24 minutos, ajustándose los valores a una recta ($R^2=0.995$ a 310nm). Esto es consistente con el hecho de que la proteína se mantiene durante su purificación y almacenado en DTT 1mM, renovando además el DTT al cambiar la solución de diálisis (durante la digestión con TEV) y durante la gel filtración. Por consiguiente, la proteína se ve forzada a permanecer con sus cisteínas reducidas, lo que resulta evidente ya que el total de ellas es susceptible al bloqueo con IAM. Si la proteína es capaz de catalizar la oxidación del DTT, no es porque parta de un estado de oxidación intermedio (lo que implicaría una velocidad de reacción concentración-dependiente, y por tanto, variable en el tiempo) sino debido a una constante tasa de oxidación espontánea, que provoca un mayor consumo de DTT.

Estas observaciones tienen consecuencias relevantes. Si bien se comprobó que el agregado de DTT 1mM evita la formación de agregados y la precipitación de la proteína PAZ-SH, también se comprobó que ésta se oxida espontáneamente, lo que conlleva a un consumo acelerado del agente reductor. Esto se vuelve crítico en ensayos de cristalogénesis, para los que la proteína se concentra por encima de 10mg/mL, lo que implica una concentración de tioles de 2.49mM, que es más del doble de la concentración de DTT utilizada. En estas condiciones, el DTT sería incapaz de detener la oxidación espontánea de la proteína, dando origen a una población heterogénea que podría inhibir por completo la formación de cristales. En este panorama, el uso de soluciones desoxigenadas y la adición de quelantes de metales como el DTPA (que den lugar a quelatos redox-inactivos) deberían acompañar la adición de DTT, cuya concentración debería aumentarse conforme se concentra a la proteína. Por otro lado, el bloqueo de las cisteínas con IAM parecería ser una alternativa prometedora de cara a la obtención de una preparación proteica homogénea, evitándose además la necesidad de adicionar agentes reductores.

Dado que era necesario a efectos comparativos, la proteína PAZ-IAM se diluyó en una solución conteniendo DTT 1mM para la obtención de los espectros de la figura 28. Sin embargo, estas no serían las condiciones de almacenamiento y trabajo con la proteína ya que su estabilidad no depende de la presencia de agentes reductores de cisteínas y puentes disulfuro. Para estudiar la estabilidad de la proteína PAZ-IAM a temperatura ambiente en condiciones de trabajo pertinentes, el ensayo se repitió diluyendo a la proteína en tampón sin DTT. El espectro de la proteína permanece prácticamente inalterado, aún después de tiempos de incubación de una hora. No se observa el aumento de señales de *scattering* que podrían ser indicativas de inestabilidad y agregación. Desde luego, este estudio se realiza a concentraciones relativamente bajas en las que las interacciones intermoleculares se ven desfavorecidas, pero indica que la proteína no es inestable *a priori* en estas condiciones.

10.2 Ensayos de cristalogénesis de las proteínas PAZ-SH y PAZ-IAM

El servicio de *screening* robótico de 384 condiciones diferentes de cristalización ofrecido por la Unidad de Cristalografía del IPMon se basa en ensayos de difusión de vapor por gota sentada, mezclando un

volumen de 0.3µL de proteína con 0.3µL de reservorio conteniendo al agente precipitante y aditivos. Para la proteína PAZ-SH, se partió de una concentración de 10mg/mL en tampón Tris 20mM pH=8; NaCl 150mM; DTT 1mM. Dado que a los días de largado el ensayo aproximadamente el 50% de las gotas presentaban la presencia de un precipitado, mientras que la mitad restante permanecían como “gotas claras”, se ratificó que la concentración utilizada fue correcta.

A los 7 días de largado el ensayo, se comprobó que una de las gotas presentaba cuatro pequeñas partículas visibles por lupa, aunque de tamaño demasiado reducido como para determinar si correspondían o no a material cristalino (figura 29, A). La condición en cuestión estaba compuesta por HEPES 0.1M pH=7.5 y jeffamina D-600 10% como agente precipitante³¹. Se dejó la placa evolucionar por un mes a 20°C, pero no se obtuvieron nuevos resultados, ni crecieron en tamaño las partículas obtenidas. Se decidió de todos modos proceder a la optimización manual de dicha condición, considerando que las pequeñas partículas mostradas en la figura 29, A podrían ser micro-cristales capaces de ser mejorados.

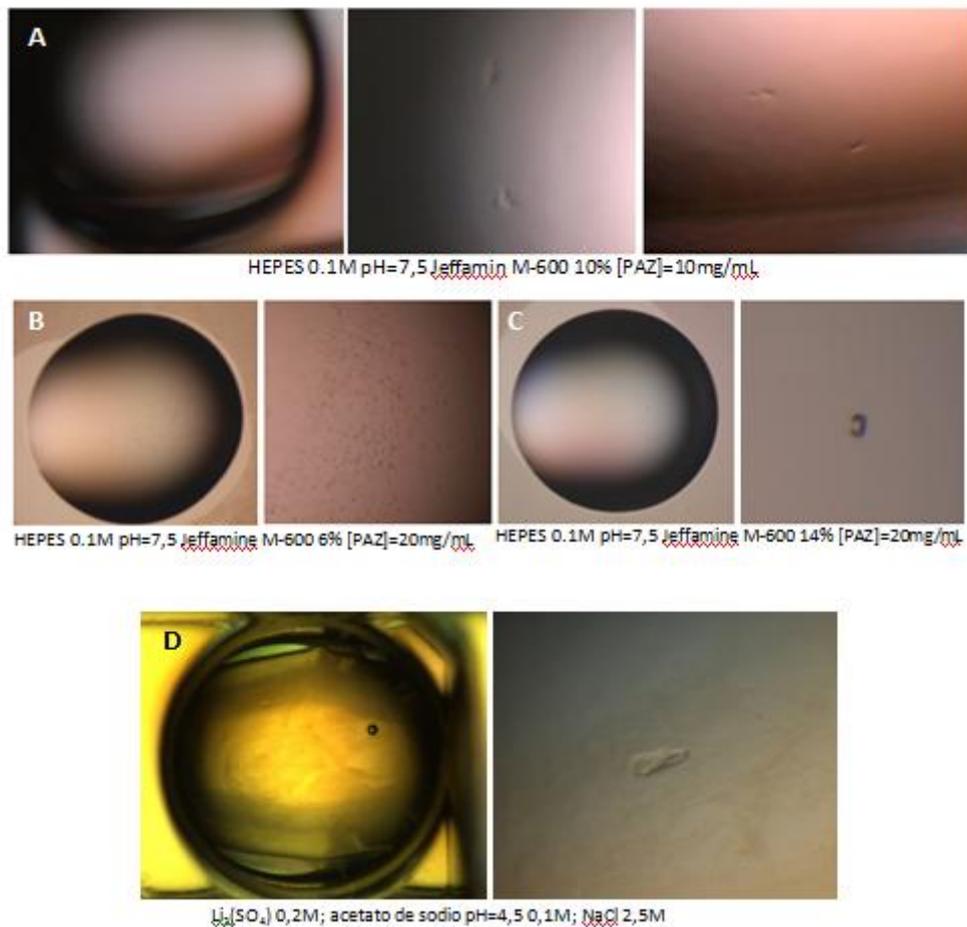


Figura 29: Ensayos de cristalogénesis de la proteína PAZ-SH (A-C) y PAZ-IAM (D). El *screening* robótico se realizó utilizando el kit Classics Lite Suite (QIAGEN, A-C), o JSCG (D) mezclando 0.3 µL de reservorio con 0.3µL de proteína PAZ-SH a 10mg/mL en Tris 20mM pH=8, NaCl 150mM, DTT 1mM.

³¹ La jeffamina D-600 (O,O'-Bis[2-aminopropyl]polypropylene glycol) es una poli-eteramina, caracterizada por un promedio de 6 unidades de oxipropileno con dos grupos amino terminales

Un primer ensayo de optimización se realizó utilizando la primera fila de una placa de 24 pocillos mediante el método de gota colgante. El diseño del ensayo se realizó de modo tal que en todos los casos el reservorio estuviese compuesto por 1mL de tampón HEPES 0.1M pH=7.5, y concentraciones crecientes de Jeffamina D-600 (6%, 8%, 10%, 12%, 14%, 16%). Se introdujeron dos gotas por pocillo, mezclando 2µL de la proteína PAZ-SH a 10mg/mL (en DTT 1mM y DTT 5mM) con 2µL del reservorio correspondiente. A los pocos minutos de incubación, se observó en algunas gotas la presencia de elementos particulados de aspecto amorfo y tamaño pequeño, pero los mismos no evolucionaron con el tiempo, aún luego de casi un mes de incubación. Es de destacar que en ninguna de las gotas se observó la formación de precipitados amorfos, implicando que si bien la concentración de proteína podría haber sido óptima para el *screening* general, tal vez fuese demasiado baja para la acción de la jeffamina como precipitante, al menos en el rango de concentraciones utilizado.

En consideración de lo anterior, se realizó un nuevo ensayo de optimización, ampliando el rango de concentraciones de la Jeffamina D-600 (6%, 10%, 14%, 18%, 22%, 26%) y subiendo la concentración de la proteína a 20mg/mL. En una de las gotas se observó la aparición de una pequeña partícula de aspecto cristalino y aparentemente birrefringente (figura 29, C) a los diez minutos de mezclado con el precipitante, pero no se observó cambios de tamaño ni aparición de nuevos núcleos en el tiempo.

Un segundo *screening* se largó para la proteína PAZ-IAM, concentrada a 8,5mg/mL en tampón Tris 20mM, NaCl 150mM (sin agente reductor). Al momento de escritura de estas líneas, existen algunas condiciones dudosas, pero no se puede afirmar aún la aparición de cristales de proteína. En una de las condiciones ensayadas, conteniendo NaCl 2,5M como agente precipitante (en sulfato de litio 0,2M y acetato de sodio 0,1M; pH=4,5), aparece una partícula amorfa, levemente birrefringente, y de tamaño pequeño pero muy superior al de los casos anteriores (figura 29, D). Se espera proceder a la optimización de esta condición en las siguientes semanas, en miras a la obtención de cristales. Si bien no se realizó un *screening* de solubilidad para la proteína PAZ-IAM, se considera que la concentración utilizada es adecuada, pues aproximadamente el 50% de las gotas presentaron precipitado.

DISCUSIÓN

- **Estudios filogenéticos de proteínas Argonauta**

- 1.1 Nomenclatura, clasificación y función de las proteínas Argonauta de tripanosomátidos (AGO/PIWI-tryp)**

En la figura 5 (Garcia Silva y Tosar et al. 2010) se muestra el único análisis filogenético presente en la literatura científica que – a nuestro entender – fue realizado con las secuencias proteicas completas de todas las proteínas Argonauta de las especies de *Trypanosoma* y *Leishmania* de las cuales se dispone de un genoma completo o en etapas avanzadas, con la excepción del genoma de *Leishmania mexicana*.

En el caso concreto de las proteínas Argonauta de eucariotas inferiores, en particular de los tripanosomátidos, la ausencia de información filogenética condujo hasta el día de hoy a una clasificación

arbitraria de las proteínas en cuestión, con nombres confusos y contradictorios que opacan cuestiones de fondo que consideramos relevantes.

La primera descripción³² de un gen Argonauta en tripanosomátidos fue realizada por el grupo de Elisabetta Ullu en el año 2004 (aceptado en 2003): “Mediante minería de bases de datos, hemos identificado un gen de copia única de *T. brucei* con los motivos característicos de la familia de proteínas Argonauta, es decir, con dominios PAZ y PIWI. Hemos denominado a este gen AGO1” (Shi et al. 2004). Pero aquí resulta necesario considerar el año de publicación: 2004, dos años antes de la publicación del trabajo de Cerutti y Casas-Mollano (2006) en los que se describió la segregación filogenética de las proteínas Argonauta eucariotas en dos grupos parálogos “Argonaute-like” y “Piwi-like”. Por tanto, el nombre “AGO1” hace referencia a que la proteína en cuestión es una proteína de la familia Argonauta. Asumir que Shi y colaboradores asignaron a la proteína AGO1 de *T. brucei* a la sub-familia “Argonaute-like” sería anacrónico. Más aún, los autores construyeron parásitos “knock-out” ago1^{-/-} y reportaron que el fenotipo obtenido “se asemeja al obtenido en moscas mutantes en la proteína Piwi”, ocasionando una acumulación de transcritos provenientes de retrotransposones, siendo el silenciamiento de elementos transponibles una función típicamente asociada a proteínas de la sub-familia “Piwi-like”. La asociación de esta proteína con el silenciamiento de elementos transponibles fue confirmada en estudios posteriores (Durand-Dubief and Bastin 2003). Sin embargo, el mencionado anacronismo se fue extendiendo en el tiempo, con ejemplos claros en artículos recientes.

En el año 2003, Durand-Dubief y Bastin publicaron un artículo donde describen “la identificación de nuevas proteínas Argonauta” en tripanosomátidos: “Búsquedas en las bases de datos de protozoarios permitieron identificar cuatro nuevos representantes de proteínas Argonauta: TbAGO1 y TbPWI1 en *T. brucei*, TcPWI1 en *T. cruzi*, LmPWI1 en *L. major* y GIAGO1 en *G. lamblia*.” (Durand-Dubief and Bastin 2003). Mencionan cuatro nuevos representantes aunque mencionan cinco proteínas, tal vez por el hecho de que TbAGO1 ya había sido previamente descrita (lo que es poco probable dado que el reporte en cuestión no estaba aún publicado) o por la presencia de la secuencia “putative PIWI protein [G. intestinalis]” (GenBank: AAO17548.1) depositada por el grupo de Hugo Luján en 2002. De nuevo, resultaría anacrónico pensar que los autores asignaron la proteína TbAGO1 a la sub-familia “Argonaute-like” y la proteína TbPWI1 a la sub-familia “Piwi-like” porque dichas subfamilias fueron descritas años más tarde. Por lo que se desprende del manuscrito, aunque no se dice explícitamente, el nombre PWI1 haría referencia a que solo contienen un dominio Piwi, no así un dominio PAZ.

Si bien Durand-Dubief y Bastin (2003) denominaron a las proteínas Argonauta de *T. brucei* como TbAGO1 y TbPWI1, depositaron las secuencias correspondientes en la base de datos del NCBI como “argonaute-like protein 1 [Trypanosoma brucei]” (GenBank: AAR10810.1) y “PIWI-like protein 1 [Trypanosoma brucei]” (GenBank: AAR10811.1) respectivamente. Llamativamente, estos son los mismos nombres que Cerutti y Casas-Mollano asignaron a las dos sub-familias de argonautas eucariotas en 2006, y aquí surge la base de la confusión. Sin embargo, Cerutti y Casas-Mollano evitaron explícitamente incluir a la proteína TbPWI1 y a las argonautas de *T. cruzi* y *L. major* en su análisis, en parte porque realizaron el mismo sobre las secuencias de los dominios PAZ y PIWI. Sí incluyeron a las proteínas TbAGO1 y GIAGO1 (que poseen un dominio PAZ, aunque muy poco conservado) y fueron estas las

³² En el párrafo siguiente se describe el artículo de Durand-Dubief y Bastin (2003). En realidad dicho artículo es posterior pues fue enviado en octubre de 2003, mientras que el de Shi et al. fue aceptado en setiembre de 2003. Sin embargo, como Durand-Dubief y Bastin publicaron en BMC Biology que es una revista on-line, la cita es del año 2003 mientras que el artículo de Shi et al. fue publicado recién en enero de 2004.

únicas dos proteínas que no lograron agrupar en ninguno de los dos grupos “Argonaute-like” ni “Piwi-like”

En una revisión del año 2004, Elisabetta Ullu y colaboradores propusieron un nombre alternativo para las proteínas Argonauta de tripanosomátidos carentes de un dominio PAZ, tal vez previendo que la anotación de las proteína TbPWI1 como “Piwi-like protein 1” podría dar lugar a problemas en el futuro. Sin embargo, la clasificación y nomenclatura anterior (PiwiS de Piwi ‘solo’) no solo no obtuvo arraigo en la literatura científica, sino que ni siquiera aparece mención a dichas proteínas en los artículos subsiguientes de los mismos autores. En 2007, Shi, Tschudi y Ullu afirman que “*T. brucei* es un modelo útil para el estudio del ARNi porque su genoma contiene una única proteína Argonauta (TbAGO1)” En un artículo del mismo grupo publicado en julio de 2011 en “*Trends in Parasitology*” que lleva por título: “*The emerging world of small silencing RNAs in protozoan parasites*”³³ tampoco aparece mención a las proteínas Argonautas de *T. cruzi*, y se habla de una única proteína Argonauta en *T. brucei*, en *L. braziliensis*, y de un pseudogen de Argonauta en *L. major*, si bien es cierto que la revisión se enfoca en el mecanismo del ARNi (Atayde et al. 2011).

De hecho, luego del artículo de Durand-Dubief y Bastin del año 2003, dónde no se detectó un compromiso del ARNi ni se pudo asignar ningún fenotipo particular a parásitos TbPWI1^{-/-}, y de los intentos fallidos por realizar ARNi en *T. cruzi* (DaRocha et al. 2004), *L. major* y *L. donovani* (Robinson and Beverley 2003), no existen otros trabajos referentes al estudio de proteínas PIWI-tryp, a excepción del nuestro (García Silva y Tosar et al. 2010). Sin embargo, la figura 5 muestra como todos los tripanosomátidos han conservado este tipo de proteínas; no así proteínas del grupo Ago-tryp. Además de su conservación en el genoma, la secuencia codificante de estas proteínas está sujeta a selección natural purificadora, a juzgar por los valores de Ka/Ks (tasa de mutación no sinónima/sinónima) menores a 0.5 en todos los casos, lo que condice con el hecho de que sabemos que al menos TcPIWI-tryp se expresa. Estas constataciones cuestionan el poco énfasis que se ha hecho en el estudio de las proteínas PIWI-tryp, a causa del sesgo en la búsqueda de vías clásicas de regulación mediadas por ARNs pequeños como miARNs y siARNs, que no parecerían depender de estas proteínas.

Es por tanto de recibo constatar que los resultados obtenidos y publicados recientemente (García Silva y Tosar et al. 2010) han logrado llamar la atención sobre este tipo de proteínas. En un artículo muy reciente del “*Journal of Proteomics*” se retoma la nomenclatura (AGO-tryp/PIWI-tryp) y enfoque propuestos en el marco de esta tesis:

“Este grupo de protistas (los tripanosomátidos) posee dos tipos distintos de proteínas Argonauta que son de algún modo más similares a sí mismas que a otras argonautas eucariotas [35]. Los tripanosomátidos que poseen proteínas AGO-tryp también poseen proteínas Dicer, que actúan juntas en las vías de silenciamiento por ARNi de ARNs de doble hebra introducidos de forma exógena. En concordancia con la ausencia de proteínas Dicer y AGO-tryp, la introducción de ARN doble hebra en T. cruzi no induce silenciamiento (...) Un segundo grupo de proteínas Argonautas (denominadas PIWI-tryp) parecen estar presentes en todos los tripanosomátidos de los cuales se dispone su genoma [35]. Las proteínas PIWI-tryp se encuentran altamente conservadas y se expresan durante todo el ciclo de vida de T. cruzi, lo que sugiere que desempeñarían un rol importante para los tripanosomátidos [35]. Las proteínas PIWI-tryp de T. cruzi y T. brucei carecen de un dominio PAZ característico de todas las proteínas Argonauta identificadas hasta la fecha. Notablemente, se ha identificado un OB-fold en las proteínas PIWI-tryp, que es estructuralmente similar a la estructura de los dominios PAZ, y que podría compensar su ausencia [35] (...) Mientras que la vía de ARNi mediada por proteínas AGO-tryp se encuentra probablemente asociada al silenciamiento de elementos transponibles, la función de las proteínas PIWI-tryp es desconocida. Los ARNs pequeños no codificantes asociados a las proteínas PIWI-tryp no

³³ “El mundo emergente de los ARNs pequeños silenciadores en protozoarios parásitos”

se conocen, pero es de notar que una potencial clase de ARNs pequeños no codificantes derivados de tARNs fue detectada en *T. cruzi* [37] (...)³⁴ (Batista and Marques 2011)

Y posteriormente: “El hecho de que la proteína PIWI-tryp de *T. cruzi* es tan distinta a todas las otras argonautas conocidas, podrían hacerla un excelente blanco terapéutico para el tratamiento de la enfermedad de Chagas. Más aún, este grupo conservado de proteínas PIWI-tryp podría estar involucrado en una vía esencial de ARNi para los tripanosomátidos, y ser un blanco potencial para la mayoría de las enfermedades causadas por los mismos” (Batista and Marques 2011)

a. Origen evolutivo de las proteínas AGO/PIWI-tryp y su presunta asociación a fragmentos derivados de ARNs de transferencia.

Como se desprende del análisis de las figuras 5 y 6, se tienen dos grupos monofiléticos dentro de la subfamilia “AGO/PIWI-tryp” y ambos contienen representantes de los géneros *Trypanosoma* y *Leishmania*. Existen tripanosomátidos con dos proteínas Argonauta en su genoma, y en dichos casos las mismas corresponden invariablemente a una proteína AGO-tryp y una proteína PIWI-tryp. La interpretación más parsimoniosa de estos resultados es que en algún momento de la evolución de los tripanosomátidos existió un evento de duplicación génica, y que el último ancestro común de todos los tripanosomátidos poseía al menos una proteína AGO-tryp y una proteína PIWI-tryp. De ser así, la proteína AGO-tryp debería haberse perdido independientemente dos veces durante la evolución: una vez durante el linaje que condujo a las especies *L. major* y *L. infantum* y después de la separación de *L. braziliensis* (subgénero *Viannia*), y una segunda vez durante la evolución de *T. cruzi*. Resulta llamativo cómo esta interpretación es análoga a la realizada por Ceruttii y Casas-Mollano (2006) para el resto de las argonautas eucariotas, sólo que aplicada específicamente a los tripanosomátidos en este caso.

La hipótesis de dos pérdidas independientes de proteínas Argonauta durante la evolución de tripanosomátidos fue también planteada por Lye et al. (Lye et al. 2010). De todos modos no pueden descartarse otras alternativas, tales como la transferencia horizontal de material genético, por poner un ejemplo. Desde luego, la necesidad de proponer dos eventos independientes de pérdida de genes AGO-tryp resulta de asumir *a priori* la monofilia de los géneros *Trypanosoma* y *Leishmania*. Esta asunción no es libre de polémica, dado que algunos autores han propuesto que el género *Trypanosoma* podría ser parafilético sobre la base de árboles construidos con la secuencia del ARN ribosomal 18S, con los tripanosomas americanos (*T. cruzi*, entre otros) agrupando junto con las especies del género *Leishmania*, y dejando por fuera a los tripanosomas africanos (*T. brucei*, *T. congolense*, entre otros) y a *T. vivax* (Hughes and Piontkivska 2003; Piontkivska and Hughes 2005). Sin embargo, reportes más recientes sobre la base de un número mucho mayor de genes (pero un número mucho menor de organismos) volverían a proponer la monofilia del género *Trypanosoma* (Leonard et al. 2011). No es nuestro objetivo contribuir a dicha polémica; simplemente señalar cómo la discusión de los resultados obtenidos guarda una relación de dependencia con el modelo evolutivo que se acepte como válido.

Respecto a la segregación de las proteínas Argonauta de tripanosomátidos en dos sub-grupos que se presumen parafiléticos (AGO-tryp y PIWI-tryp), cabe mencionar que los eventos de duplicación génica

³⁴ [35]: García-Silva y Tosar et al. (2010); [37]: García-Silva et al. (2010).

con posterior diversificación funcional y eventuales pérdidas linaje-específicas de determinados genes Argonauta han sido frecuentes en la evolución de estas proteínas (Hutvagner and Simard 2008). *C. elegans* codifica para 26 proteínas Argonauta distintas, los seres humanos para 8, por poner algunos ejemplos. Proponer un evento de duplicación génica específico de tripanosomátidos no parecería ser inverosímil, y es consistente con los resultados obtenidos.

Más especulativo sería ubicar en el tiempo dicho evento. Una posibilidad es que las proteínas AGO/PIWI-tryp se hubiesen originado a partir de una proteína “*Argonaute-like*” o “*Piwi-like*”, es decir, que el evento de duplicación génica que origina las proteínas “AGO-tryp” y “PIWI-tryp” sea posterior al evento de duplicación génica propuesto por Cerutti y Casas-Mollano (2006). En dicho caso cabría esperar que las proteínas AGO/PIWI-tryp se agrupasen junto con las proteínas “*Argonaute-like*” o “*Piwi-like*” según corresponda, lo que no sucede. Sin embargo, una alta tasa de sustituciones en la rama que conduce a los tripanosomátidos (que se evidencia en los árboles construidos con métodos basados en distancia, figura 6) podría hacer desaparecer dicho rastro.

Otra posibilidad es que el evento de duplicación génica propuesto por Cerutti y Casas-Mollano (2006) hubiese ocurrido después de la separación de los tripanosomátidos del resto de los eucariotas. En dicho escenario, las proteínas AGO/PIWI-tryp podrían retener características del gen ancestral que les diera origen, y compartir características tanto con las proteínas “*Argonaute-like*” como con las proteínas “*Piwi-like*” del resto de los eucariotas. Un ejemplo a favor de esta hipótesis es el hecho de que la proteína TbAGO-tryp participa en el silenciamiento de ARNs mensajeros por ARNi, lo que es una propiedad aparentemente exclusiva de proteínas “*Argonaute-like*”. Sin embargo, también parecería estar especializada en el silenciamiento de retrotransposones, lo cual suele ser la función de las proteínas “*Piwi-like*”.

Si se ha adentrado en la proposición de distintos escenarios evolutivos que fuesen consistentes con los resultados mostrados en las figuras 5 y 6 es porque las implicancias funcionales que podrían desprenderse de dichos escenarios son diferentes. Si el evento de duplicación génica que originó las subfamilias “*Argonaute-like*” y “*Piwi-like*” fue posterior a la separación de la rama que conduce a los tripanosomátidos del resto de los eucariotas, entonces el estudio de las proteínas AGO/PIWI-tryp podría arrojar luz sobre las funciones ancestrales de las proteínas Argonauta eucariotas. Puede ser que muchas de las funciones de las proteínas AGO/PIWI-tryp hayan sido adquiridas durante la evolución de estos organismos y sean por tanto específicas de los mismos, pero podrían ser también versiones primitivas de mecanismos más complejos perfeccionados por algunos otros eucariotas superiores; mecanismos regulatorios tal vez aún no conocidos.

El párrafo anterior permite introducir la cuestión de los fragmentos derivados de ARNs de transferencia. Dichos fragmentos parecerían ser una población de ARNs pequeños (reguladores o no de la expresión génica; eso exigiría una demostración experimental que no disponemos al día de hoy) muy representada en *T. cruzi*, sobre todo en condiciones de *stress* nutricional (Garcia-Silva et al. 2010) y en estadios infectivos del parásito (datos no publicados). De las otras clases de ARNs pequeños conocidos, se sabe que hay organismos como las plantas que sólo poseen argonautas de la sub-familia “*Argonaute-like*” y carecen por tanto de piARNs, mientras que otros organismos como el protozoario de vida libre *T. termophila* parecería codificar únicamente para proteínas “*Piwi-like*” (Cerutti and Casas-Mollano 2006; Couvillion et al. 2010) por lo que carece de microARNs y siARNs. No obstante, se han reportado recientemente poblaciones de fragmentos de tARNs clivados a nivel de la horquilla del anti-codón en la

planta *A. thaliana* (Hsieh et al. 2010) y en *T. termophila* (Couvillion et al. 2010). Más aún, los fragmentos de tARNs parecerían unirse a una de las proteínas “*Piwi-like*” de *Tetrahymena*, como consta en el citado trabajo.

Si se termina de demostrar que los fragmentos derivados de tARNs constituyen nuevas poblaciones de ARNs pequeños reguladores asociados a proteínas Argonauta (Burroughs et al. 2011; Couvillion et al. 2010; Haussecker et al. 2010), los fragmentos de tARNs podrían pasar a ser la primera población de ARNs pequeños reguladores asociados tanto a proteínas de la sub-familia “*Argonaute-like*” como a proteínas de la sub-familia “*Piwi-like*”, reflejando lo que podría llegar a ser una función ancestral de las proteínas Argonauta. De ser así, los tripanosomátidos serían importantes modelos para el estudio de este tipo de “funciones ancestrales” mediadas por la asociación entre proteínas Argonauta y fragmentos de tARN, aunque de momento esta discusión permanece en el terreno de lo especulativo.

1.3 Implicancias en el estudio de otros organismos

Es de destacar que los cuatro miembros del grupo “WAGO” incluidos en nuestro análisis (por *Worm-specific Argonaute subfamily*, sub-conjunto de proteínas Argonauta específicas de nematodos, señalados en la figura 5 con una barra gris) forman una rama dentro del grupo “*Argonaute-like*”, por lo que parecerían ser una radiación originada por eventos de duplicación génica a partir de un miembro de la familia AGO durante la evolución de los nematodos. Este grupo fue definido como “Argonautas del tipo 3” (Hutvagner and Simard 2008) e inicialmente descrito como “un grupo expandido de proteínas (argonautas) de *C. elegans*, igualmente distante de los sub-grupos *PIWI* y *AGO*” (Yigit et al. 2006). Si se permite una crítica tímida y respetuosa a un artículo publicado en Cell donde se silenciaron los 27 genes Argonauta de *C. elegans* y firmado por Craig Mello³⁵, el árbol filogenético mostrado es un árbol sin raíz donde se incluyeron únicamente proteínas “*Argonaute-like*” y “*Piwi-like*” de plantas y metazoarios, y las 27 argonautas de *C. elegans*. Si bien las proteínas definidas como WAGO forman un grupo monofilético, al no haber otras referencias que ofician de raíz, cualquier grupo monofilético podría haber sido utilizado para enraizar dicho árbol, pero esto no quiere decir que dicho grupo sea realmente externo a todos los demás.

También resulta llamativo la clasificación de la proteína Argonauta de *G. lamblia* dentro del grupo “*Piwi-like*”. En 2008, se demostró que dicho organismo utilizaba el fenómeno de silenciamiento por ARN interferente para la regulación de la expresión de sus proteínas de superficie (Prucca et al. 2008). Desde un punto de vista funcional, este tipo de función y la maquinaria molecular implicada (que incluye proteínas Dicer y RdRP) serían más propios de proteínas del tipo “*Argonaute-like*”, si bien la proteína es confidentemente asignada al grupo “*Piwi-like*” en la figura 5 (aunque es también el miembro más externo del grupo). Antes de adelantarse a sacar conclusiones al respecto, es importante aclarar que la agrupación de esta proteína fue problemática, debiendo ser eliminada de los árboles construidos únicamente con el dominio Piwi (figura 6) ya que debilitaba demasiado el sustento estadístico de los nodos que la contenían. Según el método de inferencia utilizado, la proteína podía agrupar con las argonautas procariontas o con la sub-familia “*Piwi-like*”, y en algún caso también con las proteínas AGO/PIWI-tryp. La ausencia de cualquier otro homólogo cercano en el análisis y la alta tasa de evolución molecular característica de *Girardia* (Richards and Cavalier-Smith 2005) no nos permiten

³⁵ Premio Nobel de Fisiología o Medicina ese mismo año

asegurar que su agrupación en la figura 5 como “*Piwi-like*” no esté libre de artefactos, por lo que hasta el momento sería prudente mantenerla como “inclasificada”.

- **El dominio PAZ de las proteínas PIWI-tryp**

T. cruzi y *L. major* son ejemplos paradigmáticos que abundan en la literatura científica como organismos que perdieron toda su maquinaria del ARNi, aunque retuvieron un gen que codificaría únicamente para un dominio Piwi. Esto último los diferencia de *P.falciparum* o *S. cerevisiae*, que también son ARNi-negativos (Baum et al. 2009; Suk et al. 2011) pero que carecen de todo vestigio de proteínas argonautas en su genoma. Proponer la presencia de un dominio PAZ en las proteínas Argonauta de *T. cruzi*, *L. major*, y en el resto de las proteínas PIWI-tryp, no es una cuestión tan trivial como podría parecer, dado que implica disentir con la revisión más citada en este campo de estudio (Hutvagner and Simard 2008) y con un hecho por lo general bien aceptado.

Dado que no se poseen estructuras de proteínas Argonauta eucariotas completas, y que se ha reportado únicamente la estructura de cinco dominios PAZ de argonautas eucariotas diferentes, la presencia de dominios PAZ en la inmensa mayoría de las argonautas eucariotas se realiza por homología a nivel de la secuencia de aminoácidos. La única diferencia que existe en nuestro caso es haber utilizado algoritmos bioinformáticos más sensibles, pero la presencia del dominio PAZ en las proteínas PIWI-tryp también se basa en información extraída a partir de su secuencia, con muy buen soporte estadístico a favor de la presencia del dominio (figura 8). El hecho de que el dominio PAZ no pudiese ser detectado mediante algoritmos como CDD search o Pfam no debería ser interpretado en forma restrictiva, pues se sabe que dichos algoritmos también han fallado en la predicción del dominio PAZ de argonautas procariontas, de las que luego se obtuvo su estructura completa, con la clara presencia de un dominio PAZ (Song et al. 2004)

Haber obtenido una proteína soluble luego de expresar en *E. coli* la secuencia que atribuimos correspondería al dominio PAZ, puede ser visto como una evidencia a favor de que la secuencia en cuestión corresponda a un dominio globular capaz de plegarse correctamente en bacterias. Los estudios por DLS de la proteína PAZ-SH purificada (figura 25, B) confirman una población monodispersa con un radio hidrodinámico muy similar al esperado para una proteína globular de 12 KDa. De todos modos, la prueba definitiva de que se está en presencia de un dominio PAZ sería la elucidación de la estructura cristalográfica de la proteína recombinante. Lamentablemente, no se dispone de dicha información al momento de escritura de estas líneas, aunque los ensayos correspondientes se encuentran encaminados, y otros - como la resolución de la estructura por RMN - en vías de comenzar.

Además del interés que representaría la obtención de la estructura del dominio PAZ de TcPIWI-tryp para este trabajo concreto, también sería un aporte significativo al estudio de las proteínas argonautas eucariotas, donde la información estructural es escasa, pese a la relevancia biológica evidente. No sorprende entonces que las estructuras de los dominios PAZ de las Argonautas 1 y 2 de *Drosophila* y de la Argonauta 2 humana hayan dado lugar a artículos publicados en las revistas de mayor difusión en el área (Lingel et al. 2003, 2004; Ma et al. 2004; Song et al. 2003; Yan et al. 2003). Resulta claro que los citados trabajos fueron pioneros en lo que refiere a la obtención de información estructural de

proteínas argonauta, pero dada la gran divergencia en la secuencia de los dominios PAZ de las proteínas PIWI-tryp, sería muy interesante poder constatar como ese alto grado de divergencia en la estructura primaria correlaciona con diferencias o similitudes a nivel de la estructura terciaria, así como con diferencias o similitudes funcionales.

Por último, es importante destacar que la detección de un dominio PAZ en la proteína TcPIWI-tryp así como en todas las restantes proteínas PIWI-tryp – adyacente al dominio Piwi bien conservado – posiciona a estas proteínas como miembros canónicos de la familia Argonauta. El hecho de que *T. cruzi*, *L. major* y *L. infantum* posean proteínas Argonauta en su genoma, y que las mismas se expresen y estén sujetas a selección quita validez a las afirmaciones de que estos organismos hubiesen perdido toda la maquinaria de silenciamiento mediada por ARNs pequeños.

Respecto a los problemas técnicos asociados a la oxidación y agregación del dominio PAZ, cabe señalar que la mutación de las tres cisteínas (a serinas, por ejemplo) es una posibilidad que se consideró, pero de momento la proteína PAZ-IAM parecería ser una alternativa más simple, conveniente, y aparentemente efectiva. Mientras que la modificación con iodoacetamida se realiza *a posteriori*, es decir, sobre la proteína plegada en su conformación nativa, no se puede predecir como la mutación de los residuos de cisteína afectaría la cinética de plegamiento de la proteína.

- **Obtención de la proteína TcPIWI-tryp recombinante**

Siendo que la función fisiológica concreta de las proteínas PIWI-tryp es desconocida, la obtención de proteína pura en el orden de los miligramos permitiría la realización de ensayos biofísicos, bioquímicos y estructurales, que sumado a los estudios realizados *in vivo*, significarían una fuente de información privilegiada al respecto. Dado que no se disponía de suficiente información *a priori* sobre la proteína más allá de los elementos de estructura que se desprenden del estudio de su secuencia primaria, se decidió optar por el sistema de expresión que rutinariamente ofrece mayores rendimientos y a un menor costo: *E. coli*. Tal vez la única información que sí se disponía antes de comenzar a amplificar, clonar y expresar la secuencia codificante de la proteína era que la empresa sería ambiciosa y arriesgada, considerando que numerosos grupos han intentado expresar proteínas Argonauta eucariotas en bacterias, sin éxito.

A pesar del enorme interés que reviste el estudio de estas proteínas, solamente se han reportado dos ejemplos de proteínas Argonauta eucariotas expresadas en *E. coli*: la proteína Ago2 humana expresada como fusión a la enzima glutatión S-transferasa (GST) y coexpresada con la chaperona HSP90 (Rivas et al. 2005) y la proteína Ago2 de ratón expresada como fusión a la proteína de unión a maltosa, MBP (Salvatore et al. 2010). Los autores de estos trabajos reportan que la obtención de la proteína recombinante fue posible luego de numerosos ensayos con distintos vectores y cepas bacterianas, en los que la proteína recombinante se obtenía siempre en forma insoluble. Aún en el caso de obtener proteína en la fracción soluble, las mismas presentaban problemas de estabilidad y la pureza de las preparaciones fue limitada, a juzgar por los geles mostrados.

Con estos antecedentes, parecería un tanto testarudo de nuestra parte transitar por un camino que no parece conducir a ninguna parte. Sin embargo, los estudios filogenéticos realizados permiten evidenciar que estas proteínas se encuentran poco relacionadas al resto de las argonautas eucariotas, y dada la

baja similitud a nivel de secuencia con el resto de las argonautas que se han procurado expresar, se puede decir que al menos el intento se encuentra justificado. Por otro lado, existen ejemplos más alentadores como el de las argonautas procariotas, de las cuales han podido resolverse varias estructuras (Amitsur et al. 1992; Ma et al. 2005; Parker et al. 2004, 2005; Rashid et al. 2007; Song et al. 2004; Wang et al. 2008a; Wang et al. 2009; Wang et al. 2008b; Yuan et al. 2005).

Desgraciadamente – aunque acorde a lo previsto – la proteína TcPIWI-tryp se detectó únicamente en la fracción insoluble luego de su expresión en *E. coli*. Un poco por simplicidad y porque el uso de la expresión es corriente en casos como este, se ha dicho en el capítulo de resultados que la proteína se purificó a partir de los cuerpos de inclusión. Formalmente, esto no podría afirmarse con el protocolo utilizado, ya que la insolubilidad de la proteína podría originarse debido a las condiciones de lisis (es decir, *ex vivo*). Cualquiera sea el caso, no se obtuvo proteína en la fracción soluble, al menos en cantidades detectables por tinción con azul de coomasie.

A pesar de no haber obtenido proteína soluble, la expresión de la misma permitió su purificación en urea 8M y la obtención de suero policlonal específico, que luego se utilizó en la inmunoprecipitación de la proteína, lo que permitió obtener evidencia directa de su asociación con fragmentos derivados de tARNs. Además, clonada la secuencia codificante en el vector pGEM-T Easy, se logró subclonar la misma en el vector pTEX y obtener parásitos que sobre-expresan la proteína Argonauta.

Un estudio poco menos que obligado y que debió quedar finalmente en el tintero es el uso de varias combinaciones de cepas de *E. coli* y vectores a fin de optimizar las condiciones de expresión y obtener condiciones en que la proteína se exprese de forma soluble. Más aún, la cepa de *E. coli* utilizada no es una cepa que típicamente se utilice para expresar proteínas recombinantes. Como argumentos a favor de la decisión tomada puede citarse el trabajo de Schlicke y Brakmann (Schlicke and Brakmann 2005) donde se estudió la expresión de una proteína de interés en distintas cepas de *E. coli*, y solo se obtuvo expresión soluble y a niveles adecuados con la cepa XL1Blue, expresando una copia adicional del represor lac i en *trans* (a diferencia de en *cis*, como en este caso). La ausencia de mutaciones o deleciones en las proteasas Lon y OmpT no parece haber sido un elemento determinante, pues en ningún caso se observaron signos de proteólisis excesiva. Además, para el caso del dominio PAZ la cepa utilizada no implicó mayores inconvenientes, pues se obtuvo proteína soluble en grandes cantidades, siendo la purificación de la proteína por IMAC el paso limitante en lo que refiere a los rendimientos finales.

Para el caso de la proteína completa, la alta proporción de ácidos nucleicos en la fracción insoluble y en todos los pasos de purificación subsecuentes, así como la formación de agregados solubles sensibles a la acción de la ADNasa I luego del replegamiento, permiten dudar que en otras cepas el resultado fuese diferente; la ausencia de una barrera física entre la proteína expresada y el genoma bacteriano sería un denominador común a cualquier cepa de *E. coli*, salvo que se recurriese a una expresión periplasmática. A raíz de los resultados obtenidos, un correcto plegamiento de la proteína en la bacteria y su subsecuente unión al genoma bacteriano no puede descartarse como causa de su aparente insolubilidad. Si bien se adentra en un terreno demasiado especulativo, sería una explicación atractiva de por qué las proteínas Ago2 humana y de ratón no pudieron ser expresadas de forma soluble sino como proteínas de fusión a GST y MBP respectivamente, y de por qué los problemas de insolubilidad no se encuentran asociados a argonautas procariotas, que estarían adaptadas a la ausencia de membrana nuclear. La forma más sencilla de responder a esta cuestión es una observación al microscopio

electrónico de las bacterias transformadas e inducidas, para corroborar la formación o no de cuerpos de inclusión.

Existe una explicación alternativa – y podría decirse que más factible – a la detección de grandes cantidades de ácidos nucleicos en la fracción insoluble. Dado que las proteínas Argonauta se caracterizan por ser proteínas básicas, con puntos isoeléctricos en el entorno de 9.5, y que la lisis bacteriana usualmente se realiza a valores de pH levemente alcalinos (7.5-8), es posible que los agregados de proteínas Argonauta unan fuertemente ácidos nucleicos bacterianos por simple atracción electrostática. Esto explicaría además por qué no se eliminan los ácidos nucleicos aún después de lavar las columnas de IMAC con más de 25 volúmenes de solución tampón conteniendo urea 8M, donde cualquier eventual afinidad específica debiera verse seriamente comprometida. Para someter esta hipótesis a prueba, deberían repetirse los ensayos de expresión, pero realizando la lisis bacteriana a un pH igual al punto isoeléctrico de la proteína, o levemente superior. Probablemente, la proteína seguiría siendo insoluble, pero los ensayos de replegamiento *in vitro* podrían realizarse en ausencia de ácidos nucleicos, y tal vez se minimizara así la formación de agregados solubles. También sería oportuno incluir ADNasa en el extracto y – decididos a probar este punto – sería conveniente expresar mutantes en residuos predichos como claves para la unión a ácidos nucleicos.

Por lo sencillo de los experimentos mencionados, y por el impacto que tendría la obtención de una proteína Argonauta eucariota replegada *in vitro*, probablemente sean estos los primeros ensayos que se realicen luego de retomar el trabajo de mesada. De cualquier forma, no debe olvidarse que ya se obtuvo proteína TcPIWI-tryp soluble y monomérica (figura 20, B) pero en cantidades demasiado reducidas como para realizar y optimizar estudios *in vitro*. Aún así, este resultado indica que el replegamiento *in vitro* es posible, y aunque no se obtengan miligramos de proteína como para realizar estudios estructurales, podrían realizarse ensayos de unión a oligonucleótidos, oligoribonucleótidos, ensayos de actividad ARNasa, etc. A nuestro entender, no existen reportes sobre el replegamiento *in vitro* de proteínas Argonauta, ni tampoco figura ninguna entrada a este tipo de proteínas en la base de datos Refold³⁶

Resulta llamativo comparar el replegamiento por dilución de la proteína TcPIWI-tryp completa, con el del dominio MID/PIWI en similares condiciones, que constituye la mitad C-terminal de la proteína. Mientras que el replegamiento de la proteína completa fue eficiente (aunque derivó en la formación de agregados solubles conteniendo ácidos nucleicos), el dominio MID/PIWI precipitó completamente al disminuir la concentración de urea. Esto constituye un argumento a favor de que los agregados solubles obtenidos corresponden a proteína plegada y no a una simple asociación electrostática e inespecífica con ácidos nucleicos, ya que el punto isoeléctrico del dominio MID/PIWI también es superior a 9.

3.1 Susceptibilidad de la proteína TcPIWI-tryp y del dominio PAZ a la oxidación

Poco o nada se sabe de que las proteínas Argonauta puedan estar involucradas en mecanismos redox. Inicialmente, la formación de agregados covalentes solubles de la proteína completa, o la inestabilidad del dominio PAZ purificado en ausencia de DTT, fueron juzgados como problemas técnicos a solucionar, sin adentrarse a la búsqueda de un correlato fisiológico.

³⁶ <http://refold.med.monash.edu.au/>

Sin embargo, la exploración de vías alternativas para solucionar estos problemas dejó en evidencia que ambas proteínas poseían una tendencia intrínseca a oxidarse y formar agregados covalentes, solubles en el caso de la proteína completa, e insolubles en el caso del dominio PAZ. Considerando que la cisteína es un aminoácido altamente reactivo que suele ser seleccionado en contra salvo en casos donde realmente se lo necesita, y que por esta razón es uno de los aminoácidos que se encuentra menos representado si se toma un grupo de proteínas al azar (junto con el triptófano y la metionina, dependiendo del tipo de organismos), llama poderosamente la atención que la proteína TcPIWI-tryp posea un número tan elevado de cisteínas. La proteína completa posee un total de 20 cisteínas (1,86%), de las cuales 14 (70% del total de cisteínas) se encuentran en la mitad C-terminal de la proteína, abarcando la porción final del dominio PAZ, el bucle conector con el dominio MID, y los dominios MID/PIWI. Más aún, el tramo final del dominio PAZ e inicio del bucle conector con el dominio MID posee la secuencia HCC, es decir, tres aminoácidos contiguos típicamente utilizados en catálisis o en unión a metales, incluyendo dos cisteínas adyacentes.

Desde luego, la pregunta es si es esto una característica específica de TcPIWI-tryp, o del resto de las Argonautas eucariotas. Al respecto, cabe mencionar solo algunos ejemplos de argonautas de ambas sub-familias eucariotas y de organismos representativos. La proteína Piwi-1 humana (HIWI) posee 16 cisteínas, 12 de las cuales se ubican en la mitad C-terminal (75%). Entre ellas, se distingue un motivo CxC muy frecuente en proteínas que participan de mecanismos redox tales como las tioredoxinas. La proteína Piwi-4 humana posee 23 cisteínas, de las cuales 16 (70%) se encuentran en la mitad C-terminal. La proteína Argonauta de *S.pombe* posee 17, y 12 (70%) en los dominios MID/PIWI. La emblemática proteína Ago1 de *A.thaliana* posee 19 cisteínas, y entre las 15 (79%) ubicadas en la mitad C-terminal, se encuentra un motivo CC. Para no hacer esta lista más larga, un último ejemplo: la proteína Ago4 de ratón, que posee 35 cisteínas, 25 (71%) de las cuales se ubican en la mitad C-terminal, y entre las que se incluyen dos motivos CC, al igual que en la proteína Drosha de *D. melanogaster*.

Otro argumento a favor de que las cisteínas presentes en TcPIWI-tryp pudiesen cumplir roles específicos de la presencia de un grupo tiol o tiolato sería el estudio de su conservación en el resto de los tripanosomátidos. Muchos de estos residuos se encuentran conservados en todas las proteínas PIWI-tryp.

La gran sensibilidad de la proteína TcPIWI-tryp a la oxidación y a la formación de agregados covalentes exigiría que - de estar las proteínas bien plegadas - existiesen residuos de cisteína expuestos hacia la superficie, en regiones dónde pudiesen darse interacciones proteína proteína. Como primera aproximación a esto, se tomó el modelo por homología de los dominios PAZ y MID/PIWI de la proteína TcPIWI-tryp, consuetudinario según ya fuera explicado, y se pintaron todos los residuos de cisteína en violeta, con los átomos de azufre en amarillo. Como puede observarse en la figura 30, varias cisteínas se encuentran en la superficie del modelo, incluida la cisteína 597, que forma el motivo HCC.

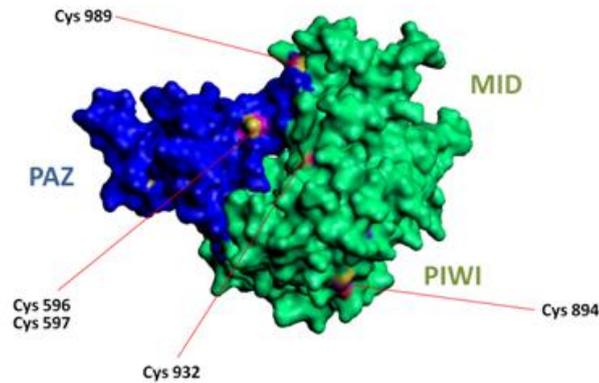


Figura 30: Conservación de los residuos de cisteína en las proteínas PIWI-tryp y predicción de residuos de cisteína que pudiesen estar expuestos en la superficie de la proteína. Se tomó el modelo por homología de la figura 9, se coloreó la construcción correspondiente al dominio PAZ en azul, y se pintaron las cisteínas de violeta, con los átomos de azufre en amarillo.

- **Función fisiológica de la proteína TcPIWI-tryp**

Lamentablemente, los resultados de este trabajo no permiten asignar fehacientemente una función biológica a estas proteínas, si bien han arrojado numerosos indicios hacia posibles funciones, a la espera de resultados experimentales que ya se han puesto en marcha por otros miembros de nuestro grupo, tales como la obtención de parásitos knock-out.

Sí resulta claro que las proteínas PIWI-tryp son argonautas canónicas en cuanto a su arquitectura de dominios, y por tanto, que deberían especializarse en la unión de fragmentos pequeños de ácidos nucleicos derivados de precursores de ARN de doble hebra. Hemos descrito en *T. cruzi* una población de fragmentos de ARN que satisface esta condición, derivados de ARN de transferencia, que podrían officiar como precursores de ARN doble hebra (García-Silva et al. 2010). Más aún, este tipo de poblaciones fueron simultánea y posteriormente descritos en numerosos organismos, incluidos eucariotas superiores, como se ha detallado en la introducción. Por consiguiente, la constatación de que la mayoría de los ARNs de menos de 35 nucleótidos que co-inmunoprecipitaron con TcPIWI-tryp hayan sido fragmentos de tARNs es un resultado que podría ser muy importante. Actualmente, se está analizando la población de ARNs que co-inmunoprecipita con la proteína por tecnologías de secuenciado de nueva generación (Illumina), de modo de corroborar los resultados preliminares, y estudiar otras poblaciones que pudiesen estar también representadas. Este estudio tal vez permitiría detectar secuencias correspondientes ya no a los fragmentos de ARN que uniría TcPIWI-tryp, sino a secuencias de ácidos nucleicos complementarias (secuencias blanco), lo que permitiría elaborar modelos funcionales factibles de ser corroborados experimentalmente.

En cualquier caso, las proteínas de la subfamilia PIWI-tryp deberían estudiarse también en otros tripanosomátidos modelos, tales como *T. brucei* (que posee otra Argonauta, e interferencia) y *L. major* (más parecido a *T. cruzi* en este contexto)

Respecto a las posibles implicancias de estas proteínas en mecanismos del tipo redox, no existe ningún reporte en la literatura que apunte en esta dirección. No obstante, los fragmentos de ARNs de transferencia sí han sido relacionados con situaciones de stress oxidativo (Thompson et al. 2008). De la

proteína TcPIWI-tryp, se sabe por trabajo de otros miembros de nuestro grupo que la misma es exportada dentro de microvesículas. Si bien la naturaleza de estas microvesículas no es aún del todo conocida, muchos tipos de vesículas son oxidantes respecto del citoplasma, por lo que la proteína podría ser sometida a ambientes con distinto potencial redox. Además, recientemente han aparecido las primeras evidencias experimentales de que algunas proteínas Argonauta son activamente secretadas al medio extracelular en complejos Argonauta/microARN libres de membrana, como es el caso de la Argonauta-2 humana (Arroyo et al. 2011). El hecho de que algunas proteínas Argonauta sean también proteínas extracelulares permite cuestionarse si la alta proporción de cisteínas en la mitad C-terminal podría corresponder a la presencia de puentes disulfuro estructurales, lo que es una posibilidad no aún explorada.

CONCLUSIÓN

En el marco de esta tesis se ha partido de una proteína que no era considerada un miembro genuino de la familia Argonauta y de la que se postulaba la pérdida de su dominio PAZ, y se ha demostrado la presencia del mismo, expresión en forma soluble en bacterias y purificación, hasta la realización incluso de estudios estructurales, actualmente en curso. La presencia de un dominio PAZ revaloriza a esta proteína y todas sus homólogas como proteínas Argonauta, y exige la búsqueda de su función biológica, hasta el momento desconocida. Se ha aportado evidencia sobre su presunta asociación con fragmentos derivados de ARNs de transferencia, que parecerían ser una nueva familia de pequeños ARNs reguladores según estudios recientes en numerosos eucariotas.

Se describió como todas las proteínas argonauta de tripanosomátidos formarían un nuevo grupo filogenético, estableciendo que las proteínas AGO-tryp y PIWI-tryp serían mucho más emparentadas de lo que podría suponerse a raíz de la nomenclatura establecida. Por ello, se propuso una nueva nomenclatura que fuese consistente con los datos filogenéticos, y la misma ya fue asumida por otros autores pese a haber sido publicada hace menos de un año.

También se obtuvieron parásitos sobre-expresantes, anticuerpos policlonales contra la proteína completa, y se pudo llevar a cabo un replegamiento *in vitro* de la misma, aunque las cantidades obtenidas fueron muy reducidas. Aún así, dada la baja cantidad de proteínas Argonauta eucariotas que pudieron ser obtenidas de forma recombinante en bacterias, este no deja de ser un resultado interesante.

Si bien este trabajo deja más preguntas abiertas de las que pudo responder, aporta herramientas, elementos y visiones novedosas en el campo de las proteínas Argonauta, y justifica la continuación de los estudios iniciados, así como de nuevas aproximaciones experimentales.

PERSPECTIVAS

El presente no es un trabajo cerrado. Por el contrario, numerosos experimentos o ensayos han quedado en etapas intermedias y deberían ser repetidos introduciendo variantes que no fueron consideradas hasta no haber digerido una visión global de todo el trabajo realizado. También existe una gama amplia de experimentos que se están actualmente realizando, en el marco de los estudios de grado, maestría y doctorado de otros miembros del grupo, y que complementan esta tesis. Referencia a algunos de estos trabajos se ha hecho cuando se creyó necesario, pero en sí las aproximaciones experimentales al tema por parte del grupo en su conjunto son considerablemente más amplias de lo que se transluce en esta tesis.

1. Variaciones en algunos de los estudios realizados

Desde luego, no existen límites teóricos (solo prácticos) para el número de experimentos y variaciones que uno quisiera introducir antes de estar en condiciones de afirmar conceptos. Se enumerarán aquí solamente algunos que se consideran relevantes, y que a la luz de los resultados obtenidos podrían tener probabilidades altas de resultar efectivos.

Respecto a la expresión de la proteína completa, es necesario un estudio empírico de numerosas combinaciones de vectores de expresión y cepas de *E. coli*, a efectos de determinar si es posible o no obtener proteína TcPIWI-tryp soluble en este sistema de expresión. Los resultados obtenidos con el replegamiento *in vitro* inducen a pensar que la proteína - pese a poseer un gran tamaño y numerosos dominios - puede adquirir una conformación globular a partir de su estructura primaria, si las condiciones son las adecuadas. El potencial redox y la presencia de ácidos nucleicos parecerían ser dos variables determinantes. El contacto con el genoma bacteriano podría evitarse expresando a la proteína en el periplasma de *E. coli*, o como proteína de fusión, de modo de interferir con la actividad de la proteína en la unión de ácidos nucleicos (lo cual es además una función esperada).

Complementariamente, deberían probarse distintos valores de pH a la hora de realizar la lisis bacteriana. Lisis a valores de pH iguales o superiores al punto isoeléctrico de la proteína podrían disminuir su afinidad por los ácidos nucleicos, en la medida que la misma dependa (específica o inespecíficamente) de atracciones electroestáticas. Aún cuando la proteína fuese insoluble en tales condiciones, podría intentarse repetir el protocolo de replegamiento *in vitro* utilizado, evitándose así el contacto con ácidos nucleicos durante el mismo.

La co-expresión de proteínas con chaperonas, o con otras proteínas con las que presuntamente pudiese interactuar, muchas veces produce buenos resultados. En la sección siguiente se describirán algunas proteínas que se cree podrían interactuar con TcPIWI-tryp en el marco de sus funciones fisiológicas.

Si todo lo anterior fallara, debería considerarse un cambio a un sistema de expresión eucariota, ya sea en *Pichia pastoris*, en sistemas de expresión en organismos similares (ej: *Leishmania*), en células de insecto, o inclusive de mamífero. En todo caso, no debe perderse de vista los ensayos a los que se pretendería destinar la proteína, ya que los niveles de expresión en muchos de estos sistemas suelen ser bajos.

En cuanto a la expresión del dominio PAZ, el protocolo optimizado en este trabajo permite la expresión hasta pureza electroforética a razón de 2,64mg por litro de cultivo, lo que es aceptable para la mayoría de las actividades que se pudieran realizar. De todos modos, si se entendiese que serían necesarios numerosos ensayos futuros que requiriesen grandes cantidades de proteína, podría ser una inversión de tiempo buscar otras cepas de *E. coli* y condiciones que permitan maximizar estos rendimientos.

Los ensayos de cristalogénesis se encuentran en curso, y resta aún realizar las pruebas de optimización de la proteína PAZ-IAM con acetato de sodio, sulfato de litio y cloruro de sodio como precipitante. También sería deseable repetir todos los ensayos de cristalogénesis, pero a 4°C. La estructura del dominio PAZ sería la prueba definitiva de que las proteínas PIWI-tryp poseen tales dominios. Además, dada la gran divergencia a nivel de la secuencia de aminoácidos que presenta este dominio, sería muy interesante observar como tales diferencias se manifiestan en su estructura, que podría estar adaptada a una función específica distinta.

El estudio de las distintas poblaciones de ARNs que co-inmunoprecipitan con TcPIWI-tryp por métodos de secuenciación de nueva generación o secuenciación masiva se encuentra actualmente en curso. Asimismo, la caracterización de los parásitos sobre-expresantes de la proteína TcPIWI-tryp, en términos de su capacidad proliferativa e infectiva. También se encuentra en curso la obtención de parásitos knock-out para la proteína TcPIWI-tryp.

2. Otras aproximaciones experimentales

Un estudio que se planificó pero que finalmente no se realizó, focalizándose primero en los ensayos de cristalogénesis, es determinar la capacidad de dominio PAZ recombinante en la unión de ácidos nucleicos. A tales efectos, se sintetizaron oligonucleótidos correspondientes a la secuencia 5' del ARN de transferencia de ácido aspártico que aparece más representado en el clonado de pequeños ARNs en *T. cruzi*, unidos a biotina en su extremo 5'. Se planea incubar la proteína recombinante con dichos oligonucleótidos, o con secuencias idénticas pero sin biotina, correr el producto en geles nativos (o en geles desnaturizantes luego de realizar un entrecruzamiento covalente), transferir a membranas de nylon y revelar mediante reacción quimioluminiscente con peroxidasa unida a estreptavidina. El propósito de este estudio es determinar si la unión a la proteína logra disminuir la movilidad electroforética de los oligonucleótidos marcados (ensayo de retardo en gel, o EMSA). Se diseñaron otros oligonucleótidos (de ADN y ARN) no marcados que permiten realizar ensayos de competencia para determinar afinidades relativas.

Habiendo determinado por EMSA la naturaleza de los ligandos más afines (ADN o ARN, simple hebra, doble hebra o doble hebra con "overhangs" 3', con o sin extremos 3'-OH, con una secuencia determinada o independiente de la misma, etc.) se podrían realizar estudios biofísicos a efectos de determinar la constante de unión y parámetros termodinámicos por calorimetría de titulación isotérmica (ITC), o constantes cinéticas de asociación y disociación por resonancia plasmónica de superficie (Biacore, GE Healthcare). Ambas tecnologías se encuentran disponibles en el IPMon.

Aún cuando no se obtuviesen cristales de proteína capaces de difractar rayos X a una resolución adecuada, tanto para la proteína PAZ-SH como para la proteína PAZ-IAM, aún podría intentarse la resolución de la estructura del dominio en solución, por resonancia magnética nuclear (RMN). De hecho,

varias de las estructuras de dominios PAZ actualmente disponibles han sido resueltas por esta técnica, dado el tamaño pequeño del dominio (12 KDa, en nuestro caso). La técnica de NMR también permite adicionar ligandos (en este caso, aquellos ácidos nucleicos determinados por EMSA como afines a la proteína) y estudiar cuáles son los residuos que se ven más afectados a partir de la unión, así como determinar constantes de unión proteína-ligando. En este sentido, nos hemos puesto en contacto con colegas del IBR de Rosario, Argentina, y se espera realizar en el corto plazo las primeras medidas de RMN del dominio PAZ, así como los ensayos de afinidad por distintos ácidos nucleicos

Volviendo al estudio funcional de las proteínas PIWI-tryp, se podría sacar provecho de que las mismas no serían necesarias para el fenómeno de interferencia en *T. brucei*, y de este modo silenciar por ARNi la expresión de la proteína TbPIWI-tryp (“*knock-down*”). Seguidamente, se podrían estudiar cambios a nivel de ARNs mensajeros por microarrays, aunque lo más conveniente sería un estudio de RNA-seq por secuenciación masiva entre la población silenciada y sin silenciar, ya que los cambios podrían darse a nivel de ARNs no codificantes (por ejemplo: retrotransposones, o los propios fragmentos derivados de ARNs de transferencia). Desde luego, la extrapolación directa de los resultados a *T. cruzi* no es posible, pero el sistema puede ser una alternativa más rápida que complemente los resultados obtenidos luego de conseguir el “*knock-out*” de TcPIWI-tryp. También existe la posibilidad de transfectar *T. cruzi* con una sonda de ARN similar a los fragmentos de tARN pero de secuencia irrelevante, de modo de saturar a la proteína TcPIWI-tryp, y generar un efecto dominante negativo.

3. Estudio de otras proteínas que podrían interactuar con TcPIWI-tryp

Las proteínas Argonautas median su función en forma de complejos de ribonucleoproteína, tales como el complejo RISC. Unas de las proteínas que forman parte de muchos de estos complejos, son necesarias para un silenciamiento efectivo por microARNs, e interactúan directamente con las proteínas Argonata, son las proteínas de la familia de GW182 (Eulalio et al. 2009; Takimoto et al. 2009). Estas proteínas son componentes característicos de los gránulos P o “*P-bodies*” (Liu et al. 2005), cuya formación es dependiente de una vía de microARNs funcional (Pauley et al. 2006). Actualmente se cree que una de las posibles formas en que estas proteínas mediarían el silenciamiento génico es por interacción directa con las proteínas citoplasmáticas de unión a poli A (PABPC1), inhibiendo de este modo la traducción (Huntzinger et al. 2010)

La proteína GW182 fue identificada como un auto-antígeno en 2002, utilizando el suero de un paciente con neuropatía sensorial y motora (Eystathioy et al. 2002). La proteína debe su nombre a su tamaño de 182KDa, y a la presencia de numerosos repetidos de glicina y triptófano (GW) en la región N-terminal. Más tarde se descubrió que dichos repetidos eran necesarios para la interacción de las proteínas GW182 con proteínas Argonata (Lian et al. 2009; Takimoto et al. 2009). Además de la región N-terminal, el silenciamiento mediado por GW182 depende también de la región media y C-terminal de la proteína (Eulalio et al. 2009), donde se ubica el dominio RRM (de “*RNA recognition motif*”, presente en numerosas proteínas de unión al ARN). Adyacente a dicho dominio se encuentra un motivo “conservado de función desconocida” (DUF), actualmente recatalogado como PAM2 (por “*Polia-A binding protein interacting motif 2*”), que posee la secuencia EFXP (Huntzinger et al. 2010). Por último, en la zona central se encuentra una secuencia rica en glutaminas, usualmente representada como “*poliQ*”.

No existen reportes que indiquen que miembros de esta familia existan en tripanosomátidos, ni hay ninguna proteína anotada como tal en las bases de datos. Sin embargo, se realizó una búsqueda en los genomas de *T. cruzi* y *T. brucei* de proteínas que contuviesen un porcentaje alto de repetidos de glicina-triptofano. Se encontró así una proteína anotada como RNA-binding protein RGGm (NCBI RefSeq: XP_814887.1; Tritydb: Tc00.1047053503419.50). La misma se encuentra conservada en *Trypanosoma* y *Leishmania* y se encuentra en loci sinténicos. Más aún, se encuentra adyacente en el genoma a las proteínas AGO-tryp, en aquellos organismos que las conservan. Por la presencia de genes de tARN, y observando datos de marcas de histonas en dicha región, se puede predecir que las proteínas TbAGO-tryp y RGGm se transcribirían en un mismo mensajero policistrónico en *T. brucei*. Esta observación no tiene más valor que lo anecdótico, pues en tripanosomátidos los genes que se transcriben en una misma unidad policistrónica no tienen por qué estar relacionados funcionalmente.

Las proteínas RGGm de tripanosomátidos son pequeñas, y en ese sentido, muy distintas de las proteínas GW182 de metazoarios. Sin embargo, poseen también numerosos repetidos GWG en su región N-terminal. Más aún, poseen también un dominio RRM bien determinado en su extremo C-terminal. El motivo EFXP que permite la unión de las proteínas GW182 a la proteína PABPC1 también se encuentra en las RGGm, modificado a EFXP. No poseen una región rica en residuos de glutamina (“*poli Q box*”, o “*Q-rich*”) pero dicho motivo sí se encuentra en las proteínas PIWI-tryp. Más aún, posee numeroso repetidos N-terminales de arginina-glicina-glicina (RGG) que también se encuentran en TbAGO-tryp y algunas otras argonautas, y facilitarían su unión a los poliribosomas (Shi et al. 2009)

Si bien estas similitudes no permiten afirmar que las proteínas RGGm sean homólogas funcionales de las proteínas GW182, sí parecería oportuno un estudio detallado de estas proteínas, como posibles proteínas asociadas a las argonautas AGO-tryp, PIWI-tryp, o ambas.

Se sabe que la proteína se expresa en epimastigotes de *T. cruzi*, pues fue detectada en estudios proteómicos (Parodi-Talice et al. 2004). En *T. brucei*, está descrito que esta proteína formaría parte de un complejo mitocondrial de proteínas de unión al ARN. El silenciamiento por ARNi de esta proteína provocó el desacoplamiento de dicho complejo, inhibió el crecimiento *in vitro* de formas procíclicas, y resultó en células con un número anormal de núcleos (Acestor et al. 2009).

AGRADECIMIENTOS

En primer lugar, un agradecimiento a los demás miembros de nuestro grupo de trabajo, co-autores de esta tesis. Parte de los resultados mostrados (ej: la sobre-expresión de TcPIWI-tryp en *T. cruzi* y la inmunoprecipitación de dicha proteína) se solapan con los resultados de la tesis de doctorado de la Mag. María Rosa García-Silva. De hecho, el cultivo de los parásitos, la incubación de los extractos de *T. cruzi* con las esferas magnéticas y el secuenciado de los pequeños ARNs co-inmunoprecipitados fueron realizados por Rosa y por Julia Sanguinetti. También un agradecimiento a Braulio Bonilla, Florencia Cabrera, Fernanda Bangueses y Catalina Güida, cuya generosidad y buena disposición en el día a día contribuyeron sin duda al desarrollo de esta tesis. Desde luego, un agradecimiento a mi orientador, Dr. Alfonso Cayota.

También quiero agradecer a todos los miembros del IPMon. Como estudiante, considero que la buena disposición de todo el personal científico del Instituto, siempre dispuestos a ayudar, engancharse con los problemas planteados y sugerir interesantes alternativas experimentales, es de incomparable valor. Si bien el agradecimiento es para todos en general, nombro algunos que participaron o se involucraron de forma más activa: Felipe Trattemberg y Nicole Larrieux (PXF), Agustín Correa (UPR), Gonzalo Obal y Federico Carrión (UBP), Madelón Portela (UByPA), Bruno Manta (G5), Martín Graña y Álvaro Pena (UBI), y todos los integrantes de UBM. También los jefes de grupo de las respectivas unidades, en particular Alejandro Buschiazzo, Otto Pritsch, Pablo Opezzo, Carlos Robello y Marcelo Comini: con todos ellos he tenido fructíferas conversaciones sobre posibilidades y alcances de esta tesis, que han sido sumamente enriquecedoras para mí. Mención aparte merece Sergio Pantano, que colaboró con nosotros para la detección del dominio PAZ por HHpred y el modelado de la proteína por homología.

Otros miembros de la comunidad académica local que se han interiorizado en este trabajo dando valiosos aportes son el Dr. Fernando Álvarez y la Dra. Adriana Parodi, de la Facultad de Ciencias (UdelaR). Del IBR de Rosario (Argentina), el Dr. Esteban Serra también fue un colaborador estrecho. También quisiera hacer mención a todos los integrantes de la Unidad de Bioquímica Analítica del CIN (Facultad de Ciencias, UdelaR), donde también trabajo. El Prof. Laíz siempre me alentó para que apostara fuerte a mi maestría como instancia formativa esencial, y esa es una actitud que se agradece mucho.

Al tribunal de esta tesis quiero agradecerles la buena disposición, y pido sinceras disculpas por la poca capacidad de síntesis, que reconozco. En especial, agradezco a la Dra. Mónica Marín que aceptó integrar el tribunal de esta tesis con poca antelación y previo a viajar al exterior

Agradezco también al PEDECIBA. En particular, a las autoridades y secretaría de PEDECIBA-biología. También a la ANII, ya que conté con beca de maestría otorgada por dicha Agencia en el período agosto 2009 – marzo 2011, y desde julio de 2011, con ingreso al SNI.

Desde fuera de lo académico, el apoyo incondicional de mi familia: los viejos, Agustín, Santiago, Diego y Magdalena, los primos, los tíos y los abuelos; sin duda contribuyeron a que pudiera terminar esta tesis con alegría, y fueron un constante cable a tierra. Un agradecimiento especial a Vale, que cargó con la parte más dura de esta tesis; lo que no aparece en los resultados ni en la discusión pero se encuentra en la base de los mismos: muchas horas de salidas de trabajar tarde, muchos planes pospuestos...

BIBLIOGRAFÍA

- Acestor, N., Panigrahi, A.K., Carnes, J., Zikova, A., Stuart, K.D., 2009. The MRB1 complex functions in kinetoplastid RNA processing. *RNA* 15(2), 277-286.
- Amitsur, M., Morad, I., Chapman-Shimshoni, D., Kaufmann, G., 1992. HSD restriction-modification proteins partake in latent anticodon nuclease. *EMBO J* 11(8), 3129-3134.
- Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., Gvozdev, V.A., 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol* 11(13), 1017-1027.
- Arroyo, J.D., Chevillet, J.R., Kroh, E.M., Ruf, I.K., Pritchard, C.C., Gibson, D.F., Mitchell, P.S., Bennett, C.F., Pogosova-Agadjanyan, E.L., Stirewalt, D.L., Tait, J.F., Tewari, M., 2011. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci U S A* 108(12), 5003-5008.
- Atayde, V.D., Tschudi, C., Ullu, E., 2011. The emerging world of small silencing RNAs in protozoan parasites. *Trends Parasitol* 27(7), 321-327.
- Batista, T.M., Marques, J.T., 2011. RNAi pathways in parasitic protists and worms. *J Proteomics* 74(9), 1504-1514.
- Baum, J., Papenfuss, A.T., Mair, G.R., Janse, C.J., Vlachou, D., Waters, A.P., Cowman, A.F., Crabb, B.S., de Koning-Ward, T.F., 2009. Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res* 37(11), 3788-3798.
- Bernstein, E., Caudy, A.A., Hammond, S.M., Hannon, G.J., 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409(6818), 363-366.
- Bochkarev, A., Bochkareva, E., 2004. From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. *Curr Opin Struct Biol* 14(1), 36-42.
- Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., Benning, C., 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J* 17(1), 170-180.
- Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., Weichenrieder, O., 2011. Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proc Natl Acad Sci U S A* 108(26), 10466-10471.
- Boland, A., Tritschler, F., Heimstadt, S., Izaurralde, E., Weichenrieder, O., 2010. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep* 11(7), 522-527.
- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., Zhu, J.K., 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 123(7), 1279-1291.
- Brennecke, J., Stark, A., Russell, R.B., Cohen, S.M., 2005. Principles of microRNA-target recognition. *PLoS Biol* 3(3), e85.
- Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Suzuki, H., Hayashizaki, Y., Daub, C.O., 2011. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol* 8(1), 158-177.
- Cerutti, H., Casas-Mollano, J.A., 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet* 50(2), 81-99.
- Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W., Green, P.J., Barton, G.J., Hutvagner, G., 2009. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* 15(12), 2147-2160.
- Collins, F.S., Green, E.D., Guttmacher, A.E., Guyer, M.S., 2003. A vision for the future of genomics research. *Nature* 422(6934), 835-847.
- Couvillion, M.T., Sachidanandam, R., Collins, K., 2010. A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo. *Genes Dev* 24(24), 2742-2747.
- Chen, C.J., Liu, Q., Zhang, Y.C., Qu, L.H., Chen, Y.Q., Gautheret, D., 2011. Genome-wide discovery and analysis of microRNAs and other small RNAs from rice embryogenic callus. *RNA Biol* 8(3).
- Chi, E.Y., Krishnan, S., Randolph, T.W., Carpenter, J.F., 2003. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* 20(9), 1325-1336.
- Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162(1), 156-159.
- Chomczynski, P., Sacchi, N., 2006. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc* 1(2), 581-585.
- DaRocha, W.D., Otsu, K., Teixeira, S.M., Donelson, J.E., 2004. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 133(2), 175-186.
- Djuranovic, S., Zinchenko, M.K., Hur, J.K., Nahvi, A., Brunelle, J.L., Rogers, E.J., Green, R., 2010. Allosteric regulation of Argonaute proteins by miRNAs. *Nat Struct Mol Biol* 17(2), 144-150.
- Durand-Dubief, M., Bastin, P., 2003. TbAGO1, an argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in *Trypanosoma brucei*. *BMC Biol* 1, 2.
- Eiberle, M.K., Jungbauer, A., 2010. Technical refolding of proteins: Do we have freedom to operate? *Biotechnol J* 5(6), 547-559.
- Elbashir, S.M., Lendeckel, W., Tuschl, T., 2001a. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* 15(2), 188-200.
- Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T., 2001b. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J* 20(23), 6877-6888.
- Ellman, G.L., 1959. Tissue sulfhydryl groups. *Arch Biochem Biophys* 82(1), 70-77.
- Emará, M.M., Ivanov, P., Hickman, T., Dawra, N., Tisdale, S., Kedersha, N., Hu, G.F., Anderson, P., 2010. Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J Biol Chem* 285(14), 10959-10968.
- Eulalio, A., Helms, S., Fritsch, C., Fauser, M., Izaurralde, E., 2009. A C-terminal silencing domain in GW182 is essential for miRNA function. *RNA* 15(6), 1067-1077.
- Eystathiou, T., Chan, E.K., Tenenbaum, S.A., Keene, J.D., Griffith, K., Fritsler, M.J., 2002. A phosphorylated cytoplasmic autoantigen, GW182, associates with a unique population of human mRNAs within novel cytoplasmic speckles. *Mol Biol Cell* 13(4), 1338-1351.

- Farazi, T.A., Juraneck, S.A., Tuschl, T., 2008. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 135(7), 1201-1214.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669), 806-811.
- Frank, F., Fabian, M.R., Stepinski, J., Jemielity, J., Darzynkiewicz, E., Sonenberg, N., Nagar, B., 2011. Structural analysis of 5'-mRNA-cap interactions with the human AGO2 MID domain. *EMBO Rep* 12(5), 415-420.
- Frank, F., Sonenberg, N., Nagar, B., 2010. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465(7299), 818-822.
- Franzen, O., Arner, E., Ferella, M., Nilsson, D., Respuela, P., Carninci, P., Hayashizaki, Y., Aslund, L., Andersson, B., Daub, C.O., 2011. The Short Non-Coding Transcriptome of the Protozoan Parasite *Trypanosoma cruzi*. *PLoS Negl Trop Dis* 5(8), e1283.
- Fu, H., Feng, J., Liu, Q., Sun, F., Tie, Y., Zhu, J., Xing, R., Sun, Z., Zheng, X., 2009. Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett* 583(2), 437-442.
- Garcia-Silva, M.R., Frugier, M., Tosar, J.P., Correa-Dominguez, A., Ronalte-Alves, L., Parodi-Talice, A., Rovira, C., Robello, C., Goldenberg, S., Cayota, A., 2010. A population of tRNA-derived small RNAs is actively produced in *Trypanosoma cruzi* and recruited to specific cytoplasmic granules. *Mol Biochem Parasitol* 171(2), 64-73.
- Garcia Silva, M.R., Tosar, J.P., Frugier, M., Pantano, S., Bonilla, B., Esteban, L., Serra, E., Rovira, C., Robello, C., Cayota, A., 2010. Cloning, characterization and subcellular localization of a *Trypanosoma cruzi* argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene* 466(1-2), 26-35.
- Getz, E.B., Xiao, M., Chakrabarty, T., Cooke, R., Selvin, P.R., 1999. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal Biochem* 273(1), 73-80.
- Ghildiyal, M., Zamore, P.D., 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10(2), 94-108.
- Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., Mello, C.C., 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106(1), 23-34.
- Grosshans, H., Filipowicz, W., 2008. Molecular biology: the expanding world of small RNAs. *Nature* 451(7177), 414-416.
- Hamilton, A.J., Baulcombe, D.C., 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286(5441), 950-952.
- Hammond, S.M., Bernstein, E., Beach, D., Hannon, G.J., 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404(6775), 293-296.
- Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z., Kay, M.A., 2010. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* 16(4), 673-695.
- Hock, J., Meister, G., 2008. The Argonaute protein family. *Genome Biol* 9(2), 210.
- Hsieh, L.C., Lin, S.I., Kuo, H.F., Chiou, T.J., 2010. Abundance of tRNA-derived small RNAs in phosphate-starved *Arabidopsis* roots. *Plant Signal Behav* 5(5).
- Hughes, A.L., Piontkivska, H., 2003. Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: evidence for paraphyly of *Trypanosoma* and six other genera. *Mol Biol Evol* 20(4), 644-652.
- Huntzinger, E., Braun, J.E., Heimstadt, S., Zekri, L., Izaurralde, E., 2010. Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing. *EMBO J* 29(24), 4146-4160.
- Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., Zamore, P.D., 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293(5531), 834-838.
- Hutvagner, G., Simard, M.J., 2008. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* 9(1), 22-32.
- Iyer, K.S., Klee, W.A., 1973. Direct spectrophotometric measurement of the rate of reduction of disulfide bonds. The reactivity of the disulfide bonds of bovine γ -lactalbumin. *J Biol Chem* 248(2), 707-710.
- Izaac, A., Schall, C.A., Mueser, T.C., 2006. Assessment of a preliminary solubility screen to improve crystallization trials: uncoupling crystal condition searches. *Acta Crystallogr D Biol Crystallogr* 62(Pt 7), 833-842.
- Izant, J.G., Weintraub, H., 1984. Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. *Cell* 36(4), 1007-1015.
- Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordseil, R., Shames, D.S., Minna, J.D., Corey, D.R., 2006. Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* 13(9), 787-792.
- Jinek, M., Doudna, J.A., 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457(7228), 405-412.
- Jochl, C., Rederstorff, M., Hertel, J., Stadler, P.F., Hofacker, I.L., Schrettl, M., Haas, H., Huttenhofer, A., 2008. Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein synthesis. *Nucleic Acids Res* 36(8), 2677-2689.
- Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., Gingeras, T.R., 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296(5569), 916-919.
- Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., Daub, C.O., Kai, C., Kawai, J., Yasuda, J., Carninci, P., Hayashizaki, Y., 2008. Hidden layers of human small RNAs. *BMC Genomics* 9, 157.
- Kelly, J.M., Ward, H.M., Miles, M.A., Kendall, G., 1992. A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*. *Nucleic Acids Res* 20(15), 3963-3969.
- Kennerdell, J.R., Carthew, R.W., 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* 95(7), 1017-1026.
- Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., Plasterk, R.H., 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev* 15(20), 2654-2659.
- Kim, V.N., 2006. Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev* 20(15), 1993-1997.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294(5543), 853-858.
- Lander, E.S., Linton, L.M., Birren, B., et al.; 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822), 860-921.

- Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294(5543), 858-862.
- Lee, R.C., Ambros, V., 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294(5543), 862-864.
- Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75(5), 843-854.
- Lee, Y.S., Shibata, Y., Malhotra, A., Dutta, A., 2009. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev* 23(22), 2639-2649.
- Leonard, G., Soanes, D.M., Stevens, J.R., 2011. Resolving the question of trypanosome monophyly: a comparative genomics approach using whole genome data sets with low taxon sampling. *Infect Genet Evol* 11(5), 955-959.
- Li, Y., Luo, J., Zhou, H., Liao, J.Y., Ma, L.M., Chen, Y.Q., Qu, L.H., 2008. Stress-induced tRNA-derived RNAs: a novel class of small RNAs in the primitive eukaryote *Giardia lamblia*. *Nucleic Acids Res* 36(19), 6048-6055.
- Lian, S.L., Li, S., Abadal, G.X., Pauley, B.A., Fritzler, M.J., Chan, E.K., 2009. The C-terminal half of human Ago2 binds to multiple GW-rich regions of GW182 and requires GW182 to mediate silencing. *RNA* 15(5), 804-813.
- Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., et al.; 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069), 803-819.
- Lingel, A., Simon, B., Izaurralde, E., Sattler, M., 2003. Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* 426(6965), 465-469.
- Lingel, A., Simon, B., Izaurralde, E., Sattler, M., 2004. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* 11(6), 576-577.
- Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., Hannon, G.J., 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305(5689), 1437-1441.
- Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R., Hannon, G.J., 2005. A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* 7(12), 1261-1266.
- Lye, L.F., Owens, K., Shi, H., Murta, S.M., Vieira, A.C., Turco, S.J., Tschudi, C., Ullu, E., Beverley, S.M., 2010. Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* 6(10), e1001161.
- Ma, J.B., Ye, K., Patel, D.J., 2004. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* 429(6989), 318-322.
- Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., Patel, D.J., 2005. Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* 434(7033), 666-670.
- Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., Doudna, J.A., 2006. Structural basis for double-stranded RNA processing by Dicer. *Science* 311(5758), 195-198.
- Makarova, K.S., Wolf, Y.I., van der Oost, J., Koonin, E.V., 2009. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct* 4, 29.
- Martens-Uzunova, E.S., Jalava, S.E., Dits, N.F., van Leenders, G.J., Moller, S., Trapman, J., Bangma, C.H., Litman, T., Visakorpi, T., Jenster, G., 2011. Diagnostic and prognostic signatures from the small non-coding RNA transcriptome in prostate cancer. *Oncogene*.
- Marton, S., Garcia, M.R., Robello, C., Persson, H., Trajtenberg, F., Pritsch, O., Rovira, C., Naya, H., Dighiero, G., Cayota, A., 2008. Small RNAs analysis in CLL reveals a deregulation of miRNA expression and novel miRNA candidates of putative relevance in CLL pathogenesis. *Leukemia* 22(2), 330-338.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., Tuschl, T., 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* 15(2), 185-197.
- Metzloff, M., O'Dell, M., Cluster, P.D., Flavell, R.B., 1997. RNA-mediated RNA degradation and chalcone synthase A silencing in *petunia*. *Cell* 88(6), 845-854.
- Murphy, D., Dancis, B., Brown, J.R., 2008. The evolution of core proteins involved in microRNA biogenesis. *BMC Evol Biol* 8, 92.
- Murzin, A.G., 1993. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J* 12(3), 861-867.
- Ngo, H., Tschudi, C., Gull, K., Ullu, E., 1998. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* 95(25), 14687-14692.
- Nowotny, M., Gaidamakov, S.A., Crouch, R.J., Yang, W., 2005. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* 121(7), 1005-1016.
- Nunes, C.C., Gowda, M., Sailsbery, J., Xue, M., Chen, F., Brown, D.E., Oh, Y., Mitchell, T.K., Dean, R.A., 2011. Diverse and tissue-enriched small RNAs in the plant pathogenic fungus *Magnaporthe oryzae*. *BMC Genomics* 12, 288.
- Ohno, S., 1972. So much "junk" DNA in our genome. *Brookhaven Symp Biol* 23, 366-370.
- Parker, J.S., Roe, S.M., Barford, D., 2004. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J* 23(24), 4727-4737.
- Parker, J.S., Roe, S.M., Barford, D., 2005. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* 434(7033), 663-666.
- Parodi-Talice, A., Duran, R., Arrambide, N., Prieto, V., Pineyro, M.D., Pritsch, O., Cayota, A., Cervenansky, C., Robello, C., 2004. Proteome analysis of the causative agent of Chagas disease: *Trypanosoma cruzi*. *Int J Parasitol* 34(8), 881-886.
- Pauley, K.M., Eystathiou, T., Jakymiw, A., Hamel, J.C., Fritzler, M.J., Chan, E.K., 2006. Formation of GW bodies is a consequence of microRNA genesis. *EMBO Rep* 7(9), 904-910.
- Piontkivska, H., Hughes, A.L., 2005. Environmental kinetoplastid-like 18S rRNA sequences and phylogenetic relationships among Trypanosomatidae: paraphyly of the genus *Trypanosoma*. *Mol Biochem Parasitol* 144(1), 94-99.
- Prucca, C.G., Slavin, I., Quiroga, R., Elias, E.V., Rivero, F.D., Saura, A., Carranza, P.G., Lujan, H.D., 2008. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* 456(7223), 750-754.
- Rashid, U.J., Paterok, D., Koglin, A., Gohlke, H., Piehler, J., Chen, J.C., 2007. Structure of *Aquifex aeolicus* argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. *J Biol Chem* 282(18), 13824-13832.
- Richards, T.A., Cavalier-Smith, T., 2005. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436(7054), 1113-1118.

- Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P.M., Nelson, F.K., Miller, P., Gerstein, M., Weissman, S., Snyder, M., 2003. The transcriptional activity of human Chromosome 22. *Genes Dev* 17(4), 529-540.
- Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., Joshua-Tor, L., 2005. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* 12(4), 340-349.
- Robinson, K.A., Beverley, S.M., 2003. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol Biochem Parasitol* 128(2), 217-228.
- Rossi, J.J., 2004. Stopping RNA interference at the seed. *Nat Genet* 43(4), 288-289.
- Salvatore, V., Potenza, N., Papa, U., Nobile, V., Russo, A., 2010. Bacterial expression of mouse argonaute 2 for functional and mutational studies. *Int J Mol Sci* 11(2), 745-753.
- Schlicke, M., Brakmann, S., 2005. Expression and purification of histidine-tagged bacteriophage T7 DNA polymerase. *Protein Expr Purif* 39(2), 247-253.
- Seidman, C.E., Struhl, K., Sheen, J., Jessen, T., 2001. Introduction of plasmid DNA into cells. *Curr Protoc Mol Biol Chapter 1, Unit1 8*.
- Shi, H., Chamond, N., Djikeng, A., Tschudi, C., Ullu, E., 2009. RNA interference in *Trypanosoma brucei*: role of the n-terminal RGG domain and the polyribosome association of argonaute. *J Biol Chem* 284(52), 36511-36520.
- Shi, H., Djikeng, A., Mark, T., Wirtz, E., Tschudi, C., Ullu, E., 2000. Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. *RNA* 6(7), 1069-1076.
- Shi, H., Djikeng, A., Tschudi, C., Ullu, E., 2004. Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. *Mol Cell Biol* 24(1), 420-427.
- Siegel, T.N., Gunasekera, K., Cross, G.A., Ochsenreiter, T., 2011. Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing. *Trends Parasitol* 27(10), 434-441.
- Simon, B., Kirkpatrick, J.P., Eckhardt, S., Reuter, M., Rocha, E.A., Andrade-Navarro, M.A., Sehr, P., Pillai, R.S., Carlomagno, T., 2011. Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure* 19(2), 172-180.
- Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., Joshua-Tor, L., 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* 10(12), 1026-1032.
- Song, J.J., Smith, S.K., Hannon, G.J., Joshua-Tor, L., 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305(5689), 1434-1437.
- Suk, K., Choi, J., Suzuki, Y., Ozturk, S.B., Mellor, J.C., Wong, K.H., MacKay, J.L., Gregory, R.I., Roth, F.P., 2011. Reconstitution of human RNA interference in budding yeast. *Nucleic Acids Res* 39(7), e43.
- Takimoto, K., Wakiyama, M., Yokoyama, S., 2009. Mammalian GW182 contains multiple Argonaute-binding sites and functions in microRNA-mediated translational repression. *RNA* 15(6), 1078-1089.
- Theobald, D.L., Mitton-Fry, R.M., Wuttke, D.S., 2003. Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct* 32, 115-133.
- Thompson, D.M., Lu, C., Green, P.J., Parker, R., 2008. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* 14(10), 2095-2103.
- Tian, Y., Simanshu, D.K., Ma, J.B., Patel, D.J., 2011. Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proc Natl Acad Sci U S A* 108(3), 903-910.
- Tolia, N.H., Joshua-Tor, L., 2007. Slicer and the argonautes. *Nat Chem Biol* 3(1), 36-43.
- Ullu, E., Tschudi, C., Chakraborty, T., 2004. RNA interference in protozoan parasites. *Cell Microbiol* 6(6), 509-519.
- Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., Zamore, P.D., 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313(5785), 320-324.
- Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gascioli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., Crete, P., 2004. Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell* 16(1), 69-79.
- Venter, J.C., Adams, M.D., Myers, E.W., et al.; 2001. The sequence of the human genome. *Science* 291(5507), 1304-1351.
- Wang, G., Reinke, V., 2008. A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol* 18(12), 861-867.
- Wang, L., Yu, X., Wang, H., Lu, Y.Z., de Rooter, M., Prins, M., He, Y.K., 2011. A novel class of heat-responsive small RNAs derived from the chloroplast genome of Chinese cabbage (*Brassica rapa*). *BMC Genomics* 12, 289.
- Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., Patel, D.J., 2008a. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* 456(7224), 921-926.
- Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., Patel, D.J., 2009. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* 461(7265), 754-761.
- Wang, Y., Sheng, G., Juranek, S., Tuschl, T., Patel, D.J., 2008b. Structure of the guide-strand-containing argonaute silencing complex. *Nature* 456(7219), 209-213.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., et al.; 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915), 520-562.
- Wightman, B., Ha, I., Ruvkun, G., 1993. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75(5), 855-862.
- Xie, Y., Wetlaufer, D.B., 1996. Control of aggregation in protein refolding: the temperature-leap tactic. *Protein Sci* 5(3), 517-523.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., Carrington, J.C., 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* 2(5), E104.
- Yamasaki, S., Ivanov, P., Hu, G.F., Anderson, P., 2009. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J Cell Biol* 185(1), 35-42.
- Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., Zhou, M.M., 2003. Structure and conserved RNA binding of the PAZ domain. *Nature* 426(6965), 468-474.
- Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., Mello, C.C., 2006. Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* 127(4), 747-757.
- Yoshikawa, M., Peragine, A., Park, M.Y., Poethig, R.S., 2005. A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*. *Genes Dev* 19(18), 2164-2175.
- Yuan, Y.R., Pei, Y., Chen, H.Y., Tuschl, T., Patel, D.J., 2006. A potential protein-RNA recognition event along the RISC-loading pathway from the structure of *A. aeolicus* Argonaute with externally bound siRNA. *Structure* 14(10), 1557-1565.

- Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., Patel, D.J., 2005. Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell* 19(3), 405-419.
- Zamore, P.D., Tuschl, T., Sharp, P.A., Bartel, D.P., 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101(1), 25-33.
- Zeng, L., Zhang, Q., Yan, K., Zhou, M.M., 2011. Structural insights into piRNA recognition by the human PIWI-like 1 PAZ domain. *Proteins* 79(6), 2004-2009.

APÉNDICE

La presente tesis de maestría se desarrolló en el Laboratorio de Genómica Funcional (Programa Cáncer) del Institut Pasteur de Montevideo, bajo la orientación del Dr. Alfonso Cayota.

En los meses previos al comienzo de esta tesis, y ya después como estudiante formal de PEDECIBA, se trabajó en el clonado y secuenciado de pequeños ARNs en *T. cruzi*, dando origen a la siguiente publicación:

- Garcia-Silva, M.R., Frugier, M., Tosar, J.P., Correa-Dominguez, A., Ronalte-Alves, L., Parodi-Talice, A., Rovira, C., Robello, C., Goldenberg, S., Cayota, A., 2010. *A population of tRNA-derived small RNAs is actively produced in Trypanosoma cruzi and recruited to specific cytoplasmic granules.* Mol Biochem Parasitol 171(2), 64-73.

La primera mitad de esta tesis se publicó en el año 2010:

- Garcia Silva, M.R.¹, Tosar, J.P.¹, Frugier, M., Pantano, S., Bonilla, B., Esteban, L., Serra, E., Rovira, C., Robello, C., Cayota, A., 2010. *Cloning, characterization and subcellular localization of a Trypanosoma cruzi argonaute protein defining a new subfamily distinctive of trypanosomatids.* Gene 466(1-2), 26-35.

¹ Los autores: María Rosa García-Silva y Juan Pablo Tosar contribuyeron equitativamente a dicho trabajo

Si bien fuera del marco de esta tesis, pero también en el período de estudiante formal de maestría por PEDECIBA-biología, se publicaron los siguientes trabajos:

- Tosar, J.P., Brañas, G., Laíz, J., 2010. *Electrochemical DNA hybridization sensors applied to real and complex biological samples.* Biosensors and Bioelectronics 26 (4) 1205-1217
- Tosar, J.P., Keel, K., Laíz, J., 2009. *Two independent label-free detection methods in one electrochemical DNA sensor.* Biosensors and Bioelectronics 24 (10) 3036-3042