Uso del modelo experimental Caenorhabditis elegans para el estudio de incorporación de selenocisteína en proteínas

> TESIS DE DOCTORADO EN QUÍMICA FACULTAD DE QUÍMICA UNIVERSIDAD DE LA REPÚBLICA

QF Lucía Otero Larre Borges

Supervisor: Dr. Gustavo Salinas Grecco Co-supervisor: Dr. Antonio Miranda Vizuete

Montevideo, 2014

Science and everyday life cannot and should not be separated. Science, for me, gives a partial explanation for life.

Rosalind Franklin, 1940

#### RESUMEN

La selenocisteína (Sec) es incorporada a la cadena polipeptídica mediante un mecanismo no canónico: una secuencia de incorporación de Sec (SECIS) presente en el ARNm de la selenoproteína reprograma al codón UGA para incorporar Sec. Si bien este mecanismo ha sido completamente elucidado en bacterias, en eucariotas aún restan aspectos por comprender. La proteína de unión al elemento SECIS (SBP2) media la incorporación de Sec, sin embargo los roles que cumplen sus dominios y el impacto de su deficiencia en la inserción de Sec no se conoce completamente. Por otra parte, existen genes que se han reportado como involucrados o presuntamente involucrados en la decodificación e incorporación de Sec cuyas funciones no han sido esclarecidas. Para estudiar algunos de estos aspectos, utilizamos Caenorhabditis *elegans* como modelo ya que se trata de un organismo eucariota simple que posee toda la maquinaria de incorporación de Sec a disposición de una única selenoproteína: la tiorredoxina reductasa citosólica (TRXR-1). Todas las SBP2 descritas hasta el inicio de esta tesis poseían un dominio de unión al ARN (RBD) y uno de incorporación de Sec (SID). Mediante análisis in silico identificamos la SBP2 de C. elegans y, sorprendentemente, encontramos que carece de SID y que consiste únicamente en un RBD. Generamos una estirpe mutante en *sbp2* que resultó incapaz de incorporar Sec, demostrando así que este gen es esencial para este fin. El análisis de genomas de otros nematodos reveló que la ausencia de SID es característica del linaje y que el mantenimiento de la incorporación de Sec se encuentra ligado a la TRXR-1. Se deduce entonces que el mecanismo de incorporación de Sec en este filo difiere, al menos en parte, del ya reportado para otros eucariotas; SID podría ser completamente prescindible, u otra proteína, no identificada aún, podría suplir sus funciones. Encontramos también que los nematodos del clado Tylenchina han perdido la capacidad de incorporar Sec. Estos son los primeros animales no artrópodos que se reportan que perdieron esta característica; a nivel genético perdieron todos los genes involucrados en su decodificación, excepto SecP43, que se presume tendría un rol adicional a la metilación de una base del anticodón del ARNt<sup>Sec</sup> y el transporte de complejos supramoleculares asociados a los ARNm de selenoproteínas. Mediante ensayos con una estirpe mutante en SecP43 encontramos que si bien no este gen no es esencial para la incorporación de Sec en condiciones óptimas de laboratorio, es necesario para la correcta expresión de TRXR-1 en condiciones desafiantes.

Análogamente, demostramos que *R13A5.9*, un gen asociado a la incorporación de Sec, pero de función desconocida, es requerido para la incorporación de Sec en condiciones desafiantes. Así, es probable que *R13A5.9*, como SecP43, tenga funciones regulatorias en el metabolismo de selenio y/o la incorporación de Sec. Globalmente, nuestros resultados indican que *C. elegans* y el linaje nematodo proveen indicios claves para comprender el mecanismo de incorporación de Sec y la evolución de la utilización de Sec y de los genes asociados a su incorporación, así como de las selenoproteínas y los selenoproteomas.

Finalmente nuestros estudios mostraron evidencia de inicio no canónico de la traducción en *C. elegans*, mecanismo que no había sido previamente descrito en este organismo modelo. Se identificaron dos genes (uno de ellos *sbp2*) cuya traducción no iniciaría en AUG sino en AUU. Se prevé que el número de transcriptos que tengan inicio no canónico de la traducción sea aún mayor.

# AGRADECIMIENTOS

Quisiera agradecer a las agencias que financiaron y apoyaron esta tesis y las estancias realizadas en el Centro Andaluz de Biología del Desarrollo: AECI, ANII, AUIP, CSIC, PEDECIBA, Universidad de la República-CAP. A la Cátedra de Inmunología de las Facultades de Química y Ciencias y al Institut Pasteur por brindarme espacio para llevar a cabo esta tesis.

Agradezco también a los integrantes del tribunal por aceptar evaluar este trabajo.

Quiero agradecer a mucha gente que me ha acompañado durante el desarrollo de mi tesis de doctorado:

En primer lugar quiero agradecer a Gustavo. Jefe, muchas gracias por confiar en mí, allá por 2007, para ayudarte a arrancar con el primer laboratorio de *elegans* en Uruguay; y por seguir apoyándome y creyendo en mí, incluso cuando ni yo confié en mí misma. Gracias por acompañarme durante estos años, tanto desde el punto de vista científico como personal, por tenerme paciencia en momentos difíciles. Y también gracias por aceptarme como posdoc!

En segundo lugar quiero agradecer al co-supervisor de mi tesis, Antonio. Muchas gracias por haberme recibido tantas veces en tu laboratorio, donde siempre me sentí como en casa, y por enseñarme los secretos del gusano. Gracias por comprometerte tanto con mi trabajo y por el avance del mismo.

Quiero agradecer también a los "otros jefes" de inmuno: Ceci, Anita y Álvaro. Por su ayuda, por sus aportes, por sus preguntas, por sus sugerencias, así como por su cariño durante tanto tiempo que nos tocó compartir laboratorio. Muy en particular a Ceci, por tener siempre un minuto (o veinte) para escucharme hablar de mis problemas, de mis alegrías, de mis preocupaciones, en fin, de mí.

A mis compañeros: a los de inmuno, a los del IP, a los de Sevilla. Muy particularmente a Mariana, Laura, Hugo (*sí, estás en los agradecimientos!!!*), Inés, Cecilia, Martín, Fernanda, Leo, Vivian, Álvaro, Paula, Lucía, María, Maite, Valeria, Ana Lía, Mariana. A Briseida, María y María, Ana, Nando, Fran, Machupi. Todos ustedes fueron mucho más que compañeros de trabajo para mí. A mis abuelos que siempre se mostraron muy orgullosos de mí y de mi trabajo. A toda mi familia, que siempre me dio para adelante y creyó en mí aunque no entendiera de lo que les hablaba.

A mis ahijados pequeños Feli y Belu, que les tocó soportar ausencias de su madrina durante largas temporadas debido a viajes, a experimentos y a la escritura eterna de esta tesis. A sus padres, por aguantarme así como soy y seguir queriéndome...

A mis amigos por estar, muchos sin entender nada. Parece poco, pero es mucho. Tenerlos conmigo me hace fuerte.

A mis padres. Por ser grandes ejemplos para mí, por tenerme y darme confianza. Por enseñarme que siempre se puede dar un poquito más, por enseñarme a amar mi trabajo y a trabajar con amor y no por dinero. Por ser ejemplos académicos y de vida.

A mis hermanos y sobrino. Ustedes son todo para mí. Lo que más amo y quienes siempre van a estar. Gracias por acompañarme, por escuchar, por llorar conmigo, por buscar soluciones conmigo y para mí y sobre todo por hacerme reír más que nadie en este mundo.

Finalmente a Ed, no sólo por haber ayudado a que exista el capítulo 5 de esta tesis, también gracias por intentar enseñarme a programar, a usar office, a reconocer un *bug*. Gracias por estar a mi lado (incluso literalmente a mi lado mientras escribo esta tesis), por darme ánimos y fuerza, por ser tan honesto, por buscar la forma de acomodar tus proyectos con los míos. Gracias por cambiar mi vida, ito.

# INDICE

RESUMEN	3
AGRADECIMIENTOS	5
ABREVIATURAS	9
INTRODUCCIÓN	11
1.1 Selenocisteína, selenoproteínas y selenoproteomas	12
1.1.1 Selenocisteína: el vigesimoprimer aminoácido	12
1.1.2 Selenoproteínas y selenoproteomas	15
1.2 Incorporación de Sec a la cadena polipeptídica	16
1.2.1 Biosíntesis de Sec	17
1.2.2 Decodificación	18
1.2.3 Genes adicionales	21
1.3 Caenorhabditis elegans	23
1.3.1 El gusano modelo	23
1.3.2 Ciclo de vida	25
1.3.3 Anatomía	28
1.3.3.1 Hermafrodita	28
1.3.3.2 Macho	34
1.3.4 Genética directa y reversa	36
1.3.5 ARN de interferencia	38
1.3.6 Transgénesis	39
1.3.7 Nomenclatura	40
1.3.8 Otras ventajas	41
1.4 <i>C. elegans</i> y el linaje nematodo	42
1.5 ¿Por qué Sec en <i>C. elegans</i> ?	44
1.6 Antecedentes	45
OBJETIVOS	46
21 Objetivo general	/7
2.1. Objetivos específicos:	
2.2. Objetivos especificos	4/
GENERACION DE HERRAMIENTAS PARA EL ANALISIS DE GENES IN INCORPORACIÓN DE SEC	VOLUCRADOS EN LA
21 Curado do mutantos simplos	 /0
211 Caparagión y montonimiento do mochos	47
212 Curado do estiros mutantos	47
2.2 Encourse preliminares con los estimos mutantes en los	47
5.2 Ensayos preliminares con las estirpes mutantes en los (	genes presuntamente
asociados a la incorporación de Sec	33 
3.2.1 Ensayos de Akini con <i>gsr-1</i>	55
3.2.1.1 AKINI por alimentacion en <i>C. elegans</i>	55
3.2.1.2 Resultados	56
3.2.2 Ensayos de incorporación de "Se	56
3.3 Discusion	58

3.4	SBP2 da para más:	60	
AJUSTE NEMAT	S, EXTINCIÓN Y VESTIGIOS DE LA INCORPORACIÓN DE SELENOCISTEÍNA EN EL LIN ODO	AJE 61	
4.1	Primer anexo	76	
4.1.	1 Reviewer 1 Comments for the Author	76	
4.1.	2 Reviewer 2 Comments for the Author	79	
4.1.	3 Reviewer 3 Comments for the Author	83	
4.2	Segundo anexo al capítulo 4:	85	
4.3	Tercer anexo al capítulo 4	88	
4.3.	1 Generación de doble mutante <i>sbp2; rrf-3</i>	88	
4.3.	2 Ensayo de ARNi con <i>gsr-1</i> sobre la estirpe doble mutante generada	89	
4.3.	3 Intento de análisis del patrón de expresión de <i>scbp-2</i>	90	
4.3.	Intento de rescate de fenotipo de arresto larvario generado por el mutante er	n SBP2	
		91	
4.3.	Discusión sobre ensayos adicionales de SBP2	92	
4.3.	6 Análisis del genoma de <i>Panagrellus redivivus</i>	93	
INICIO	NO CANÓNICO DE LA TRADUCCIÓN EN C. elegans	95	
5.1.	Métodos	96	
5.1.	1. Programas para identificar candidatos a tener inicio no canónico de la traduc	cción 96	
5.1.	2. Análisis de genes candidatos a tener inicio no canónico de la traducción	99	
5.2.	Resultados	101	
5.2.	1. Transcriptos identificados y candidatos descartados	101	
5.2.	2. Análisis de F01D5.8	104	
5.3.	Discusión	106	
5.4.	Primer anexo al capítulo 5: Programas	110	
5.4.	1. 01codingCDSgeneFinder.py	110	
5.4.	2. 02patternFilter.py	112	
5.4.	<ol> <li>03concatenatePatternFoundWithCDS.py</li> </ol>	114	
DISCUS	IÓN Y PERSPECTIVAS	116	
6.1	Incorporación de Sec en nematodos	117	
6.2	SBP2 muy particulares	120	
6.3	Otros genes involucrados en la incorporación de Sec en C. elegans	122	
6.4	Inicio no canónico de la traducción en <i>C. elegans</i>	123	
CONC	USIONES	126	
REFERE	REFERENCIAS 12		
		137	
		10/	

# ABREVIATURAS

ADN	Ácido desoxiribonucleico
ADNc	ADN copia
ARN	Ácido ribonucleico
ARNi	Ensayo de ARN de interferencia
ARNm	ARN mensajero
ARNt	ARN de transferencia
ATP	Adenosin trifosfato
BLAST	Herramienta de búsqueda de alineamiento local básico (del inglés
	Basic Local Alignment Search Tool)
C. elegans	Caenorhabditis elegans
CDS	Secuencia de ADN codificante (del inglés <i>coding DNA sequence</i> )
CGC	Caenorhabditis Genetics Center
Cys	Cisteína
elF4a3	Factor de iniciación eucariota 4a3
EMS	Etilenmetanosulfano
IPTG	Isopropil-β-D-1-tiogalactopiranósido
IR	Cebador <i>inner right</i>
kb	kilo bases
L1	Estadio larvario 1
L2	Estadio larvario 2
L3	Estadio larvario 3
L4	Estadio larvario 4
Met	Metionina
NBRP	National Bioresource Project of Japan
OL	Cebador <i>outer left</i>
OR	Cebador <i>outer right</i>
ORF	Marco abierto de lectura (del inglés open reading frame)
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa (del inglés polymerase
	chain reaction)
RBD	Dominio de unaión al ARN (del inglés RNA-binding domain)
Sec	Selenocisteína
SecARNt <sup>[Ser]Sec</sup>	ARN de transferencia de selenocisteína
SECIS	elemento de incorporación de selenocisteína (del inglés
	selenocysteine incorporation sequence)
SBP2	Proteína de unión al elemento SECIS (del inglés <i>SECIS-binding protein</i> )

SID	Dominio de incorporación de Sec (del inglés Sec incorporation
	domain)
SNP	Polimorfismo de un único nucleótido (del inglés single nucleotide
	polymorphism)
SPS	Selenofosfato sintetasa
UTR	Región no traducida (del inglés untranslated region)

# INTRODUCCIÓN

# 1.1 Selenocisteína, selenoproteínas y selenoproteomas

## 1.1.1 Selenocisteína: el vigesimoprimer aminoácido

La selenocisteína (Sec) es un aminoácido análogo a la cisteína (Cys) que contiene un átomo de selenio (Se) en lugar de azufre (S) (Figura 1.1).



Selenocisteína (Sec) Cisteína (Cys)

Figura 1.1: Estructura de selenocisteína (Sec) y cisteína (Cys). Sec y Cys son aminoácidos análogos que difieren en el heteroátomo de su cadena lateral: Sec contiene selenio en lugar de azufre

Conocida como el vigesimoprimer aminoácido, Sec se encuentra genéticamente codificada por un codón UGA (usualmente un codón de finalización de la traducción) y una secuencia de inserción de Sec, denominada elemento SECIS, presente en el ARNm de las selenoproteínas (Figura 1.2). El elemento SECIS es el responsable de la reprogramación del codón UGA como Sec y permite expandir el código genético a 21 aminoácidos [1]. Su inserción en proteínas requiere de una serie factores adicionales, respecto a los otros típicos 20 aminoácidos. El mecanismo de inserción de Sec se explica más adelante.



**Figura 1.2: Esquema del ARNm de una selenoproteína.** Sec se encuentra genéticamente codificada por en codón UGA y un elemento SECIS. En eucariotas SECIS se localiza en el 3'UTR del ARNm.

Dependiendo del tipo de reacción catalizada y del rol de Sec en la misma, se han postulado una serie de explicaciones acerca de las posibles diferencias de reactividad de Sec y de Cys. En las reacciones en las cuales Sec interviene en el ataque, una de las ventajas que presenta es una mayor nucleofilia del selenolato como atacante respecto al tiolato (Figura 1.3 A). En el caso de las reacciones que involucran intercambios tiol(selenol)- disulfuro(selenil-sulfuro), la etapa resolutiva puede llevarse a cabo mediante una reacción de intercambio donde el tiol resolutivo ataque al selenio de la Sec y éste actúe como electrófilo, o que el tiol resolutivo ataque al azufre de la Cys contigua a la Sec, y que el selenio actúe como grupo saliente. Estas dos situaciones son excluyentes y en ambos Sec presenta ventajas sobre Cys en las dos, tanto por ser un mejor electrófilo (Figura 1.3 B) como por ser un mejor grupo saliente (Figura 1.3 C) [2]. Según el mecanismo por el cuál se desarrolle la reacción, tendrá más importancia la electrofilia (ataque al Se) o la capacidad como grupo saliente (ataque al S) del Se [3], además de otros factores que puedan influir como el impedimento estérico.



Figura 1.3 El selenio en reacciones de intercambio de tiol/disulfuro en comparación con azufre. La constante de velocidad de la reacción de intercambio se compone de tres constantes de velocidad individuales: la constante para el ataque nucleofílico ( $K_{Nuc}$ ), la constante que rige la velocidad a la que el átomo central (Xc) acepta electrones ( $K_{EI}$ ), y la constante que regula la velocidad a la cual el enlace  $S_c$ -X<sub>Lg</sub> se rompe ( $K_{Lg}$ ). La velocidad global de la reacción se puede aumentar mediante el aumento de cualquiera de estos tres parámetros: nucleofilia, electrofilia y la capacidad del grupo saliente. A) La velocidad de la reacción aumenta si X = Se, ya que la nucleofilia de un selenolato es mayor que la de un tiolato:  $k_1$ >  $k_2$ . B) La velocidad de la reacción aumenta si Xc = Se porque el selenio es más electrofílico de azufre:  $k_3$ >  $k_4$ . C) La velocidad de la reacción aumenta si X<sub>Lg</sub> = Se porque el selenol tiene un pKa inferior al del tiol y  $k_5$ >  $k_6$ . El Se también es mejor grupo saliente que el S porque es más polarizable que el S. Modificado de [2].

Otra característica que presenta Sec (llamada la paradoja del selenio) es que el Se puede ser oxidado (a ácido selenénico -RSeOH- o a ácido selenínico -RSeO<sub>2</sub><sup>-</sup>) más fácilmente que el S (a ácido sulfénico -RSOH- o a ácido sulfínico -RSO<sub>2</sub><sup>-</sup>-). A su vez, RSeOH y RSOH pueden ser reconvertidos a selenol (RSeH) y tiol (RSH) por la acción de otro tiol; pero también pueden continuar oxidándose hasta RSeO<sub>2</sub><sup>-</sup>y RSO<sub>2</sub><sup>-</sup>. Sin embargo, es extremadamente difícil reducir el RSO<sub>2</sub><sup>-</sup> nuevamente a RSOH, lo que lleva a la inactivación irreversible de la enzima; el RSeO<sub>2</sub><sup>-</sup>, en cambio, puede ser reducido de forma reversible a RSOH (y por tanto a RSeH). A su vez, el S puede ser sobreoxidado de forma irreversible a RSO<sub>3</sub><sup>-</sup>, mientras que el Se no se sobreoxida a RSeO<sub>3</sub><sup>-</sup>. Se ha sugerido que esta característica química puede ser una de las principales ventajas que presentan las enzimas que contienen Sec, respecto a las que contienen Cys [3]. Estos procesos se representan en la figura 1.4



**Figura 1.4 Explicación de la paradoja del Se.** RSeH se oxida a RSeOH con más facilidad que RSH a RSOH (k1>k4), además estos compuestos pueden continuar oxidándose a RSeO<sub>2</sub><sup>-</sup> y RSO<sub>2</sub><sup>-</sup>, respectivamente. Sin embargo, RSeO<sub>2</sub><sup>-</sup> se recicla nuevamente a su forma activa RSeH mucho más fácilmente que RSO<sub>2</sub><sup>-</sup> a RSH (k3>>k6); a la vez que RSO<sub>2</sub><sup>-</sup> puede reaccionar con otro equivalente de oxígeno de forma irreversible para dar RSO<sub>3</sub><sup>-</sup>, mientras que RSeO<sub>2</sub><sup>-</sup> no se sobreoxida [3]

## 1.1.2 Selenoproteínas y selenoproteomas

La mayor parte de las selenoproteínas de función conocida son enzimas óxidorreducatasas, donde Sec es parte del sitio redox activo. Al ser la Sec más reactiva que Cys, las selenoproteínas usualmente tienen superioridad catalítica respecto a sus homólogos que contienen cisteína [4]. A pesar de esta ventaja, el uso de Sec también puede estar sujeto a selección negativa debido a la poca abundancia de selenio en la corteza terrestre (1:1000 con respecto al azufre).

Las selenoproteínas pueden clasificarse arbitrariamente en tres grupos: i) las que contienen Sec hacia el extremo N-terminal (usualmente con plegamiento tipo tiorredoxina), ii) las que contienen Sec hacia el extremo C-terminal (hasta ahora sólo descritas en eucariotas) y iii) las que utilizan Sec para coordinar metales redox activos [5].

La mayoría de las selenoproteínas son enzimas y contienen un único residuo de Sec. La selenoproteína P (SeIP) es una excepción: no es una enzima, sino que funciona como transportador plasmático de Se en vertebrados y contiene entre 7 y 17 residuos de Sec, dependiendo de la especie [6].

Cada especie posee un conjunto discreto de proteínas que se denomina selenoproteoma. El tamaño de los selenoproteomas varía significativamente entre los eucariotas. Los organismos con mayor cantidad de selenoproteínas son acuáticos: peces y algas. Algunas algas son especialmente ricas en selenoproteínas: *Aureococcus anophagefferens* contiene al menos 56 genes que codifican para selenoproteínas [7], el mayor selenoproteoma eucariota descrito hasta el momento; mientras que los peces codifican para entre 30 y 37 selenoproteínas, dependiendo de la especie. La disponibilidad de un suministro constante de Se en el agua de mar podría haber hecho posible aumentar el uso de este elemento para diversas óxidorreductasas [8].

El cambio de hábitat de acuático a terrestre planteó un desafío para plantas y animales, tanto por la disminución en la disponibilidad de algunos elementos traza como por la exposición a mayores niveles de oxígeno. Como resultado muchos organismos terrestres perdieron sus selenoproteínas o remplazaron Sec por Cys [9]. El tamaño de los selenoproteomas de organismos terrestres varía significativamente.

15

El genoma humano es uno de los más ricos en selenoproteínas: codifica para 25 selenoproteínas; incluye tiorredoxina reductasas, glutatión peroxidasas, deiodinasas de la hormona tiroidea, metionina sulfóxido reductasa, selenofosfato sintetasa y selenoproteínas de función desconocida [10]. El selenio en cantidades traza es esencial para el ser humano y otros mamíferos, debido a la función que cumplen las selenoproteínas: los ratones *knockout* en el ARNt de Sec y en las selenoproteínas tiorredoxinas reductasas 1 y 2 (o tiorredoxina reductasa citosólica y mitocondrial, respectivamente) mueren durante la implantación embrionaria [11, 12]. El selenoproteoma animal más reducido es el de los gusanos *Caenorhabditis elegans y Caenorhabditis briggsae:* la maquinaria de biosíntesis e incorporación de selenocisteína está al servicio de una única selenoproteína, la tiorredoxina reductasa mitocondrial pero que codifica para Cys en la posición homóloga a Sec.

Si bien Sec se encuentra en los tres dominios de lo viviente: bacteria, archaea y eukarya, no todos los organismos incorporan Sec. Hasta el momento no se han identificado selenoproteínas ni su maquinaria de incorporación en plantas superiores ni en hongos. Dentro de los animales se han encontrado algunas especies de insectos como *Apis mellifera, Tribolium castaneum* y *Bombyx mori* que no incorporan Sec [15]. En estos organismos se encuentran proteínas homólogas que contienen Cys en lugar de Sec.

#### 1.2 Incorporación de Sec a la cadena polipeptídica

Como se mencionó previamente, el mecanismo de incorporación de Sec es diferente al de los otros 20 aminoácidos convencionales. En bacterias la biosíntesis y la incorporación de Sec fue dilucidada por el trabajo pionero del grupo de August Böck, que combinó enfoques y estrategias genéticas y bioquímicas para entender la recodificación de Sec en *Escherichia coli* [16]. En eucariotas este proceso es más complejo y si bien se ha avanzado en lo sustancial, aún restan varios aspectos por terminar de dilucidar. Este trabajo se centrará en el mecanismo de decodificación e incorporación de Sec en eucariotas.

#### 1.2.1 Biosíntesis de Sec

La biosíntesis de Sec se da enteramente en su ARNt (Sec-ARNt<sup>[Ser]Sec</sup>). Sec-ARNt<sup>[Ser]Sec</sup> ha sido descrita como una molécula clave [17] y el componente central de la biosíntesis de Sec [18]. En un primer paso la seril ARNt sintetasa (SerS) canónica aminoacila el ARNt<sup>[Ser]Sec</sup> en presencia de serina, ATP y Mg<sup>+2</sup>[19]. En un segundo paso la fosfoseril ARNt kinasa (PSTK) cataliza la transformación de Ser-ARNt<sup>[Ser]Sec</sup> en PSer-ARNt<sup>[Ser]Sec</sup> (Sep-ARNt<sup>[Ser]Sec</sup>) [20]. Finalmente la selenocisteína sintasa (SecS) cataliza la conversión de Sep-ARNt<sup>[Ser]Sec</sup> en Sec-ARNt<sup>[Ser]Sec</sup>, utilizando selefosfato (SeP) como precursor [21]. A su vez, SeP es generado en la célula a partir de selenuro y ATP por la selenofosfato sintetasa 2 (SPS2) [21]. Todos los pasos implicados en la biosíntesis de Sec se representan en la Figura 1.5.



Figura 1.5: Biosíntesis de Sec en eucariotas. Esquema que representa la biosíntesis de Sec sobre su propio ARNt y las enzimas involucradas (ver detalles y abreviaciones en el texto). Abreviaciones Ser: Serina, PSer: fosfoserina, Sep: selenofosfato, SerS: seril ARNt sintetasa, PSTK: fosfoseril ARNt kinasa, SPS2: selenofosfato sintetasa 2, SecS: Sec sintasa. Figura modificada de [22].

Cabe destacar que en mamíferos e insectos existen dos SPS. La SPS2, es indispensable para la incorporación de Sec y responsable de la conversión de selenuro en SeP. En algunos organismos SPS2 es una selenoproteína que se propone autorregularía del proceso de biosíntesis de Sec [23, 24]. SPS1 se encuentra presuntamente involucrada en la vía del reciclaje del selenio [25] y en la regulación de la síntesis de vitamina B6 [26]. Se ha postulado que, con excepción de los protozoarios, la presencia de SPS2 se correlaciona perfectamente con la presencia de linajes, selenoproteínas, embargo SPS1 distintos sin se encuentra en independientemente de su capacidad o no de incorporar Sec [27].

# 1.2.2 Decodificación

Sec es incorporada co-traduccionalmente a las selenoproteínas mediante un proceso complejo que requiere de otras proteínas actuando en trans, de Sec-ARNt<sup>[Ser]Sec</sup> y del elemento SECIS presente en el propio ARNm de la selenoproteína. Cuando el ribosoma se encuentra con un codón UGA en el ARNm de una selenoproteína, la maguinaria de incorporación de Sec interactúa con la maguinaria canónica de la traducción para aumentar el potencial codificador de los codones UGA y prevenir la finalización prematura de la traducción [22]. El elemento SECIS es el factor que redirige la recodificación de UGA como Sec [23, 26]. En presencia del elemento SECIS en el ARNm de la selenoproteína, Sec-ARNt<sup>[Ser]Sec</sup>, que tiene un anticodón complementario a UGA, traduce UGA como Sec. Se requiere de al menos otros dos factores que actúan en trans: la proteína de unión al elemento SECIS (SBP2\*) y el factor de elongación específico de Sec (EFSec). SBP2 se une de forma estable al elemento SECIS y recluta a EFSec que tiene unido al Sec-ARNt<sup>[Ser]Sec</sup>. Este complejo interactúa con el ribosoma, donde SBP2 es intercambiada con la proteína ribosomal L30. Esto lleva a cambios conformacionales en el elemento SECIS que son transmitidos a EFSec y que desencadenan la hidrólisis de GTP y la liberación de Sec-ARNt<sup>[Ser]Sec</sup>, incorporando Sec a la cadena polipeptídica en formación [28, 29]. Este proceso se representa en la figura 1.6.

\*Nota: Originalmente se describieron dos proteínas de unión al elemento SECIS [30, 31], llamadas SBP1 y SBP2. Estudios posteriores revelaron que únicamente SBP2 es la responsable de unir SECIS [32]. Para evitar conflictos en la literatura SBP2 mantuvo su nombre, aunque técnicamente no existe SBP1.



Figura 1.6 Representación gráfica del modelo del proceso de la recodificación de UGA en eucariotas. El elemento SECIS se une a la proteína de unión al elemento SECIS (SBP2), que recluta al factor de elongación específico de selenocisteína (EFSec) y el ARNt de selenocisteína (Sec-ARNt<sup>[Ser]Sec</sup>). El complejo interacciona con el ribosoma, intercambia SBP2 por la proteína ribosomal L30 y un cambio conformacional del SECIS lleva a la liberación de Sec-ARNt<sup>[Ser]Sec</sup> y a la hidrólisis de GTP. Modificado de [22].

Los elementos SECIS son estructuras, del tipo tallo bucle, presentes en el 3'UTR del ARNm de las selenoproteínas [33]. Están formados por dos hélices separadas por un bucle interno, un cuarteto GA (núcleo del SECIS) y un bucle apical [34] (Figura 1.7). El cuarteto GA se ubica en la base de la hélice 2 y se compone cuatro pares de bases con interacción no Watson-Crick, incluyendo dos pares de bases G·A/A·G en tándem, que son características de los motivos *kink-turn* [35, 36]. El cuarteto GA es el elemento funcional principal del SECIS y requerido para la interacción con SBP2. Se encontró que SBP2 se une a SECIS hacia el 5'UTR del bucle interno y la parte superior de la hélice I [37]. En algunos SECIS el bucle apical forma un pequeño tallo adicional o *"mini tallo"*, que se utiliza para clasificar a los SECIS en dos tipos. Los elementos SECIS cuyo bucle apical carece de mini tallo se clasifican como tipo 1 y los que lo tienen como tipo 2 [38, 39]. Para la incorporación de Sec se requiere también de un motivo AAR en la región apical del SECIS [40]. La función del motivo AAR no ha sido identificada aún.



Figura 1.7 Estructura de los elementos SECIS eucariotas. Los SECIS tipo I están formados por una hélice 1, un bucle interno, una hélice 2 y un bucle apical. Los tipo dos muestran un pequeño tallo adicional en el bucle apical. Se destacan el cuarteto GA en la base de la hélice 2 y el motivo AAR en el bucle apical. Modificado de [22].

SBP2 se compone de tres porciones diferentes: un módulo N-terminal, un dominio de incorporación de Sec (SID) y un dominio de unión al ribosoma (RBD) en el C-terminal (Figura 1.8). El módulo N-terminal no es necesario para la incorporación de Sec, incluso algunos organismos carecen del mismo, sugiriendo que puede cumplir un rol regulatorio [41, 42]. RBD pertenece a la familia L7Ae, que une ARN con motivos del tipo *kink-turn*, y es responsable de la unión de SBP2 a SECIS [43, 44]. Sin embargo, según los resultados de [45, 46] para que esta unión sea lo suficientemente estable como para mediar la incorporación de Sec a la cadena polipeptídica, requiere de la presencia de SID. Como SID no tiene motivos de unión al ARN se cree que SID puede fortalecer la unión RBD-SECIS. Estudios de [46-48] sugieren que SID sería esencial para la decodificación de Sec.



Figura 1.8 Esquema de las SBP2 de humano y *Drosophila*. Las SBP2 identificadas hasta el momento se componen de tres módulos: uno N-terminal de rol presuntamente regulatorio (presente en la de humanos, ausente en la de *Drosophila melanogaster*), un dominio de incorporación de Sec o SID (amarillo) y un dominio de unión al ARN o RBD (verde). Al lado de los esquemas se indica el largo de cada proteína.

Sec tiene su propio factor de elongación, EFSec, responsable de reclutar Sec-ARNt<sup>[Ser]Sec</sup> y, junto con SBP2, incorporar a Sec a la cadena polipeptídica en formación [38, 49]. Al igual que el factor de elongación canónico EF1A, involucrado en la incorporación de los otros 20 aminoácidos, EFSec tiene actividad GTPasa. A diferencia de EF1A, tiene alta afinidad por Sec-ARNt<sup>[Ser]Sec</sup> aunque no une Sep-ARNt<sup>[Ser]Sec</sup> ni otros ARNt [20, 49]. EFSec tiene una estructura tipo "cáliz", compuesta por cuatro dominios [50]. Los dominios I, II y III forman la copa del cáliz y son similares a EF1A. EFSec y EF1A se diferencian en la extensión C-terminal de EFSec, llamada dominio IV, que se ubica en la base del cáliz. Se ha propuesto que EFSec interacciona con SBP2 y con el brazo variable de ARNt<sup>[Ser]Sec</sup> a través del dominio IV [51, 52]. *In vitro*, se encontró que EfSec forma un complejo estable junto con SBP2 (al cual se une a través del RBD) y SECIS y que esta unión es estabilizada por Sec-ARNt<sup>[Ser]Sec</sup> [47, 52].

La proteína ribosomal L30 interactúa con el elemento SECIS en parte a través del mismo sitio de unión de la SBP2. L30 promueve la decodificación de Sec y compite con SBP2 en presencia de magnesio. Se cree que el magnesio induce un giro llamado *kink-turn* en la estructura del SECIS que aumenta la afinidad L30-SECIS, mientras atenúa la interacción SBP2-SECIS. Se ha propuesto que L30 desplaza a SBP2 del ARNm de la selenoproteína, por la inducción de un cambio conformacional en el elemento SECIS y la inmovilización del aparato de decodificación de Sec a la subunidad ribosómica grande. La disociación de SBP2 del SECIS liberaría Sec-ARNt<sup>[Ser]Sec</sup> en el sitio ribosomal A y promueve la hidrólisis de GTP, reciclando de este modo estos factores para la siguiente ronda de decodificación Sec [28, 53].

#### 1.2.3 Genes adicionales

Existen otros genes involucrados o posiblemente involucrados en la incorporación Sec, presentes en las especies que incorporan Sec pero ausentes en las que no incorporan [21]; aunque sus funciones precisas no están totalmente dilucidadas.

Uno de estos genes codifica para SecP43. Se postula que podría mediar varias interacciones proteína-proteína y proteína-ácidos nucleicos. *In vitro* e *in vivo*, SecP43 se asocia a SecS, a SPS2 y a Sec-ARNt<sup>[Ser]Sec</sup> y podría facilitar el tránsito núcleo-citosol de este complejo; en el núcleo SecP43 formaría otro complejo molecular diferente, con EF-Sec y SBP2, que permitiría limitar la degradación del ARNm debido a la presencia

de un codón de terminación en un marco abierto de lectura (fenómeno denominado *non-sense mediated decay*). Se encuentra involucrado también en la metilación de una base del anticodón del Sec-ARNt<sup>[Ser]Sec</sup>. [54, 55].

La nucleolina, una de las proteínas mayoritarias del núcleo, crítica para la biogénesis de ribosomas, se encuentra también involucrada en la regulación de la expresión de las selenoproteínas. Según los resultados de [56] la nucleolina parece ser un regulador positivo selectivo de la expresión selenoproteína, ya que se une a los ARNm de algunas selenoproteínas, incrementando sus niveles de expresión. El mecanismo por el cual este incremento ocurre no está claro y existen varias hipótesis al respecto. Una de ellas propone que, debido a su participación en interacciones proteína-proteína [57], podría reclutar directamente a los factores que intervienen en la incorporación Sec; también que podría facilitar el reclutamiento del complejo EFSec-Sec-ARNt<sup>[Ser]Sec</sup>, por su capacidad de interactuar con factores de elongación [58].

El factor de iniciación eucariota 4a3 (eIF4a3) tiene un rol opuesto al de la nucleolina: la interacción de eIF4a3 con el ARNm de las selenoproteínas impide la unión de SBP2, inhibiendo la síntesis de la selenoproteína. A su vez la expresión de eIF4a3 se encuentra regulada en respuesta al selenio, por lo que se propone que eIF4a3 vincula el nivel de selenio con la expresión diferencial de selenoproteínas, dando jerarquía para la expresión de determinadas selenoproteínas, a expensas de otras, en caso de deficiencia de selenio. [59]. eIF4a3 evita la unión de SBP2 al elemento SECIS al unirse él mismo a los bucles interno y apical y una uridina en el 3' del SECIS, lo que le da mayor afinidad por los SECIS de tipo I, que suele coincidir con las selenoproteínas categorizadas como no esenciales [60].

Adicionalmente, en vertebrados existe una proteína paráloga a SBP2, Ilamada SBP2L (del inglés *SECIS-binding protein like*). Se ha postulado que SBP2L sería una SBP2 ancestral ya que la SBP2 de invertebrados se encuentra más relacionada con ella que con la SBP2 de vertebrados [61]. A diferencia de SBP2, SBP2L no media la incorporación Sec y, si bien no se conoce su función, se postula que estaría implicada en la regulación postranscripcional de la expresión de las selenoproteínas [42].

Finalmente existen otros cuatro genes que se encuentran presentes en organismos que incorporan Sec, pero ausentes en todos los organismos que no incorporan: dos proteínas hipotéticas (Q5LJR2 y Q8T3V7), un presunto transportador de nitrato/nitrito

(Q7KUF9) y un inhibidor de ribonucleasa tipo-RI (Q9VEH4, los nombres de los genes corresponden a los números de acceso de *Drosophila melanogaster*). Se presume que estos genes estarían de alguna forma involucrados en la incorporación de Sec o en la regulación de su expresión pero sus funciones son aún desconocidas [21].

# 1.3 Caenorhabditis elegans

# 1.3.1 El gusano modelo

*C. elegans* es un nematodo de vida libre que habita en el suelo, sobre todo en frutas y raíces en descomposición donde se alimenta de bacterias, y que se encuentra distribuido a lo largo de casi todo el planeta [62, 63]. Este pequeño organismo (Figura 1.9) de poco más de un milímetro de largo fue propuesto como organismo modelo en la década de los 60 por Sydney Brenner por su simplicidad para realizar estudios genéticos, por su ciclo de vida corto con progenie numerosa, por el bajo costo y la sencillez para la manipulación experimental, además de contar con una anatomía simple y ser transparente [64].



Figura 1.9: El gusano modelo *C. elegans.* Imágenes de estereomicroscopio de organismos A) macho adulto y B) hermafrodita adulto. Modificado de WormAtlas

En el laboratorio se mantiene entre 16 y 25°C y se crece en placas de agar o en cultivo líquido, con *E. coli* como fuente de alimento; típicamente se utiliza la estirpe OP50, auxótrofa para uracilo. Por ser transparente en todos sus estadios, puede ser examinado vivo al microscopio a nivel celular. Se ha completado la descripción anatómica de todo el animal a nivel de microscopía electrónica y se ha establecido su linaje celular completo, el cual es invariante entre organismos (Figura 1.10) [65-69].



**Figura 1.10 Linaje celular completo de** *C. elegans.* El cigoto a partir de un linaje celular invariante genera todos los tejidos somáticos. Además dos células madre dan lugar a la línea germinal (marcadas en rojo). Las células de la línea germinal se divide de forma variable durante el desarrollo larvario para dar lugar al tejido de la línea germinal, el cual es mantenido en el adulto. Modificado de www.stembook.org

*C. elegans* es un organismo androdioico: existen hermafroditas y machos. Los hermafroditas tienen 959 células somáticas y los machos 1031 y pueden diferenciarse de forma sencilla a la lupa cuando son adultos. Los hermafroditas han sido definidos como hembras capaces de autofecundarse cuyo único caracter masculino es la habilidad de producir una cantidad limitada de esperma con el solo fin de la autofertilización interna. Esta "hembra modificada" es, por lo tanto, capaz de autorreproducirse en ausencia de machos. Sin embargo, en caso de existir la posibilidad, copulan con machos para utilizar su esperma de forma preferente con el fin de generar mayor progenie [70]. La autofecundación de los hermafroditas permite a los organismos homocigotas generar progenie genéticamente idéntica (exceptuando

la diversidad derivada de la mutación germinal y la meiosis); el cruce con machos facilita el aislamiento y el mantenimiento de estirpes así como mover mutaciones entre estirpes [71].

Ambos poseen 5 pares de autosomas, los hermafroditas poseen además dos cromosomas sexuales X (XX) y los machos únicamente un cromosoma sexual X (X0). Los machos ocurren naturalmente por errores en la disyunción del cromosoma X durante la meiosis, con una tasa de 0.1% [72].

# 1.3.2 Ciclo de vida

El ciclo de vida de este nematodo en condiciones de laboratorio a 20°C es de unos tres días y medio. En ese período atraviesa un estadio embrionario, cuatro estadios larvarios (L1-L4) y alcanza la adultez (Figuras 1.11 y 1.12). Cada estadio larvario finaliza con una muda, durante la cual sintetiza una nueva cutícula y se despoja de la vieja. Durante este período el bombeo de la faringe cesa y el animal entra en una fase de letargo [71].

El gusano adulto es fértil durante cuatro días y luego vive entre 10 y 15 días adicionales. Durante el período de fertilidad puede generar por sí mismo hasta unos 300 embriones, siendo la limitante la cantidad de esperma producido. Como se mencionó previamente, en caso de aparearse con un macho el número de su progenie puede aumentar.



Figura 1.11: Esquema del ciclo de vida de *C. elegans* a 22°C. Se representan las distintas fases embrionarias y larvarias por las que pasa un gusano hasta llegar a adulto. Los números azules sobre las flechas indican el tiempo que el nematodo está en determinado estadio antes de mudar al siguiente. La primera división celular ocurre aproximadamente 40 minutos después de la fecundación. La puesta del huevo se da unos 150 minutos después de la fertilización, durante la fase de gástrula cuando cuenta con 30 células. Modificado de WormAtlas.



Figura 1.12: Estadios larvales del desarrollo. Imágenes DIC de embrión, de cada uno de los estadios larval y adultos hermafrodita y macho. Modificado de WormBook.

Hacia el final del estadio L1 si las condiciones ambientales no son favorables para su crecimiento, puede entrar en un estado de arresto y resistencia llamado larva dauer. Factores ambientales como la presencia de feromonas que indican alta densidad de población, ausencia de comida o altas temperaturas actúan como señales que pueden llevar a la formación de una larva L2 particular llamada L2d. La larva L2d mantiene el potencial de formar tanto una larva dauer como L3, dependiendo de la persistencia de las condiciones ambientales adversas. Si las condiciones desfavorables se mantienen, L2d muda a larva dauer. El tiempo de permanencia en este estadio no afecta la longevidad postdauer, pero en caso que las condiciones ambientales no se tornen favorables, mueren al cabo de unos meses. Durante este estadio la alimentación está arrestada de forma indefinida y la locomoción marcadamente reducida. El estadio dauer finaliza cuando experimenta condiciones favorables: luego de una hora con acceso a alimento sale de dauer, luego de 2-3 horas comienza a alimentarse y después de 10 horas muda a L4. Morfológicamente, las larvas dauer son finas con una cutícula gruesa. La cavidad bucal es sellada por un bloque cuticular, las células del intestino adquieren apariencia oscura, el lumen de faringe e intestino se encoge y la gónada se arresta en L2 [73].

#### 1.3.3 Anatomía

#### 1.3.3.1 Hermafrodita

*C. elegans* es un nematodo de cuerpo cilíndrico que se estrecha en los extremos, con la típica organización de los nematodos compuesto por un tubo exterior y uno interior separados por el pseudoceloma. El tubo exterior (pared del cuerpo) se compone de la cutícula, hipodermis, sistema excretor, neuronas y músculos; y el tubo interior comprende la faringe, intestino, y, en el adulto, las gónadas y el aparato ovopositor. Todos estos tejidos están bajo una presión hidrostática interna, regulada por un sistema osmorregulatorio. Las principales estructuras anatómicas se representan en la figura 1.13. [73]

Una cutícula principalmente compuesta por colágeno, secretada por el epitelio subyacente (hipodermis), rodea el gusano en el exterior y también recubre la faringe y el recto. Varios tejidos se abren hacia el exterior a través de la cutícula. El poro excretor se encuentra en la línea media en el lado ventral de la cabeza. La vulva es otra gran

abertura en la parte ventral del cuerpo central, y el ano forma otra abertura ventral, justo antes de la cola.



Figura 1.13: *C. elegans* hermafrodita adulto. A) Imagen DIC de un hermafrodita adulto, vista lateral; barra de escala 0,1 mm. B) Esquema representando las principales estructuras anatómicas, vista lateral. C) Esquema representando las principales estructuras anatómicas, corte transversal del cuerpo. Modificada de WormAtlas

Las células hipodérmicas (Figura 1.14) son las que secretan la cutícula. Existen once células hipodérmicas: cinco células polinucleadas concéntricas en la cabeza (hyp 1 a hyp 5), dos polinucleadas en el cuerpo (hyp 6 e hyp 7, que se fusionan en el estadio L3) y cuatro en la cola, tres mononucleadas y una polinucleada. En el adulto hyp 7 constituye el sincicio del organismo. En los lados laterales, la hipodermis se interrumpe por la fila sincicial de las células de la costura que forman el *alae* (un conjunto de crestas cuticulares que sobresalen y se extienden longitudinalmente a lo largo de los dos lados del animal sobre las células de costura) en la superficie de la cutícula durante ciertas etapas de desarrollo. La hipodermis y los tejidos internos que se abren hacia el exterior están conectados entre sí por las células interfaciales especializadas [74].



**Figura 1.14. Hipodermis en L1.** Esquema representando la vista lateral izquierda de una larva L1. En este estadio hyp 7 hace dos anillos completos que envuelven la parte posterior de la cabeza y la región postanal. Las células hyp 1 a 5 se localizan en la cabeza, mientras que las células hyp 8 a 11 se localizan en la cola. Los círculos oscuros representan la posición de los núcleos. Modificada de WormAtlas

El sistema nervioso es el sistema más complejo de *C. elegans*: 302 de las 959 células somáticas son neuronas (Figura 1.15). Éstas se organizan distribuidas principalmente en ganglios en la cabeza y en la cola. La mayoría de las neuronas de *C. elegans* se encuentran en la cabeza alrededor de la faringe. En el cuerpo, una fila continua de los cuerpos celulares de neuronas se encuentra en la línea media, adyacente a la hipodermis ventral. Además, hay dos pequeños ganglios laterales posteriores en los lados, así como algunas neuronas dispersas a lo largo del lateral del cuerpo. Los procesos de la mayoría de las neuronas viajan ya sea en el cordón nervioso ventral o dorsal y proyectan hacia el anillo nervioso en la cabeza que constituye el principal neuropilo del animal. La mayor parte de las neuronas se desarrollan durante la embriogénesis, pero 80 neuronas (sobre todo motoneuronas) se desarrollan en fases

post-embrionarias. La estructura del sistema nervioso ha sido descrita en detalle por [75] mediante reconstrucción por microscopía electrónica. La alta resolución de estas imágenes permitieron a White y colaboradores identificar todas las sinapsis (unas 5000 sinapsis químicas, 2000 junciones neuro-musculares y unas 600 junciones gap), mapear todas las conexiones y descifrar todos los circuitos neuronales [76].



Figura 1.15. Sistema nervioso de *C. elegans.* Identificación de los somas de las neuronas por marcado con proteína fluorescente amarilla (foto invertida) mediante la expresión de *rab-3::YFP*. Se muestra el ganglio de la cabeza, el conrdón ventral y el ganglio de la cola. Tomada del laboratorio del Dr. Oliver Hober [http://hobertlab.org/]

Las neuronas y la hipodermis se separan de la musculatura por una lámina basal delgada. Los músculos reciben el aporte de las neuronas mediante proyecciones musculares a los procesos de las neuronas motoras que se encuentran a lo largo de los cordones nerviosos o del anillo nervioso. Los músculos de la pared del cuerpo oblicuamente estriados están dispuestos en bandas en cuatro cuadrantes, dos dorsales y dos ventrales, a lo largo de toda la longitud del animal (Figura 1.16). Los músculos no estriados, más pequeños, se encuentran en la faringe y alrededor de la vulva, el intestino y el recto [71].



Figura 1.16 Organización somática de los músculos de *C. elegans* hermafrodita adulto. Identificación de las células musculares por marcado con proteína fluorescente verde mediante la expresión de *unc-27::GFP* (marcador de músculos). Los músculos de la pared del cuerpo se organizan en cuatro cuadrantes, en esta foto se muestran los dos cuadrantes dorsales y la línea media dorsal que divide a los cuadrantes. La barra representa 50 µm. Modificada de WormAtlas

Cuatro células situadas en el lado ventral de la parte posterior de la cabeza forman el sistema excretor, que funciona en la osmorregulación y la eliminación de residuos. El sistema excretor se abre al exterior a través del poro excretor [71].

Tres pares de celomocitos, situados en la cavidad pseudocelómica funcionan como células barrenderas que endocitan fluido del pseudoceloma y se proponen como un sistema inmunológico primitivo en *C. elegans* [71].

*C. elegans* se alimenta a través de una faringe bilobulada, que es casi un órgano autónomo con su propio sistema neuronal, músculos y epitelio. La faringe se separa de los tejidos del tubo exterior y del pseudoceloma por su propia lámina basal. El lumen de la faringe es continuo con el lumen del intestino; la faringe pasa los alimentos, que toma del suelo, hacia el intestino a través de la válvula faríngeointestinal. El tejido que conforma al intestino, que es el único tejido somático derivado de un único linaje, está hecho de 20 células dispuestas para formar un tubo con el lumen central. Las superficies apicales de las células intestinales tienen numerosas microvellosidades. El contenido intestinal de desecho se excreta al exterior a través de una válvula rectal que conecta el intestino con el recto y el ano. Los cuatro músculos entéricos que contribuyen a la defecación se encuentran alrededor del recto y del intestino posterior [77] (Figura 1.17)



**Figura 1.17. Sistema alimentario de** *C. elegans.* Esquema representado los principales componentes del sistema alimentario de un hermafrodita adulto: cavidad bucal, faringe, válvula faríngeo-intestinal, intestino, recto y ano. Modificado de WormAtlas.

El aparato reproductor consiste en la gónada somática, la línea germinal y el aparato de la ovopositor. En el hermafrodita hay dos gónadas en forma de U, bilateralmente simétricas, que están conectadas a un útero central a través de la espermateca. La línea germinal, localizada en los brazos distales de los ovarios, es sincicial con núcleos de la línea germinal rodeados de un citoplasma central. A medida que se acercan a la zona proximal, las células germinales pasan secuencialmente a través de las etapas de mitosis, profase meiótica y diacinesis. En la curva del brazo de la gónada (oviducto), los oocitos se agrandan, se desprenden del sincicio, y maduran a medida que se mueven hacia la zona más proximal. Los espermatozoides son producidos por la línea germinal de las gónadas (150 espermatozoides cada una) previo a la generación de los ovocitos, durante la fase larvaria L4, y se almacenan en la espermateca. Los oocitos son fecundados al atravesar la espermateca y los cigotos diploides resultantes se almacenan en el útero y luego son expulsados a través de la vulva, que sobresale en la línea media ventral [78] (Figura 1.18).



Figura 1.18. Sistema reproductor de *C. elegans.* Arriba, vista lateral del hermafrodita adulto, mostrando la ubicación del sistema reproductor en el animal. El sistema reproductor es bilateralmente simétrico y consta de dos gónadas en forma de U, unidas a un útero común. Debajo, representación ampliada de la mitad del sistema reproductor. En la porción más distal de la gónada se encuentra el sincicio de la línea germinal, cuyos núcleos van madurando a oocitos a medida que se aproximan a la porción proximal. La espermateca (Sp) aloja al esperma en su lumen y es dónde se da la fecundación de los oocitos. Los cigotos resultantes de la fecundación atraviesan la válvula espermateca-útero y se alojan en el útero hasta ser puestos a través de la vulva. Modificado de WormAtlas.

#### 1.3.3.2 Macho

Las larvas macho de *C. elegans* inicialmente muestran la misma forma cilíndrica simple de los hermafroditas, pero a partir del estadio L2, la forma de la parte posterior cambia y sus órganos sexuales comienzan a desarrollarse. Con la excepción de tal vez la faringe y el sistema excretor, prácticamente todos los tejidos muestran algún grado de dimorfismo sexual, aunque no sean fácilmente perceptibles. Las diferencias más importantes se ven en los tejidos de la parte posterior, que forma el aparato copulador masculino [71] (Figura 1.19).



**Figura 1.19: Macho** *C. elegans.* A) Esquema representando las principales estructuras anatómicas de un macho, vista lateral. B) Imagen DIC de un macho adulto, vista lateral, la barra de escala representa 0,1mm. C) La gónada distal magnificada del gusano en B. D) Cola del macho adulto, vista ventral: la flecha muestra la cloaca, la punta de flecha el abanico, los asteriscos muestran los rayos 1-9 del lado derecho. E) Cola del macho L3 (abajo) y L4 (arriba): el extremo comienza a abultarse en L3 y la hipodermis de la cola se retrae en L4. Modificado de Wormatlas.

Con respecto al hermafrodita, el sistema muscular del macho contiene 41 células musculares sexo-específicos adicionales y el sistema nervioso tiene 89 neuronas adicionales que incluyen varios tipos de sensillias de la cola. El aparato reproductivo consiste en una sola gónada que se abre al exterior en la cloaca (ano) a través de una cámara epitelial rectal modificada llamado el proctodeo. El proctodeo incluye dos espículas sensoriales escleróticas utilizadas por el macho durante el apareamiento para localizar la vulva del hermafrodita, y mantenerla abierta durante la transferencia de esperma. El aparato copulador, ubicado en la cola del gusano, se compone de una extensión lateral de la cutícula llamada abanico donde se ubican nueve pares bilaterales de rayos sensoriales, el gancho y la sensilla post-cloacal (ubicados a cada lado de la cloaca) y las espículas [79].

#### 1.3.4 Genética directa y reversa

Buena parte del atractivo inicial de *C. elegans* como organismo experimental, se debió a los poderosos métodos que podían ser desarrollados para realizar ensayos genéticos. Su popularidad aumentó en 1998 cuando se convirtió en el primer organismo multicelular (y el segundo eucariota, luego de *Saccharomyces cerevisiae*) en tener su genoma completamente secuenciado. La anotación de genoma de unas 100 Mb es de las más completas entre los organismos eucariotas [80].

Una gran ventaja que presenta *C. elegans*, desde el punto de vista de análisis genético, es la facilidad para generar mutaciones. Muchos compuestos químicos, en particular etilmetanosulfonato (EMS, un agente alquilante que induce modificaciones químicas en los nucleótidos que lleva a desapareamientos y cambios en las bases nucleotídicas), son muy efectivos agentes mutagénicos para este gusano. También es posible generar mutantes mediante el uso de radiación ionizante o de transposones. Al ser un organismo diploide muchas mutaciones pueden ser inducidas y propagadas sin matar al animal. A su vez, como su principal forma de reproducción es mediante autofecundación de animales hermafroditas, los efectos de una mutación en homocigosis pueden estudiarse automáticamente como resultado de la segregación mendeliana [80].

Recientemente también se ha comenzado a generar mutantes mediante CRISPR-Cas9 (del inglés *clustered, regularly interspaced, short palindromic repeats-associated endonuclease Cas9*) que permite generar modificaciones sitio-específicas (tanto deleciones como inserciones, mutaciones puntuales, fusiones a GFP, entre otras). Esta técnica implica la microinyección de un plásmido que codifica para Cas9 (endonucleasa responsable del clivaje del ADN) y otro para el ARN guía que dará lugar a la modificación deseada en el sitio del genoma especificado por la secuencia guía, y la posterior identificación de los gusanos en los cuales se dio el evento de corte y recombinación (por PCR y/o por coinyección de un marcador) [81-83]. A pesar del uso creciente de esta metodología, una de la técnica más utilizadas para la generación de estirpes mutantes en un gen específico, es aún mutagenizar el genoma de forma aleatoria (mediante EMS, por ejemplo) y luego analizar si se mutagenizó el gen de interés mediante el método de reacción de polimerasa en cadena (PCR) (aunque esta técnica está siendo desplazada por CRISPR-Cas9). [84]. Dos consorcios, el *C. elegans*
Knockout Consortium y el *C. elegans* National Bioresource Project of Japan (NBRP) buscan generar mutantes en cada uno de los genes de *C. elegans*. Entre los dos habían aislado e identificado, para 2012, más de 6.700 alelos mutantes de deleción, los cuales están disponibles para la comunidad científica. Las estirpes provenientes de del Knockout Consortium se encuentran disponibles a través del *Caenorhabditis* Genetics Center (CGC) y sus alelos se identifican con los prefijos *ok* o *gk*. Las estirpes generadas por el NBRP se designan con el prefijo *tm* y están disponibles luego del envío de un acuerdo de transferencia de material [85]. Una vez que se recibe una estirpe mutante, la misma debe ser cruzada con la estirpe de referencia N2 para remover mutaciones de fondo no deseadas y siempre conviene verificar por secuenciación la presencia de la mutación en la estirpe [86].

Los estudios genéticos implican el análisis de la función de determinados productos génicos al estudiar organismos a los cuales se les ha alterado esos genes. En los estudios convencionales de genética directa los organismos típicamente se tratan con agentes mutagénicos para inducir lesiones en el ADN y luego se buscan los mutantes que muestran el fenotipo de interés. Una vez que se identifican mutantes con el fenotipo deseado se identifica al gen mutado mediante técnicas de biología molecular. La técnica de preferencia para el mapeo de mutaciones ha sido la de mapeo a través de polimorfismos de un único nucleótido (SNPs, del inglés single nucleotide polymorphisms), que se basa en utilizar los SNPs que hay entre dos estirpes de C. *elegans* (típicamente las estirpes silvestres N2 y CB4856 proveniente de Hawaii) como marcadores genéticos [87]. La técnica de ARN de interferencia (ARNi) también puede usarse a escala global como herramienta de genética directa, silenciando la expresión de todos los genes para inducir o rastrear el fenotipo de interés. Estudios pormenorizados del fenotipo mutante junto con el mapeo del gen permite la elucidación de los procesos en los que puede estar involucrado dicho gen y la función del producto para el cual codifica. La genética directa ha sido responsable de la elucidación de muchos procesos biológicos y es un excelente método para identificar genes que participan de procesos particulares.

En genética reversa, típicamente luego de la mutagénesis se buscan los mutantes en el gen de interés y luego se analiza el fenotipo (es decir, se rastrean organismos que llevan una mutación en lugar de buscar los organismos que manifiestan un determinado fenotipo). Utilizando diferentes técnicas, la función de un gen puede ser alterada para luego analizar el efecto en el desarrollo o comportamiento del organismo. La genética reversa es un complemento importante de la genética directa. Por ejemplo, mediante aproximaciones de genética reversa se puede investigar la función de todos los genes de una familia de genes, algo difícil de hacer mediante genética directa. Además, se puede estudiar el rol de un gen involucrado en un proceso de interés en otro organismo, para el cual aún no se hayan identificado mutantes. La disponibilidad de las secuencias completas del genoma combinada con herramientas de genética reversa puede permitir el estudio de todos los genes.

Existen principalmente dos métodos para realizar ensayos de genética reversa en *C. elegans:* ARNi y la creación de mutantes de deleción (mediante mutagénesis aleatoria o dirigida). Ambas técnicas pueden aplicarse para el estudio de genes individuales. Mediante ARNi se obtiene en pocos días organismos con la expresión de un gen disminuida (*knockdown*) sin afectar su ADN. En contraste, llevar a cabo la deleción de un gen (o parte de un gen), mediante mutagénesis aleatoria, lleva al menos un mes de trabajo y aunque puede llegar a remover de forma permanente toda función del gen. La decisión de qué técnica utilizar depende de la naturaleza de los experimentos que se desean realizar. Las técnicas también pueden ser combinadas, utilizando primero ARNi para un análisis rápido de pérdida de función y luego generando mutantes de deleción para estudiar los genes de particular interés [88].

#### 1.3.5 ARN de interferencia

Como se mencionó previamente, los ensayos de ARNi se utilizan tanto para estudios de genética directa como para genética reversa. En el primer caso, utilizando bibliotecas de ARNi, se silencian individualmente una colección de genes hasta encontrar los organismos que manifiestan el fenotipo de interés y luego se ve cuál fue el gen cuya interferencia en la expresión es responsable de dicho fenotipo. En el segundo caso se silencia la expresión de un gen de interés y luego se estudian los posibles fenotipos que ese silenciamiento puede dar.

Para llevar a cabo los ensayos de ARNi existen tres opciones: inyectar en jóvenes adultos ARNdh producido in vitro [89], embeber a los gusanos en una solución conteniendo ARNdh [90] o alimentar a los gusanos con bacterias que expresen ARNdh de la secuencia que se desea silenciar [91]. La sensibilidad al ARNi varía entre estirpes

y tejidos. Dependiendo del experimento que se desee realizar se pueden utilizar estirpes supersensibles al ARNi como son los mutantes en *rrf-3* [92] o en *eri-1* [93]. Estos mutantes pueden revelar fenotipos que pasen desapercibidos en estirpes silvestres ya que esos genes se encuentran involucrados en la inhibición y regulación negativa del ARNi.

Actualmente existen dos bibliotecas de ARNi para *C. elegans* para el método de alimentación. Una es la del laboratorio de la Dra. Ahringer, hecha a partir del clonado de fragmentos genómicos gen específico entre dos promotores de T7 [94, 95]. Los insertos contienen intrones y exones y sus tamaños varían entre los 500 pb y 2.5 kb; actualmente esta biblioteca tiene clonas para todos los genes. La otra biblioteca fue generada en el laboratorio del Dr. Vidal, que contiene 11.511 clonas, hechas por *Gateway cloning* de marcos abiertos de lectura (ORF) de largo completo de ADNc en vector con doble T7 [96]. Ambas bibliotecas utilizan la estirpe bacteriana HT115 de *E. coli* transformada con el plásmido de la clona de ARNi. HT115 posee polimerasa T7 inducible por IPTG y una disrupción del gen de RNAsa III (una RNAsa específica de ARNdh), esta última marcada con resistencia a tetraciclina [97]. Existe cierto solapamiento entre las dos bibliotecas y entre ellas cubren aproximadamente el 94% de los genes de *C. elegans*.

#### 1.3.6 Transgénesis

Una ventaja importante que presenta este organismo es la capacidad de generar estirpes transgénicas de forma relativamente sencilla. Los transgénicos se suelen generar para rescatar mutantes, sobreexpresar genes endógenos o exógenos, para expresar proteínas reporteras, estudiar estructura, función o localización de proteínas o dominios proteicos y para estudiar elementos regulatorios de ADN o ARN. El método más habitual para esto es la microinyección de ADN en la zona distal de la gónada (Figura 1.20). Esta técnica genera grandes arreglos extracromosomales de ADN que se heredan en tasas constantes (aunque no necesariamente toda la progenie hereda estos arreglos). Existen otras técnicas más complejas para generar transgénicos estables como el bombardeo de micropartículas de oro con ADN [98] o la generación de líneas trasgénicas estables mediante CRISPR-Cas9 o MosSCI (del inglés *Mos1*-

*mediated Single Copy Insertion*) que inserta un única copia de un transgen en un sitio definido [99].



Figura 1.20: Microinyección en la gónada de *C. elegans.* Se representa la posición óptima de inyección de la aguja en el centro citoplasmática de la línea germinal distal. La solución de inyección debe fluir en ambas direcciones tanto a través de la línea germinal distal y proximal (flechas). Tomado de WormBook

## 1.3.7 Nomenclatura

La comunidad de científicos que trabaja en *C. elegans* adoptó en a finales de los años 70 un sistema de nomenclatura para que resulte no ambiguo [100]. Esta convención ha sido complementada posteriormente por la WormBase, teniendo en cuenta los avances en ingeniería genómica [101]. Esta convención vuelve la nomenclatura inequívoca y rígida: así los nombres de los genes en *C. elegans* pueden diferir a los nombres en otros organismos de acuerdo a otras convenciones (HUGO Gene Nomemenclature Committee, por ejemplo) y con nombres "convencionales" de los genes.

En el anexo de nomenclatura se detallan las guías de nomenclatura establecidas. A modo de resumen, se comenta que los genes se nombran en minúscula, cursiva con entre tres y cuatro letras seguidas de un guión y un número; los alelos entre paréntesis luego del nombre del gen, también en minúscula cursiva con una a tres letras que designan al laboratorio que lo generó o identificó y un número; y las estirpes con dos o tres letras en mayúscula que identifican al laboratorio que la generó (diferentes de las de identificación de alelos) y un número. Las proteínas llevan el nombre del gen que las codifican pero se escriben en mayúscula y sin cursiva.

### 1.3.8 Otras ventajas

Además de las ventajas ya presentadas, *C. elegans* posee otras ventajas que hacen de este organismo un modelo ampliamente utilizado a nivel mundial (existen aproximadamente unos cuatro mil laboratorios que trabajan con *C. elegans* registrados). Por ejemplo, las estirpes pueden mantenerse congeladas a -80°C por largos períodos de tiempo. La formación de larva dauer, que puede tolerar condiciones ambientales adversas le permite sobrevivir hasta ocho veces su tasa de sobrevida normal de unas tres semanas, es otra gran ventaja [102].

A pesar de su anatomía simple, este pequeño animal muestra un gran repertorio de comportamientos como locomoción, forrajeo, alimentación, defecación, puesta de huevos, formación de larva dauer, respuestas sensoriales al tacto, a olores y sabores y a temperatura; y algunos comportamientos complejos como apareamiento entre hermafroditas y machos, comportamiento social, aprendizaje y memoria [103, 104].

# 1.4 *C. elegans* y el linaje nematodo

El filo *Nematoda* es uno de los más diversos, no sólo en morfología, sino también en tamaño (que varía de menos de un milímetro a más de seis metros), ciclo de vida (desde partenogéneticos a ciclos complejos que alternan sus estrategias sexuales), ecología (existen organismos de vida libre así como parásitos de prácticamente todos los organismos multicelulares), entre otros aspectos. Aunque sólo se han descrito unas 23.000 especies de nematodos, las estimaciones actuales calculan que deben existir más de un millón de especies de nematodos en la Tierra [105].

Las relaciones filogenéticas aceptadas de los nematodos han ido cambiando. Su clasificación según los datos genómicos muestran tres grandes ramas: Enolia, Dorylaima y Chromadoria [106, 107] como se representa en la figura 1.21. Chromadoria se subdivide en tres clados: clado III o Spirurina compuesto por parásitos de animales como *Brugia malayi*, clado IV o Tylenchina que incluye importantes parásitos de plantas como *Meloidogyne hapla*, parásitos de animales y gusanos de vida libre; y clado V o Rhabditina al que pertenece *C. elegans*, junto con otros nematodos de vida libre y parásitos de animales. En Dorylaimia están especies depredadoras terrestres que juegan un papel clave en las cadenas alimentarias y parásitos de insectos y mamíferos. Enoplia se compone mayormente por nematodos marinos, que incluyen predadores y microvíboros y un grupo de herbívoros terrestres [105].



Figura 1.21: Filogenia de nematodos. Representación filogenética de los nematodos basada en el análisis del gen de la subunidad pequeña del ribosoma. Se muestran los nombres sistemáticos dados por De Ley y Blaxter [106, 107] y la convención de clados [108]. Se muestra también, debajo de las ramas, la clasificación numérica según Helder y colaboradores [109]. Los modos de alimentación, parasitismo de animales o plantas y asociación con vectores se representan con un ícono y se nombran especies representativas de algunos grupos. Las especies con genoma secuenciado se marcan con un asterisco Tomado de [105].

# 1.5 ¿Por qué Sec en *C. elegans*?

Una peculiaridad interesante de *C. elegans* es que su presunta SBP2 (K04G2.11/ SCBP-2 según la versión de la WormBase; identificada por el grupo de trabajo previo al inicio de esta tesis) es significativamente más corta que la de los otros eucariotas (85 aminoácidos, en tanto la SBP2 más corta de cualquier otro eucariota es de aproximadamente 400 residuos). Esta probable SBP2 carecería de SID y no completaría el RBD (Figura 1.21), lo que podría llevar a revisar cuáles son los requerimientos mínimos para la unión SBP2-SECIS y para su funcionalidad; dejando entrever que existe la posibilidad de que la maquinaria de decodificación de Sec sea ligeramente diferente en *C. elegans*. Por otra parte, *C. elegans* posee dos de los genes presentes en organismos que incorporan Sec, y ausentes en los que no lo incorporan (ver antecedentes).



Figura 1.21 Esquema comparativo de las SBP2 de humano y *Drosophila* y de la SBP2 identificada de *C. elegans.* La SBP2 de *C. elegans* identificada previo al inicio de esta tesis codificaría para una proteína de 85 aminoácidos sin completar el RBD y careciendo de SID

A su vez, de forma interesante y como se mencionó previamente, este nematodo posee toda la maquinaria de incorporación de Sec dedicada a un único residuo en todo su proteoma, ubicado en el sitio activo de la tiorredoxina reductasa (TRXR-1) [13, 48]. *trxr-1* se expresa en hipodermis, faringe y en algunas células del sistema nervioso [110]. No es esencial en condiciones de laboratorio, pero es necesaria para la muda cuando la función de la glutatión reductasa (*gsr-1*) está comprometida, ya que al interferir la expresión de *gsr-1* en estirpes mutantes en *trxr-1* o EFSec (*efsc-1* en *C. elegans*), las mismas manifiestan fenotipo de arresto larvario [110]. *C. elegans* también codifica también para TRXR-2, una tiorredoxina reductasa mitocondrial que posee Cys en la posición homóloga a la de Sec en el sitio activo.

Finalmente, la evolución de la incorporación de Sec no ha sido exhaustivamente estudiada en eucariotas. La evolución de su utilización ha sido caracterizada en vertebrados [8] y se ha estudiado detalladamente en artrópodos, un linaje que tiene varias especies con genomas completamente secuenciados. Estos últimos estudios llevaron a la identificación por primera vez de un linaje animal que contiene organismos que carecen de selenoproteínas [15, 111]. Sin embargo la evolución de la incorporación y maquinaria de biosíntesis de Sec no ha sido extensamente estudiada en *Nematoda*, un filo diverso que incluye tanto organismos de vida libre como parásitos de plantas y animales, con genomas secuenciados.

Estos elementos hacen de *C. elegans* un excelente modelo para estudiar la incorporación y utilización de Sec y de *Nematoda* un linaje excepcional para el estudio de la incorporación y biosíntesis de Sec, así como de la evolución de su uso y de los selenoproteomas.

### 1.6 Antecedentes

Previo al inicio de esta tesis, se identificaron los genes de *C. elegans* involucrados o presuntamente involucrados en la incorporación de Sec y se realizaron ensayos de ARNi con todos estos genes sobre la estirpe N2 y sobre las estirpes mutantes (más sensibles al ARNi) rrf-3(pk1426) y eri-1(mg366). Los resultados se describen dentro de los resultados del capítulo 4, bajo el subtítulo "Identification and silencing of genes involved in Sec incorporation in C. elegans" y pueden resumirse de la siguiente manera: i) los ensayos de ARNi con los genes involucrados tanto en la traducción canónica como en la decodificación de Sec (Ser-ARNt sintetasa canónica - sars-1- y de la proteína ribosomal L30 - rpl-30-) dieron fenotipo de letalidad; ii) los ensayos de ARNi con los genes presuntamente esenciales o involucrados en la biosíntesis e incorporación de Sec (tRNA[<sup>Ser]Sec</sup> -K11H12.t1-, SPS2 -seld-1-, PSTK -pstk-1-, SecS secs-1-, EfSec -efsc-1-, SecP43 - T07F10.3-), así como los presuntamente asociados pero de función desconocida (R13A5.9 y tag-321) no manifestaron fenotipo evidente al ser interferidos. Estos resultados fueron consistentes con la idea, previamente propuesta, de que la síntesis e inserción de Sec, así como la expresión de trxr-1, no son esenciales en condiciones de laboratorio.

# **OBJETIVOS**

# 2.1. Objetivo general.

Entender aspectos claves del mecanismo de incorporación de selenocisteína en eucariotas utilizando al nematodo *Caenorhabditis elegans* como modelo experimental y la evolución de la utilización de selenocisteína en el linaje *Nematoda* como modelo evolutivo.

# 2.2. Objetivos específicos:

- Estudiar el rol de SecP43 y de un gen presuntamente involucrado en la incorporación de selenocisteína en *C. elegans.*
- Estudiar la relevancia de la proteína de unión al elemento SECIS y de sus dominios de unión al ribosoma y de incorporación de selenocisteína para la incorporación de selenocisteína en *C. elegans*.
- Estudiar la evolución del uso de selenocisteína y de los genes involucrados en su incorporación en el filo *Nematoda*.
- Estudiar la ocurrencia de inicio no canónico de la traducción en C. elegans.

GENERACIÓN DE HERRAMIENTAS PARA EL ANÁLISIS DE GENES INVOLUCRADOS EN LA INCORPORACIÓN DE SEC La realización de esta tesis implicó una primera etapa de familiarización con las técnicas de manejo de *C. elegans* y la generación de una plataforma de estirpes mutantes. Una vez establecida esta plataforma, se llevó a cabo una serie de ensayos con las estirpes generadas. La generación de las estirpes se describe a continuación, así como parte de los ensayos realizados con las mismas junto con sus resultados y discusión. Los materiales y métodos se describen en el capítulo 4.

El mantenimiento de las estirpes se hizo según [112].

# 3.1 Curado de mutantes simples

# 3.1.1 Generación y mantenimiento de machos

La generación espontánea de machos ocurre naturalmente a una tasa muy baja (0,05 a 0,1%) [72]. Sin embargo, la misma puede aumentarse exponiendo larvas L4 a altas temperaturas, ya que un aumento en la temperatura se ve asociado a un aumento en la probabilidad de errores en la disyunción del cromosoma X, lo que lleva a la generación de gametos aneuploides. La unión de uno de estos gametos a otro portador de X dará lugar a un organismo macho X0.

En este trabajo se optó por colocar tres placas con 10 hermafroditas L4 y exponerlos a 30°C por 6 horas [113] cuando se requirió de la generación de machos. Luego del choque térmico, se los volvió a colocar en el incubador a 20°C y se analizó su progenie en busca de machos (se espera un aumento de la tasa de generación de machos, pudiendo alcanzar hasta el 10% de la progenie). Los machos se mantuvieron por cruces consecutivos con hermafroditas de su misma estirpe y por congelación. En este último caso, se tomaron placas llenas de gusanos con alta proporción de machos y se los congeló siguiendo el protocolo descrito en [112].

# 3.1.2 Curado de estirpes mutantes

Se solicitó a los consorcios NBRP-Japan y Oklahoma Consortium que generaran estirpes mutantes en los genes *scbp-2* (presunta SBP2), *T07F10.3* (presunta SecP43), *tag-321* (presuntamente asociada a la incorporación de Sec) y *R13A5.9* (presuntamente asociada a la incorporación de Sec).

# GENERACIÓN DE HERRAMIENTAS PARA EL ANÁLISIS DE GENES INVOLUCRADOS EN LA INCORPORACIÓN DE SEC

El consorcio NBRP-Japan produjo organismos mutantes por deleción en los genes *R13A5.9* y *scbp-2* y un mutante por deleción/inserción en *T07F10.3*, utilizando EMS. Ninguno de los consorcios fue capaz de generar organismos mutantes en *tag-321* durante el transcurso de esta tesis. El mutante en *R13A5.9* lleva una deleción de 277 pb en el sexto y mayor exón y genera un codón de terminación en fase lo que lo lleva a codificar para un polipéptido de 214 aminoácidos en lugar de 630. El mutante en *scbp-2* lleva una deleción de 143 pb que elimina parte de los últimos dos exones, que si bien no genera nuevos codones de terminación en marco daría lugar a una proteína de 121 aminoácidos en lugar de una de 153, perdiendo el motivo funcional L7Ae. El mutante en *T07F10.3* tiene una deleción de 261 pb entre el primer exón y el segundo intrón y una inserción de siete pares de bases, lo que elimina parcialmente al primer exón y completamente al segundo y genera un codón de terminación en el marco de lectura, por lo que codificaría para un polipéptido de 52 aminoácidos en lugar de 371. Todas estos alelos se presume que darían pérdida de función del gen. El mapeo de las mutaciones se representa en la Figura 3.1.

Todas estas estirpes mutantes fueron enviadas sin previo "curado" de otras mutaciones de fondo (no deseadas), producto del proceso mutagénico aleatorio (por regla general una estirpe mutante que se recibe contiene un genoma con 300 mutaciones). El correcto y exhaustivo curado o limpieza de todas las mutaciones excepto la de interés es un aspecto crucial antes de comenzar a trabajar con los organismos mutantes en los experimentos específicos, ya que un fenotipo resultante de un animal no curado puede ser engañoso.

El proceso de curado implica seis cruces consecutivos con la estirpe salvaje de referencia N2, en tres rondas de dos cruces cada una. En la figura 3.2 se esquematiza una ronda de curado. En un primer paso se cruzan hermafroditas mutantes con machos N2 (generalmente en proporción 1:3); de la progenie (F1) se toman machos (ya que necesariamente provienen del cruce del hermafrodita mutante con el macho silvestre) y se los cruza con hermafroditas silvestres.

## GENERACIÓN DE HERRAMIENTAS PARA EL ANÁLISIS DE GENES INVOLUCRADOS EN LA INCORPORACIÓN DE SEC



**Figura 3.1. Representación gráfica de la estructura génica de genes de interés y mapeo de sus mutaciones** A) Asociado a Sec (*R13A5.9*) B) SBP2 (*scbp-2*)\* y C) SecP43 (*T07F10.3*). Deleción significa la eliminación del fragmento que se indica en rojo. Deleción/inserción significa la eliminación del fragmento indicado en azul y la adición de siete bases. El número romano arriba a la izquierda de cada panel representa el número de cromosoma en el que está presente el gen y la regla indica la posición del fragmento de cromosoma mostrado. Las rectángulos representan exones y las líneas que conectan los exones representan los intrones; en celeste se indican las CDS y en gris los 5' y 3' UTRs. \*Estructura del gen en las versiones de la wormbase anteriores a WS243

Se separan 10 larvas de la progenie de ese cruce (F2) en placas individuales, se espera que alcancen la adultez y pongan huevos. Luego se retiran los gusanos de la F2, ahora adultos, y se los analiza por PCR (se espera un 50% de heterocigotas). De la descendencia de organismos heterocigotas (F3) producto de la autofecundación de F2, se separan 20 larvas. Se espera que alcancen la adultez y que pongan huevos, luego se retiran los gusanos de la F3, ahora adultos, y se los analiza por PCR. Finalmente se seleccionan aquellos que provienen de organismos homocigotos mutantes (se espera un 25% de homocigotas mutantes). Este proceso se repite dos

veces más partiendo de organismos homocigotos mutantes generados en la ronda de curado anterior, completando así seis cruces mediante 3 rondas.



**Figura 3.2 Esquema del proceso de una ronda de curado.** Una ronda de curado implica dos cruces de la estirpe mutante (m/m) con la estirpe silvestre de referencia (N2 o +/+) (marcados con cuadros punteados naranja). En un primer paso se cruzan hermafroditas mutantes con machos N2, de la progenie (F1) se toman machos (m/+) y se los cruza con hermafroditas N2. Se separan organismos hermafroditas heterocigotas de la F2, los cuales se autofecundan y, de su progenie (F3) se seleccionan organismos hermafroditas homocigotas mutantes (marcados en cuadro de línea continua naranja).

Cuando, como en estos casos, la presencia del alelo mutante no se asocia a la manifestación de un fenotipo evidente, el seguimiento de cada uno de los cruces se realiza mediante PCR. Para esto se diseñan 3 cebadores: dos que flanquean la deleción (OL, del inglés *outer left* -5' de la deleción- y OR *outer right* -3' de la deleción-) y otro que hibrida en parte de la secuencia deletada (IR, del inglés *inner right*) como se representa en la figura 3.2a. Utilizando combinaciones de cebadores OL/OR y OL/IR se pueden identificar organismos homocigotas silvestres, homocigotas mutantes y heterocigotas. Con la combinación OL/OR los homocigotas silvestres dan una única banda; los homocigotas mutantes dan también una única banda, pero de menor

# GENERACIÓN DE HERRAMIENTAS PARA EL ANÁLISIS DE GENES INVOLUCRADOS EN LA INCORPORACIÓN DE SEC

tamaño que la de los silvestres (para los alelos de deleción) y los heterocigotas dan las dos bandas anteriores. Con la combinación OL/IR los homocigotas silvestres y heterocigotas dan una única banda (del mismo tamaño) y los homocigotas no dan producto.

Se curaron las estirpes mutantes recibidas dando lugar a las estirpes IH1 *R13A5.9(tm3356) III*, IH2 *scbp-2(tm3417) I* e IH3 *T07F10.3(tm3512) V*\*, mutantes en el gen asociado a la incorporación de Sec, SBP2 y SecP43, respectivamente.

\* Nota: Si bien la nomenclatura que se utiliza en *C. elegans* es explicada con detenimiento en el anexo Nomenclatura, cabe recordar que las estirpes se nombran con dos o tres letras mayúsculas, correspondientes a la nomenclatura del laboratorio donde fue generada (IH en nuestro caso), y un número. Opcionalmente el nombre de la estirpe puede ir seguido de su genotipo, como ser el nombre del gen, seguido por el alelo entre paréntesis y luego el cromosoma donde se encuentra dicho gen, todo esto en cursiva. Por ejemplo IH1 *R13A5.9(tm3356) III*, es la primer estirpe generada en nuestro laboratorio, tiene el alelo *tm3356* del gen *R13A5.9* localizado en el cromosoma III.

# GENERACIÓN DE HERRAMIENTAS PARA EL ANÁLISIS DE GENES INVOLUCRADOS EN LA INCORPORACIÓN DE SEC



**Figura 3.3 Seguimiento de curado por PCR.** A) Diseño de cebadores OL y OR externos a la deleción e IR interno a la deleción B) Ejemplo de resultado de amplificación con cebadores OL/OR en organismos heterocigotas, homocigotas mutantes y homocigotas silvestres C) Ejemplo de resultado de amplificación con cebadores OL/IR en organismos heterocigotas u homocigotas silvestres y homocigotas mutantes.

# 3.2 Ensayos preliminares con las estirpes mutantes en los genes presuntamente asociados a la incorporación de Sec

3.2.1 Ensayos de ARNi con gsr-1

# 3.2.1.1 ARNi por alimentación en C. elegans

Para los ensayos de ARNi se utilizaron clones pertenecientes a la biblioteca de la Dra. Ahringer [95], amablemente cedidos por el Dr. Peter Askjaer (Universidad Pablo de Olavide, España). En esta biblioteca los genes de interés se encuentran clonados en el plásmido pL4440, que contiene dos promotores T7 para esa ARN polimerasa flanqueando el sitio de clonado (Figura 3.4). Estas clonas se transforman en la cepa de *E. coli* HT115 (DE3), la cual es mutante para la ARNasa III, lo que aumenta la estabilidad del ARN, y posee una ARN polimerasa T7 inducible por IPTG, lo que hace que el ARN clonado se transcriba en ambos sentidos en presencia del inductor. Como control negativo de la técnica se utilizaron bacterias transformadas con *el vector* pL4440[*unc-22*] que producen fenotipos de animales anchos y cortos (*dumpy*) y de movimiento descoordinado (*uncoordinated*), respectivamente. En estos experimentos, además, se utilizaron como controles positivos las estirpes *trxr-1(sv47)* y *efsc-1(sv36)* y como control negativo la estirpe N2.



**Figura 3.4 Esquema de bacteria** *E. coli* **HT115(DE3) transformada con vector L4440.** Para los ensayos de ARNi por alimentación se transforman bacterias *E. coli* HT115(DE3) (mutante para la ARNasa III y que codifica para polimerasa de T7 inducible por IPTG) con vector pL4440, al cual se le clona el gen de interés en un sitio de clonado flanqueado por dos promotores de polimerasa T7. Modificado de [91].

# 3.2.1.2 Resultados

Como se mencionó en la introducción, las estirpes incapaces de incorporar Sec (mutantes en *trxr-1* o en EFSec *-efsc-1-*) muestran un fenotipo de arresto larvario cuando se les interfiere la expresión del gen *gsr-1* (que codifica para la única glutatión reductasa de *C. elegans*) [99]. Con el fin de identificar posibles roles de los genes identificados como presuntamente involucrados en la incorporación de Sec, se realizó este ensayo de ARNi sobre las estirpes mutantes obtenidas, mediante alimentación [109]. Esto implicó la alimentación de gusanos con bacterias que expresan ARNdh de *gsr-1* por hasta seis generaciones.

Se encontró que todas las estirpes analizadas reproducen el fenotipo de arresto larvario típico de los gusanos que no incorporan Sec, cuando se interfirió la expresión de *gsr-1*. Cabe destacar que la estirpe mutante en *sbp2* (*scbp-2(tm3417)*) evidenció este fenotipo en la tercera generación, mientras que la estirpe mutante en el gen asociado a la incorporación de Sec (*R13A5.9(tm3356)*) y la estirpe mutante en secP43 (*T07F10.3(tm3512)*) lo hicieron en la cuarta. A su vez, el mutante en *sbp2* mostró un fenotipo más penetrante que los demás mutantes: de todas las estirpes interferidas, los gusanos de esta estirpe fueron los que mostraron el fenotipo de arresto larvario en mayor proporción.

Un dato interesante es que con las consecutivas rondas el número de gusanos que alcanzaba la adultez iba disminuyendo debido al arresto, disminuyendo de esta forma también el tamaño de la población de la siguiente generación. En el caso del mutante en *sbp2*, al cabo de la quinta generación el experimento no podía continuarse debido a que ya no quedaban más gusanos.

# 3.2.2 Ensayos de incorporación de <sup>75</sup>Se

Este ensayo fue llevado a cabo por el Dr. Anton Turanov en el Brigham and Women's Hospital, Harvard Medical School, Estados Unidos.

Para analizar la capacidad de incorporar Sec de las estirpes mutantes, se realizó un ensayo de incorporación metabólica de <sup>75</sup>Se. Para esto se creció *E. coli* OP50 en medio suplementado con Na<sub>2</sub>SeO<sub>3</sub> marcado con <sup>75</sup>Se. Con estas bacterias se alimentaron las

estirpes de gusanos a analizar y los controles N2 (control positivo) y las estirpes mutantes en *trxr-1* y en EFSec (controles negativos).

Luego de crecerlos durante 7 días se evaluaron los extractos de proteínas por electroforesis (SDS-PAGE), transfiriendo a membrana de PVDF y detectando posteriormente señal radioactiva en un PhosphorImager.

Los resultados obtenidos, mostraron que la estirpe mutante en *sbp2* no incorpora Sec, en tanto que el mutante en el gen asociado a la incorporación de Sec *R13A5.9* y el mutante en SecP43 sí incorporan (Figura 3.5).



Figura 3.5. Ensayo de incorporación de selenio marcado radiactivamente. Análisis de extracto de proteínas de gusanos N2 (control positivo), *trxr-1(sv47)* (control negativo mutante en TRXR-1), *scbp-2(tm3417)* mutante en SBP2, *efsc-1(sv36)* (control negativo, mutante en EFSec), *T07F10.3(tm3512)* mutante en SeCP43 y *R13A5.9(tm3356)* mutante en el gen asociado a la incorporación de Sec. La flecha llena marca TRXR-1, única selenoproteína de *C. elegans*, se destaca que el mutantes en SBP2 no es capaz de incorporar Sec en estas condiciones. La flecha punteada marca la formiato deshidrogenasa O de *E. coli*.

# 3.3 Discusión

Los ensayos de ARNi con *gsr-1* sobre las estirpes mutantes en el gen asociado a la incorporación de Sec (*R13A5.9(tm3356)*) y en SecP43 (*T07F10.3(tm3512)*) dieron fenotipo de arresto larvario evidente en la cuarta generación, sin embargo el ensayo de incorporación metabólica de <sup>75</sup>Se mostró que esas estirpes incorporan Sec. Si bien los resultados obtenidos parecen ir en direcciones opuestas, los mismos no necesariamente son contradictorios. Estas diferencias pueden deberse a las condiciones en que se lleva a cabo cada ensayo o el rol que cumplen dichos genes en la incorporación de Sec. El ensayo de ARNi requiere de al menos tres semanas (cuatro generaciones) para manifestar el fenotipo en un contexto desafiante; mientras que el ensayo de incorporación metabólica de <sup>75</sup>Se se realizó en unos pocos días sobre gusanos de diferentes estadios, en condiciones óptimas de laboratorio. De hecho estos ensayos se complementan y pueden poner en evidencia la relevancia de la correcta expresión de estos genes en diferentes contextos: óptimo (ensayo de incorporación de <sup>75</sup>Se) y desafiante (ensayo de ARNi con *gsr-1*).

SecP43 fue identificada inicialmente por [114] como una proteína que se une al ARNt<sup>[Ser]Sec</sup>. En [54], se vio que la disminución de sus niveles de expresión mediante ARNi se acompaña de una disminución de los niveles de expresión de la selenoproteína Gpx1 en células NIH3T3 y TCMK-1. Encontraron también que SecP43 se encuentra involucrada en la metilación del uracilo del anticodón de ARNt<sup>[Ser]Sec</sup>. En [55] se encontró que forma complejos supramoleculares con ARNt<sup>[Ser]Sec</sup>-EfSec-SBP2 y se propuso como involucrada en el mantenimiento de la integridad de los ARNm de las selenoproteínas, evitando el *non-sense mediated decay* (decaimiento o degradación de ARNm mediado por codones stop en marcos abiertos de lectura) que puede darse por tener codones UGA en marco. En el caso de *C. elegans*, descartamos que sea ese el rol que cumple ya que el residuo Sec es el penúltimo de la cadena polipeptídica, separado sólo un codón del verdadero codón de finalización.

La información de la que se dispone del gen asociado a la incorporación de Sec (*R13A5.9*) es mucho más limitada. Su ortólogo de *Drosophila* fue reportado en [21], donde fue identificado como un gen presente en organismos que incorporan Sec, pero ausente en los que no incorporan Sec. Este gen codificaría para una proteína con dominio perteneciente a la superfamilia de facilitadores principales (MFS), una de las

# GENERACIÓN DE HERRAMIENTAS PARA EL ANÁLISIS DE GENES INVOLUCRADOS EN LA INCORPORACIÓN DE SEC

dos familias más grandes de transportadores de membrana, las cuales pueden funcionar como uniport, simport o antiport de moléculas tales como azúcares simples, oligosacáridos, aminoácidos, ésteres organofosfatados, metabolitos del ciclo de Krebs y una gran variedad de iones orgánicos e inorgánicos [115] (no se han identificado transportadores de selenio hasta el momento). De mediar la captación celular de selenio, es de esperar que en el mutante los niveles intracelulares de selenio se vean disminuidos, afectando la biosíntesis de Sec y, por lo tanto, su incorporación en TRXR-1; lo que explicaría el fenotipo de arresto larvario en el ensayo de ARNi con gsr-1. En este sentido cabe comentar que la señal radiactiva dada por el 75Se incorporado a TRXR-1 es menor que la estirpe silvestre. Asimismo, la captación celular de selenio podría darse por diferentes mecanismos que medien el transporte de diferentes especies químicas de selenio y que el selenio que entra por otro mecanismo sea suficiente para incorporar Sec en la TRXR-1 en condiciones óptimas (y por lo tanto se obtenga señal radiactiva en el ensayo de incorporación de <sup>75</sup>Se), pero que los niveles de incorporación de selenio y, por lo tanto de enzima activa, que se alcancen en condiciones desafiantes (ARNi con *gsr-1*) no sean suficientes. Cabe resaltar que las bacterias de las cuales se alimentan los gusanos metabolizan selenio, que puede derivar en diferentes especies químicas (selenuro o selenofosfato, por ejemplo) que ingiere el gusano.

En suma, es probable que estos dos genes estén implicados en la incorporación de Sec, si bien no serían imprescindibles para la misma en condiciones óptimas de laboratorio. Sin embargo, podrían ser importantes para la incorporación de Sec; podrían tener un rol regulatorio de la expresión y/o regulación de *trxr-1*, por ejemplo manteniendo elevados los niveles de ARNm, ser importantes para su expresión en un determinado estadio (lo cual podría ser inadvertido en el ensayo de incorporación de Se) o estar implicados en la incorporación de selenio y que este rol se vuelva evidente y necesario bajo condiciones desafiantes. Cabe destacar que la marca de selenio en estos mutantes es menos intensa (particularmente en el mutante en el gen asociado a la incorporación de Sec) que la del control N2, esto refuerza la hipótesis de que estos genes estén involucrados en la regulación de la expresión de selenoproteínas y/o en la captación celular de Se.

# 3.4 SBP2 da para más:

Teniendo en cuenta que los resultados obtenidos con la estirpe mutante en *scbp-2(tm3417)* fueron prometedores, se decidió profundizar en ellos. De particular interés es el hecho que la SBP2 de *C. elegans* identificada es notoriamente diferente de las SBP2 de eucariotas hasta ahora secuenciadas. Los resultados obtenidos son mostrados y discutidos en el capítulo 4, en [116] y anexos.

AJUSTES, EXTINCIÓN Y VESTIGIOS DE LA INCORPORACIÓN DE SELENOCISTEÍNA EN EL LINAJE NEMATODO Los resultados de este capítulo se resumen en el artículo:

# "Adjustments, extinction, and remains of selenocysteine incorporation machinery in the nematode lineage"

Lucía Otero, Laura Romanelli-Cedrez, Anton A. Turanov, Vadim N. Gladyshev, Antonio Miranda-Vizuete and Gustavo Salinas

RNA, 2014. 20(7): p. 1023-34

Todos los experimentos mostrados en este artículo fueron realizados por mí excepto la identificación de los genes asociados a la incorporación de Sec y los ensayos de ARN de interferencia con los mismos y el ensayo de incorporación de <sup>75</sup>Se.

# Adjustments, extinction, and remains of selenocysteine incorporation machinery in the nematode lineage

# LUCÍA OTERO,<sup>1,5</sup> LAURA ROMANELLI-CEDREZ,<sup>1,5</sup> ANTON A. TURANOV,<sup>2</sup> VADIM N. GLADYSHEV,<sup>2</sup> ANTONIO MIRANDA-VIZUETE,<sup>3,4,6</sup> and GUSTAVO SALINAS<sup>1,5,6</sup>

<sup>1</sup>Cátedra de Inmunología, Facultad de Química, Instituto de Higiene, Universidad de la República, Montevideo 11600, Uruguay

<sup>2</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

<sup>3</sup>Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, 41013 Sevilla, Spain <sup>4</sup>Departamento de Fisiología, Anatomía y Biología Celular, Centro Andaluz de Biología del Desarrollo (CABD-CSIC), Universidad Pablo de Olavide, 41013 Sevilla, Spain

#### ABSTRACT

Selenocysteine (Sec) is encoded by an UGA codon with the help of a SECIS element present in selenoprotein mRNAs. SECISbinding protein (SBP2/SCBP-2) mediates Sec insertion, but the roles of its domains and the impact of its deficiency on Sec insertion are not fully understood. We used *Caenorhabditis elegans* to examine SBP2 function since it possesses a single selenoprotein, thioredoxin reductase-1 (TRXR-1). All SBP2 described so far have an RNA-binding domain (RBD) and a Secincorporation domain (SID). Surprisingly, *C. elegans* SBP2 lacks SID and consists only of an RBD. An *sbp2* deletion mutant strain ablated Sec incorporation demonstrating SBP2 essentiality for Sec incorporation. Further in silico analyses of nematode genomes revealed conservation of SBP2 lacking SID and maintenance of Sec incorporation linked to TRXR-1. Remarkably, parasitic plant nematodes lost the ability to incorporate Sec, but retained SecP43, a gene associated with Sec incorporation. Interestingly, both selenophosphate synthetase (SPS) genes are absent in plant parasitic nematodes, while only Cys-containing SPS2 is present in Sec-incorporation and the evolution of Sec utilization trait, selenoproteomes, selenoproteins, and Sec residues. Finally, our study provides evidence of noncanonical translation initiation in *C. elegans*, not previously known for this wellestablished animal model.

Keywords: selenocysteine; SECIS-binding protein; SBP2; Caenorhabditis elegans; selenoprotein; translation initiation

#### INTRODUCTION

Selenocysteine (Sec) is the twenty-first amino acid and is inserted into selenoproteins at in-frame UGA codons by a specific tRNA<sup>Sec</sup> (Leinfelder et al. 1988; Hatfield and Gladyshev 2002). Insertion of Sec requires translational recoding, which requires a Sec insertion sequence (SECIS) element, a *cis*-acting stem–loop structure present in selenoprotein mRNAs (Berry and Larsen 1993). In eukaryotes, SECIS element binds SECIS-binding protein 2 (SBP2) and then recruits the Secspecific elongation factor (EfSec), which binds tRNA<sup>Sec</sup>. After association with the ribosome, SBP2 exchanges for the ribosomal protein L30. This leads to a conformational change in the SECIS element that triggers the release of tRNA<sup>Sec</sup> and the hydrolysis of GTP (Chavatte et al. 2005; All-

<sup>6</sup>Corresponding authors

mang et al. 2009). In mammals, nucleolin and eIF4a3 also bind the SECIS element of some selenoprotein mRNAs and regulate Sec incorporation (Budiman et al. 2009; Miniard et al. 2010). SBP2 is likely an essential and limiting factor for Sec incorporation in mammals (Mehta et al. 2004). Invertebrate and vertebrate SBP2 differ significantly in length (~850 and 350 amino acids long, respectively), however the N-terminal half of mammalian SBP2 is not required for Sec incorporation (Donovan and Copeland 2009). Invertebrate SBP2 and the C-terminal half of mammalian SBP2 possess an RNA-binding domain (RBD) and a Sec-incorporation domain (SID), which are presumed to be essential for decoding Sec (Takeuchi et al. 2009; Donovan and Copeland 2010). Sec synthesis is the additional part of the pathway required for biosynthesis of selenoproteins. It takes place on

<sup>&</sup>lt;sup>5</sup>**Present address:** Institut Pasteur de Montevideo, Montevideo 11400, Uruguay

E-mail gsalin@fq.edu.uy

E-mail amiranda-ibis@us.es

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.043877.113.

<sup>© 2014</sup> Otero et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/ by-nc/4.0/.

#### Otero et al.

tRNA<sup>Sec</sup>, which is first aminoacylated with serine by a canonical servl-tRNA synthetase (SerS) (Leinfelder et al. 1988; Lee et al. 1989). The servl-tRNA<sup>Sec</sup> is phosphorylated to form Ophosphoseryl-tRNASec by O-phosphoseryl-transfer tRNASec kinase (PSTK) (Carlson et al. 2004), and then Sec synthase (SecS) catalyzes the formation of selenocysteinyl-tRNASec from O-phosphoseryl-tRNA<sup>Sec</sup> and selenophosphate (Xu et al. 2007b). In mammals, selenophosphate synthetase 2 (SPS2) is responsible for the synthesis of selenophosphate from selenide and ATP (Bock et al. 1991), while SPS1 has been recently postulated to function in a pathway unrelated to Sec biosynthesis (Xu et al. 2007a). Finally, other genes are present only in Sec-incorporating species but their precise function is not known; one of these genes is SecP43, thought to participate in the nucleus-cytosol transport of tRNA<sup>Sec</sup> (Small-Howard et al. 2006).

The evolution of Sec incorporation has not been fully explored across eukaryotes. Evolution of Sec utilization in vertebrates has been characterized (Mariotti et al. 2012). In addition, Sec incorporation evolution has been examined in detail in arthropods, a lineage that contains numerous species with sequenced genomes. These studies have led to the identification of the only known animal lineage in which some representatives lack selenoproteins (Chapple and Guigó 2008; Lobanov et al. 2008). However, the Sec incorporation and biosynthesis machinery has not been fully examined in Nematoda, a diverse phylum that includes free-living as well as plant and animal parasitic species with sequenced genomes. Interestingly, the nematode Caenorhabditis elegans possesses the entire Sec insertion system dedicated to a single Sec residue in its proteome, located in the active site of thioredoxin reductase-1 (TRXR-1) (Buettner et al. 1999; Gladyshev et al. 1999). trxr-1 is not an essential gene under standard laboratory conditions, but it is required for molting when glutathione reductase (GSR-1) function is compromised (Stenvall et al. 2011). These facts along with its powerful genetics make C. elegans an excellent experimental model to study Sec incorporation and utilization. Furthermore, several Caenorhabditis species and additional nematodes from other clades inhabiting very diverse environments have been completely sequenced. Therefore, Nematoda is an exceptional lineage to study the evolution of Sec incorporation and biosynthesis, selenoproteomes, and Sec residues.

In this study, we identified the genes involved or putatively involved in Sec biosynthesis and incorporation in *C. elegans* and silenced their expression, generated an SBP2 deletion mutant strain, and examined the nematode lineage for the genes involved in Sec utilization. We found that (i) nematodes, including *C. elegans*, possess the shortest eukaryote SBP2 that lacks SID, (ii) *C. elegans* SBP2 is essential for inserting Sec in TRXR-1, (iii) plant nematode parasites have lost the ability to incorporate Sec, lack both SPS genes, but retained SecP43, a gene previously associated exclusively with selenium incorporation.

#### 1024 RNA, Vol. 20, No. 7

#### RESULTS

#### Identification and silencing of genes involved in Sec incorporation in *C. elegans*

C. elegans and C. briggsae code for only one selenoprotein that contains a single Sec (Buettner et al. 1999; Gladyshev et al. 1999; Taskov et al. 2005). Therefore, we hypothesized that adjustments in Sec incorporation machinery may have occurred in representatives of this lineage (e.g., due to relaxed SECIS-binding constraints). To examine this possibility, we first searched and analyzed the nematode genes involved in Sec incorporation. For this purpose, we used human amino acid sequences as queries except for SBP2, as homology analyses revealed no clear ortholog of the human SBP2. Using a multiple query consisting of the C-terminal half of human SBP2, human SBP2-like, and invertebrate SBP2, we identified the gene *scbp-2* (hereafter referred to as *sbp2*) as the putative C. elegans SBP2, which is significantly shorter than any known eukaryotic SBP2. We further identified 10 out of 11 genes involved in Sec decoding in humans: (i) the essential and presumptively essential genes for Sec incorporation: tRNA<sup>Sec</sup> (K11H12.t1), SPS (seld-1), seryl-tRNA synthetase (sars-1), phosphoseryl-tRNA (Ser/Sec) kinase (pstk-1), SectRNA synthase (secs-1), EfSec (efsc-1, formerly named in WormBase as selb-1) a gene previously identified by Fagegaltier et al. (2000), and ribosomal protein L30 (rpl-30); (ii) involved in Sec insertion, but not essential for this pathway: SecP43 (T07F10.3); and (iii) two of the three genes present in organisms that incorporate Sec, but are absent in those that do not incorporate Sec (i.e., putatively involved only in Sec incorporation): R13A5.9 and tag-321 (sequences are shown in Supplemental Fig. 1). Interestingly, in contrast to other metazoa, C. elegans possesses only a Cys-containing sps gene.

In order to identify a phenotype associated with Sec incorporation, we knocked down the expression of genes involved in this process by RNA interference (RNAi). These experiments were carried out with three C. elegans strains: N2 (wild type), rrf-3(pk1426), and eri-1(mg366). The latter two strains are more sensitive to RNAi (Simmer et al. 2002; Kennedy et al. 2004). These experiments showed the same outcome in all strains assayed: (i) Interference of the expression of any gene involved in both canonical translation and Sec decoding (*rpl-30* and *sars-1*) showed a lethal phenotype, (ii) interference of the expression of genes involved or putatively involved only in Sec decoding did not exhibit any apparent phenotype. These results were consistent with the known model of Sec synthesis and insertion and the fact that trxr-1 is not essential in this model under laboratory conditions. Indeed, both trxr-1 RNAi down-regulation and the null alleles *trxr-1(sv47)* and *efsc-1(sv36)* did not exhibit any apparent phenotype (Stenvall et al. 2011). Interference of the expression of genes associated with Sec synthesis but without a known function (R13A5.9 and tag-321) did not generate



В

A

ggtttaattacccaagtttgagcttcagaaaaattcccgaaacaataaatgtaacggctaac V - L P K F E L Q K I S E T I N V T A N attgataacttggctttcgagcttatcaagcgcttgaaaagccaacaagacaagaatcgc I D N L A F E L I K R L K S Q Q D K N R aaaggaaatgcggatgaaaagaactttctgaagagacgagtaatatgtggagctcacgaa K G N A D E K N F L K R R V I C G A H E acactaaagtttgcacagtctggaagagtaaagctcatcttctacgcgaaaaatatggac T L K F A Q S G R V K L I F Y A K N M D E N N L R A A S N F H S L O I V C P L A ggaataccgatgatcgaggtgttaacgagaaaggaattgtctcgaattatgaacaaattt IPMIEVLTRKELSRIM Κ F G Ν Y V A VVGVIDFS G F Ε R Е Т E D atcgtgaagaactggaaaatacacgattcatgcaagctgtttcacatcgatcaatcttccV K N W K I H D S C K L F Н Т D S S 0 ctcttgccgtgaatcaaccatcttgtaaaataaagtatttga L Ρ L



**FIGURE 1.** cDNA, protein, and genomic organization of *C. elegans sbp2* gene coding for a SECIS-binding protein. (*A*) Amplification of the full-length sbp2 mRNA with SL1 as forward primer and a reverse specific primer. A 493-bp product corresponding to the full-length cDNA was obtained (lane 1), 100-bp gene ruler (lane 2). (*B*) The conceptual translation of the amplified mRNA shows that the reading frame remains open upstream of the first in-frame ATG codon (dotted frame) and that there is an ATT codon at the beginning of the cDNA (bold frame) in a favorable *Caenorhabditis* Kozak context. (*C*) The gene model proposed by WormBase (black and white rectangles correspond to coding and noncoding exons, respectively), codes for an 85-amino-acid-long protein and has a long 5' UTR and a truncated RBD (gray). The first in-frame ATG codon is indicated in gray. (*D*) The gene model put forward in the present work codes for a 153-amino-acid-long protein with a short 5' UTR and a complete RBD. The ATT codon that serves for translation initiation is indicated in gray.

any evident phenotype, indicating that their functions are not essential for *C. elegans* under laboratory conditions.

#### C. elegans SBP2 is the first C. elegans protein with a non-AUG initiation of translation codon and the first SBP2 without SID

As described above, we identified *scbp-2* as the putative *C. elegans* SBP2. The annotated gene was significantly shorter than any known SBP2 whether from vertebrates or invertebrates. Mammalian SBP2 are  $\sim$ 800 amino acids long, while insect SBP2 are  $\sim$ 400 amino acids long (Donovan and Copeland 2009). Surprisingly, the predicted *C. elegans* gene codes for an 85-residue-long protein, corresponding to the C-terminal end of eukaryote SBP2.

In order to determine if the identified gene was correctly annotated, we amplified the full-length cDNA from embryos using a reverse gene specific primer and the splice leader 1, a mini exon transspliced at the 5' end of certain mRNAs, as a forward primer (Blumenthal and Steward 1997; Hastings 2005). We obtained a 493-bp product (Fig. 1A,B) identical to the transcript sequence annotated in WormBase. The first in-frame ATG codon was 231 nucleotides (nt) downstream from the beginning of the amplified sequence and it would encode an 85-amino-acid protein with a truncated RBD (Fig. 1C). Further analysis of the sequence showed that upstream of the first ATG, the reading frame continued for an additional 228 nt (Fig. 1B). We found that near the beginning of the full-length cDNA sequence there is an ATT codon preceded by AAA, which provides a favorable Kozak context for potential noncanonical translation initiation in C. elegans (Fig. 1B; Blumenthal and Steward 1997). Importantly, translation from this first AUU would complete the functional RBD (Fig. 1D). Also consistent with a non-AUG initiator codon is the fact that the amino acid homology with other eukaryotic SBP2 extends upstream of the first AUG to the first AUU (Fig. 2). In addition, C. elegans Ribo-seq data indicate translation upstream of the first AUG

#### Otero et al.

H_sapiens D_rerio D_melanogaster C_elegans	RIETPKFQSKQQPQDNFKNNVKKSQLPVQLDLGGMLTALEKKQHSQHAKQSSKPVVVSVG KKSKVPVQLDLGNMLAFLEQKQQSQKSKQDQRPVTLTVG MTEKIRRSRNQDVQDTAIKITPRSSKYKNQHRKREQQTS	480 180 39
	Sec incorporation domain	
H_sapiens D_rerio D_melanogaster C_elegans	-AVPVLSKECASGERGRRMSQMKTPHNPLDSSAPLMKKGKQREIPKAKKPTSLKKIILKE GALPVVHKEPSVQKKQWQQEKTPHNPLDSTCPSVKKGKQREVPKAKKPTPLKRVILRE LLDFVIKPRPKTQRQTKAHKLQKTHLAITRGSYIVYKPKGKTRLDPKKKITRLKKSVRVY	539 238 99
H_sapiens D_rerio D_melanogaster C_elegans	RQERKQRLQENAV SPAFT SDDTQDGE SGGDDQ FPEQAEL SG PEGMDELI ST PSVEDKS REERKQNRLLEERDP SRVATEAGE KETSENP SEA SL PEQEHLES KI EQ PAS PS DD KL RTS KKAERE VAENDLEG VP VV GQ DI NPNAI PLEQ QV	597 295 135
H_sapiens D_rerio D_melanogaster C_elegans	EEPPGTELQRDTEASHLAPNHTTFPKTHSRRFRDYCSQMLSKEVDACVTDLLKEVVRFOD DDGLSENPTLQTEIPNSNP-ISKVPKTHSRKFRDYCNQVLSKDVDECVSNLLKEVVRFOD QNLSLSKTPAPNNPSQAKTVHATHSRRFRSYCDNCTRPRLKELSTOLLRDFDRFK VTANIDNLAFELIKRKSQOD	657 354 191 27
	RNA-binding domain	
H_sapiens D_rerio D_melanogaster C_elegans	RMYQKDP-VRAKTKRRIVIGIREVIKHIKIKKIRCVIISPNCEKIQSKGGLDDIHTIID RLYQKDP-MRARMKRRIVMGIREVIKHIKIKKVRCVIISPNCERIQSKGGLDEALHNIID RAFAKNE-IRARAHPRIVIGVREALARIRINKVRLIFIATDCEICPGESGLDATIEGIKF KNRKGNADERNFIKRRVICGAHETIKFAQSGRVRIIFYAKNMDENNIRAASNFHSIQI	716 413 250 85
H_sapiens D_rerio D_melanogaster C_elegans	YACEQNIEFVFAUNRKAUGRSLNKAVFVSVVGIFSYDCAQDQFHKMVELTVAARQAYK TCRDQSVEFVFALSRKAUGRCVNKAVFVSLVGIFNYDCAQDHYHKMIELSAEARKSYE QCQQQQVEYCFFELRREDSYALQKRAQICCVAILDFDCANATYADLLNEIEDARAEYK VCFLAGIEMIEVLTRKELSRIMNKFFYVAVVGVIDFSCFERETEDIVKNWKIHDSCKLFH	774 471 308 145
H_sapiens D_rerio D_melanogaster C_elegans	TMLENVQQELVGEPRPQAPPSLLTQGPSCPAEDGPPALKEKEEPHYIEIWKKHLEA-YSG VMIANLDGDSQEDLQEDQQEEEPPGETSGTEEPEYIKIWKKMLEKGYSH RRTASIDQSSLLP	833 520 313 153
H_sapiens D_rerio D_melanogaster C_elegans	CTLELEESLEASTSQMMNLNL- 854 PFLNFEDQFSSLSIGSEQLLDDEGS 545	

**FIGURE 2.** Sequence alignment of SBP2 from *Homo sapiens, Danio rerio, Drosophila melanogaster*, and the identified *C. elegans* SBP2. The reading frame of *C. elegans* SBP2 remains open upstream of the first Met, and translation from the first Ile completes the RBD (overlined in black). *C. elegans* SBP2 lacks the Sec incorporation domain (overlined in gray). The full-length *D. melanogaster* and *C. elegans* SBP2s are shown; in the case of vertebrate SBP2s, the first 400 amino acids (absent in invertebrate SBP2s) are not shown. Black blocks indicate positions which have a single, fully conserved residue. Dark gray and light gray blocks indicate conservation between groups of strongly or weakly similar properties, respectively. Numbers indicate the amino acid residue in the polypeptide chain.

(Michel et al. 2014). Translation initiation from a non-AUG codon is a striking and conspicuous feature of SBP2 mRNA since it has not been previously described for any *C. elegans* mRNA. Our results indicate that for *sbp2* there is noncanonical initiation of translation and that an AUU codon in a favorable translation initiation context likely signals the beginning of translation.

Another remarkable feature of *C. elegans* SBP2 is the absence of the SID. SBP2-flanking genes have neither a related function nor a SID. Even more, SBP2 is not a "split gene" in *C. elegans*, since the SID is absent in the entire genome. No hits were found in BLAST searches using several eukaryotic SIDs (independent or concatenated) or conserved regions of the SID as queries. This finding suggests that the *C. elegans* Sec-incorporating machinery differs from that in other eukaryotes.

# Sec biosynthesis and incorporation in the *Caenorhabditis* lineage

The peculiar *C. elegans* SBP2 prompted us to analyze the *Caenorhabditis* lineage taking advantage of several *Caenorhabditis* species that have been completely sequenced. We analyzed *C. briggsae, C. brenneri, C. remanei, C. japonica, C. sp5,* and *C. sp11* genomes annotated in WormBase as of July 2013 (WS238). We found that all these species possess the machinery to incorporate Sec. All of them encode a single Sec residue in the entire proteome: the Sec residue of *trxr-1*. Interestingly, unlike other metazoa, all *Caenorhabditis* species encode a single SPS. Even though SPS displays slightly higher identity to SPS1 than to SPS2, the *C. elegans* gene is phylogenetically an SPS2 (M Mariotti, pers. comm.) with a Cys residue at the active site.

We further identified and analyzed *Caenorhabditis sbp2*. All analyzed Caenorhabditis species lacked the SID and coded for a short sbp2, i.e., 152-155 amino acid long (Fig. 3A), with an identical gene structure: four exons and three introns. We then examined signals for translation initiation in Caenorhabditis sbp2 and found that they all shared the putative AUU initiation codon, preceded by a highly conserved Kozak sequence <sup>-4</sup>AAA<sup>-1</sup> (Fig. 3B). They also shared a conserved cytosine in the +5 position, suggesting that this position may also be important for non-AUG translation initiation. In addition, we found that all Caenorhabditis SBP2 displayed high levels of similarity at the amino acid sequences from the first isoleucine residue (encoded by the proposed AUU initiator codon) including conservation of basic residues at the beginning of the RBD (Fig. 3A). Strikingly, at the homologous position of the first in-frame methionine of C. elegans SBP2, a leucine residue was found in all other Caenorhabditis spp., providing further evidence that C. elegans sbp2 translation initiation is not at this AUG codon (Fig. 3A). It is also important to mention that, of all analyzed *Caenorhabditis* SBP2, only the *C. japonica* protein has a methionine upstream of the homologous position of the first methionine of *C. elegans* SBP2. Together, these results strongly support that noncanonical translation initiation occurs in *Caenorhabditis* SBP2. Although described in other species, noncanonical translation initiation has not been previously reported for any *Caenorhabditis* (Rhoads et al. 2006).

With regard to Sec incorporation, the results of our analysis strongly indicate that *Caenorhabditis* SBP2 lack SID. The absence of a SID may relate to the number and type of SECIS elements encoded by *Caenorhabditis*. All *Caenorhabditis* genomes encode a single selenoprotein (TRXR-1). The analysis of *trxr-1* SECIS elements using the SECISearch3 program (Mariotti et al. 2013) revealed the presence of the unusual type I SECIS element typical of nematodes with a guanine immediately upstream of the non-Watson-Crick base-pairing quartet (Fig. 4; Grundner-Culemann et al. 1999; Taskov et al. 2005).

Finally, it is worth mentioning that the *Caenorhabditis* SBP2 contain a short C-terminal tripeptide which is highly conserved in this lineage but absent in vertebrates.



**FIGURE 3.** *Caenorhabditis* SBP2s. (*A*) Alignment of SBP2 proteins in all sequenced *Caenorhabditis* species. The RBD is overlined; the first methionine encoded in *C. elegans* SBP2 and amino acids corresponding to the homologous position in the other *Caenorhabditis* SBP2 are marked with an asterisk. Black blocks indicate positions which have a single, fully conserved residue. Dark gray and light gray blocks indicate conservation between groups of strongly or weakly similar properties, respectively. Numbers indicate the amino acid residue in the polypeptide chain. (*B*) Frequency of nucleotides found at the beginning of the cDNA of *Caenorhabditis sbp2*. Adenine +1 corresponds to the first base of the coding sequence. A complete conservation of a favorable *Caenorhabditis* Kozak context is observed at positions –1 to –3.



**FIGURE 4.** SECIS elements of *Caenorhabditis* Sec-containing thioredoxin reductase. (*A*) *C. brenneri*, (*B*) *C. briggsae*, (*C*) *C. elegans*, (*D*) *C. japonica*, (*E*) *C. remanei*, (*F*) *C. sp5*, (*G*) *C. sp11*. All *Caenorhabditis* encode a type I SECIS element with a guanine immediately upstream of the non-Watson-Crick base-pairing quartet. The non-Watson-Crick base-pairing quartet (SBP2-binding site) and the AAR triplet are highlighted.

# C. *elegans* SBP2 is a functional and essential gene for Sec incorporation

We performed metabolic incorporation of radioactive <sup>75</sup>Se in the *scbp-2(tm3417)* mutant to analyze the relevance of *sbp2* for Sec incorporation. We used the N2 wild-type strain as a Sec incorporation positive control and both *trxr-1(sv47)* and *efsc-1(sv36)* mutants as Sec incorporation negative controls. This experiment showed that the *scbp-2(tm3417)* was unable to incorporate Sec (Fig. 5). In a previous study, it was shown that the *trxr-1* and *efsc-1* deletion mutant strains showed larval arrest phenotype when the expression of *gsr-1* gene was silenced (Stenvall et al. 2011). Importantly, *gsr-1* RNAi in *scbp-2(tm3417)* mutant also reproduced the larval arrest phenotype.

These results indicate that the identified *sbp2* is an essential gene for Sec incorporation in *C. elegans* and therefore provides support to the concept that lineage-specific adjustments may have occurred in the Sec insertion machinery particularly regarding requirements for SECIS element decoding.

#### Analysis of transgenic SBP2 strain provides further evidence for a noncanonical start of translation and function of the identified gene

To gain further information, we immobilized the *trxr-1* SECIS element (wtSECIS) in order to pull down the native SBP2 from a total wild-type worm aqueous extract (see Materials and Methods). The proteins retained by the wtSECIS column were analyzed by SDS-page and MS/MS. We isolated the eukaryotic translation elongation factor 1a (*eef-1A.2*) and several ribosomal proteins: *rpl-3*, *rpl-4*, *rpl-22*, and *Y37E3.8*; however, SBP2 was not identified. This result can be explained considering that ribosomal proteins and general translation factors are abundant, whereas SBP2 expression in *C. elegans* is expected to be low considering that it is used for insertion of just one Sec residue. Furthermore, SBP2 is thought to be the limiting factor for Sec incorporation (Mehta et al. 2004).

To overcome the low endogenous levels of SBP2, we generated the transgenic strain IH11, ciuEx3[pLO109(Psur-5:: *scbp-2::polyHis*); *pRF4(rol-6(su1006))*] that overexpresses SBP2 with a C-terminal poly-histidine extension under the control of the sur-5 promoter to assure constitutive expression. Aqueous extracts of adult *ciuEx3* and wild-type control animals obtained from liquid culture were chromatographed on Ni-NTA magnetic agarose beads. Two differential bands, present in the *ciuEx3* extract, but absent in the wild-type extract, were retained by the metal affinity magnetic beads (Fig. 6). One of them corresponded to the expected size of the full-length SBP2, starting from the AUU codon in the appropriate Kozak context. The other poly-His-enriched protein present in the transgenic strain may be associated with SBP2 and copurify with it or be a truncated form of SBP2 that retained the poly-His tail. Soluble mutated trxr-1 SECIS



**FIGURE 5.** *C. elegans* SBP2 deletion mutant strain is unable to incorporate Sec. Metabolic incorporation of radioactive <sup>75</sup>Se showed that the deletion mutant strain *scbp-2(tm3417)* was unable to incorporate Se in TRXR-1 (lane 3) as well as the deletion mutant strains *trxr-1 (sv47)* and *efsc-1(sv36)* (lanes 2 and 4, respectively). The wild-type strain N2 incorporates Se in the only selenoprotein expressed in *C. elegans*: TRXR-1 (lane 1), shown by the bold arrow. The dashed arrow shows *E. coli* formate dehydrogenase. The bands at the low part of the gel show unbound Se and Se-labeled tRNA in *E. coli*.



**FIGURE 6.** Isolation of an 18-kDa band by SECIS affinity chromatography. Sequential metal and SECIS affinity chromatography of aqueous extracts of the wild-type strain N2 (as control) and the transgenic strain IH11, *ciuEx3[pLO109(Psur-5::scbp-2::polyHis); pRF4(rol-6 (su1006))]* overexpressing *sbp2*. An 18-kDa protein band is purified from the IH11 extract (arrows), but not from the control. This mass corresponds to the expected to SBP2 if translation starts from the proposed AUU codon. (Lane 1) Molecular weight marker (MWM). (Lanes 3,4) Eluates from the Ni<sup>+2</sup> column. (Lanes 6,7) Flow through the wtSECIS element column. (Lanes 9,10) Eluates from the wtSECIS element column from wild type and *ciuEx3*, respectively (containing soluble mutSECIS). One asterisk shows the streptavidin monomer and two asterisks show the streptavidin dimer.

element (mutSECIS, having mutations at the non-Watson-Crick quartet, required for SBP2 binding [Fletcher et al. 2001; Allmang et al. 2002], and at the conspicuous AAR triplet present at the apical loop; see Materials and Methods for sequence) was added to the proteins eluted from the Ni-NTA magnetic beads, in order to decrease binding of RNA-binding proteins to wtSECIS in the following chromatographic step. The mixes were applied to a wtSECIS column as previously described, and retained proteins were analyzed by SDS-PAGE (Fig. 6). The enriched 18- and 16-kDa bands present in the transgenic strain, but absent in the N2 strain, were bound to wtSECIS. These results support the noncanonical initiation of translation of SBP2.

#### Selenoprotein extinction in plant nematode parasites

Since *Caenorhabditis* lineage has a single Sec residue in its proteome and its SBP2 lacks SID, we examined the Sec incorporation machinery in additional nematode lineages with completely sequenced genomes. These lineages corresponded to the clade III (*Spirurina*, a clade that includes several animal parasitic nematodes) and clade IV (*Tylenchina*, a clade that includes plant parasitic nematodes). We first searched for the Sec incorporation trait using SecS and PSTK genes as signatures of the trait. If the trait was present, we then searched for their SBP2 and other genes involved in Sec incorporation.

For the *Spirurina* lineage, we analyzed *Ascaris suum*, *Brugia* malayi, *Dirofilaria immitis*, and *Loa loa* genomes. These nematodes fully conserved the ability to incorporate Sec. However, they encoded a longer SBP2 than the *Caenorhabditis* 

proteins, between 211 and 228 amino acids long, but shorter than other invertebrate SBP2 (Fig. 7). The N-terminal extension of *Spirurina* SBP2 is rich in basic amino acid residues. Similar to *Caenorhabditis* SBP2, *Spirurina* SBP2 consist of an RBD and a C-terminal motif conserved in nematodes but not in other eukaryotes, but they lack the SID. All analyzed spirurina *sbp2* had a canonical initiation of translation. Interestingly, the data from transcriptome for additional nematode lineages indicate that they possess more than one selenoprotein (Taskov et al. 2005) (data not shown).

We then examined *Meloidogyne hapla*, *Meloidogyne incognita*, and *Globodera pallida* genomes of the *Tylenchina* clade. These plant parasitic nematodes have lost the machinery for Sec incorporation: They lack *secs-1* and *pstk-1* (Fig. 8). Furthermore, they lacked both Sec and Cyscontaining *trxr* genes; instead, they appear to rely on *gsr* as the core enzyme for disulfide reduction pathways. Inter-

estingly, plants neither incorporate Sec nor encode large thioredoxin reductases. These results reinforce the idea that thioredoxin reductase is the selenoprotein that maintains the selenium incorporation trait in the nematode lineage and suggest adaptations of parasite metabolism governed by the host metabolism. Finally, the analysis of the other genes involved in Sec incorporation in plant parasitic nematode genomes indicated the loss of all genes involved in Sec incorporation, including SPS, but the presence of SecP43. This finding suggests that SecP43 might be involved in processes other than Sec incorporation, or that in this lineage this gene acquired a new function.

#### DISCUSSION

SBP2 is a key player in dictating Sec incorporation; however, its essentiality for Sec incorporation has not been conclusively demonstrated. *C. elegans* SBP2 had not been previously identified, remaining as an elusive gene. We identified *C. elegans* SBP2 by sequence comparison analysis and proved that the identified gene is essential for Sec incorporation in TRXR-1, the only Sec-containing protein in *C. elegans*. Furthermore, the larval arrest phenotype observed when *gsr-1* expression was silenced on *trxr-1* or *efsc-1* mutants was also observed in the SBP2 deletion mutant strain. Caution should be taken to avoid generalizations regarding SBP2 essentiality in Sec incorporation for all selenoproteins in any lineage.

*C. elegans* SBP2 is the shortest eukaryotic SBP2 described so far: It lacks a SID and comprises exclusively an RBD. Other *Caenorhabditis* spp. also have a short SBP2 without

C_elegans C_remanei C_sp5 C_briggsae C_priggsae		
C_spii C_brenneri		
C iaponica		
L loa	MKRFESDEKDIFIRNPLKRKPLLGEFITPKSLATKEEGAKHOKGKOF	47
B malayi	MESEEKESRRKKNFVRNRLDHKPLFGQFISPKLLATEQKDAKNQKEKOE	49
Dimmitis	MSEFEFYKKKKFA-NKKNHDLLLADFITPNLLDAKPKDAKHQEVKQE	46
Āsuum	MVVRGGTMGSPNAISQQQNINSEIKKPVERKAKKGISPLKKAILRDKSERCFNEVGGAQD	60
C elegans	ISET INVTAN IDNLAFEL IKRLKSQQDKNRKGNADEKNFLKRRVICA	48
Cremanei	IPHFINKPSEIDNLVFEFLKCLKNEQDLKRNGKSDEKNFQNRRIVCCA	48
C sp5	IASSINERFAIDNLVFAFLKSLKNQQDKNRVDDKKEKNAQNRRIVCCA	48
C_briggsae	IAEPTNERSKIDNLAYAFIKNLKNQQDKKRVADKKEKNNQGRILLCCA	48
C_sp11	IAGKSPETEKI <mark>D</mark> NLAFEFLKKL <mark>K</mark> TQQEKKNEENTKEKNFQKRRVVC <b>C</b> A	48
C_brenneri	ITPQANECSKI <mark>D</mark> KLAAELLTKL <mark>K</mark> ERQDKTRKENSLEKSLQKKRLVFCA	48
C_japonica	IAINCSE-SEIDILVSQLLRKLKRQQDKKSQGNTEEKNFQKRRIVCCA	47
L_loa	KNDTTVESDKFDPFNLTAKDADVAVIRFLKKVKKFQDSAWKKNPIKAK-SKHRLVYCL	104
B_malayi	KDDRSFESDKLDPSSLIKSNADTAIIRFLKKVKKFQDSAWKKNPIKAK-SKPRLVYCL	106
D immitis A_suum	SNSNAVDLDKSEPFNLTIKDTDAAVIEFLEKINKFQDRAWKKNPTKAK-AKRRLVYGL GGASSGNSNQICGSVTAPSPTLLDTAILKLLKKLNSFYDRAYKRNPVKAK-SKRRFVYGL	103 119
C_elegans C remanei	HDTLKFAQSGRVKLIFYAKNMDENNLRAASNFHSLQIVOPLAGIPMIEVLTRKELSRI HDTLKNLERGKIKLIFYAKNLNVEENNVRAASNFHSLQKLOPIEKIPLVEALTRKEMSRI	106 108
C_sps	HIGHLAND AND AND AND AND AND AND AND AND AND	106
C_priggsae		106
C_spii C_bronneri		106
C japonica	HEALKYAOTGKURMUTHIKNLDEKNURASSNEHSLORIOPINEIPLUEAMTKKELSRI	105
L loa	REVEKOLLETVECVIIARDIETNKC-IKLAAEIDAIKEMOASNHINIWISSKNDLGRA	163
B malavi	REIRKOLLLETVOCVILARDIEIEKTKLATKVDTIREIGASNHINILWISSKHDLAKA	164
D immitis	KOVRKOLSLETVRCVILARDIDTNKS-IELSTDVDMIKEVCASNHIDILWISSKHNLAKV	162
A_suum	H <mark>e</mark> vrrílvlstarcvvl <b>a</b> kdievvsekeklgsevatvrkl <mark>e</mark> eennvpllefskkhklgna	179
C elegans	MNKFRYWAVVGVIDFSGFERETEDIWKNWKIHDSCKLFRIMOSSILP	153
C remanei	MNKFPYVGVVGVMDFOHFEKGAEKIVDNWRRSDSFKLFHIDOSSLLP	155
C sp5	MNKFEYVGVVGVLNWSGFEKEAEGIIRAWKMSDSYKQFHTDQSSLLP	153
C_briggsae	VNKFEYVGVIGVFNFHGFDDEAKEIIQEWRTSSSFKQFHTDQSSLLP	153
C spl1	MNKFEYWGVVAVTDFSGFEKEAQEIIKLWRQSDSFKLHHIDQSSILP	153
C brenneri	ANKYEYWGVIGVLDVSLVESIANEIVTEWRSSDSYKLYHIDQSSLLP	153
C_japonica	VNKFEYWSVIGVINYEGFESEFDEVVREWRSSDSFRNFHIDQSSILP	152
L_loa	VNKWEVWSAVALLDYRGAEDEYSAAVQTLADFMACCRIADHADQSSLFA	212
B_malayi	VDKWEIVSAIAILDYRGAEDAYSEAMHTLTNFMACCRIVEHTDQPSLLA	213
D_immitis	VKKWPIVSAVAILDYSGAEDAYNLAIQTLANFTACCRIADHVDQPSFSV	211
A suum	I.NAREFUSAVAVI.DYSGAEDI.YFVVMEVHKOKERSBRIGDER OPSTI.A	228

**FIGURE 7.** Sequence alignment of the identified nematode SBP2s. *Spirurina* SBP2 are longer than *Caenorhabditis* SBP2s; however, all identified nematode SBP2 lack the SID. The N-terminal extension of *Spirurina* SBP2s is rich in basic residues. Black blocks indicate positions which have a single, fully conserved residue. Dark gray and light gray blocks indicate conservation between groups of strongly or weakly similar properties, respectively. Numbers indicate the amino acid residue in the polypeptide chain.

SID. Remarkably, SID is completely absent from the genomes of these organisms. The absence of SID emphasizes that adjustments in the Sec incorporation machinery occurred in some eukaryotes and questions the proposed minimal requirements for Sec incorporation in eukaryotes. We compared other key proteins involved in Sec decoding to test whether additional and/or compensatory adjustments occurred in these sequences, but found no differences with these proteins in other lineages. Although speculative, the possibility that SBP2 could interact with the SECIS indirectly, via protein–protein interactions with another genuine SECIS-binding protein, cannot be ruled out. The minimal requirement of an RBD for SECIS binding in nematodes provides a potential clue to identify the archaeal missing link in Sec decoding. Indeed, archaeal selenoprotein mRNAs contain the SECIS element in the 3' UTR, but the SECIS-binding protein has not been identified yet (Rother et al. 2001).

The *C. elegans* SBP2 mRNA does not begin translation at an AUG codon. This is a striking feature since noncanonical translation initiation has not been previously described for this well-studied animal model. The evidence we have obtained is compelling: Translation from the first AUG codon leads to an incomplete RBD, the ORF and the amino acid sequence homology extends well beyond the first methionine, a



**FIGURE 8.** Sec incorporation and SBP2 in the nematode lineage. Nematodes from the clades III, IV, and V were examined for the presence of Sec incorporation machinery. Nematodes from the clade IV have lost the ability to incorporate Sec. Nematode SBP2s are the shortest eukaryote SBP2. For comparison *D. melanogaster* and human SBP2s are included in the phylogenetic analysis. RBDs are represented in light upward diagonal and Sec incorporation domains in wide upward diagonal.

different amino acid is present at the homologous position in other *Caenorhabditis* spp., and the Ribo-seq data available indicate translation upstream of the first AUG. The evidence for translation initiation from an AUU is also compelling: This codon is absolutely conserved together with the appropriate *Caenorhabditis* Kozak consensus  $^{-4}AAA^{-1}$  sequence, amino acid conservation extends until this codon, and the biochemical evidence obtained from the transgenic strain expressing SBP2 is consistent with the length expected for translation initiation from this codon. As this is the first gene that has been shown to have a non-AUG initiation of translation, we anticipate that examination of full-length mRNAs with long 5' UTRs will lead to correction of additional gene models and would provide further clues regarding noncanonical initiation of translation in *C. elegans*.

Analyzing the genome sequences of other nematodes, we found that the Spirurina lineage also possessed short SBP2 that lack SID, reinforcing the idea that the RBD may suffice for Sec incorporation. Interestingly, the RBD of Spirurina SBP2 is preceded by a K/R rich region. This fact may reflect the need for recognizing more than one SECIS element. Another remarkable finding of our study is that nematode plant parasites that belong to the Tylenchina clade have lost the Sec trait. Sec extinction in these nematodes does not appear to be a recent event: All essential genes in Sec incorporation were lost. We could only identify SecP43 as the only gene predicted to be involved in Sec incorporation, suggesting that it might be involved in some other cellular processes in this lineage. Furthermore, rather than a Sec $\rightarrow$ Cys conversion, we observed selenoprotein gene loss, which may also indicate a distant and irreversible event in selenoprotein extinction. In this context, it is interesting to note that plants do not incorporate Sec, and thus selenoproteinless nematodes likely evolved as an adaptation of this lineage to its selenium-incompatible niche. So far, the only reported Metazoa unable to incorporate Sec were a few insect species (Chapple and Guigó 2008; Lobanov et al. 2008). This finding enlarges the group of selenoproteinless animals, stressing a highly dynamic evolutionary process and highlighting that in the nematode lineage Sec incorporation is maintained by thioredoxin reductase.

Interestingly, in some selenoproteinless insects SBP2, SecP43, and SPS1 were still conserved (Chapple and Guigó 2008). Although SPS are responsible for the synthesis of selenophosphate from selenide and ATP, it has been suggested that SPS1 functions in a pathway unrelated to Sec biosynthesis (Lobanov et al. 2008). Our results portray a slightly different scenario in nematodes: Both *sps* genes are absent in plant parasite nematodes, and only one *sps* gene is present in Secincorporating nematodes. This SPS is the ortholog to metazoan SPS2 (M Mariotti, pers. comm.) but contains Cys instead of Sec at its active site.

Our results indicate that SID is not required for Sec decoding in nematodes and that nematodes provide an interesting scenario to study Sec incorporation, selenoproteomes, selenoproteins, and Sec/Cys evolution. Finally, the demonstration that a non-AUG translation initiation occurs in *C. elegans* is an important observation that should lead to revision in protein sequences, beyond SBP2, in this organism.

#### MATERIALS AND METHODS

#### C. elegans strains and culture conditions

Bristol-N2 strain was obtained from the CGC Center; VB1414 *trxr-1* (*sv47*) IV, VB0740 *efsc-1(sv36*) I, VB1646 *rrf-3(pk1426*) II; *trxr-1* (*sv47*) IV, and VB1286 *rrf-3(pk1426*) II; *efsc-1(sv36*) I strains were kindly provided by Dr. Simon Tuck (Umeå University, Sweden); TM3417 *scbp-2(tm3417*) I was obtained from the *C. elegans* National Bioresource Project of Japan. *Escherichia coli* strains

OP50 and HT115 were obtained from the CGC. RNAi clones were kindly provided by Dr. Peter Askjaer (Universidad Pablo de Olavide, Spain). The standard methods for culturing and maintaining *C. elegans* were used as described in Brenner (1974). All experiments were performed at 20°C. To obtain IH2 *scbp-2(tm3417)* I strain, TM3417 *scbp-2(tm3417)* I was backcrossed six times with N2 strain.

#### Amplification of C. elegans sbp2 cDNA

RNA from N2 embryos was kindly provided by Dr. Peter Askjaer (Universidad Pablo de Olavide, Spain). Reverse transcription was performed using the gene-specific reverse primer (TCACGGCAA GAGGGAAG). Amplification of *sbp2* cDNA was carried out using the same specific reverse primer and a forward primer (GGTTT AATTACCCAAGTTTGAG) derived from the SL1 exon *trans*spliced to the 5' end of the full-length mRNAs. The PCR product was analyzed by electrophoresis, purified from 1.5% agarose gel, cloned into pGEM-T Easy (PROMEGA), and sequenced.

#### Identification of proteins involved in Sec decoding in *C. elegans* and nematodes

Searches for genes and proteins involved in Sec decoding in *C. elegans* were performed using as queries human tRNA<sup>Sec</sup>, EfSec, PSTK, L30, SecS, SecP43, SerS, SPS1, SPS2, and SBP2. Additionally, *D. melanogaster* sequences were used as queries for some genes and proteins: SPS1, SPS2, SBP2-like and the protein sequences encoded by genes present in Sec-incorporating organisms and absent in nonincorporating ones: Q7KUF9, Q8T3V7, and Q9VEH4. These sequences were used in blastn, tblastn, or blastp searches against WormBase (WS238). In all cases, the hit sequence was retroblasted to obtain the best reciprocal hit. For identifying *C. elegans* SBP2, additional searches were carried out using several invertebrate SBP2 sequences independently and concatenated.

SecS and PSTK genes were used as signature for maintenance of the Sec incorporation trait in other members of the *Caenorhabditis* genus and in other nematodes. *C. elegans* genes were used as queries to identify nematode orthologs searching by blastp and tblastn using WormBase (WS238). To confirm the absence of the trait in plant nematode parasites, the following databases were also examined: *M. hapla* 10× (Plant Nematode Genomics Group, North Carolina State University), *M. incognita* genome, proteins and dbEST (*Meloidogyne incognita* resources, INRA), and *G. pallida* predicted genes and contigs (Wellcome Trust Sanger Institute).

*C. brenneri, C. briggsae, C. japonica, C. remanei, C. sp5*, and *C. sp11* SBP2 were obtained by tblastn using *C. elegans* SBP2 as query and WS238 databases. Other nematode SBP2 were obtained by tblastn using *C. elegans* SBP2 as query and WormBase (WS238), nematode.net v3.0 (http://nematode.net/NN3\_frontpage.cgi), UniProtKB (http://www.uniprot.org/), 959 Nematode Genomes (http://www. nematodes.org/nematodegenomes/index.php/Main\_Page) databases. Once a new nematode SBP2 sequence was obtained, it was also used as query for searching for SBP2 of closely related nematodes.

#### Multiple sequence alignments

Multiple sequence alignments were performed using clustalw program and its default parameters at the European Bioinformatics Institute website: http://www.ebi.ac.uk/Tools/msa/clustalw2/. All alignments were manually refined.

#### SECISearch analyses of nematode sequences

The 3' UTR sequences of the Sec-containing *trxr-1* of *Caenorhabditis* species were obtained from WormBase (WS238) via blastn tool, using *C. elegans* TRXR-1 sequence as query and *C. brenneri*, *C. briggsae*, *C. japonica*, *C. remanei*, *C. sp5*, and *C. sp11* genome databases. The obtained 3' UTR sequences were then analyzed using SECI-Search3 tool for prediction of SECIS elements (Mariotti et al. 2013) at http://seblastian.crg.es/ using infernal method and threshold 10.

# Identification of homologs of known selenoprotein genes

By blastp and tblastn similarity searches, we identified homologs to the following eukaryotic selenoproteins in *Caenorhabditis* spp.: glutathione peroxidases (Gpx), selenoprotein 15 (Sel15), selenophosphate synthetase (SPS), selenoprotein K (SelK), selenoprotein R (SelR), selenoprotein T (SelT), selenoprotein U (SelU), thioredoxin reductases (TRXR-1 and TRXR-2). No homologs were found to other eukaryotic selenoproteins. In all cases, except TRXR-1, these homologs contained Cys at the homologous Sec position in mammals.

#### Metabolic labeling with <sup>75</sup>Se

*E. coli* OP50 strain was grown overnight at 37°C in 5 mL Luria Broth (LB) medium supplemented with cysteine and Na<sub>2</sub>SeO<sub>3</sub> radiolabeled with <sup>75</sup>Se (20 µCi of <sup>75</sup>Se). Two hundred microliters of the overnight culture was spread onto a Nematode Growth Media (NGM) agar plate and allowed to grow for 24 h at room temperature (~20°C). Then, the plate was seeded with 10 L4 hermaphrodite worms of the appropriate strain. The worms were harvested shortly before they had consumed all of the bacteria (~7 d). Worms were then collected, washed, and boiled in sample buffer. The total protein extracts were analyzed by SDS-PAGE and then transferred to a PVDF membrane. Radioactive signal was detected with a PhosphorImager.

#### **RNAi** assays

HT115 *E. coli* strain was transformed with the plasmid pL4440 containing the sequence of the gene to be down-regulated or empty vector as a control. The bacteria were first grown overnight at 37°C in LB medium containing 100  $\mu$ g/mL ampicillin. Then, they were seeded in plates containing NMG agar with 100  $\mu$ g/mL ampicillin and 1 mM IPTG and incubated overnight at 37°C to induce the expression of dsRNA. Worms were then seeded onto plates at 20°C and phenotypes scored from the first generation onward by allowing interfered gravid hermaphrodites to lay eggs during 2 h on fresh RNAi plates.

#### Sequential Ni<sup>+2</sup> SECIS chromatography

Adult N2 and IH11 *ciuEx3[pLO109(Psur-5::scbp-2::polyHis); pRF4* (*rol-6 (su1006))]* worms were harvested from liquid culture and extensively washed. Worms were ground to a powder under liquid
nitrogen using a mortar and a pestle, resuspended in buffer containing 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, 0.05% Tween and 1 mM PMSF, sonicated, and the suspension centrifuged at 20,000g for 20 min. Six milliliters of the aqueous protein extract of 15.71 mg/mL of N2 and 16.60 mg/mL of IH11 were added to 120  $\mu$ L Ni-NTA Magnetic Agarose Beads (Qiagen), washed, and eluted in 150  $\mu$ L following manufacturer's recommendations and protocol.

The buffer of the eluates obtained from the Ni<sup>+2</sup> column was exchanged to PBS using an Amicon Ultra-0.5 (Millipore) cartridge in order to decrease imidazole concentration (<10 µM). To 85 µL of the resulting protein solution we added 2 µL Ribolock RNase Inhibitor (Thermo) to avoid RNA degradation and 0.5 µL of 100 mM mutSECIS: trxr-1 SECIS element that has the critical SBP2binding site (the non-Watson-Crick quartet), and the typical AAR triplet mutated (uucacacgaccuuuggcuccacuccaucgugagcgccucuggu cugaugc) to minimize unspecific binding to the wild-type SECIS element (wtSECIS) (uugugacgaccuuuggcuaaacuccaucgugagcgccucu ggucugaugc). The mixes were incubated with biotinylated wtSECIS immobilized on streptavidin magnetic beads (Thermo), following manufacturer's recommendations and protocol. After binding and washing, bound proteins were eluted adding SDS-PAGE sample buffer and boiled for 5 min. The N2 and IH11 eluates and the flow through fractions were analyzed by 15% SDS-PAGE and silver stained according to the protocol described in Shevchenko et al. (1996).

#### Generation of transgenic C. elegans strain

The *sbp2* sequence was amplified using genomic DNA (extracted from N2 adults), a specific forward primer (ACGTGCTAGCCTC TTCCTGTTTTATGAAATC), and a specific reverse primer that adds a C-terminal poly-His tag (TCGGTACCATGGTTAGTGAT GATGGTGATGATGCGGCAAGAGGGAAGATTG). The amplified sequence was analyzed and purified from 1% agarose gel and then cloned into pPD49.26 vector that has already cloned the *sur-5* promoter (kindly provided by Dr. Ana Bratanich, UBA) and sequenced. This construct was microinjected into N2 L4 larvae gonads along with pRF4 plasmid—that carries the *rol-6(su1006)* dominant transformation marker—to generate stable transgenic lines (50 ng/µL each construct).

#### SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

#### ACKNOWLEDGMENTS

We thank Dr. Peter Askjaer (Universidad Pablo de Olavide, Spain) for provision of *C. elegans* embryos mRNA and helpful discussions, Dr. Simon Tuck (Umeå University, Sweden) for kindly providing stains, Dr. Ana Bratanich (UBA, Argentina) for providing the pPD49.26 vector, and Dr. Briseida Cacho-Valadez for experimental assistance. We are grateful to the Caenorhabditis Genetics Center (CGC) and the Japanese National Bioresource Project for the Experimental Animal Nematode *C. elegans* for provision of *E. coli* and *C. elegans* strains. We also thank the instructors and assistants of the CSHL *C. elegans* course attended by G.S. This work was sup-

ported by Universidad de la República, Uruguay (Grant Number 418 to G.S., PhD fellowship to L.O.); Asociación Española de Cooperación Internacional (C/7646/07 to A.M.-V. and G.S.; A/016083/08 to A.M.-V. and G.S.); Asociación Universitaria Iberoamericana de Posgrado and Agencia Nacional de Innovación e Investigación (BE\_POS\_2009\_183 and BE\_POS\_2010\_2160 to L.O.), and was partially funded by FOCEM (MERCOSUR Structural Convergence Fund), [COF 03/11].

Received December 9, 2013; accepted March 6, 2014.

#### REFERENCES

- Allmang C, Carbon P, Krol A. 2002. The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *RNA* 8: 1308–1318.
- Allmang C, Wurth L, Krol A. 2009. The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated. *Biochim Biophys Acta* 1790: 1415–1423.
- Berry MJ, Larsen PR. 1993. Recognition of UGA as a selenocysteine codon in eukaryotes: a review of recent progress. *Biochem Soc Trans* 21: 827–832.
- Blumenthal T, Steward K. 1997. RNA processing and gene structure. In *C. elegans* II (ed. Riddle DL, et al.), pp. 117–145. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, Zinoni F. 1991. Selenocysteine: the 21st amino acid. *Mol Microbiol* **5**: 515–520.
- Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77**: 71–94.
- Budiman ME, Bubenik JL, Miniard AC, Middleton LM, Gerber CA, Cash A, Driscoll DM. 2009. Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation. *Mol Cell* 35: 479–489.
- Buettner C, Harney JW, Berry MJ. 1999. The *Caenorhabditis elegans* homologue of thioredoxin reductase contains a selenocysteine insertion sequence (SECIS) element that differs from mammalian SECIS elements but directs selenocysteine incorporation. *J Biol Chem* 274: 21598–21602.
- Carlson BA, Xu XM, Kryukov GV, Rao M, Berry MJ, Gladyshev VN, Hatfield DL. 2004. Identification and characterization of phosphoseryl-tRNA<sup>[Ser]Sec</sup> kinase. *Proc Natl Acad Sci* **101**: 12848–12853.
- Chapple CE, Guigó R. 2008. Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. *PLoS One* **3**: e2968.
- Chavatte L, Brown BA, Driscoll DM. 2005. Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat Struct Mol Biol* **12:** 408–416.
- Donovan J, Copeland PR. 2009. Evolutionary history of selenocysteine incorporation from the perspective of SECIS binding proteins. *BMC Evol Biol* **9**: 229.
- Donovan J, Copeland PR. 2010. Threading the needle: getting selenocysteine into proteins. *Antioxid Redox Signal* **12**: 881–892.
- Fagegaltier D, Hubert N, Yamada K, Mizutani T, Carbon P, Krol A. 2000. Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J* 19: 4796–4805.
- Fletcher JE, Copeland PR, Driscoll DM, Krol A. 2001. The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA* **7**: 1442–1453.
- Gladyshev VN, Krause M, Xu XM, Korotkov KV, Kryukov GV, Sun QA, Lee BJ, Wootton JC, Hatfield DL. 1999. Selenocysteine-containing thioredoxin reductase in *C. elegans. Biochem Biophys Res Commun* 259: 244–249.
- Grundner-Culemann E, Martin GW III, Harney JW, Berry MJ. 1999. Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. *RNA* **5:** 625–635.

Hastings KE. 2005. SL *trans*-splicing: easy come or easy go? *Trends Genet* **21**: 240–247.

- Hatfield DL, Gladyshev VN. 2002. How selenium has altered our understanding of the genetic code. *Mol Cell Biol* 22: 3565–3576.
- Kennedy S, Wang D, Ruvkun G. 2004. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans. Nature* 427: 645–649.
- Lee BJ, Worland PJ, Davis JN, Stadtman TC, Hatfield DL. 1989. Identification of a selenocysteyl-tRNA<sup>Ser</sup> in mammalian cells that recognizes the nonsense codon, UGA. J Biol Chem 264: 9724– 9727.
- Leinfelder W, Zehelein E, Mandrand-Berthelot MA, Bock A. 1988. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* 331: 723–725.
- Lobanov AV, Hatfield DL, Gladyshev VN. 2008. Selenoproteinless animals: selenophosphate synthetase SPS1 functions in a pathway unrelated to selenocysteine biosynthesis. *Protein Sci* **17:** 176–182.
- Mariotti M, Ridge PG, Zhang Y, Lobanov AV, Pringle TH, Guigo R, Hatfield DL, Gladyshev VN. 2012. Composition and evolution of the vertebrate and mammalian selenoproteomes. *PLoS One* 7: e33066.
- Mariotti M, Lobanov AV, Guigo R, Gladyshev VN. 2013. SECISearch3 and Seblastian: new tools for prediction of SECIS elements and selenoproteins. *Nucleic Acids Res* **41**: e149.
- Mehta A, Rebsch CM, Kinzy SA, Fletcher JE, Copeland PR. 2004. Efficiency of mammalian selenocysteine incorporation. *J Biol Chem* **279:** 37852–37859.
- Michel AM, Fox G, M Kiran A, De Bo C, O'Connor PBF, Heaphy SM, Mullan JPA, Donohue CA, Higgins DG, Baranov PV. 2014. GWIPSviz: development of a ribo-seq genome browser. *Nucleic Acids Res* 42 (Database issue): D859–D864.
- Miniard AC, Middleton LM, Budiman ME, Gerber CA, Driscoll DM. 2010. Nucleolin binds to a subset of selenoprotein mRNAs and regulates their expression. *Nucleic Acids Res* 38: 4807–4820.
- Rhoads RE, Dinkova TD, Korneeva NL. 2006. Mechanism and regulation of translation in *C. elegans*. In *WormBook* (ed. The *C. elegans*

Research Community), pp. 1–18. doi: 10.1895/wormbook.1.63.1, http://www.wormbook.org.

- Rother M, Resch A, Gardner WL, Whitman WB, Bock A. 2001. Heterologous expression of archaeal selenoprotein genes directed by the SECIS element located in the 3' non-translated region. *Mol Microbiol* 40: 900–908.
- Shevchenko A, Wilm M, Vorm O, Mann M. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 68: 850–858.
- Simmer F, Tijsterman M, Parrish S, Koushika SP, Nonet ML, Fire A, Ahringer J, Plasterk RH. 2002. Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr Biol* 12: 1317–1319.
- Small-Howard A, Morozova N, Stoytcheva Z, Forry EP, Mansell JB, Harney JW, Carlson BA, Xu XM, Hatfield DL, Berry MJ. 2006. Supramolecular complexes mediate selenocysteine incorporation in vivo. *Mol Cell Biol* 26: 2337–2346.
- Stenvall J, Fierro-Gonzalez JC, Swoboda P, Saamarthy K, Cheng Q, Cacho-Valadez B, Arner ES, Persson OP, Miranda-Vizuete A, Tuck S. 2011. Selenoprotein TRXR-1 and GSR-1 are essential for removal of old cuticle during molting in *Caenorhabditis elegans*. Proc Natl Acad Sci 108: 1064–1069.
- Takeuchi A, Schmitt D, Chapple C, Babaylova E, Karpova G, Guigo R, Krol A, Allmang C. 2009. A short motif in *Drosophila* SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins. *Nucleic Acids Res* 37: 2126–2141.
- Taskov K, Chapple C, Kryukov GV, Castellano S, Lobanov AV, Korotkov KV, Guigo R, Gladyshev VN. 2005. Nematode selenoproteome: the use of the selenocysteine insertion system to decode one codon in an animal genome? *Nucleic Acids Res* 33: 2227–2238.
- Xu XM, Carlson BA, Irons R, Mix H, Zhong N, Gladyshev VN, Hatfield DL. 2007a. Selenophosphate synthetase 2 is essential for selenoprotein biosynthesis. *Biochem J* 404: 115–120.
- Xu XM, Carlson BA, Mix H, Zhang Y, Saira K, Glass RS, Berry MJ, Gladyshev VN, Hatfield DL. 2007b. Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol* 5: e4.

**ERRATUM** 

RNA 20: 1023–1034 (2014)

# Adjustments, extinction, and remains of selenocysteine incorporation machinery in the nematode lineage

# LUCÍA OTERO, LAURA ROMANELLI-CEDREZ, ANTON A. TURANOV, VADIM N. GLADYSHEV, ANTONIO MIRANDA-VIZUETE, and GUSTAVO SALINAS

In the above-noted article, on p. 1027, column 2, the statement, "The analysis of *trxr-1* SECIS elements using the SECISearch3 program (Mariotti et al. 2013) revealed the presence of the unusual type I SECIS element ...," should instead state, "The analysis of *trxr-1* SECIS elements using the SECISearch3 program (Mariotti et al. 2013) revealed the presence of the type II SECIS element ...,"

The authors note that at the time that they had performed the analysis, the program's output provided a general-type I-like structure, instead of the actual SECIS-type structure. Thus, they have reanalyzed the *Caenorhabditis* SECIS elements in the updated and corrected version of SECISearch, marking the new box "predict SECIS-type." Accordingly, the third sentence in the Figure 4 legend (p. 1028) should be replaced by: "All *Caenorhabditis* encode a type II SECIS element with a guanine immediately upstream of the non-Watson-Crick base-pairing quartet. Critical binding sites for SBP2 are highlighted."

The authors thank Dr. Marco Mariotti (Centre de Regulació Genòmica, Universitat Pompeu Fabra, Barcelona, Spain), developer of the SECISearch3 program, for pointing out this error in Figure 4 of this paper.

## 4.1 Primer anexo

Comentarios de los revisores al manuscrito presentado a RNA y respuestas de los autores:

### 4.1.1 Reviewer 1 Comments for the Author

This manuscript describes a thorough analysis of selenocysteine utilization in the nematode lineage based on a combination of bioinformatics and experimental studies. This work is very interesting and reports many novel observations, including the discovery of nematode SBP2s, the finding that these proteins lack the SID domain (previously thought to be essential for Sec insertion), the observation that nematode SBP2 synthesis is initiated from a non-canonical start codon (representing the first such case in nematodes), and the discovery of nematodes that lost the entire selenocysteine trait. Presentation of the data could be improved in several places, as detailed below.

#### Figures

<u>**Reviewer**</u>: Fig. 3. The left and right parts of the figure legend are missing, and the figure is also cropped on one side.

**Reply**: Corrected as indicated.

**<u>Reviewer</u>**: Fig. 1. The isoleucine ATT codon predicted to initiate protein synthesis is not in a favorable Kozak sequence, whereas the next ATT is. A leucine codon downstream is also in a favorite Kozak, although conservation of sequences shown in Fig. 3 suggests that the second ATT is the likely initiator codon. Also, the bold squire in Figure 1 is not a squire.

**<u>Reply</u>**: Both isoleucines are fully conserved. However the proposed first lle codon is located in a more favorable translation initiation context (see below the nucleotide sequence alignment for *Caenorhabditis* species) [117]. The optimal translation initiation context for *Caenorhabditis* is aaaaAUG, (lowercase letters indicate -1 to -4 nucleotides, see refence in the manuscript). This context is different than vertebrates Kozak sequence, yet it is also named "Kozak context". Below, we show and compare the nucleotide context of both lle codons for all *Caenorhabditis* species:

#### Sequence context of the manuscript proposed lle initiation codon:

Kozak initiation context for C. elegans

- C. elegans
- C. sp11
- C. japonica
- C. briggsae
- C. sp5
- C. brenneri
- C. remanei

C. brenneri C. remanei

а	aaaAUG
u	aaaAU <mark>U</mark>
а	aaaAU <mark>U</mark>
u	aaaAU <mark>U</mark>
а	aaaAU <mark>U</mark>
g	aaaAU <mark>U</mark>
g	aaaAU <mark>U</mark>
а	aaaAU <mark>U</mark>

#### Sequence context of the lle initiation codon suggested by the reviewer:

Kozak ignition context for *C. elegans C. elegans C. spl1 C. japonica C. briggsae C. sp5* 

aaaaAUG uaacAUU aaaaAUC agaaAUU uaaaAUU cgctAUC caaaAUC ugaaAUC

Bold square changed to Bold frame.

**<u>Reviewer</u>**: Fig. 4. SECIS elements, not SECIS element. Two sentences are shown in bold as the title. Combine them. Some C. for *Caenorhabditis* is not shown in italics in the legend.

**Reply**: Corrected as indicated.

**<u>Reviewer</u>**: Fig. 5. Improve the title of this figure. Also, lane designations should be made clear (for instance, showing them diagonally). The entire figure legend should also be improved. For instance, commas could separate lane 1 and its description (this applies to other lanes too).

Reply: Corrected as indicated.

<u>**Reviewer**</u>: Figure 6. Like figure 1, this figure legend is cut on the sides. The description should be improved.

Reply: Corrected as indicated

#### Introduction

<u>**Reviewer</u>**: The discovery of Sec-containing thioredoxin reductase in *C. elegans* is in Buettner et al., JBC, 1999 and Gladyshev et al. BBRC, 1999, but these studies are not cited.</u>

**<u>Reply:</u>** Corrected as indicated. Thank you for pointing this error.

#### Results

<u>Reviewer</u>: Page 6, line 1. Kinase should not be with a capital letter. <u>Reply:</u> Done.

<u>**Reviewer**</u>: Page 7. Riboseq should be Ribo-seq. In the same sentence, parenthesis should be corrected.

Reply: Corrected as indicated.

**Reviewer**: Page 8. The fact that *C. elegans* has only a single SPS is interesting, but not striking. I suggest removing this word.

<u>**Reply:**</u> Corrected as indicated.

**<u>Reviewer</u>**: In many places in the manuscript, either SBP2s contain or SBP2 contains should be used.

**<u>Reply:</u>** Corrected as indicated.

**<u>Reviewer</u>**: Page 10. Radiolabeled should be radioactive in the first sentence. On the same page, GSR-1 is abbreviated, but this abbreviation is already defined in the Introduction.

Reply: Corrected as indicated.

**Reviewer**: In addition, on page 12 as well as in Discussion, the full name of this protein is again used. Other abbreviations are used or not used freely in the manuscript, such as thioredoxin reductase and selenocysteine. This should be corrected.

**Reply:** Corrected as indicated in most places. In some instances it should be noted that the peculiar nomenclature of *C. elegans* genes does not allow the usual generic name to be used. In those cases when we refer specifically to the *C. elegans* gene we used the allowed nomenclature, but when we refer to the same gene in a generic manner we prefer to use the conventional name (eg when we referred to TRXR-1 we are referring to the worm gene since it has this name; however, when we refer to thioredoxin reductase in general we do not use that nomenclature). Another example is precisely SECIS binding protein 2: this protein is usually abbreviated as SECISBP2 or SBP2, but the gene that encodes *C. elegans* SBP2 is called *scbp-2*, so in some instances we must refer to it as SCBP-2 (*sbp* stands for sterol binding protein in *C. elegans*)

Thank you very much for pointing all unnoticed errors that generate nomenclature inconsistencies.

#### Discussion

**<u>Reviewer</u>**: The last sentence of the first paragraph should be corrected grammatically. Strikingly and remarkably are used too often in the manuscript. In most places, the observations are interesting, but this choice of words is too strong.

**<u>Reply</u>**: Corrected as indicated. Thank you very much for pointing the overuse of adjectives in the manuscript

# 4.1.2 Reviewer 2 Comments for the Author...

This is a well described study of selenoprotein synthesis in nematodes. The authors argue that because most nematodes express a single selenoprotein gene (TRXR-1), the lack of which is suprisingly non-lethal, these organisms constitute an optimal model system for studies of eukaryotic selenoprotein synthesis. Because that synthesis is most complex, as it involves its own dedicated elongation factor, a specific structure in the mRNA, redefinition of a UGA stop codon, and more, this is a compelling topic with general interest. The authors present a good study with a number of novel insights: SBP2 (or, rather, SCBP-2) was identified in the nematode to be an essential component for selenoprotein synthesis even if expressed in a shorter form than in higher eukaryotes, the SBP2 protein was found to be translated from a non-canonical ATT initiation codon (the first such shown in nematodes), and some nematode species were discovered to lack selenoproteins. These findings are all novel and interesting. Issues that require further attention include the following:

# Major issue

**Reviewer**: 1. The experimental validation of SECIS binding by the nematode SBP2 as shown in Fig 6 is not sufficient. In the materials section the authors state that a mutated SECIS variant did not bind SBP2 (p. 20). That important control and experimental result should be shown. Moreover, what is the major protein band at around 24 kDa seen in the SECIS eluates of both nematode species in Fig 6? How is the 75Se labeling of the *C. elegans* TRXR-1 affected by the overexpression of SBP2 in the IH11 strain, considering that stoichiometry of the selenoprotein synthesis machinery players has previously been shown to be important for full efficiency? Overall, better experimental validation of SBP2 function should be possible to provide, especially considering that it is surprising that full function should be maintained although the SID domain found in other species is apparently missing.

For the sake of clarity we divide this paragraph:

**<u>Reviewer</u>**: The experimental validation of SECIS binding by the nematode SBP2 as shown in Fig 6 is not sufficient. In the materials section the authors state that a mutated

SECIS variant did not bind SBP2 (p. 20). That important control and experimental result should be shown.

**<u>Reply</u>**: The mutated SECIS (mutSECIS) has specific mutations at two known key sites: aa $\rightarrow$ cc at the apical loop critical for Sec incorporation, and gug $\rightarrow$ cac at the non-Watson and Crick quartet that forms the stem and binds SBP2. Theoretically the former mutation should prevent SBP2 binding and the latter mutation would prevent the formation of the typical SECIS secondary structure. Therefore *C. elegans* SBP2 would be unable to recognize this oligoribonucleotide as a SECIS element. Consistent with this, the SECISearch program does not predict any SECIS element when mutSECIS is analyzed. Below we highlight the differences of the wild-type and mutated SECIS used in the experiment.

In the experiment presented in the manuscript we added soluble mutSECIS to avoid unspecific binding to RNA, since in a previous experiment we isolated from wild-type worms several RNA-binding proteins using either immobilized mutSECIS or wtSECIS. The result of that experiment showed that immobilized SECIS, *but not immobilized mutSECIS*, bound a protein that in 2D gels has a molecular weight around 18 kDa and an isoelectric point around 10, which is consistent with the predicted Mr and pl of the full length native SBP2 and with the results obtained in experiment with the transgenic strain (this latter has an additional C-terminal extension containing a polyHis tail, and therefore the Mr of the purified protein in that case is slightly higher). Unfortunately we could not obtained proteomic information for the isolated dot. We show below the 2D gels corresponding to this experiment. Considering this result and the nature of mutations we stated that the mutSECIS was unable to bind SBP2. Since formally we have rephrase this sentence to avoid misleading statements.

The evidence strongly suggests that *C. elegans* SBP2 binds wtSECIS: protein bands/spots of the expected molecular weight and isoelectric point are purified with the wtSECIS. Nevertheless, we concur with the reviewer that our experiments do not completely validate that SBP2 binds SECIS element. Unfortunately we could not obtain direct evidence of the identity of the isolated band/spot. Thus, we have rephrased the last sentence of page 11 from "These results support the non-canonical initiation of translation of SCBP-2, and also indicate that the transgenic strain produced an SBP2 that bound trxr-1 SECIS element" to "These results support the non-canonical initiation

of translation of SBP2, and also suggest that the transgenic strain produced an SBP2 that bound *trxr-1* SECIS element.".



<u>**Reviewer**</u>: Moreover, what is the major protein band at around 24 kDa seen in the SECIS eluates of both nematode species in Fig 6?

**<u>Reply</u>**: The bands at 13 kDa in lanes 9 and 10 correspond to streptavidin chain A of Streptavidin magnetic beads. The bands at 26 kDa in the same lanes correspond to streptavidin dimer. We add, for the reviewer, a 15% polyacrylamide gel with the same molecular weight marker as in figure 6, and the magnetic beads alone in SDS-page sample buffer. In the gel shown below both streptavidin bands can be seen and are marked with arrows. We have included this explanation in the figure legend. This is why those bands appeared in both *C. elegans* strains after the SECIS affinity purification. This result was confirmed by MS/MS of the 26 kDa bands from the gel showed in figure 6.



**<u>Reviewer</u>**: How is the 75Se labeling of the *C. elegans* TRXR-1 affected by the overexpression of SBP2 in the IH11 strain, considering that stoichiometry of the selenoprotein synthesis machinery players has previously been shown to be important for full efficiency?

**<u>Reply</u>**: The generation of transgenic strains by microinjection generates large extrachromosomal DNA arrays that are inherited and expressed variably among cells and among individuals. This mosaicism leads to the loss of the stoichiometry between SBP2 and TRXR-1, and therefore quantification in transgenic worms is usually not advised. In any case, we measured the total thioredoxin reductase activity (arising from both TRXR-1 and TRXR2, the mitochondrial Cys-containing enzyme) as a comparative estimation of TRXR-1 expression in extracts from SBP2 transgenic and wild type strains. No significant differences were found between strains. Taking into account this result and the mosaicism generated in transgenic strains we consider that metabolic incorporation of 75Se will not provide further information.

**<u>Reviewer</u>**: Overall, better experimental validation of SBP2 function should be possible to provide, especially considering that it is surprising that full function should be maintained although the SID domain found in other species is apparently missing.

**<u>Reply</u>**: We have provided additional indirect evidence for the reviewer, and modified the text of the manuscript to state clearly that, despite the indirect evidence, we do not demonstrate directly the binding SECIS to SBP2. In any case, we would like to emphasize that this experiment is not a key one for the manuscript, rather it was thought as a confirmatory one. The key experiments that allow us to conclude that SBP2 is a functional and required gene for Sec incorporation are those performed with the deletion mutant strain; which showed the absence of selenium incorporation in TRXR-1 and the larval arrest phenotype.

#### Minor issues

**<u>Reviewer</u>**: The authors switch back and forth the nomenclature of either "SCBP-2" or "SBP2" for the same protein, which is confusing. Please state at its first use what nomenclature should be used, and then use it consistently.

**<u>Reply</u>**: As we answered to reviewer 1, *C. elegans* has a peculiar nomenclature. In some instances we use the generic SBP2 name, but in some other instances, when we refer to the *C. elegans* gene we must refer to it as *scbp-2*.

<u>**Reviewer**</u>: p. 5, last sentence of introduction, please write "...a gene previously associated..." (add "previously"). <u>**Reply**</u>: Corrected as indicated

<u>**Reviewer</u>**: p. 6, middle, please write "...putatively involved only in Sec decoding..." (add "only") for increased clarity. <u>**Reply**</u>: Corrected as indicated</u>

**<u>Reviewer</u>**: Please rephrase the last sentence of the first paragraph in the discussion for better clarity.

Reply: Corrected as indicated

#### 4.1.3 Reviewer 3 Comments for the Author

Selenocysteine synthesis and its incorporation into selenoproteins require in eukaryotes a number of gene products among which SBP2, the protein that binds the SECIS element in the 3'UTR of selenoprotein mRNAs. Only the mammalian and drosophila SBP2 had been characterized so far, its existence in *C. elegans* remaining elusive. In this manuscript, Otero et al. identified and characterizedSBP2 in *C. elegans* and other nematodes. Surprisingly, it is a very short protein lacking the SID domain thought to be essential for Sec incorporation in other eukaryotes; translation of its mRNA starts with an AUU codon, a unique case in this animal. The authors identified the other genes required for Sec synthesis and incorporation. Another interesting finding in this manuscript is that the plant parasite nematodes do not contain the specialized genes required for Sec synthesis and incorporation at all, in line with land plants not having selenoproteins. The manuscript brings novelty and is of great interest in many respects.

**Reviewer**: The only important remark I have is to amend the content of the second paragraph and Fig 6, page 11. (i) the 2 bands are not marked on Fig 6; (ii) the use of mutSECIS is not explained, nor is the type of mutation introduced.

**<u>Reply</u>**: (i) The bands corresponding to the SBP2 size have been marked with an arrow, there are two bands that corresponds to the monomeric (13 kDa) and dimeric (26 kDa) streptavidin chain in lanes 9 and 10 (see also reply to reviewer 2). Those bands are now shown with one and two asterisks, respectively.

(ii) Corrected as suggested in text and in figure 6

#### Other points

#### Introduction

**<u>Reviewer</u>**: According to published data, the SECIS/SBP2/EFSec-tRNASec complex is not guided by L30. Please change the sentence.

**<u>Reply</u>**: We have amended the sentence. We changed from "The SECIS/SBP2/EfSectRNA<sup>®</sup> macromolecular complex interacts with the ribosome, guided by the ribosomal protein L30, and mediates Sec incorporation during protein synthesis" to "The SECIS/SBP2/EfSec-tRNA<sup>®</sup> macromolecular complex interacts through the SECIS element with the ribosomal protein L30, and mediates Sec incorporation during protein synthesis"

**<u>Reviewer</u>**: Further below: the sentence starting with 'being the N-terminal half...' is awkward.

Reply: Corrected as indicated

**<u>Reviewer</u>**: Page 4. 2nd paragraph, 'These data along its....'. It should read 'along with its'. **<u>Reply</u>**: Corrected as indicated. Thank you very much for pointing this error

**<u>Reviewer</u>**: Page 6. The *C. elegans* EFSec was already identified in the paper by Fagegaltier et al. EMBO J. 2000 (Krol lab). Please make the necessary changes so that the finding does not appear as a novelty.

<u>**Reply**</u>: Corrected as indicated

<u>Reviewer</u>: Below. *tag-321* is not shown in Suppl. Fig 1.

<u>**Reply**</u>: *tag-321* sequence was shown but was named only as C33H5.19. This was corrected and now is named with both common and *C. elegans* names (TAG-321 and C33H5.19, respectively)

<u>**Reviewer</u>**: Page 8. Even though SPS1 and SPS2 have been mentioned in the Introduction, it might be helpful for the reader to briefly describe their role. <u>**Reply**</u>: Corrected as indicated</u>

### 4.2 Segundo anexo al capítulo 4:

Luego de publicado el artículo Otero et al, *RNA*, 2014 el Dr. Marco Mariotti nos hizo notar un error en el texto y en la figura 4 del mismo en el mail que se transcribe a continuación:

#### De: Marco Mariotti

Fecha: 17 de julio de 2014, 6:06 Asunto: SECIS type Para: Gustavo Salinas

#### Hola Gustavo,

¿qué tal? Te escribo porque he notado que en el paper en RNA sobre extinciones tenéis un pequeño error, y por mi culpa.

Decís:

The analysis of *trxr-1* SECIS elements using the SECISearch3 program (Mariotti et al. 2013) revealed the presence of the unusual type I SECIS element typical of nematodes with a guanine immediately upstream of the non-Watson-Crick base-pairing quartet (Fig. 4).

Y también la figura dice lo mismo. El error es que SECISearch3 de base no predice la stem3 de los SECISes, o sea, no predice el tipo de SECIS. Sin saberlo, todos parecen de tipo I (sin la stem3).

En realidad, los SECIS de nematodos (por lo menos lo que está en *trxr-1* de *C. elegans*) son de tipo II.

No es la primera vez que alguien hace este error. Pues, he añadido una opción al webserver (http://seblastian.crg.es/) para predecir el tipo de SECIS. No había incluido esta opción en el webserver original porque no me gusta al 100% como funciona, e incluso las imágenes salen un poco más feas.

Ya avisé Vadim del error, junto a Roderic, también para comentarle del cambio en el webserver.

No sé si vale la pena hacer algo más, como enviar un comentario o un errata al journal... ¿qué opinas?

En función del comentario del Dr. Mariotti, se volvieron a analizar los elementos SECIS de los *Caenorhabditis* utilizando la nueva herramienta disponible en SECISearch "predict SECIS-type", se generó una nueva figura para sustituir a la figura 4 y se envió una nota al editor de la revista, que se adjuntan en la siguiente página.

#### ERRATUM

Dr Marco Mariotti (Centre de Regulació Genòmica, Universitat Pompeu Fabra, Barcelona,Spain), the developer of SECISearch3, has brought to our attention that there is an error in Figure 4. At the time that we performed the analysis the program's output provided a general type I-like structure, instead of the actual SECIS type structure. Thus, we re-analyzed the Caenorhabditis SECIS elements in the updated and corrected version of SECISeach, marking the new box "predict SECIS-type". We corrected accordingly Fig 4 that should read: All Caenorhabditis encode a type II SECIS element with a guanine immediately upstream of the non-Watson and Crick base pairing quartet. Critical binding sites for SBP2 are highlighted.



Figure 1: This is the corrected Figure 4 of the original paper

In page 1027 where it reads "The analysis of *trxr-1* SECIS elements using the SECISearch3 program (Mariotti et al. 2013) revealed the presence of the unusual type I SECIS element...", it should read "The analysis of *trxr-1* SECIS elements using the SECISearch3 program (Mariotti et al. 2013) revealed the presence of the unusual type II SECIS element...

Este fue el erratum enviado al editor de la revista RNA; la forma editada se encuentra adjunta al artículo (ver página 74)

#### 4.3 Tercer anexo al capítulo 4

Algunos experimentos con SBP2 quedaron fuera del artículo [116], por no aportar resultados significativos para el mismo; sin embargo consideramos pertinente incluirlos en la tesis, en el presente anexo. Por otra parte, el genoma de *Panagrellus redivivus*, nematodo de vida libre del clado IV *Tylenchina*, fue publicado luego de enviar el artículo para revisión, por lo que su análisis no pudo presentarse en el artículo y se incluye también en este anexo.

#### 4.3.1 Generación de doble mutante *sbp2; rrf-3*

*rrf-3* codifica para un homólogo de la ARN polimerasa dirigida por ARN que inhibe el proceso de ARNi en células somáticas. Los gusanos mutantes en *rrf-3* son hipersensibles al ARNi somático [92, 118]. Para profundizar en el análisis del fenotipo de arresto larvario al interferir la expresión de *gsr-1* en la estirpe mutante en *sbp2* y buscar otros posibles fenotipos asociados a la incorporación de Sec, se generó un doble mutante IH5 *scbp-2(tm3417) I; rrf-3(pk1426) II*, más sensible al ARNi que la estirpe mutante sólo en *sbp2* (IH2).

La generación de este doble mutante implicó el cruce de dos hermafroditas de la estirpe mutante en *sbp2* con seis machos de la estirpe mutante en *rrf-3*. De la progenie de este cruce (F1), heterocigota para ambos genes, se separaron diez larvas individualmente (cada una en una placa) y se les permitió autofecundarse. Se separaron cincuenta larvas de la progenie de la F1 (F2) en placas individuales para que se autofecundaran y, una vez puestos los huevos (F3), se analizó el genotipo de la F2. En este paso se buscaban homocigotas mutantes para ambos genes. De encontrar individuos homocigotas mutantes en la F2, se selecciona la placa con gusanos provenientes de la autofecundación del mismo. Sin embargo, dada la baja probabilidad de encontrar un organismo homocigota mutante en la F2 (1:16), se separaron placas provenientes de gusanos homocigota mutante para uno de los genes y heterocigota para el otro (probabilidad 4:16), se dejó que se autofecundaran y luego se buscó en esa progenie (F3) organismos homocigotas mutantes en ambos genes (Figura 4.3.1).

En este caso, los cruces también se siguieron mediante PCR utilizando los cebadores OL/OR y OL/IR, para cada uno de los genes de interés.



**Figura 4.3.1. Esquema del proceso de generación de un doble mutante.** La generación de un doble mutante implica el cruce de hermafroditas homocigota mutante (n/n) en un gen con machos homocigota mutante en otro gen (m/m). Su progenie (F1) será heterocigota para ambos genes (m/n), y de su autofecundación se pueden obtener gusanos (F2) homocigotas mutantes para ambos genes (organismos de interés, recuadro naranja, m/m; n/n). Como éstos se dan en baja proporción (1:16), también es posible separar de la F2 organismos homocigotas mutantes para un gen y heterocigotas para otro (m/m; n/+ o m/+; n/n, que se dan en proporción 4:16), y buscar en su progenie (F3) organismos homocigota mutante para ambos genes (recuadros naranja) (en proporción 1:2)

#### 4.3.2 Ensayo de ARNi con gsr-1 sobre la estirpe doble mutante generada

Se realizó el ensayo de ARNi con *gsr-1* sobre la estirpe doble mutante (en SBP2 y en *rrf-3*) generada, utilizando controles estirpes con fondo genético *rrf-3* (como control negativo se utilizó la estirpe mutante en *rrf-3* y como controles positivos se utilizaron las estirpes mutantes doble mutantes en *trxr-1* y *rrf-3(pk1426)* y en EFSec y *rrf-3)*. Como control negativo de la técnica se utilizaron bacterias transformadas con el vector

pL4440 vacío y como controles positivos bacterias transformadas con *pL4440[dpy-11]* o *pL4440[unc-22]* que dan fenotipo regordete y descoordinado, respectivamente.

La estirpe mutante en *sbp2* en fondo mutante para *rrf-3* reprodujo el fenotipo de arresto larvario a partir de la segunda generación con mayor penetrancia y expresividad que la sólo mutante en *sbp2*: entre la tercera y cuarta generación ya no había gusanos que alcanzaran la adultez. En la figura 4.3.2 se muestran dos gusanos sincronizados





**Figura 4.3.2 ARNi de** *gsr-1* **da fenotipo de arresto larvario en gusanos mutantes en** *sbp2.* Ensayo de ARNi de *gsr-1* en gusanos mutantes en *rrf-3* (a la izquierda) que no presentan fenotipo aparente y en gusanos mutantes en *rrf-3* y *sbp2* (a la derecha) que muestra fenotipo de arresto larvario. Los gusanos, sincronizados, son parte de la segunda generación interferida. Las barras representan 80 μm.

#### 4.3.3 Intento de análisis del patrón de expresión de scbp-2.

Se intentó generar estirpes transgénicas que expresaran SCBP-2 fusionada a GFP para estudiar su patrón de expresión. Para esto se generó el plásmido pLO111, el cual contiene una construcción traduccional (promotor y gen) de *scbp-2* fusionada a la secuencia que codifica para GFP en el vector pPD95.77. Este tipo de transgénicos se utilizan para estudiar el patrón de expresión de un gen ya que al estar fusionado a la GFP se puede analizar por fluorescencia en qué estadio, bajo qué circunstancias, en qué células (incluso en qué organelos) se expresa dicho gen.

Dicho plásmido se microinyectó en la las gónadas de gusanos L4 el plásmido junto con pRF4 (*rol-6(su1006)*) como marcador, que da un fenotipo tipo *roller* (de giro sobre sí mismos) según [98]. La microinyección se realiza en la zona distal de la gónada, que

contiene los núcleos de la línea germinal, rodeados por un citoplasma común formando un sincicio. Así, el ADN inyectado puede ser transmitido a la progenie bajo la forma de arreglo extracromosomal en número de copias variable [119].

El marcador se inyectó a una concentración de 50 ng/µL y la concentración del plásmido se varió entre 0,5 y 100 ng/µL. La construcción pLO111 resultó tóxica para los gusanos en todo el rango de concentración usado: no se encontró ninguna larva con fenotipo *roller* y solamente se vio fluorescencia en algunos embriones que llegaban a desarrollarse o muertos. Lamentablemente no se pudo obtener ninguna línea transgénica que expresara la fusión traduccional SCBP-2::GFP.

4.3.4 Intento de rescate de fenotipo de arresto larvario generado por el mutante en SBP2

Se realizó un intento de rescate del fenotipo de arresto larvario generado por el mutante en *sbp2* cuando se le interfiere la expresión de *gsr-1*, con las herramientas de las que se disponía. Para esto se cruzaron gusanos machos mutantes en SBP2 (IH2 *scbp-2(tm3417)*) con hermafroditas transgénicos (IH11 *ciuEx3\**) que llevan a la construcción pLO109 (que codifica para SBP2 bajo la regulación del promotor *sur-5* de expresión ubicua y una marca de polihistidina N-terminal -ver [Otero rna 2014]-) junto con un marcador *roller*. De esta forma se obtuvieron gusanos homocigotas mutantes para SBP2 y que sobrexpresan *sbp2* transgénica y de fenotipo *roller*: IH16 *scbp-2(tm3417) l; ciuEx3[pLO109(Psur-5::scbp-2::polyHis); pRF4(rol-6(su1006))]*. Se realizó el ensayo de ARNi con *gsr-1* sobre esta estirpe, utilizando como control positivo la estirpe mutante en *sbp2* y una estirpe mutante en *trxr-1* y como control negativo gusanos de fondo genético silvestre que llevan la construcción pLO109 como arreglo extracromosomal (estirpe transgénica IH11) además de la estirpe silvestre N2.

Se esperaba que la estirpe IH16 revirtiera *parcialmente* el fenotipo de arresto larvario al interferirle la expresión de *gsr-1* -se destaca el *parcialmente*, ya que no todos los gusanos de la estirpe mutante transgénica son portadores del plásmido pLO109, por el tipo de herencia y de expresión que tienen estos arreglos extracromosomales y aquellos que no llevan el arreglo no serían capaces de complementar al alelo mutante-. Sin embargo, este ensayo no funcionó como se esperaba, ya que los gusanos con fenotipo *roller* (de los mutantes transgénicos que se deseaban evaluar y

#### AJUSTES, EXTINCIÓN Y VESTIGIOS DE LA INCORPORACIÓN DE SELENOCISTEÍNA EN EL LINAJE NEMATODO

del control negativo) mostraron problemas de desarrollo al interferirles la expresión de *gsr-1,* lo que dificultó la observación del fenotipo de arresto larvario que se pretendía evaluar. *rol-6* codifica para una proteína de cutícula, expresada en la hipodermis durante las mudas; en la misma línea, *gsr-1* es requerida para una correcta muda; por lo tanto es probable estos problemas se deban a la combinación de bajos niveles de expresión de *gsr-1* al mismo tiempo que una proteína ROL-6 no funcional.

Para realizar este ensayo correctamente se debería microinyectar, en primer lugar, otro marcador (podría ser uno de fluorescencia, por ejemplo) que no presente fenotipo al interferirle la expresión de *gsr-1* y, en segundo lugar, otra construcción donde la expresión de *sbp2* esté regulada por su propio promotor, ya que asegura que la misma se exprese en la células y en la etapa en es requerida. Si bien sabíamos de estos posibles inconvenientes, de todos modos procedimos de esta forma ya que no contábamos con la posibilidad de microinyectar una nueva construcción, pero sí con la estirpe que lleva los arreglos extracromosomales con las construcciones que sobreexpresa *sbp2* y marcador *roller*.

\**ciuEx3* es el nombre del alelo que lleva el transgénico IH11 *ciuEx3[pLO109(Psur-5::scbp-2::polyHis); pRF4(rol-6 (su1006))]*, que lleva al plásmido pLO109 como arreglo extracromosomal que sobreexpresa *scbp-2* y como marcador al plásmido pRF4 que codifica para una proteína ROL-6 mutada que da fenotipo *roller*.

#### 4.3.5 Discusión sobre ensayos adicionales de SBP2

Se intentaron distintas aproximaciones para obtener mayor información respecto a *scbp-2* pero, como se mencionó al inicio de este anexo, las mismas resultaron poco esclarecedoras.

El fenotipo de arresto larvario, previamente observado en la estirpe mutante *scbp-2(tm3417)*, se vio exacerbado en la estirpe doble mutante *scbp-2(tm3417)*; *rrf-3(pk146)*, más sensible al ARNi, llegando a ser letal rápidamente. Esto se debe a que ni TRXR-1 ni GSR-1 por sí mismas son capaces de proveer suficiente poder reductor durante la muda y, por lo tanto, ambas proteínas son necesarias para este proceso [110]. En el caso de gusanos mutantes en *rrf-3*, la sensibilidad al ARNi es mayor, por lo que se

logra disminuir aún más el nivel de expresión de *gsr-1,* acentuando el fenotipo de arresto larvario. Esto refuerza la idea de que *scbp-2* es la SBP2 de *C. elegans,* imprescindible para la incorporación de Sec y para la expresión de un TRXR-1 funcional.

Se intentó sin éxito revertir el fenotipo de arresto larvario mediante complementación con pLO109, ya que el fondo genético que da fenotipo *roller* interfiere en la visualización del arresto larvario.

También se intentó estudiar el patrón de expresión de *scbp-2* con el fin de compararlo con el de *trxr-1* y ver si co-localizan. Lamentablemente la expresión de la construcción creada con este fin resultó tóxica para *C. elegans*.

#### 4.3.6 Análisis del genoma de Panagrellus redivivus

De forma análoga a la descrita en [116] se analizó el genoma de *P. redivivus*. Se encontró que esta especie no codifica para SecS, PSTK ("marcadores" de la incorporación), ni para selenoproteínas o tiorredoxinas reductasas, aunque sí codifica para SecP43; al igual que los otros nematodos del clado IV. Este resultado sugiere que la pérdida de la maquinaria de incorporación de Sec es una característica de ese clado y no una adaptación al parasitismo de plantas (Figura 4.3.2). Si bien esta hipótesis resultaba atractiva y por ello fue planteada en la discusión del artículo, los nuevos resultados indican que muy probablemente no sea correcta: el escenario más parsimonisoso es que la incorporación de Sec se haya perdido en el ancestro del clado IV, luego de la divergencia con los clados IV y V. Este ejemplo ilustra acerca del sesgo que en alguna medida impone la información existente a la hora realizar análisis genómicos comparativos y la cautela extrema que se debe tener a la hora de extraer conclusiones. Retrospectivamente vemos que la información genómica disponible en determinado momento nos llevó a plantear hipótesis interesantes, pero incorrectas.

AJUSTES, EXTINCIÓN Y VESTIGIOS DE LA INCORPORACIÓN DE SELENOCISTEÍNA EN EL LINAJE NEMATODO





Figura 4.3.3. Pérdida de Sec en nematodos del clado IV. Esquema representando el mantenimiento y pérdida de Sec entre nematodos. Los círculos azules representan el mantenimiento y los círculos blancos la pérdida. Se representa también el tipo de ecología de los nematodos representados según esquema en recuadro. La línea roja representa el "momento" en que postulamos que se perdió la capacidad de incorporar Sec en nematodos.

# INICIO NO CANÓNICO DE LA TRADUCCIÓN EN C. elegans

Tanto en procariotas como en eucariotas, la traducción de proteínas comienza, salvo excepciones, con metionina, usualmente codificada por AUG. Este codón debe encontrarse en un contexto determinado para ser reconocido como iniciador de la traducción por los ribosomas, como ser la secuencia Shine-Dalgarno o Kozak para bacterias y eucariotas, respectivamente. En eucariotas este contexto ha sido estudiado en detalle para vertebrados por Kozak; para los cuales la secuencia GCC(A/G)CC<u>AUG</u>G es la óptima [120]. En *C. elegans* el contexto de la iniciación también ha sido estudiado, siendo AAAA<u>AUG</u> el contexto óptimo para inicio de la traducción en esta especie [117]. Sin embargo, ocasionalmente, otros codones, pueden llegar a señalizar el inicio a la traducción [121]. El inicio no AUG de la traducción ha sido descrito tanto para procariotas [122] como para varias especies de eucariotas, incluyendo plantas [123], Drosophila [124] y humanos [125].

Como se mencionó en el capítulo anterior, hasta el momento, a pesar de haber sido el primer organismo multicelular en ser completamente secuenciado [126] y anotado, no se había reportado inicio no AUG de la traducción en *C. elegans*. En este trabajo se encontró que la SBP2 de *C. elegans* inicia su traducción en un codón AUU, al igual que las SBP2 de todas las especies del género *Caenorhabditis* analizadas. Se encontró también que los ARNm de las SBP2 de *Caenorhabditis* spp. inician en un codón AUU ubicado en un contexto <sup>-4</sup>AAA<u>AUU</u>NCN<sup>+6</sup> (se subraya al codón iniciador de la traducción), que señalizaría el inicio de la traducción.

Como es probable que el inicio no canónico de la traducción de SBP2 no sea un caso aislado, y con el fin de encontrar otros genes que también pudieran tener inicio no canónico de la traducción se analizaron, de forma informática, todos los transcriptos codificantes (transcriptos procesados, es decir sin intrones) con 5'UTR largos.

#### 5.1. Métodos

5.1.1. Programas para identificar candidatos a tener inicio no canónico de la traducción

Con la colaboración de Edgardo Crovetto se crearon tres programas en Python usando la biblioteca de Biophyton (adjuntos en el primer anexo a este capítulo) que utilizan como entrada los transcriptos codificantes (transcriptos procesados, incluye región codificante y regiones no traducidas) y las secuencias codificantes (CDS -del AUG al stop-) disponibles en la versión 241 de la wormbase (WS241). En la figura 5.1 se esquematizan los pasos seguidos en cada uno de los programas; el código de los mismos se adjunta en el primer anexo al capítulo 5.



Figura 5.1: Estrategia para la identificación de transcriptos candidatos a tener inicio no canónico de la traducción en *C. elegans.* A) Esquema representando los principales pasos seguidos en los programas representados en utilizados para identificar candidatos a tener inicio no AUG de la traducción en *C. elegans.* B). (en página siguiente) Diagrama de flujo representando los pasos seguidos en cada uno de los programas representados en A).

В



\*Expresión regular: r'aaaatt[atgc]c[atgc](?:[atgc]{3}){50,}'

En caso de que haya un codon de STOP buscar de nuevo: r'(?:[atgc]{3})\*[(?:tag)](?:tag)](?:[atgc]{3})\*aaaatt[atgc]c[atgc](?:[atgc]{3}){50,}' Secuencias con codon de STOP en marco son descartadas Nota: en WS241 se encuentran los archivos c\_elegans.PRJNA13758.WS241.cds\_transcripts.fa.gz (contiene las secuencias de los transcriptos que son traducidas, según las anotaciones -CDS-, en formato fasta) y c\_elegans.PRJNA13758.WS241.coding\_transcripts.fa.gz (contiene las secuencias de los transcriptos codificantes procesados: CDS, 5' y 3'UTRs, en formato fasta) pero no hay disponible un archivo con los 5'UTRs

El primer programa crea un archivo con los 5'UTR de los transcriptos, en formato fasta, por comparación de los mismos con los CDS.

El segundo programa analiza los 5'UTRs generados en el primer programa, buscando cuáles de éstos cumplen con el patrón encontrado que podría dar lugar a un inicio no canónico de la traducción: <sup>-4</sup>AAA<u>AUU</u>NCN<sup>+6</sup>. Esta búsqueda se realiza utilizando una expresión regular en python; teniendo en cuenta que el codón AUU debe estar en el mismo marco de lectura que el AUG que daría inicio a la traducción según la anotación y que no debe haber codones de terminación en marco. Luego crea un archivo en formato fasta con las porciones de los 5'UTRs que cumplen con el patrón. Se utilizaron dos expresiones regulares diferentes: la primera busca que haya al menos 10 codones de distancia entre el posible iniciador AUU y el canónico AUG, mientras que la segunda, más exigente, que haya al menos 50

El tercer programa concatena la porción de los 5'UTRs que cumplen con el patrón con sus CDS correspondientes y crea un archivo fasta con las secuencias concatenadas de los transcriptos candidatos a tener inicio no canónico de la traducción desde el AUU. Aprovechando que se tiene cada una de las secuencias que luego se quieren traducir a proteínas, se crea otro archivo fasta con las mismas secuencias traducidas.

#### 5.1.2. Análisis de genes candidatos a tener inicio no canónico de la traducción

Una vez obtenida la lista de candidatos, se inspeccionó a cada uno de ellos individualmente con el fin de determinar si era factible ese inicio no canónico. En una primera instancia se analizó si las nuevas secuencias se encontraban contenidas dentro de otras CDS más largas ya anotadas. En ese caso, no se profundizó en el análisis ya que no se contaba con herramientas adecuadas para poder discernir si ese

inicio no canónico encontrado podía estar dándose (ver figura 5.2). Como se explica a continuación, las herramientas que se utilizaron para analizar las nuevas secuencias identificadas fueron datos de Ribo-seq y blast y si la nueva secuencia se encuentra incluida dentro de otra CDS, ambas herramientas dan hits positivos.

Con las secuencias que no se encontraban incluidas dentro de otras CDS más largas, se realizó una búsqueda por alineamiento contra todos los organismos utilizando las herramientas blastp y tblastn (se usaron las bases de datos non-redundant protein sequences (nr) y nucleotide collection (nr/nt), respectivamente), con particular interés en los resultados obtenidos con otros nematodos. Los alineamientos se realizaron con las partes nuevas de las secuencias más los primeros 15 aminoácidos de la CDS anotada y luego con la secuencia de largo completo. Se buscaba encontrar si las nuevas porciones se encontraban conservadas en otros organismos, contenían o completaban dominios o completaban proteínas que, por su anotación, parecían estar incompletas.

Finalmente, se analizó si existe evidencia de que las nuevas porciones identificadas serían traducidas mediante al análisis de los datos publicados de Ribo-seq en el sitio web http://gwips.ucc.ie [127].



Figura 5.2 Esquema de análisis de transcriptos candidatos a tener inicio no canónico de la traducción. Una vez identificados los transcriptos candidatos a tener inicio no AUG de la traducción, en un primer paso se inspecciona si la nueva secuencia que se generaría se encontraba incluida dentro de otra CDS del mismo gen ya anotada. De ser así, no se continua el análisis; de no ser así, se realiza una búsqueda por alineamiento (sigue en página siguiente)

(viene de pie de figura de página anterior) contra todos los organismos y se analiza si existen datos de Ribo-seq que sugieran que esa porción, anotada como 5'UTR, pudiera estar siendo traducida.

#### 5.2. Resultados

#### 5.2.1. Transcriptos identificados y candidatos descartados

Se obtuvieron 12 transcriptos que cumplen con el patrón y que éste dista al menos 50 codones del inicio anotado y 330 que cumplen con el patrón y distan al menos 10 codones del inicio anotado. Los primeros se muestran en la tabla 5.1 junto con un esquema de su análisis; los segundos no fueron analizados aún y no se muestran en esta tesis. Se destaca que con las dos aproximaciones se obtuvieron los dos transcriptos anotados de *scbp-2*.

De los 10 nuevos transcriptos identificados, cinco están incluidos dentro de otras CDS (ver ejemplo en Figura 5.3A), tres no dieron hits en el análisis por blast y los datos de Ribo-seq indican que no habría traducción hacia el 5' de la CDS anotada (ver ejemplo en Figura 5.3B y C) y dos transcriptos (pertenecientes al mismo gen) dieron altos niveles de similitud con proteínas de otras especies (particularmente de otros *Caenorhabditis*) y los datos de Ribo-seq indican que habría traducción desde el patrón encontrado (ver análisis en apartado 5.2.2).

Transcripto identificado	¿Incluido en otra CDS más larga?	Blastp/tblastn con 5'UTR + primeros 15 aa	¿Traducción de 5'UTR? (Ribo-seq)	¿Probable inicio no AUG?
<i>B0412.1b</i>	sí nc		nc	?; ?
C31H5.6b	SÍ	nc	nc	?;
C50E10.2a	SÍ	nc	nc	<u>?</u> خ
F01D5.8.1	no	ortólogos de otros <i>Caenorhabditis</i>	SÍ	SÍ
F01D5.8.2	no	ortólogos de otros <i>Caenorhabditis</i>	SÍ	sí
F13B10.1e.2	SÍ	nc	nc	?;
F35B12.5c.1	SÍ	nc	nc	;?
K04G2.11.1	no	ortólogos de otros nematodos	SÍ	SÍ
K04G2.11.2	no	ortólogos de otros nematodos	SÍ	SÍ
Y18H1A.7b	SÍ	sin hits	no	no
Y47H9C.4d	no	sin hits	no	no
Y56A3A.3.2	<i>56A3A.3.2</i> no		no	no

Tabla 5.1: Transcriptos de *C. elegans* identificados como con posible inicio no canónico de la traducción mediante análisis *in silico*. Cuadro comparativo de los resultados de los distintos análisis de los transcriptos con posible inicio no canónico de la traducción según los programas generados. Nótese que para el gen *F01D5.8* se identificaron dos transcriptos codificantes que difieren en el 3'UTR y que se denotan como *F01D5.8.1* y *F01D5.8.2*. *K04G2.11.1* y *K04G2.11.2* corresponden a los transcriptos de *sbp2*. En rojo se recuadran a los transcriptos con probable inicio no canónico de la traducción. nc: no corresponde.



**Figura 5.3 Ejemplos de transcriptos candidatos descartados.** A) Por una posible CDS incluida dentro de otra y B) por no dar hits en blast ni en Ribo-seq. El número romano arriba a la izquierda de cada panel representa el número de cromosoma en el que está presente el gen y la regla indica la posición del fragmento de cromosoma mostrado. Las rectángulos representan exones y las líneas que conectan los exones representan los intrones; en rosado se indican las CDS y en gris los 5' y 3' UTRs. Tomado de www.wormbase.org. Las flechas rojas indican al transcripto identificado como candidato y la cabeza de flecha verde la ubicación del AUU que cumple con el patrón y fue identificado como posible iniciador de la traducción. C) Datos de Ribo-seq del gen Y56A3A.3. Los gráficos rojos muestran los resultaods de Ribo-seq de [128] (fila superior) y de [129] (fila intermedia). En azul se representa la estructura de la secuencia codificante según wormbase y en bordeau los transcriptos anotados en la wormbase (línea guresa para las regiones codificantes, línea fina para las regiones no traducidas y en punteado para los intrones). El recuadro verde destaca la falta de hits de Ribo-seq en el 5'UTR del transcripto.

#### 5.2.2. Análisis de F01D5.8

*F01D5.8.1* y *F01D5.8.2* son los dos transcriptos del gen *F01D5.8* que sólo difieren en su 3'UTR, por lo que se los trató como si fuera uno; su estructura se representa en la figura 5.2.



**Figura 5.2 Estructura del gen** *F01D5.8* **según WS243.** Este gen tiene dos transcriptos que difieren en 22 nucleótidos en el 3'UTR. Los rectángulos representan exones y las líneas que conectan los exones representan los intrones; en rosado se indican las CDS y en gris los 5' y 3' UTRs. Las cabezas de flecha verde indican la ubicación del AUU que cumple con el patrón y fue identificado como posible iniciador de la traducción. Tomado de www.wormbase.org.

Los análisis por blastp y tblastn, utilizando como secuencias consultas la traducción del nuevo fragmento más los primeros quince primeros aminoácidos del CDS por un lado y la nueva secuencia aminoacídica completa por otro, devolvieron como principales resultados las secuencias CRE\_04040 de *Caenorhabditis remanei* (Figura 5.3 A) y CBG20828 de *Caenorhabditis briggsae* (Figura 5.3B); ortólogos de F01D5.8, con un *E-value* de cero (Figura 5.3A), indicando una alta conservación a nivel aminoacídico. Con CRE\_04040 la similitud comienza en el aminoácido 20 (tomando como 1 al codificado por el AUU identificado) y con CBG20828 desde el primer aminoácido; esto refuerza la hipótesis de inicio no canónico de la traducción.

hypothetical protein CRE\_04040 [Caenorhabditis remanei] Sequence ID: ref[XP\_003102531.1] Length: 364 Number of Matches: 1 > See 1 more title(s)

Range 1: 10 to 363 GenPept Graphics Vext Match 🔺 Previous Match						Previous Match	
Score		Expect	Method		Identities	Positives	Gaps
651 bi	ts(167	79) 0.0	Composition	nal matrix adjust.	305/359(85%)	338/359(949	%) 5/359(1%)
Query	21	VFRRDPIAA		FLDVLRACGLICYVA	CPPVPSVITRKLAF	HPPEKGMTYRIV	80
Sbjct	10	LFQREPVA-	- ENQPGCCDV	FIDLLRACGLICYVA	CPPVPSIVTRKLAF	HPPEKGMTYRIA	67
Query	81		KNIRACKDEP	MOMVVRNINNGADY	HPEQDVEVFSVKTA		140
Sbjct	68	LKSDPEKRF	KNIRGCRDEP	QLVVRNISNGADY1	HSEKEVEVFSVTTA	NNNDLVCIKCTP	127
Query	141	DSYSSNPAV		NSSDLGGFLQPNSM			200
Sbjct	128	DSYSSNPAV	SDQVVLFCQP	SSDLGGFLQPSSM	IFVTYANVFETDFYA	FDYSGYGFSSGT	187
Query	201	QGEKNVYAD	VRAVYEKILE	MRPDKKIVVMGYSIC	TTAAVDLAATNPDR	LAGVVLIAPETS	260
Sbjct	188	QGEKNMYAD	IRAVYDKIRE	TRPDKKIVVMGYSIC	TTAAVDLASSNPEG	LAGVVLIAPFTS	247
Query	261			FKSFDKINNIDTRVL			320
Sbjct	248	GLRLFSSKP	DKPDTCWADS	TSFDKVNRIETRVL	ICHGDLDEVIPLAH	GMALYEKLKNPV	307
Query	321	PPLIVHGAN	HHTILSGKYI	HVFTRIANFLRNETL	VSCRSAEVDSSQQQ	QULSSKRTENE	379
Sbjct	308	PPLIVHGAN	HHTILSGKYI	VFTRIAGFLRHETL	VSCRSIEVDSQ	SSSKKKTESE	363

Bownload v GenPept Graphics

Hypothetical protein CBG20828 [Caenorhabditis briggsae]

Sequence ID: ref[XP\_002631639.1] Length: 382 Number of Matches: 1

Range	1: 5 to	377 GenPept	Graphics					Vext Match	A Previous Match
Score		Expect	Method			Identities		Positives	Gaps
644 b	its(166	52) 0.0	Composit	ional matri	x adjust.	309/376(8	32%)	343/376(91	%) 3/376(0%)
Query	1	ISDKPAKSY	QVFPSGTE				RACGL		60
Sbjct	5	ISDEKPTKF	RIFPNVWE	VSKVLSREPI	LDDQP	GCCDAFLGLL	RACGL	MCYVACPPVPS	61
Query	61	VITRKLAF			KRFKNIRA				120
Sbjct	62	VITRKLAFF	IPPEKGMTY	RIAVKSDPE	KRFKNIRG		RNMS	IGADYVH E++V	121
Query	121	EVFSVKTAN	INNDLVCVK	CTPDSYSSN	PAVAEQVV	LFCOPNSSDL	GGFL	PNSMNFVTYAN	180
Sbjct	122	EVFSV	INNELVCIK	CTPDTYSAN	PAVAEQVV	LFCQPNSSDL	GGFLC	PNSMNFVTYAN PNSMNFVTYAN	181
Query	181	VFETDLYAF	DYSGYGFS	SGTOGEKNV	YADVRAVY	EKILEMRPDK	KIVV	IGYSIGTTAAVD	240
Sbjct	182	VFETDEYAF	DYSGYGFS	SGTQGEKNV	YAD+RAVY YADIRAVY	+KI E RPDK DKIRETRPDK	KIVV	IGYSIGTTAAVD IGYSIGTTAAVD	241
Query	241	LAATNPDRL	AGVVLIAP	FTSGLRLFS	SKPDKPDT	CWADSFKSFD	KINNI	DTRVLICHGDV	300
Sbjct	242	LA++NP+ L LASSNPEGL	AGVVLIAP	FTSGLRLFS	REPOREDT	CWADSE SED	K+N J KVNRJ	LDTRVLICHGD+ LDTRVLICHGDL	301
Query	301	DEVIPLSHO	LALYEKLK	NPVPPLIVH	GANHHTIL	SGKYIHVFTR	IANFL	RNETLVSCRSA	360
Sbjct	302	DEVIPL+H0	HALYEKLK MALYEKLK	NPVPPLIVHO	SANHHTIL	SGKYIHVFIR SGKYIHVFTR	IA FL	RHETLVS RS	361
Query	361	EVDSS000	QLSSKRT	376					
Sbict	362	EV+S QQ Q EVES000S0	NTSSKAT	377					

Figura 5.3 Alineamientos de los primeros dos hits con mayor valor E de blastp A) Alineamiento con la proteína hipotética CRE\_04040 de *C. remanei.* B) Alineamiento con la proteína hipotética CBG20828 de *C. briggsae.* Se observan altos niveles de conservación entre estas secuencias, incluso en la región en la que se extiende el ORF hacia el 5' del presunto inicio canónico de la proteína.

El análisis de traducción por Ribo-seq de *F01D5.8* mostró evidencias de traducción desde el inicio del transcripto y en toda la secuencia anotada como 5'UTR (Figura 5.3) tanto en los experimentos de [129] como en los de[128].



**Figura 5.4 Análisis de datos de Ribo-seq de** *F01D5.8* **Datos de Ribo-seq del gen Y56A3A.3. Los gráficos rojos muestran los resultados de Ribo-seq de [128] (fila superior) y de [129] (fila intermedia). En la fila inferior se representa la estructura de los modelos del gen según wormbase: los rectángulos representan exones y las líneas que conectan los exones representan los intrones; en rosado se indican las CDS y en gris los 5' y 3' UTRs. Las cabezas de flecha verde indican la ubicación del AUU que cumple con el patrón y fue identificado como posible iniciador de la traducción. El recuadro verde destaca los hits de Ribo-seq en la porciónanotada como 5'UTR del transcripto que cumple con el patrón.** 

#### 5.3. Discusión

Luego de haber identificado que SBP2 tiene inicio no AUG de la traducción, nos planteamos que probablemente éste no fuera un caso aislado en *C. elegans*, sino que era probable que también se diera en otros transcriptos. Así, nos propusimos analizar transcriptos con 5'UTR largo. Para esto generamos tres programas mediante los cuales identificamos, de los 30.500 transcriptos anotados de *C. elegans*, 12 transcriptos que tienen el patrón AAAAUUNCN hacia el 5'UTR de la CDS anotada, en marco con ésta y que donde el AUU dista, al menos, 50 codones del AUG anotado como iniciador. De estos doce transcriptos, dos corresponden a *K04G2.11.1* y *K04G2.11.2*, los dos transcriptos identificados de *sbp2* que dieron lugar al patrón que se utilizó para la búsqueda. Es importante haber encontrado estos transcriptos ya que funcionan como control del algoritmo utilizado. De los restantes diez transcriptos identificados, dos

corresponden al gen *F01D5.8* y los otros ocho a otros genes, todos con más de un transcripto identificado. Los únicos transcriptos que tienen indicios claros de tener inicio no canónico de la traducción son los correspondientes al gen *F01D5.8*: el 5'UTR se extiende 74 codones hasta el AUU, mantiene altos niveles de identidad con ortólogos de otras especies (incluidos otros *Caenorhabditis*) y los resultados de Riboseq muestran que esa región estaría siendo traducida. El proceso realizado para llegar a identificar candidatos y los resultados del análisis de los mismos se resume en la figura 5.5



Figura 5.5 Resumen de la búsqueda e identificación de transcriptos con posible inicio no canónico de la traducción. Este esquema representa la aproximación utilizada para identificar nuevos transcriptos candidatos a tener inicio no canónico de la traducción y los resultados obtenidos, donde se identifican a los dos transcriptos del gen *F01D5.8* como con probable inicio no canónico de la traducción.
Sólo se pudo identificar el presunto inicio no canónico de la traducción para un gen adicional a SBP2, ya que el método utilizado presenta varias limitaciones, dadas por las restricciones impuestas. Una de ellas es que se busca un patrón determinado sin flexibilidad, que no da lugar a variaciones que podrían existir en relación al motivo identificado para SBP2. Una forma de flexibilizarla sería eliminar la condición de tener una citosina en la posición +5. Otra limitación fue que para reducir el número de secuencias a inspeccionar individualmente se estipuló una distancia de 50 codones entre el primer AUG en marco y el AUU en contexto favorable, la cual es una condición muy exigente. Esto quedó demostrado al comparar la cantidad de secuencias que cumplen con el patrón y que distan al menos diez codones del primer AUG en marco, con las que distan al menos 50 codones: para la primera condición hay 328 transcriptos que cumplen con el patrón, mientras que para la segunda hay 10, sin contar a *scbp-2* (aún resta analizar esos 328 transcriptos). Estas limitaciones pueden dar lugar a falsos negativos. Finalmente, otra restricción impuesta fue que los inicios no canónicos estuvieran en marco con la CDS anotada, asumiendo también que éstas están correctamente anotadas.

Una forma de solucionar este inconveniente podría ser analizar todos los transcriptos que mantengan el marco de lectura abierto hacia el 5' como mínimo 10 codones, hasta el primer codón de terminación en marco, para luego analizar por blast los transcriptos identificados y finalmente analizar los datos de Ribo-seq. Las causa más común que dio lugar a "falsos positivos" fue la presencia de transcriptos con más de un posible inicio de la traducción y por lo tanto con más de una CDS. Los transcriptos que tienen inicio anotado más hacia el 3' pueden dar falsos positivos si entre los dos posibles inicios (AUG) tienen AAAAUUNCN en marco. Para solucionar este inconveniente se deberían modificar secuencias de comando del segundo programa, verificando que la secuencia generada no se encuentre incluida dentro de otra CDS del mismo gen.

#### 5.4. Primer anexo al capítulo 5: Programas

#### 5.4.1. 01codingCDSgeneFinder.py

```
*******
# Este programa crea un archivo fasta con los 5'UTR del coding en función de las #
# CDS
                                                                       #
# la variable fnmCoding es el nombre del archivo fasta de Coding transcripts
                                                                       #
# la variable fnmCDS es el nombre del archivo fasta de CDS transcripts
                                                                       #
# la variable fnmOutput es el nombre del archivo de salida del programa con los ·#
# 5'UTRs
                                                                       #
# Para ejecutarlo en Sublime Text 2 es Ctrl+B o Tools->Build
                                                                       #
******
                import SeqI0
from Bio
from Bio.Seq
                import Seq
from Bio.SeqRecord import SeqRecord
import os
import re
#Esta función recorre la lista de la CDS cargada por un id que empiece por
# partialName o que sea igual, si no encuentra nada devuelve ERROR
def genePartialNameFinder(partialName,where):
   for seg record in where:
       if ( seq record.id.startswith(partialName) or (seq record.id ==
partialName)):
          return seq_record
   print ("ERROR: " + partialName + " startswith="
+str(seq record.id.startswith(partialName) ) )
   return "ERROR"
#Si dentro del archivo coding hay por ejemplo un id K04G2.11.1 y K04G2.11.2 en el
#archivo CDS solo va a aparecer el K04G2.11 ya que tienen CDS en común.
#Esta función se queda con el nombre que debería estar en la CDS
def geneGetPartialName (fullSeqId):
   fullSeqId = fullSeqId.split(" ")[0]
   fullSeqId.split(".")
   return fullSeqId.split(".")[0]+"."+fullSeqId.split(".")[1]
#Cargo fasta Coding transcripts en memoria para poder trabajarlo
           = 'coding_transcripts.dna'
fnmCoding
codingLoaded = SeqIO.parse(fnmCoding, "fasta")
#Cargo fasta CDS transcripts en memoria para poder trabajarlo
           = 'c elegans.PRJNA13758.WS241.cds transcripts.fa'
fnmCDS
           = SeqIO.parse(fnmCDS, "fasta")
CDSLoaded
#Nombre del archivo de salida
           = "fivePrimeUTR.fa"
fnmOutput
#Inicializo contadores para saber al final cuantos genes tienen 5'UTR
contadorTotal = 0
contadorMatch = 0
#Para iterar paso a variable lista el transformado de CDSLoaded como una lista
```

```
lista = list ( CDSLoaded )
#inicializo mi lista resultado
records = []
print ("Start")
#itero entre todos los registros del coding cargado en memoria para buscar si
#tienen 5'UTR en caso de que tengan se agregan a la lista resultado que luego va a
#crear el archivo de salida
for seq_recCoding in codingLoaded:
    partialGeneName = geneGetPartialName(seq recCoding.id)
    CDSofGene = genePartialNameFinder ( partialGeneName, lista )
    if (CDSofGene!="ERROR"):
        if (not (str(seq_recCoding.seq).startswith(str(CDSofGene.seq)))):
            fivePrimeUTR = str(seq_recCoding.seq).split(str(CDSofGene.seq))[0]
records.append(SeqRecord(Seq(fivePrimeUTR,seq recCoding.seq.alphabet),id=seq recCo
ding.id,description=seq_recCoding.description,
name=seq recCoding.name,dbxrefs=seq recCoding.dbxrefs))
            contadorMatch += 1
    contadorTotal += 1
print ("De " + str(contadorTotal) + " secuencias, " + str(contadorMatch) + "
tienen 5'UTR")
SeqIO.write(records, fnmOutput, "fasta")
```

#### 5.4.2. 02patternFilter.py

```
# Este programa crea un archivo fasta con las partes de los 5'UTRs que cumplen
                                                                            #
# con el patrón de la variable pattern que es una expresión regular en python.
                                                                            #
# No agrega al fasta cualquier parte de secuencia que contenga señal de STOP.
                                                                            #
#
                                                                            #
# la variable fnmInput es el nombre del archivo fasta de 5'UTRs
                                                                            #
# la variable fnmOutput es el nombre del archivo de salida del programa con la
                                                                            #
# parte de los 5'UTRs que cumplen el patron.
                                                                            #
# Al terminar imprime en consola cuantas partes de 5'UTR cumplen el patron en
                                                                            #
# función del total.
                                                                            #
# Para ejecutarlo en Sublime Text 2 es Ctrl+B o Tools->Build
                                                                            #
from Bio
                  import SeqIO
from Bio.Seq
                  import Seq
from Bio.SegRecord import SegRecord
from Bio.Alphabet import IUPAC
import os
import re
#Nombre de archivo fasta de entrada con los 5'UTR
fnmInput = "fivePrimeUTR.fa"
#Nombre de archivo fasta de salida con solo la parte que cumple el patrón MENOS el
#primer codón
#Se puede cambiar para generar archivos distintos, si se hacen cambios y se vuelve
#a correr el programa se sobreescribe archivo anterior
fnmOutput = "PatternMatchFivePrimeUTR.fa"
#Patrón de filtro otra posibilidad a
r'aaaatt[atgc]c[atgc](?:[atgc][30]){50,}'puede ser #r'aaaatt(?:[atgc]{3}){50,}'
#Entre el inicio del patrón hasta el final de la cadena para que cumpla tiene que
#haber por lo menos 50 codones a menos que se cambie el 50 en {50,} por otro valor
#es una expresión regular se pueden buscar otro tipo de patrones para buscar
# usando las reglas de éstas para python.
pattern
               = r'aaaatt(?:[atgc]{3}){50,}'
#Patrón igual que pattern pero busca de nuevo el pattern con alguna señal de STOP
#anterior.
patternwithStop =
r'(?:[atgc]{3})*[(?:tag)|(?:taa)|(?:tga)](?:[atgc]{3})*aaaatt(?:[atgc]{3}){50,}'
#Se carga en memoria archivo fasta de 5'UTRs
fivePrimeUTRLoaded = SeqIO.parse(fnmInput, "fasta")
#Inicializo lista de registros con mi resultado luego a escribir en el archivo
#fasta de salida
records = []
#Inicializo contadores para llevar cuenta al final del programa cuantas secuencias
#se encontraron con los patrones queridos.
contadorTotal = 0
contadorMatch = 0
print ("Start")
for seq rec in fivePrimeUTRLoaded:
    contadorTotal += 1
    if len(re.findall(pattern,str(seg rec.seg)))>0:
       newSeq = re.findall(pattern,str(seq rec.seq))[0]
       #Si se encuentra dentro de los que matchean el gen K04G2.11.1 se imprime
       #en consola (es el gen de donde sale el patrón)
       if (seq_rec.id == "K04G2.11.1"):
```

```
print ("K04G2.11.1")
       #si encuentro que la parte que cumple el patrón pattern además cumple el
       #patrón patternwithStop
        #busco pattern de nuevo desde después del primer aaaatt
        if len(re.findall(patternwithStop, str(newSeq)))>0:
            newSeq = re.findall(pattern,str(seq_rec.seq)[6:])[0]
       #Traduzco la secuencia a aminoácidos para buscar que no haya STOP usando
       #una búsqueda en la secuencia traducida de '*'
       newSeqTranslated = str(Seq(newSeq[3:],IUPAC.unambiguous dna).translate())
       if len(re.findall("\*", newSegTranslated))==0 and len(re.findall("M",
       newSeqTranslated))==0:
             contadorMatch += 1
             #Se agrega a la lista resultado la parte que cumple patrón MENOS el
             #primer codón por eso el [3:] para agregar el codón nuevamente
             #cambiar a [0:]
            records.append(SeqRecord(Seq(newSeq[3:], seq_rec.seq.alphabet),
            id=seq_rec.id, description=seq_rec.description, name=seq_rec.name,
            dbxrefs=seq_rec.dbxrefs))
print ("De " + str(contadorTotal) + " 5'UTRs, " + str(contadorMatch) + " cumplen
patron")
#Se escribe archivo de salida
SeqIO.write(records, fnmOutput, "fasta")
```

#### 5.4.3. 03concatenatePatternFoundWithCDS.py

```
# Este programa crea un archivo fasta con la parte de los 5'UTR concatenados con #
# su CDS correspondiente, busca por cada secuencia del archivo con la parte
                                                                        #
# 5'UTRs en el archivo fasta de CDSs la del mismo seq ID (Nombre del gen) y crea #
# una nueva secuencia con ambas y luego las inserta como una nueva secuencia que #
# queda como resultado en el archivo de salida.
                                                                        #
# Aprovechando que se tiene cada una de las secuencias que luego se quieren
                                                                        #
# traducir, se crea un archivo fasta resultado también con las mismas secuencias
                                                                        #
# traducidas
                                                                        #
#
                                                                        #
# La variable fnmFivePrimeUTRs es el nombre del archivo fasta con los pedazos de #
# 5'UTR
                                                                        #
# La variable fnmCDS es el nombre del archivo fasta con el CDS entero.
# La variable fnmOutput es el nombre del archivo fasta de salida con las partes #
# concatenadas.
                                                                        #
# La variable fnmOutTranslated es el nombre del archivo fasta de salida con cada #
# secuencia de fnmOutput, pero traducidas a proteínas.
                                                                        #
# Para ejecutarlo en Sublime Text 2 es Ctrl+B o Tools->Build
                                                                        #
******
from Bio
                 import SeqI0
from Bio.Seq
                 import Seq
from Bio.SeqRecord import SeqRecord
from Bio.Alphabet import IUPAC
import os
import re
#Esta función recorre la lista de la CDS cargada por un id que empiece por
#partialName o que sea igual, si no encuentra nada devuelve ERROR
def genePartialNameFinder(partialName,where):
   for seq record in where:
       if ( seq_record.id.startswith(partialName) or (seq_record.id ==
partialName)):
          #print (seq record.id + " vs " + partialName)
          return seq_record
   print ("ERROR: " + partialName + " startswith="
+str(seq_record.id.startswith(partialName) ) )
   return "ERROR"
#Si dentro del archivo coding hay por ejemplo un id K04G2.11.1 y K04G2.11.2 en el
#archivo CDS solo va a aparecer el K04G2.11 ya que tienen CDS en común.
#Esta función se queda con el nombre que debería estar en la CDS
def geneGetPartialName (fullSeqId):
   fullSeqId = fullSeqId.split(" ")[0]
   fullSeqId.split(".")
   #print (fullSeqId.split(".")[0]+"."+fullSeqId.split(".")[1])
   return fullSeqId.split(".")[0]+"."+fullSeqId.split(".")[1]
#nombre del archivo de entrada fasta con las partes de 5'UTRs a concatenar con la
#CDS correspondiente.
fnmFivePrimeUTRs = "PatternMatchFivePrimeUTR.fa"
#cargo en memoria el archivo
fpmatchingLoaded = SeqIO.parse(fnmFivePrimeUTRs, "fasta")
```

```
#nombre del archivo de entrada fasta con toda la CDS
                 = 'c elegans.PRJNA13758.WS241.cds transcripts.fa'
fnmCDS
#cargo en memoria el archivo
CDSLoaded
                = SeqIO.parse(fnmCDS, "fasta")
#Nombre de archivo de salida con la concatenación de parte de 5'UTRs y CDS
fnmOutput
                = "PatternMatchFivePrimeUTRWithCDS.fa"
#Nombre de archivo de salida con la concatenación de parte de 5'UTRs y CDS
#traducidas
fnmOutTranslated = "PatternMatchFivePrimeUTRWithCDS translated.fa"
#la CDS cargada en memoria la manejo como una lista para iterar sobre ella
lista
                 = list ( CDSLoaded )
#inicializo variables que luego se escribirán en los archivos de salida
records = []
recordsT = []
#imprimo Start para que se vea en consola que inicio la operación
print ("Start")
#inicializo el contador para saber al final en cuantas secuencias se hizo la
#operación
contadorTotal = 0
#Por cada secuencia del archivo de partes de 5'UTR se busca su CDS
#correspondiente, se le concatena con ella y se agrega a las listas resultados una
#lista sin traducir y otra lista con la concatenación traducida.
for seq rec in fpmatchingLoaded:
   contadorTotal += 1
    partialGeneName = geneGetPartialName (seq_rec.id)
   CDSrec = genePartialNameFinder(partialGeneName,lista)
    newSeq = str(seq_rec.seq) + str(CDSrec.seq)
records.append(SeqRecord(Seq(newSeq,seq_rec.seq.alphabet),id=seq_rec.id,descriptio
n=seq_rec.description, name=seq_rec.name,dbxrefs=seq_rec.dbxrefs))
recordsT.append(SegRecord(Seg(str(Seg(newSeg,IUPAC.unambiguous dna).translate()),I
UPAC.protein),id=seq rec.id,description=seq rec.description,
name=seq rec.name,dbxrefs=seq rec.dbxrefs))
#Cuando termina de revisarse todo el archivo se imprime cuántas secuencias se
#procesaron y se escriben los archivos fasta de salida.
print ("Se concatenaron " + str(contadorTotal) + " secuencias")
SeqIO.write(records, fnmOutput,
                                       "fasta")
SeqIO.write(recordsT, fnmOutTranslated, "fasta")
```

# **DISCUSIÓN Y PERSPECTIVAS**

# 6.1 Incorporación de Sec en nematodos

Desde el descubrimiento de las selenoproteínas en los años setenta [130-132], mucho se ha avanzado en la caracterización del mecanismo de incorporación de Sec y en la evolución de su utilización en distintos linajes. Se sabe que Sec está presente en todos los dominios de lo viviente, aunque aún no se han identificado en algunos reinos (como *Fungi* o *Plantae*), así como algunos animales, particularmente algunas especies de insectos [15, 111].

Nematoda es uno de los filos más diversos, compuesto por más de 25.000 especies sumamente variadas en tamaño, ciclo de vida y distribución; así existen nematodos de vida libre, parásitos de plantas y de animales. Los nematodos parásitos de humanos son agentes etiológicos de numerosas patologías que afectan predominantemente a los países subdesarrollados y significan un importante problema sanitario [133]. Los nematodos parásitos de animales y plantas representan importantes pérdidas para la producción ganadera y agropecuaria [134, 135]. Como los nematodos son de gran interés sanitario y económico y la secuenciación es cada vez más accesible, la disponibilidad de datos genómicos y transcriptómicos de este linaje ha aumentado significativamente en los últimos años [136]. Sin embargo no se ha estudiado la evolución del uso de Sec y de la maguinaria asociada a su incorporación en este filo, que por sus características resulta muy interesante. En esta tesis se analizó la capacidad de incorporar Sec en los nematodos con los genomas mejor anotados, representantes de los clados III (Spirurina), IV (Tylenchina) y V (Rhabditina). Se analizó también el mantenimiento o no de los genes asociados a la incorporación de Sec, con especial interés en SBP2, cuya esencialidad para este proceso ha sido cuestionada recientemente [137].

El clado IV se compone de un grupo diverso de nematodos: parásitos de plantas, parásitos de vertebrados e invertebrados y de vida libre. Originalmente se analizaron los genomas de algunos parásitos de plantas (*Meloidogyne hapla, Meloidogyne incognita* y *Globodera pallida*) ya que esos eran los datos de los que se disponía. Se encontró que estos nematodos habían perdido la capacidad de incorporar Sec, así como prácticamente todos los genes involucrados en esta vía. Como ya mencionamos, TRXR-1 es la única selenoproteína expresada en *C. elegans* [13, 14]. Se

destaca que los genomas de nematodos parásitos de plantas no sólo no codifican para Sec, sino que tampoco poseen genes que codifiquen para tiorredoxinas reductasas. Es decir no es que el residuo de Sec haya sido remplazado por Cys; sino que estos genomas no codifican para tiorredoxina reductasas: han perdido esta vía. Esto nos llevó a proponer que la pérdida de dichas vías en estos nematodos podía deberse a una adaptación a sus hospederos, ya que las plantas no incorporan Sec ni poseen tiorredoxinas reducatasas de alto peso molecular [138].

En estudios posteriores, analizamos la presencia de genes asociados a la incorporación de Sec en *Panagrellus redivivus* (un nematodo de vida libre perteneciente al clado IV, que habita diversos ambientes [139]) y la conservación o no de la tiorredoxina reductasa. Se encontró que este nematodo no incorpora Sec y que perdió la mayor parte de los genes involucrados en la decodificación de Sec, así como las tiorredoxinas reductasas. Esto nos llevó a reconsiderar la idea de asociar la pérdida de la capacidad de incorporar Sec a la adaptación al parasitismo de plantas para pasar a considerar que se trata de un ajuste evolutivo, tal vez dado por la adaptación a hábitats deficientes en selenio o a la posibilidad de que las reacciones que llevaban a cabo las tiorredoxinas reductasas puedan ser llevadas a cabo de forma exitosa por otras enzimas, como por ejemplo la glutatión reductasa; los sistemas tiorredoxina y glutatión son vías metabólicas con amplio solapamiento de funciones [140].

Se vio también que el único gen asociado a la decodificación de Sec que mantienen los nematodos del clado IV es SecP43, aunque con bajos niveles de identidad. Es posible que este gen se encuentre involucrado en otros procesos, más allá de la decodificación de Sec o que haya adquirido una nueva función, ya que también se mantiene en otros organismos que no incorporan Sec como *Drosophila willistoni.* Es interesante mencionar, además, que SecP43 se perdió en algunos organismos que sí incorporan Sec como *Anopheles* spp. [27]. SecP43 es pues un gen interesante a estudiar: si bien participa en la incorporación de Sec, no es indispensable en todos los linajes que incorporan Sec, al tiempo que se mantiene en otros linajes que no incorporan. En este sentido la estirpe mutante en SecP43 de *C. elegans* generada en esta tesis podría servir como herramienta para ulteriores estudios.

Los nematodos analizados del clado III incorporan Sec en varias selenoproteínas y mantienen los genes asociados a la incorporación de Sec. Entre las selenoproteínas para las que codifican se encuentra la tiorredoxina reductasa. Estos organismos presentan una SBP2 peculiar: se trata de una proteína corta (entre 211 y 228 aminoácidos) que consta únicamente de un dominio de unión a ARN o RBD con una extensión N-terminal rica en aminoácidos básicos (lisina y arginina) y que carece del dominio de incorporación de Sec o SID, propuesto como esencial para la funcionalidad de SBP2 previo a los resultados de esta tesis.

Dentro del clado V, se analizaron varias especies de *Caenorhabditis*. Se encontró que, así como se había descrito previamente para *C. elegans* y *C. briggsae* en [14], todas las especies analizadas codifican para una única selenoproteína: TRXR-1. Además se identificó a la SBP2 de estos organismos: una proteína aún más corta que las del clado III, que posee únicamente un RBD y carece de SID. Estas SBP2 poseen altos niveles de identidad entre ellas, mantienen la misma estructura génica y tienen largos similares (entre 152 y 155 aminoácidos).

Globalmente *los resultados antes mencionados indican que el mantenimiento de Sec en nematodos se asocia a la conservación de trxr-1.* Éste es el único gen que codifica para una selenoproteína común a todos los nematodos que incorporan Sec (incluso el único que codifica para una selenoproteína en *Caenorhabditis spp*); a la vez, que está ausente en los nematodos que no incorporan Sec. Estos resultados muestran que *Nematoda* es un linaje variado y muy interesante para estudiar la incorporación de Sec, así como la evolución de su utilización. Los resultados respecto a los genes asociados a la incorporación de Sec y a las selenoproteínas en los nematodos se resumen en la figura 6.1. El caso particular de la SBP2 de nematodos se discute con mayor detalle en la siguiente sección.



Figura 6.1 Mantenimiento y pérdida de genes asociados a la incorporación de Sec y de selenoproteínas en nematodos. Esquema representando el mantenimiento (círculo lleno) o pérdida (círculo vacío) de genes involucrados exclusivamente en la incorporación de Sec (PSTK y SecS, en azul), gen anteriormente asociado a la incorporación de Sec (SPS1, en violeta), gen involucrados en la incorporación de Sec y probablemente en otros procesos (SecP43, en rojo) y genes que codifican para selenoproteínas (tiorredoxina reductasa u otras, naranja)

# 6.2 SBP2 muy particulares

En [47] se demostró que, en mamíferos, SBP2 requiere de los dos dominios SID y RBD para lograr la correcta incorporación de Sec. SBP2 se une al SECIS y posteriormente interactúa con el ribosoma a través de RBD [44]. La función de SID no se conoce con exactitud pero en [47] se encontró que mutantes en la porción C-terminal de SID unen SECIS con menor afinidad y se propone que SID interactúa con EfSec para facilitar la liberación de ARNt<sup>[Ser]Sec</sup> y mediar de esa forma con la incorporación de Sec a la cadena polipeptídica. En ese mismo trabajo, Donovan y colaboradores demostraron que, en ausencia de SBP2, RBD y SID (expresados individualmente) pueden actuar en conjunto para mediar la incorporación de Sec.

Se encontró que las SBP2 de nematodos tiene la particular característica de carecer de SID; es más, SID se encuentra ausente en sus genomas. Hasta el momento todas las SBP2 descritas contaban con ese dominio y se presumía esencial para su funcionalidad. A su vez, se encontró que los nematodos del clado III codifican para varias selenoproteínas y los del clado V para una. Esto lleva a pensar que la extensión N-terminal al RBD de la SBP2 del clado III puede estar implicado en el reconocimiento de varios SECIS, mientras que su ausencia (como en *Caenorhabditis spp.*) se deba a que la necesidad de reconocer tan solo un SECIS. Refuerza esta idea el hecho que la extensión N-terminal de las SBP2 del clado III son ricas en aminoácidos básicos, los cuales darían alta afinidad por los grupos fosfatos de la cadena nucleotídica del ARNm. Sin embargo no se puede descartar que otra proteína, con poca similitud de secuencia con SID, no identificada aún, pueda llegar a cumplir la función de SID e interactúe en *trans* con SBP2 vía interacciones proteína-proteína.

Futuros experimentos podrían establecer si la función de SID en nematodos es suplida por otra molécula o si la incorporación de Sec requiere únicamente de RBD para mediar la interacción entre el elemento SECIS y EFSec. Igualmente, resulta claro que *el mecanismo de decodificación e incorporación de Sec en este filo difiere, al menos en parte, del ya reportado para otros eucariotas.* Habiendo encontrado que la SBP2 de *C. elegans* se une a SECIS y que es esencial para la decodificación de Sec en las condiciones ensayadas, postulamos dos posibles escenarios en estos organismos: uno en el cual SID sea completamente prescindible (y en el caso de los nematodos del clado III la extensión N-terminal de la SBP2 colabore a la estabilización de la unión SBP2-SECIS), y otro donde otra proteína pueda mediar la interacción SBP2-EFSec y estabilizar la unión RBD-SECIS (Figura 6.2). Un potencial tercer escenario donde otra proteína pueda mediar la interacción del complejo SBP2-EFSec con el ribosoma fue descartado ya que se ha reportado previamente que SID no contribuye a la interacción SECIS-ribosoma [47]



**Figura 6.2. Posibles mecanismos de incorporación de Sec en nematodos propuestos en esta tesis.** A) Mecanismo donde SID es completamente prescindible y B) mecanismo donde otra proteína medie en la interacción SBP2-EFSec y estabilice la unión RBD-SECIS.

Por otra parte, resulta interesante especular que los resultados de esta tesis podrían colaborar a esclarecer el mecanismo de incorporación de Sec en el dominio archaea. En estos organismos el mecanismo de incorporación de Sec se asemeja más al de eucariotas que al de bacterias: Sec es codificada por un codón UGA y un SECIS localizado en el UTR (habitualmente en el 3'UTR, aunque se ha reportado un caso en que se localiza en el 5'UTR) y se describió un factor de elongación específico de Sec que media la incorporación del aminoácido pero que, a diferencia de lo que ocurre en *Bacteria* y similar a lo que ocurre en *Eukarya*, no une SECIS. Sin embargo hasta el momento no se ha identificado a la proteína responsable de reconocer al SECIS, el equivalente funcional a SBP2 [141]. Teniendo en cuenta los resultados obtenidos respecto a las SBP2 de nematodos, se plantea que tal vez el mecanismo de incorporación de Sec en archaea pueda asemejarse al de los nematodos, y que pueda tener una "SBP2 mínima", formada por RBD o simplemente un motivo L7Ae, lo que ha dificultado su identificación.

### 6.3 Otros genes involucrados en la incorporación de Sec en C. elegans

Al codificar para una única selenoproteína, que a su vez no es esencial para el desarrollo del gusano en condiciones óptimas de laboratorio, *C. elegans* resulta un excelente modelo para estudiar el rol que pueden cumplir los genes potencialmente involucrados en la incorporación de Sec. La expresión de TRXR-1 puede revelarse mediante incorporación metabólica de <sup>75</sup>Se y, si bien los gusanos mutantes en TRXR-1 no presentan defectos en su desarrollo ni manifiestan fenotipos evidentes en condiciones habituales de laboratorio, sí presentan fenotipo de arresto larvario cuando se les interfiere la expresión de *gsr-1*, debido a la imposibilidad de mudar la cutícula. En [110] se postula que en su hábitat natural, TRXR-1 y GSR-1 por sí solas no son capaces de proveer suficiente poder reductor para llevar a cabo la muda, siendo necesarias ambas proteínas. Destacan también que en los dos tejidos en los cuales TRXR-1 se ve involucrada en la muda (hipodermis y faringe) están expuestos a oxígeno molecular, por lo que la muda puede requerir poder reductor adicional para sobrellevar el ambiente oxidante de la atmósfera.

Se estudiaron, además de SBP2, dos genes presuntamente involucrados en la incorporación de Sec que resultaban interesantes abordar utilizando este modelo: i)

*R13A5.9*, presente en organismos que incorporan Sec y ausente en los que no lo hacen, de función desconocida; ii) *T07F10.3*, ortólogo de SecP43, de función no completamente elucidada. Los ensayos de ARNi con *gsr-1* mostraron que *R13A5.9* y *T07F10.3* estarían involucrados en la incorporación de Sec ya que los mutantes en estos genes reprodujeron el fenotipo de arresto larvario al cabo de varias generaciones. Al mismo tiempo se encontró que los mutantes en estos genes mantienen la capacidad de incorporar Sec en las condiciones del ensayo de incorporación metabólica de <sup>75</sup>Se. Estos resultados pueden interpretarse como que estas proteínas posiblemente se encuentren involucradas en el mecanismo de incorporación de Sec, por ejemplo a nivel de la regulación de la expresión, o pueden ser necesarios para la incorporación de selenio y, si bien no serían necesarias en condiciones ambientales óptimas, como las utilizadas durante la incorporación de Sec, las mismas se vuelvan necesarias en condiciones desafiantes, como por ejemplo frente a bajos niveles de GSR-1.

Para profundizar en el análisis de su función se podrían cuantificar y comparar los niveles del ARNm de *trxr-1* que expresan los mutantes en los genes de interés, tanto en condiciones óptimas de laboratorio como en condiciones desfavorables (por ejemplo las del ensayo de ARNi con *gsr-1*) o realizar ensayos de incorporación de <sup>75</sup>Se en estirpes mutantes interferidas. Si bien, en teoría, se podría medir y comparar actividad tiorredoxina reducatasa, en la práctica se vio que los extractos proteicos de gusanos tienen bajos niveles de esta actividad. También se podrían generar anticuerpos contra las proteínas para las que codifican para aislar e identificar complejos supramoleculares a los que se encuentran asociados y generar estirpes transgénicas donde se marquen estos genes con proteínas fluorescentes para analizar si colocalizan con TRXR-1.

#### 6.4 Inicio no canónico de la traducción en *C. elegans*

Se encontró que el transcripto de la SBP2 no comenzaría a ser traducido en el primer AUG en marco, sino 201 bases secuencia arriba en un codón AUU, completando un RBD como único dominio funcional. Si bien este inicio no pudo confirmarse experimentalmente de forma definitiva, la información que se obtuvo revela que es altamente probable que la traducción se inicie en este codón ya que el marco de lectura permanece abierto hacia el 5' del primer AUG en marco, tiene altos niveles de similitud con SBP2 de otros organismos, completa el RBD y los ensayos de ribo-seq muestran traducción secuencia arriba del primer.

El inicio no canónico de la traducción ha sido descrito para varios organismos (por ejemplo GUG como iniciador de la traducción en *Candida albicans* [142] y en mamíferos [143], o AUU en *E. coli* [122]), aunque no se había encontrado transcriptos en los que se diera este fenómeno de forma natural en *C. elegans*. En [144] se había descrito que estirpes mutantes en la subunidad beta del factor 2 de iniciación de la traducción (elF2b) eran capaces de dar lugar al inicio no canónico de la traducción en *C. elegans*. Para demostrar esto generaron construcciones reporteras de GFP con modificaciones en el codón de inicio de la traducción que microinyectaron en gusanos mutantes en elF2b y en gusanos silvestres como control. Estos autores encontraron que los transgénicos que llevaban las construcciones con GFP con inicio AUU y CUG mostraban fluorescencia los gusanos mutantes en elF2b, y también (aunque a bajo nivel) en los silvestres. Este fue un primer indicio de que el inicio no canónico de la traducción de la traducción podría ocurrir naturalmente en *C. elegans*.

En base los resultados obtenidos sobre el inicio de la traducción de la SBP2 en este organismo, se decidió analizar las SBP2 de otros *Caenorhabditis*. Se encontró que todos mantienen el codón AUU en la misma posición. Incluso mantienen el contexto en el cual se encuentra dicho codón: <sup>-4</sup>AAA<u>AUU</u>NCN<sup>+6</sup> (se subraya el codón de inicio propuesto), siendo <sup>-4</sup>AAA<sup>-1</sup> un buen contexto de inicio de la traducción para *C. elegans* [117][Blumenthal] (no se tiene datos respecto al contexto favorable para el inicio de la traducción en otros *Caenorhabditis*, pero se supone que ha de ser similar). La posición +5 también está completamente conservada, a pesar de que ese segundo codón codifica para diferentes aminoácidos en diferentes organismos (serina, prolina, alanina y treonina), por lo que creemos que su conservación podría ser relevante.

Con estos resultados, y previendo que este inicio no canónico de la traducción de SBP2 no habría de ser un caso aislado, se decidió analizar todos los 5'UTRs de *C. elegans*, en búsqueda de posibles secuencias codificantes (CDS) mal anotadas. Mediante la aproximación utilizada (que combinó análisis de los 5'UTRs, comparación por BLAST con otros organismos y análisis de datos de Ribo-seq) se encontró un gen con dos transcriptos similares con probable inicio no canónico de la traducción. Se

identificaron otros 4 transcriptos adicionales de los que no se puede descartar (ni confirmar) que su inicio de la traducción sea no canónico.

Cabe destacar que la aproximación elegida tiene limitaciones importantes. En primer lugar exige que el patrón AAA<u>AUU</u>NCN se cumpla de forma estricta, sin permitir ningún tipo de variación ni en el codón iniciador ni en su contexto. A su vez, supone que la CDS anotada es correcta y que el inicio no canónico estaría en fase con dicha CDS, no contemplando la posibilidad de errores de anotación debido, justamente, a no encontrar el inicio de la traducción. Finalmente, por motivos prácticos, se puso un punto de corte arbitrario de 50 codones de distancia entre el iniciador AUG anotado y el posible iniciador AUU.

Es probable que el número de transcriptos que tengan inicio no canónico de la traducción sea mayor a los ya identificados en esta tesis. Por este motivo, se prevé volver a analizar los 5'UTRs obtenidos, cambiando de estrategia. En principio se piensa analizar todos los 5'UTR que mantengan el marco de lectura abierto secuencia arriba del AUG anotado como iniciador; esas secuencias se traducirán y (junto con los primeros 15 aminoácidos anotados) se compararán mediante TBLASTN con genomas de organismos completamente secuenciados para analizar su conservación y en base a esos resultados se buscarán nuevos candidatos a tener inicio no canónico de la traducción, de forma independiente del patrón ya encontrado.

Finalmente, se prevé intentar una estrategia experimental, similar a la utilizada en [144], generando transgénicos que expresen GFP con el patrón encontrado para estudiar la relevancia que pueda tener la citosina en la posición +5 para el inicio no canónico de la traducción. En caso que se identificaran mediante análisis *in silico* nuevos posible patrones también se estudiaran.

CONCLUSIONES

En forma global, los resultados de esta tesis muestran que *scbp-2* es la *sbp2* de *C. elegans*, funcional y esencial para la incorporación de Sec en este organismo. Se trata de la SBP2 eucariota más corta descrita hasta el momento, compuesta únicamente por un RBD; a la vez que el SID está ausente en todo el genoma. Este escenario es compartido por todos los *Caenorhabditis* analizados en este trabajo; lo que sugiere que el mecanismo de incorporación de Sec en *Caenorhabditis spp.* sería, al menos, parcialmente diferente al descrito para todos los otros eucariotas.

Se encontró que los nematodos del clado III (*Spirurina*), que codifican para varias selenoproteínas, poseen una SBP2 corta, compuesta por un dominio RBD con una extensión N-terminal rica en aminoácidos básicos, y que no codifican para SID. Ese extremo N-terminal podría estar implicado en la necesidad de reconocer varios SECIS y/o suplir de alguna manera la función de SID.

Los nematodos del clado IV (*Tylenchina*) perdieron la capacidad de incorporar Sec. Estos son los primeros animales no artrópodos que se reportan que perdieron esta característica, al igual que todos los genes involucrado en su decodificación, excepto SecP43. Se presume, por lo tanto, que SecP43 debe tener un rol adicional al ya atribuido en la decodificación de Sec que resta por explorar. Los nematodos de este clado perdieron también las vías de la tiorredoxina y basan su metabolismo redox en las vías del glutatión. Sin embargo los del clado III mantienen ambas vías. Esto sugiere que, en nematodos, el mantenimiento de Sec se encuentra ligado al mantenimiento de la tiorredoxina reductasa.

Se encontró que en *C. elegans* se da inicio no canónico de la traducción evento no reportado hasta el momento en este organismo modelo. Se identificaron dos genes (*scbp-2* y *F01D5.8*) en los cuales la traducción comenzaría en un codón AUU, así como otros transcriptos también con posible inicio no canónico de la traducción.

- 1. Hatfield, D.L. and V.N. Gladyshev, *How Selenium Has Altered Our Understanding of the Genetic Code.* Mol. Cell. Biol., 2002. **22**(11): p. 3565-3576.
- 2. Hondal, R.J., S.M. Marino, and V.N. Gladyshev, *Selenocysteine in thiol/disulfidelike exchange reactions*. Antioxid Redox Signal, 2013. **18**(13): p. 1675-89.
- 3. Hondal, R.J. and E.L. Ruggles, *Differing views of the role of selenium in thioredoxin reductase*. Amino Acids, 2011. **41**(1): p. 73-89.
- 4. Kim, H.Y., et al., *Catalytic advantages provided by selenocysteine in methionine-S-sulfoxide reductases.* Biochemistry, 2006. **45**(46): p. 13697-704.
- Gladyshev, V.N., Selenoproteins and selenoproteomes, in Selenium: Its Molecular Biology and Role in Human Health, D.L. Hatfield, Editor 2012, Springer Science+Business Media. p. 109-123.
- 6. Lobanov, A.V., D.L. Hatfield, and V.N. Gladyshev, *Reduced reliance on the trace element selenium during evolution of mammals.* Genome Biol, 2008. **9**(3): p. R62.
- 7. Gobler, C.J., et al., *Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics.* Proc Natl Acad Sci U S A, 2011. **108**(11): p. 4352-7.
- 8. Mariotti, M., et al., *Composition and evolution of the vertebrate and mammalian selenoproteomes.* PLoS One, 2012. **7**(3): p. e33066.
- 9. Lobanov, A.V., et al., *Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life.* Genome Biol, 2007. **8**(9): p. R198.
- Kryukov, G.V., et al., *Characterization of mammalian selenoproteomes*. Science, 2003. 300: p. 1439 1443.
- Bosl, M.R., et al., *Early embryonic lethality caused by targeted disruption of the mouse selenocysteine tRNA gene (Trsp).* Proc Natl Acad Sci U S A, 1997. 94(11): p. 5531-4.
- Conrad, M., et al., *Essential role for mitochondrial thioredoxin reductase in hematopoiesis, heart development, and heart function.* Mol Cell Biol, 2004.
   24(21): p. 9414-23.
- 13. Buettner, C., J.W. Harney, and M.J. Berry, *The Caenorhabditis elegans* homologue of thioredoxin reductase contains a selenocysteine insertion sequence (SECIS) element that differs from mammalian SECIS elements but directs selenocysteine incorporation. J Biol Chem, 1999. **274**(31): p. 21598-602.
- 14. Gladyshev, V.N., et al., *Selenocysteine-containing thioredoxin reductase in C. elegans.* Biochem Biophys Res Commun, 1999. **259**(2): p. 244-9.
- 15. Lobanov, A.V., D.L. Hatfield, and V.N. Gladyshev, *Selenoproteinless animals: selenophosphate synthetase SPS1 functions in a pathway unrelated to selenocysteine biosynthesis.* Protein Sci, 2008. **17**(1): p. 176-82.
- 16. Bock, A., et al., *Selenocysteine: the 21st amino acid.* Mol Microbiol, 1991. **5**(3): p. 515-20.

- 17. Bock, A., et al., *Selenoprotein synthesis: an expansion of the genetic code.* Trends Biochem Sci, 1991. **16**: p. 463 - 467.
- Hatfield, D.L.C., I. S.; Ohama, T.; Jung, J. E.; Diamond, A. M., Selenocysteine tRNA[Ser]Sec Isoacceptors as Central Components in Selenoprotein Biosynthesis in Eukaryotes, in Selenium in Biology and Human Health, R.F. Burk, Editor 1994. p. 25-44.
- 19. Leinfelder, W., et al., *Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine.* Nature, 1988. **331**: p. 723 725.
- 20. Carlson, B.A., et al., *Identification and characterization of phosphoseryltRNA[Ser]Sec kinase.* Proc Natl Acad Sci U S A, 2004. **101**: p. 12848 - 12853.
- 21. Xu, X.M., et al., *Biosynthesis of selenocysteine on its tRNA in eukaryotes.* PLoS Biol, 2007. **5**(1): p. e4.
- 22. Labunskyy, V.M., D.L. Hatfield, and V.N. Gladyshev, *Selenoproteins: molecular pathways and physiological roles.* Physiol Rev, 2014. **94**(3): p. 739-77.
- 23. Guimaraes, M.J., et al., *Identification of a novel selD homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism?* Proc Natl Acad Sci U S A, 1996. **93**(26): p. 15086-91.
- 24. Kim, I.Y., et al., *Fetal mouse selenophosphate synthetase 2 (SPS2): characterization of the cysteine mutant form overproduced in a baculovirusinsect cell system.* Proc Natl Acad Sci U S A, 1997. **94**(2): p. 418-21.
- Tamura, T., et al., Selenophosphate synthetase genes from lung adenocarcinoma cells: Sps1 for recycling L-selenocysteine and Sps2 for selenite assimilation. Proc Natl Acad Sci U S A, 2004. 101(46): p. 16162-7.
- 26. Lee, K.H., et al., *Drosophila selenophosphate synthetase 1 regulates vitamin B6 metabolism: prediction and confirmation.* BMC Genomics, 2011. **12**: p. 426.
- 27. Mariotti, M., *Computational genomics of selenoproteins*, in *Departament Bioinformatics and Genomics, Centre de Regulacio Genomica (CRG)*2013, Universitat Pompeu Fabra.
- Chavatte, L., B.A. Brown, and D.M. Driscoll, *Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes.* Nat Struct Mol Biol, 2005. 12(5): p. 408-16.
- Allmang, C., L. Wurth, and A. Krol, *The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated.* Biochim Biophys Acta, 2009. 1790(11): p. 1415-23.
- 30. Shen, Q., P.A. McQuilkin, and P.E. Newburger, *RNA-binding proteins that specifically recognize the selenocysteine insertion sequence of human cellular glutathione peroxidase mRNA.* J Biol Chem, 1995. **270**(51): p. 30448-52.

- Shen, Q., et al., Identification and molecular cloning of a human selenocysteine insertion sequence-binding protein. A bifunctional role for DNA-binding protein B. J Biol Chem, 1998. 273(10): p. 5443-6.
- 32. Copeland, P.R. and D.M. Driscoll, *Purification, redox sensitivity, and RNA binding properties of SECIS-binding protein 2, a protein involved in selenoprotein biosynthesis.* J Biol Chem, 1999. **274**(36): p. 25447-54.
- 33. Low, S.C. and M.J. Berry, *Knowing when not to stop: selenocysteine incorporation in eukaryotes.* Trends Biochem Sci, 1996. **21**: p. 203 208.
- 34. Walczak, R., et al., *A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs.* RNA, 1996. **2**(4): p. 367-79.
- 35. Goody, T.A., et al., *The kink-turn motif in RNA is dimorphic, and metal iondependent.* RNA, 2004. **10**(2): p. 254-64.
- 36. Matsumura, S., Y. Ikawa, and T. Inoue, *Biochemical characterization of the kinkturn RNA motif.* Nucleic Acids Res, 2003. **31**(19): p. 5544-51.
- Fletcher, J.E., et al., *The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2.* RNA, 2001. 7(10): p. 1442-53.
- 38. Fagegaltier, D., et al., *Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation.* EMBO J, 2000. **19**: p. 4796 4805.
- 39. Grundner-Culemann, E., et al., *Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes.* RNA, 1999. **5**(5): p. 625-35.
- 40. Martin, G.W., 3rd, J.W. Harney, and M.J. Berry, *Functionality of mutations at conserved nucleotides in eukaryotic SECIS elements is determined by the identity of a single nonconserved nucleotide.* RNA, 1998. **4**(1): p. 65-73.
- 41. Copeland, P.R., et al., *A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs.* EMBO J, 2000. **19**(2): p. 306-14.
- 42. Donovan, J. and P.R. Copeland, *Selenocysteine insertion sequence binding* protein 2L is implicated as a novel post-transcriptional regulator of selenoprotein expression. PLoS One, 2012. **7**(4): p. e35581.
- 43. Allmang, C., P. Carbon, and A. Krol, *The SBP2 and 15.5 kD/Snu13p proteins* share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. RNA, 2002. **8**(10): p. 1308-18.
- 44. Copeland, P.R., V.A. Stepanik, and D.M. Driscoll, *Insight into mammalian selenocysteine insertion: domain structure and ribosome binding properties of Sec insertion sequence binding protein 2.* Mol Cell Biol, 2001. **21**(5): p. 1491-8.
- 45. Bubenik, J.L. and D.M. Driscoll, *Altered RNA binding activity underlies abnormal thyroid hormone metabolism linked to a mutation in selenocysteine insertion sequence-binding protein 2.* J Biol Chem, 2007. **282**(48): p. 34653-62.

- Takeuchi, A., et al., A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins. Nucleic Acids Res, 2009. 37(7): p. 2126-41.
- Donovan, J., et al., A novel protein domain induces high affinity selenocysteine insertion sequence binding and elongation factor recruitment. J Biol Chem, 2008. 283(50): p. 35129-39.
- 48. Donovan, J. and P.R. Copeland, *Threading the needle: getting selenocysteine into proteins.* Antioxid Redox Signal, 2010. **12**(7): p. 881-92.
- 49. Tujebajeva, R.M., et al., *Decoding apparatus for eukaryotic selenocysteine insertion*. EMBO Rep, 2000. **1**(2): p. 158-63.
- 50. Leibundgut, M., et al., *Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors.* EMBO J, 2005. **24**(1): p. 11-22.
- 51. Itoh, Y., et al., *Decameric SelA\*tRNA(Sec) ring structure reveals mechanism of bacterial selenocysteine formation.* Science, 2013. **340**(6128): p. 75-8.
- 52. Zavacki, A.M., et al., *Coupled tRNA(Sec)-dependent assembly of the selenocysteine decoding apparatus.* Mol Cell, 2003. **11**(3): p. 773-81.
- Squires, J.E. and M.J. Berry, *Eukaryotic selenoprotein synthesis: mechanistic insight incorporating new factors and new functions for old factors.* IUBMB Life, 2008. 60(4): p. 232-5.
- Xu, X.M., et al., Evidence for direct roles of two additional factors, SECp43 and soluble liver antigen, in the selenoprotein synthesis machinery. J Biol Chem, 2005. 280(50): p. 41568-75.
- 55. Small-Howard, A., et al., *Supramolecular complexes mediate selenocysteine incorporation in vivo.* Mol Cell Biol, 2006. **26**(6): p. 2337-46.
- 56. Miniard, A.C., et al., *Nucleolin binds to a subset of selenoprotein mRNAs and regulates their expression*. Nucleic Acids Res, 2010. **38**(14): p. 4807-20.
- 57. Hanakahi, L.A., Z. Bu, and N. Maizels, *The C-terminal domain of nucleolin accelerates nucleic acid annealing.* Biochemistry, 2000. **39**(50): p. 15493-9.
- 58. Barel, M., et al., A novel receptor ligand pathway for entry of Francisella tularensis in monocyte-like THP-1 cells: interaction between surface nucleolin and bacterial elongation factor Tu. BMC Microbiol, 2008. **8**: p. 145.
- 59. Budiman, M.E., et al., *Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation.* Mol Cell, 2009. **35**(4): p. 479-89.
- 60. Budiman, M.E., J.L. Bubenik, and D.M. Driscoll, *Identification of a signature motif for the eIF4a3-SECIS interaction*. Nucleic Acids Res, 2011. **39**(17): p. 7730-9.
- 61. Donovan, J. and P.R. Copeland, *Evolutionary history of selenocysteine incorporation from the perspective of SECIS binding proteins.* BMC Evol Biol, 2009. **9**: p. 229.

- 62. Blaxter, M. and D.R. Denver, *The worm in the world and the world in the worm.* BMC Biol, 2012. **10**: p. 57.
- 63. Barriere, A. and M.A. Felix, *Isolation of C. elegans and related nematodes.* WormBook, 2006: p. 1-9.
- 64. Brenner, S., *The genetics of Caenorhabditis elegans.* Genetics, 1974. **77**(1): p. 71-94.
- 65. Brenner, S., *The genetics of behaviour.* Br Med Bull, 1973. **29**(3): p. 269-71.
- Byerly, L., R.C. Cassada, and R.L. Russell, *The life cycle of the nematode Caenorhabditis elegans. I. Wild-type growth and reproduction.* Dev Biol, 1976. 51(1): p. 23-33.
- 67. Sulston, J.E., et al., *The embryonic cell lineage of the nematode Caenorhabditis elegans.* Dev Biol, 1983. **100**(1): p. 64-119.
- 68. Wood, W.B., *Determination of pattern and fate in early embryos of Caenorhabditis elegans.* Dev Biol (N Y 1985), 1988. **5**: p. 57-78.
- 69. Lewis, J.A. and J.T. Fleming, *Basic culture methods.* Methods Cell Biol, 1995. **48**: p. 3-29.
- 70. Riddle, D.L., et al., *Introduction to C. elegans*, in *C. elegans II*, D.L. Riddle, et al., Editors. 1997: Cold Spring Harbor (NY).
- 71. Altun, Z.F.H., D. H., *Introduction. In WormAtlas*, 2009.
- 72. Wood, W.B., *Introduction*, in *The Nematode Caenorhabditis elegans*, W.B. Wood, Editor 1988, Cold Spring Harbor. p. 1-16.
- 73. Altun, Z.F.H., D. H., *Handbook of C. elegans Anatomy*, in *WormAtlas*2012.
- 74. Altun, Z.F.H., D. H., *Epithelial system*, in *WormAtlas*2009.
- White, J.G., et al., *The structure of the nervous system of the nematode Caenorhabditis elegans.* Philos Trans R Soc Lond B Biol Sci, 1986. **314**(1165): p. 1-340.
- 76. Altun, Z.F.H., D. H., *Nervous system, general description*, in *WormAtlas*2011.
- 77. Altun, Z.F.H., D. H., *Alimentary system, overview*, in *WormAtlas*2009.
- 78. Lints, R.H., D. H., *Reproductive system*, in *WormAtlas*2009.
- 79. Lints, R.H., D. H., *Male introduction*, in *WormAtlas*2009.
- 80. Hodgkin, J., *Introduction to genetics and genomics*, in *WormBook*2005.
- 81. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
- 82. Friedland, A.E., et al., *Heritable genome editing in C. elegans via a CRISPR-Cas9 system.* Nat Methods, 2013. **10**(8): p. 741-3.
- 83. Kim, H., et al., *A Co-CRISPR Strategy for Efficient Genome Editing in Caenorhabditis elegans.* Genetics, 2014. **197**(4): p. 1069-1080.
- 84. Kutscher, L.M. and S. Shaham, *Forward and reverse mutagenesis in C. elegans.* WormBook, 2014: p. 1-26.

- 85. Mitani, S., *Nematode, an experimental animal in the national BioResource project.* Exp Anim, 2009. **58**(4): p. 351-6.
- 86. *large-scale screening for targeted knockouts in the Caenorhabditis elegans genome.* G3 (Bethesda), 2012. **2**(11): p. 1415-25.
- 87. Wicks, S.R., et al., *Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map.* Nat Genet, 2001. **28**(2): p. 160-4.
- 88. Ahringer, J., *Reverse genetics*, in *WormBook*2006.
- 89. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.
- 90. Tabara, H., A. Grishok, and C.C. Mello, *RNAi in C. elegans: soaking in the genome sequence.* Science, 1998. **282**(5388): p. 430-1.
- Timmons, L. and A. Fire, *Specific interference by ingested dsRNA*. Nature, 1998.
   395(6705): p. 854.
- 92. Simmer, F., et al., *Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi.* Curr Biol, 2002. **12**(15): p. 1317-9.
- Kennedy, S., D. Wang, and G. Ruvkun, *A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans.* Nature, 2004. 427(6975): p. 645-9.
- 94. Fraser, A.G., et al., *Functional genomic analysis of C. elegans chromosome I by systematic RNA interference.* Nature, 2000. **408**(6810): p. 325-30.
- 95. Kamath, R.S. and J. Ahringer, *Genome-wide RNAi screening in Caenorhabditis elegans.* Methods, 2003. **30**(4): p. 313-21.
- 96. Rual, J.F., et al., *Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library.* Genome Res, 2004. **14**(10B): p. 2162-8.
- 97. Timmons, L., D.L. Court, and A. Fire, *Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans.* Gene, 2001. 263(1-2): p. 103-12.
- 98. Evans, T.C., *Transformation and microinjection*, in *WormBook*2006.
- 99. Frokjaer-Jensen, C., et al., *Single-copy insertion of transgenes in Caenorhabditis elegans.* Nat Genet, 2008. **40**(11): p. 1375-83.
- 100. Horvitz, H.R., et al., *A uniform genetic nomenclature for the nematode Caenorhabditis elegans.* Mol Gen Genet, 1979. **175**(2): p. 129-33.
- 101. <u>http://www.wormbase.org/about/userguide/nomenclature</u>.
- 102. Cassada, R.C. and R.L. Russell, *The dauerlarva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans.* Dev Biol, 1975. **46**(2): p. 326-42.
- 103. Rankin, C.H., *From gene to identified neuron to behaviour in Caenorhabditis elegans.* Nat Rev Genet, 2002. **3**(8): p. 622-30.
- 104. de Bono, M., *Molecular approaches to aggregation behavior and social attachment.* J Neurobiol, 2003. **54**(1): p. 78-92.

- 105. Baxter, S.W., et al., *Linkage mapping and comparative genomics using nextgeneration RAD sequencing of a non-model organism.* PLoS One, 2011. **6**(4): p. e19315.
- 106. De Ley, P.B., M., *A new system for Nematoda: combining morphological characters with molecular trees, and translating clades into ranks and taxa*, in *Nematology monographs and perspectives*, R.H. Cool, D. J., Editor 2004, Brill, E. J.: Leiden. p. 633-653.
- 107. De Ley, P.B., M., *Systematic position and phylogeny*, in *The biology of nematodes*, D. Lee, Editor 2002, Taylor & Francis: London. p. 1-30.
- 108. Blaxter, M.L., et al., *A molecular evolutionary framework for the phylum Nematoda*. Nature, 1998. **392**(6671): p. 71-5.
- 109. Holterman, M., et al., *Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades.* Mol Biol Evol, 2006. **23**(9): p. 1792-800.
- Stenvall, J., et al., Selenoprotein TRXR-1 and GSR-1 are essential for removal of old cuticle during molting in Caenorhabditis elegans. Proc Natl Acad Sci U S A, 2011. 108(3): p. 1064-9.
- Chapple, C.E. and R. Guigó, *Relaxation of Selective Constraints Causes* Independent Selenoprotein Extinction in Insect Genomes. PLoS ONE, 2008. 3(8): p. e2968.
- 112. Stiernagle, T., *Maintenance of C. elegans*, in *WormBook*2006.
- 113. Koelle, M. *Generating males by heat shock*. 1998; Available from: <u>http://medicine.yale.edu/lab/koelle/protocols/745\_180559\_Generatingmalesbyh</u> <u>eatshock.pdf</u>.
- 114. Ding, F. and P.J. Grabowski, *Identification of a protein component of a mammalian tRNA(Sec) complex implicated in the decoding of UGA as selenocysteine.* RNA, 1999. **5**(12): p. 1561-9.
- 115. Pao, S.S., I.T. Paulsen, and M.H. Saier, Jr., *Major facilitator superfamily.* Microbiol Mol Biol Rev, 1998. **62**(1): p. 1-34.
- 116. Otero, L., et al., *Adjustments, extinction, and remains of selenocysteine incorporation machinery in the nematode lineage.* RNA, 2014. **20**(7): p. 1023-34.
- 117. Blumenthal, T. and K. Steward, *RNA Processing and Gene Structure*, in *C. elegans II*, D.L. Riddle, et al., Editors. 1997: Cold Spring Harbor (NY).
- 118. Sijen, T., et al., *On the role of RNA amplification in dsRNA-triggered gene silencing.* Cell, 2001. **107**(4): p. 465-76.
- Mello, C.C., et al., *Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences.* EMBO J, 1991. 10(12): p. 3959-70.
- 120. Kozak, M., *An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs.* Nucleic Acids Res, 1987. **15**(20): p. 8125-48.

- 121. Cooper, G.M., *Protein Synthesis, Processing, and Regulation*, in *The Cell: A Molecular Approach*, G.M. Cooper, Editor 2000, Sinauer Associates: Massachusets.
- Sacerdot, C., et al., Sequence of a 1.26-kb DNA fragment containing the structural gene for E.coli initiation factor IF3: presence of an AUU initiator codon. EMBO J, 1982. 1(3): p. 311-5.
- 123. Schmitz, J., et al., *Non-canonical translation mechanisms in plants: efficient in vitro and in planta initiation at AUU codons of the tobacco mosaic virus enhancer sequence.* Nucleic Acids Res, 1996. **24**(2): p. 257-63.
- 124. Sugihara, H., V. Andrisani, and P.M. Salvaterra, *Drosophila choline* acetyltransferase uses a non-AUG initiation codon and full length RNA is inefficiently translated. J Biol Chem, 1990. **265**(35): p. 21714-9.
- 125. Hann, S.R., et al., *A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas.* Cell, 1988. **52**(2): p. 185-95.
- 126. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science, 1998. **282**(5396): p. 2012-8.
- 127. Michel, A.M., et al., *GWIPS-viz: development of a ribo-seq genome browser.* Nucleic Acids Res, 2014. **42**(Database issue): p. D859-64.
- 128. Stadler, M., et al., *Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of C. elegans heterochronic miRNA targets.* Genome Res, 2012. **22**(12): p. 2418-26.
- 129. Stadler, M. and A. Fire, *Wobble base-pairing slows in vivo translation elongation in metazoans.* RNA, 2011. **17**(12): p. 2063-73.
- 130. Turner, D.C. and T.C. Stadtman, *Purification of protein components of the clostridial glycine reductase system and characterization of protein A as a selenoprotein*. Arch Biochem Biophys, 1973. **154**(1): p. 366-81.
- 131. Andreesen, J.R. and L.G. Ljungdahl, *Formate dehydrogenase of Clostridium thermoaceticum: incorporation of selenium-75, and the effects of selenite, molybdate, and tungstate on the enzyme.* J Bacteriol, 1973. **116**(2): p. 867-73.
- 132. Flohe, L., W.A. Gunzler, and H.H. Schock, *Glutathione peroxidase: a selenoenzyme.* FEBS Lett, 1973. **32**(1): p. 132-4.
- Hotez, P.J., et al., *Control of neglected tropical diseases*. N Engl J Med, 2007.
   357(10): p. 1018-27.
- Chitwood, D.J., Research on plant-parasitic nematode biology conducted by the United States Department of Agriculture-Agricultural Research Service. Pest Manag Sci, 2003. 59(6-7): p. 748-53.
- 135. Coles, G.C., *The future of veterinary parasitology.* Vet Parasitol, 2001. **98**(1-3): p. 31-9.

- 136. Martin, J., et al., Nematode.net update 2011: addition of data sets and tools featuring next-generation sequencing data. Nucleic Acids Res, 2012.
  40(Database issue): p. D720-8.
- Seeher. S.; Arassi, T.M., Y.; Carlson, B. A.; With, E. K.; Klein, M. O.; Reix, N.; Miniard,
   A. C.; Schomburg, L.; Hatfield, D. L; Driscoll, D. M.; Scheweizer, U., *Targeted deletion of Secisbp2 in the mouse*, in *10th International Symposium on Selenium in Biology and Medicine*2013: Berlin, Germany.
- 138. Williams, C.H., et al., *Thioredoxin reductase two modes of catalysis have evolved.* Eur J Biochem, 2000. **267**(20): p. 6110-7.
- 139. Hechler, H.C., *Taxonomic notes on four species of panagrellus thorne (nematoda: cephalobidae).* J Nematol, 1971. **3**(3): p. 227-37.
- 140. Winyard, P.G., C.J. Moody, and C. Jacob, *Oxidative activation of antioxidant defence.* Trends Biochem Sci, 2005. **30**(8): p. 453-61.
- 141. Stock, T. and M. Rother, *Selenoproteins in Archaea and Gram-positive bacteria.* Biochim Biophys Acta, 2009. **1790**(11): p. 1520-32.
- Abramczyk, D., M. Tchorzewski, and N. Grankowski, Non-AUG translation initiation of mRNA encoding acidic ribosomal P2A protein in Candida albicans. Yeast, 2003. 20(12): p. 1045-52.
- 143. Takahashi, K., et al., *Evolutionarily conserved non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2)*. Genomics, 2005. **85**(3): p. 360-71.
- 144. Zhang, Y. and L.L. Maduzia, *Mutations in Caenorhabditis elegans elF2beta permit translation initiation from non-AUG start codons.* Genetics, 2010. **185**(1): p. 141-52.

Este anexo hace referencia a la nomenclatura utilizada en C. elegans.

#### Genes

La wormbase considera que un gen es una región del genoma que es expresada o una que ha sido expresada y ahora es un pseudogen. Puede expresar uno o más ARN no codificantes (ARNnc) o codificar para proteínas (CDS, del inglés *coding DNA sequence*).

Todos los genes de la wormbase tienen un **identificador** único (*Gene ID*), del tipo *WBGene00006415*, que hace referencia al locus. De esta forma se busca garantizar el seguimiento de un gen más allá de los cambios que se puedan hacer a su estructura. Esta nomenclatura no es "amigable para humanos", su fin es identificar al locus dentro de la base de datos de la wormbase.

A su vez todos los genes de *C. elegans* de la wormbase también tienen **nombre de secuencia**, que deriva del fósmido, cósmido o YAC en el que fueron clonados originalmente en el proyecto de secuenciación. Por ejemplo el gen *F38H4.7*, se encuentra en el cósmido *F38H4* (y en ese cósmido hay al menos seis genes más).

Si un gen produce una proteína que puede ser clasificada como miembro de una familia, al gen también se le puede asignar un **nombre aprobado**, que consiste en tres o cuatro letras en cursiva, un guion y un número arábigo también en cursiva. Por ejemplo el gen *unc-30* es el trigésimo miembro de la familia de genes unc.

El origen del nombre asignado puede ser debido al fenotipo que presenta un mutante en ese gen (descoordinado – *uncoordinated - unc*), basado en similitud de secuencia o características de la secuencia (superóxido dismutasa – *superoxide dismutase - sod*) o puede hacer referencia a un gen ortólogo de otros organismo modelo (en Drosophila gen *runt –* en *C. elegans rnt*), entre otros casos. Estos nombres deben ser aprobados por el curador de la wormbase antes de ser publicados.

### Alelos y mutaciones

Cada mutación tiene una designación única. Las mutaciones se nombran con una o dos letras seguidas de un número, todo en cursiva; por ejemplo *e61* o *mn138*. Las letras hacen referencia al laboratorio donde se identificó la mutación.

Cuando se hace referencia a una mutación o a un alelo que difiera del reportado en la estirpe N2 de un gen, este nombre se incluye entre paréntesis; por ejemplo *dpy-5(e61)*. Cuando la referencia a esa mutación resulta no ambigua (única mutación conocida o a la que se hace referencia en la publicación y es correctamente citada) puede utilizarse únicamente el nombre del gen (*dpy-5* en lugar de *dpy-5(e61)*)

# <u>CDS</u>

Las CDS son las partes de la estructura del gen que se cura manualmente en la wormbase. La estructura del gen y sus transcriptos derivan de la estructura de la CDS. Tienen un nombre de secuencia que deriva del nombre de secuencia del gen que las codifica; así el gen *F38H4.7* tiene una CDS llamada *F38H4.7*.

Las CDS especifican los exones codificantes desde el codón de inicio hasta el codón de finalización (incluido). Los genes pueden codificar para múltiples proteínas, resultado de *splicing* alternativo. Estas isoformas tienen por nombre el nombre de secuencia seguido de una única letra. Por ejemplo, el gen *bli-4* tiene seis CDS conocidas llamadas *K04F10.4a*, *K04F10.4b*, *K04F10.4c*, *K04F10.4d*, *K04F10.4e* and *K04F10.4f*.

# **Transcriptos**

Los transcriptos de un gen derivan automáticamente del mapeo de todos los ARNm y ADNc disponibles con los modelos de las CDS. Si no hay ARNm o ADNc disponibles, entonces el transcripto tendrá exactamente la misma estructura que la CDS, sin embargo usualmente incluyen UTRs flanqueando la CDS.

Los transcriptos se nombran en base a las CDS utilizadas para crearlos, por ejemplo *F38H4.7* o *K04F10.4a.* Sin embargo, en caso que haya empalme alternativo que no cambie la secuencia de la proteína, los transcriptos alternativamente empalmados se nombrarán agregando un dígito al final; por ejemplo: *K04F10.4a.1* y *K04F10.4a.2.* Si no hay isoformas codificantes, pero hay empalme alternativo en los UTRs, también se les agrega un número al final; por ejemplo la CDS *AC3.5* no tiene isoformas pero tiene las variantes de empalme de transcriptos *AC3.5.1* y *AC3.5.2*.

# <u>Proteínas</u>

Las proteínas son nombradas haciendo referencia al gen que las codifica, pero en mayúsculas y sin cursiva. Así, por ejemplo la proteína codificada por *unc-13* se llama UNC-13. Cuando se predice más de un producto proteico a partir de un mismo gen, éstos se distinguen agregando una letra mayúscula al final en orden alfabética. Por ejemplo TRA-1A, TRA-1B.

# **Transgénicos**

La transformación de *C. elegans* con ADN exógeno suele llevar a la formación de arreglos extracromosomales transmisibles que contienen varias copias del ADN inyectado. A los arreglos extracromosomales se los nombra en cursiva con el nombre de alelo que corresponde al laboratorio que lo generó (estos nombres están regulados por el *Caenorhabditis* Genomics Center -CGC-), seguido por las letras *Ex* y un número. De forma opcional puede ser seguido de información genotípica o molecular que describa al transgen entre paréntesis rectos, como *stEx5* o *stEx5* [*sup-7(st5) unc-22(+)*].

# <u>Estirpes</u>

Una estirpe es un conjunto de individuos con un genotipo particular con la capacidad de producir más individuos con idéntico genotipo. Las estirpes se nombran con dos o tres letras mayúsculas seguidas de un número, sin cursiva. Las letras refieren al laboratorio donde se originó la estirpe. De forma opcional puede ser seguido de información genotípica o molecular. Como ejemplo la estirpe CB1833 tiene fenotipo *dpy-5(e61) unc-13(e51)*, originalmente generada en el laboratorio del Dr. Brenner en el Laboratorio de Biología Molecular del MRC (estripe CB, prefijo de alelo *e*). También puede nombrarse con el nombre de la estirpe seguido del fenotipo CB1833 *dpy-5(e61); unc-13(e51)*, o incluyendo el número del cromosoma donde se localizan los alelos: CB1833 *dpy-5(e61) l; unc-13(e51) l.* 

In the book of life, the answers are not in the back!

Charlie Brown