DEPARTAMENTO DE GENÉTICA

FACULTAD DE BIOLOGÍA

UNIVERSIDAD DE SEVILLA

Caracterización de las UTR 5´y 3´de las lipocalinas de mamífero.

Estudios predictivos sobre su papel en la regulación post-transcripcional

TESIS DOCTORAL

ANDRÉS MEJÍAS ROMERO

LICENCIADO EN CIENCIAS BIOLÓGICAS

Directores:

DR. GABRIEL GUTIERREZ POZO[1]

DR. DIEGO SÁNCHEZ ROMERO[2]

(1) DEPARTAMENTO DE GENÉTICA, UNIV. SEVILLA

(2) DEPARTAMENTO DE BIOQUÍMICA, BIOLOGÍA MOLECULAR Y FISIOLOGÍA, UNIV. VALLADOLID

Agradecimientos

A los directores Gabriel Gutierrez Pozo y Diego Sánchez Romero, por haberme dado la oportunidad de realizar esta tesis y por su continua ayuda y ánimos para que la misma vea la luz. A la Dra. María D. Ganfornina por su apoyo y colaboración durante la realización de la tesis. Al Profesor Luis Corrochano que amablemente cedió su laboratorio para la realización de la parte experimental de la tesis.

A mis padres y hermana, que en todo momento me han apoyado y animado. A mi hermano, que aunque no se encuentra ya entre nosotros fue siempre fuente de inspiración y creatividad científica y de esfuerzo y tesón.

A mi esposa e hijos, por su apoyo y porque que han sabido comprender la falta de tiempo que en algunos momentos les he dedicado, les prometo que les compensaré.

Índice de contenido

Resumen

Las lipocalinas son generalmente proteínas de pequeño tamaño que son secretadas extracelularmente que aunque en un principio han sido clasificadas como proteínas típicamente transportadoras de pequeñas moléculas hidrófobas, la visión actual es que las lipocalinas cubren un abanico de funciones diverso y todavía no totalmente comprendido. Diversos miembros de esta familia son proteínas de gran interés. Por ejemplo la apolipoproteína D (ApoD) y la sintetasa de la prostaglandina D (Ptgds) son proteínas que se expresan principalmente en sistema nervioso, interviniendo en procesos claves.

La expresión de las citadas lipocalinas, entre otras, se encuentra muy regulada en diferentes tejidos o condiciones fisiológicas. Así por ejemplo ApoD es sobreexpresada en ciertos tipos de cáncer y subexpresada en otros, encontrándose además sobreexpresada en Alzheimer. Estos hechos hacen necesario conocer como es llevada a cabo la regulación de la expresión génica de esta lipocalina, así como de otros miembros de interés de la familia. Aunque algunos aspectos del control de la transcripción de algunas lipocalinas son conocidos, los conocimientos sobre su regulación postranscripcional son muy escasos. Debido a que esta regulación es ejercida principalmente por las UTRs 5´y 3´ es del máximo interés conocer como son estas regiones y dilucidar cual es el papel regulador que ejercen en la expresión génica de esta familia.

Este ha sido el propósito de la presente tesis cuyos resultados han aportado algo de luz sobre estos mecanismos de regulación. Se ha encontrado que ciertas lipocalinas, las más ancestrales, hacen uso de UTRs 5´alternativas, mientras que las evolutivamente más recientes no. Así mismo la organización genómica y los mecanismos que controlan la producción de estas alternativas son de cierta complejidad y muestran signos de estar finamente regulados. Respecto a las regiones UTRs 3´ también se ha encontrado, principalmente en las lipocalinas ancestrales, que existe cierta variabilidad, encontrándose formas cortas y largas. Siendo en este caso más fácilmente explicable el origen de esta diversidad mediante señales de poliadenilación-corte alternativo.

Los experimentos realizados con la lipocalina ApoD han demostrado la realidad biológica de los transcritos alternativos seleccionados en las bases de datos, portadores de diferentes UTRs 5´. Los resultados ponen de manifiesto además que hay diferencias en la expresión de las formas alternativas en diferentes tejidos e incluso en diferentes condiciones fisiológicas. Por lo que es de esperar que las distintas UTRs 5´alternativas ejerzan diferentes tipos de regulación

postranscripcional.

Los estudios evolutivos realizados ponen de manifiesto que, en las lipocalinas más ancestrales, parte de la arquitectura genómica de la región UTR 5´ se ha conservado en los mamíferos, mientras que en parte se ha producido cierta divergencia entre los diferentes linajes de mamíferos, seguramente como resultado de las diferentes necesidades fisiológicas de cada uno de ellos . En las lipocalinas más recientes no se han encontrado indicios de conservación en las respectivas regiones UTR 5´. Para las regiones UTRs 3´se han encontrado resultados semejantes.

Los análisis predictivos realizados sobre las UTRs 5´y 3´ han permitido identificar en dichas regiones potenciales elementos que pueden afectar a la eficiencia de la traducción o de la estabilidad del ARNm. En las lipocalinas más ancestrales estos elementos son abundantes, muestran diversidad en las formas alternativas y cierto grado de conservación en lo mamíferos, demostrando que la regulación postranscripcional en estas lipocalinas es importante. Sin embargo los resultados obtenidos sugieren que las lipocalinas más recientes sufren una regulación postranscripcional escasa. Las predicciones en estas últimas sugieren que sus transcritos son traducidos por lo general de forma eficiente.

Motivación de la tesis

Los miembros de la familia proteica de las lipocalinas son generalmente proteínas de pequeño tamaño que son secretadas extracelularmente. Aunque en un principio han sido clasificadas como proteínas típicamente transportadoras de pequeñas moléculas hidrófobas, la visión actual es que las lipocalinas cubren un abanico de funciones diverso y todavía no totalmente comprendido. Entre algunas de sus funciones podemos citar el transporte de retinol, la implicación en funciones olfativas, en el transporte de feromonas, en la síntesis de prostaglandina y su participación en diversos procesos de homeostasis celular.

Si hay una lipocalina que nos sirve de modelo para ilustrar la multiplicidad de funciones es ApoD (*apolipoprotein D*). De las funciones que desempeña dicha proteína destacan las que ejerce en el sistema nervioso, interviniendo en la formación de mielina y en procesos de reinervación, así como en la protección frente al daño cerebral [1]. Está proteína se encuentra especialmente regulada en diversas fases de formación del sistema nervioso durante el desarrollo, así mismo se encuentra sobreexpresada en ciertos procesos patológicos como el Alzheimer e ictus [1]. Respecto a cáncer ApoD es sobreexpresada en ciertos tipos de cáncer como de pulmón, ovario y piel entre otros. Sin embargo es subexpresada en otros, como en el carcinoma esofágico de células escamosas [1].

Podemos mencionar a otros miembros de interés de la familia con funciones diversas y que muestran variaciones en su expresión génica. Entre estos miembros podemos destacar a PTGDS (*prostaglandin D synthase)* y NGAL (*neutrophil gelatinase-associated lipocalin*, también denominada como LCN2). PGDS es expresada principalmente en sistema nervioso y muestra diferencias de expresión en diferentes tejidos y tipos celulares [2]. NGAL también se muestra como una lipocalina multifuncional, implicada en la respuesta de inmunidad innata y que puede actuar como factor de crecimiento [3, 4]. Esta lipocalina sufre variaciones en su expresión, elevando sus niveles cuando se producen daños renales [5] y alterándose su expresión en algunos cánceres, por lo que ha sido propuesta como biomarcador [6, 7].

Por lo anteriormente expuesto parece fundamental conocer cuales son los mecanismos responsables de la regulación de la expresión de estas lipocalinas, entre otras. Algunos aspectos de la regulación de la expresión génica de lipocalinas como ApoD y PTGDS son conocidos a nivel de la transcripción, habiéndose identificado ciertos elementos reguladores en su región promotora [8, 9]. Además de la relevancia que el control de la transcripción tiene en la regulación de la expresión génica, es bien conocido el importante papel que desempeña la regulación a nivel

postranscripcional. Dicha regulación es ejercida principalmente por elementos que se encuentran en las regiones no traducidas 5´y 3´del ARNm (UTR 5´y UTR 3´) [10]. El conocimiento que hay sobre estas regiones en las lipocalinas y la regulación postranscripcional que las mismas puedan ejercen es muy escaso. Así por ejemplo hay indicios de un miARN que actuaría en la UTR 3´de NGAL e inhibiría su expresión, teniendo esto importancia para la progresión de tumores, ya que esta lipocalina es sobreexpresada en los mismos [11]. Se desconoce, entre otros muchos aspectos, la relevancia que pueda tener la expresión de UTRs alternativas, fenómeno frecuente en genes eucariotas, en la regulación de la expresión génica en esta familia proteica.

El propósito de esta tesis es aportar algo de luz sobre los mecanismos de regulación postranscripcional ejercidos por las UTRs de las lipocalinas. Para llevar a cabo esta investigación se han seleccionado una serie de lipocalinas de mamíferos. Los motivos de limitarlo a este taxón son: que las lipocalinas son abundantes en este grupo, que existe amplia información para las diferentes especies de mamíferos en las bases de datos de secuencias, que hay muchas lipocalinas ortólogas entre diferentes especies del taxón, lo que facilita las comparaciones y por último y no menos importante se acota la magnitud del tema para hacerlo abordable en una tesis. En la selección de las lipocalinas se ha tenido en cuenta que estas representen bien las variadas funciones que realizan, así como la historia evolutiva de las mismas.

Bibliografía

[1] Akerstrom, B. et al. Plasma Lipocalins A1agp, ApoD, ApoM, C8GC. Lipocalins Book Review. Landes Bioscience (2006)

[2] Akerstrom, B. et al. PGDS. Lipocalins Book Review. Landes Bioscience (2006)

[3] Yang, J. *et al.* An Iron Delivery Pathway Mediated by a Lipocalin. *Molecular Cell* **10,** 1045–1056 (2002).

[4] Schmidt-Ott, K. M. *et al.* Dual Action of Neutrophil Gelatinase–Associated Lipocalin. *Journal of the American Society of Nephrology* **18,** 407–413 (2007).

[5] Bennett, M., Dent, C. L., Ma, Q., Dastrala, S., Grenier, F., Workman, R., … Devarajan, P. . Urine NGAL Predicts Severity of Acute Kidney Injury After Cardiac Surgery: A Prospective Study. *Clinical Journal of the American Society of Nephrology* **3**, 665–673 (2008)

[6] Bauer, M. *et al.* Neutrophil gelatinase-associated lipocalin (NGAL) is a predictor of poor prognosis in human primary breast cancer. *Breast Cancer Res. Treat.* **108,** 389–397 (2008).

[7] Moniaux, N., Chakraborty, S., Yalniz, M., Gonzalez, J., Shostrom, V. K., Standop, J., Batra, S. K. . Early diagnosis of pancreatic cancer: neutrophil gelatinase-associated lipocalin as a marker of

pancreatic intraepithelial neoplasia. *British Journal of Cancer* **98**, 1540–1547 (2008).

[8] Fujimori, K. *et al.* Regulation of Lipocalin-type Prostaglandin D Synthase Gene Expression by Hes-1 through E-box and Interleukin-1β via Two NF-κB Elements in Rat Leptomeningeal Cells. *J. Biol. Chem.* **278,** 6018–6026 (2003).

[9] Zhang, P.-X. *et al.* Regulation of neutrophil gelatinase-associated lipocalin expression by C/EBPβ in lung carcinoma cells. *Oncol Lett* **4,** 919–924 (2012).

[10] Barrett, L. W. & Fletcher, S. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* **69,** 3613–3634 (2012).

[11] Lee, Y. C., Tzeng, W.-F., Chiou, T.-J. & Chu, S. T. MicroRNA-138 Suppresses Neutrophil Gelatinase-Associated Lipocalin Expression and Inhibits Tumorigenicity. *PLoS One* **7**, (2012).

Objetivos y organización de la tesis

Los objetivos que se pretenden alcanzar con esta tesis son:

1. - Obtener una base de datos de secuencias de transcritos de lipocalinas ortólogas de mamíferos y caracterizar las regiones no traducidas (UTRs 5´y 3´) de los mismos en cuanto a longitud y composición de nucleótidos.

2. - Determinar la existencia de UTRs alternativas en las lipocalinas de mamíferos, tratando de dilucidar los mecanismos que las originan.

3. - Llevar a cabo una confirmación experimental de las UTRs que "in sílico" muestren tener una mayor variabilidad, para confirmar su relevancia biológica.

4. - Llevar a cabo un estudio comparativo de las UTRs 5´y 3´ ortólogas de lipocalinas entre los diferentes grupos de mamíferos, para conocer en que medida se ha producido conservación y conocer el proceso evolutivo que ha originado la diversidad de estas regiones.

5. - Una vez determinadas las regiones UTRs 5´y 3´, realizar estudios predictivos, tanto de estructura primaria como secundaria "2D", para identificar potenciales elementos que intervengan en la regulación postranscripcional.

6. - Realizar una integración del conjunto de datos obtenidos (identificación de motivos de secuencia primaria, identificación de elementos estructurales y elementos conservados) para elaborar un modelo de como estas regiones ejercen su regulación postranscripcional en la expresión de las lipocalinas.

La tesis se ha organizado de la siguiente manera:

- En una primera parte se expone un ***Material Introductorio*** con dos capítulos. Un primer capítulo donde se describe detalladamente a la familia de las lipocalinas y un segundo capítulo donde se hace una revisión de los mecanismos de regulación de la expresión génica en eucariotas. Centrando el tema en la regulación de la transcripción y en

los mecanismos de regulación postranscripcional.

- En una segunda parte se exponen los R***esultados Obtenidos,*** organizados en diferentes capítulos. En cada uno de los seis capítulos se incluyen los siguientes apartados: objetivos específicos del capítulo, material y métodos, resultados y discusión.

- En una tercera parte se exponen unas C***onclusiones Finales,*** donde se integran los resultados de los diferentes capítulos para obtener una visión de conjunto.

INTRODUCCIÓN

I. Las lipocalinas

II. La regulación de la expresión génica en eucariotas

I

LAS LIPOCALINAS

1.- Concepto

Los miembros de la familia proteica de las Lipocalinas son generalmente proteínas de pequeño tamaño y son secretadas extracelularmente. Aunque las Lipocalinas muestran gran diversidad en sus secuencias, se caracterizan por una serie de propiedades comunes: su habilidad para unirse a moléculas hidrofóbicas de pequeño tamaño, su unión a receptores en la superficie celular y la posibilidad de formación de complejos macromoleculares. A pesar de la escasa conservación de sus secuencias, presentan un patrón general de plegamiento que si está altamente conservado.

Aunque en un principio han sido clasificadas como proteínas transportadoras, la visión actual es que las Lipocalinas cubren un abanico de funciones diverso. Entre ellas podemos citar el transporte de retinol, la implicación en funciones olfativas, en el transporte de feromonas, en la síntesis de prostaglandina e incluso la participación en la respuesta inmunitaria y en la homeostasis celular.

La investigación de las Lipocalinas se caracteriza por el continuo descubrimiento de nuevos miembros. Desde que fueron descubiertas en 1981 [1] esta familia de proteínas ha crecido rápidamente y actualmente hay identificadas más de 40 proteínas, en bacterias, plantas y animales. Los análisis filogenéticos de las Lipocalinas son complejos y han dado como resultado la clasificación de las mismas hasta en 14 clados o grupos diferentes [2a].

2.- La secuencia proteica de la familia de las Lipocalinas

Los genes de las lipocalinas son transcritos, salvo excepciones, en ARNm de un tamaño que oscila entre 0.6 y 1 Kb. Estos son traducidos a polipéptidos de un tamaño entre 160 y 230 aminoácidos, la mayoría de los cuales presenta un péptido señal que permite su exportación extracelular [2b]. Una clara excepción a esta exportación se presenta en lipocalinas procariotas, que son ligadas mediante lípidos a las membranas celulares, o bien se encuentran solubles en el espacio periplásmico bacteriano. Entre los ecucariotas también encontramos excepciones [2b], así la lipocalina Lazarillo de saltamontes se encuentra ligada mediante GPI (glicosifosfatidilinositol) a las membranas neuronales y en el caso de la Probasina de rata su localización parece ser nuclear.

El polipéptido maduro de las lipocalinas tiene un peso molecular medio estimado (sin considerar las

modificaciones postraducccionales) de 19.4 kDa, pero sus valores oscilan entre 17.7 y 21.7 KDa.

Esta variación se debe a la diferente extensión que muestran los extremos carboxilo o amino terminal, característica propia de ciertas Lipocalinas que pertenecen a diferentes clados [2b].

En las lipocalinas se da frecuentemente un bajo nivel de conservación, con valores de identidad entre secuencias de proteínas parálogas en el rango de 20% - 30 %. Valores que están dentro de la "zona de penumbra" a la hora de asignar una proteína a una familia determinada. Sin embargo todas las lipocalinas comparten suficiente semejanza en forma de cortos motivos conservados (SCRs). El alinemaiento múltiple de lipocalinas llevó a considerar inicialmente [2b] a tres de estos motivos (SCR1, SCR2 y SCR3) como propios de las lipocalinas genuinas y a todas las que contenían a estos tres SCRs se las consideró lipocalinas centrales o principales. Las que carecían de alguno o algunos de ellos fueron consideradas lipocalinas secundarias o periféricas. El descubrimiento de nuevas lipocalinas y el análisis de sus estructuras y los alineamientos mútltiples ha llevado a relajar los criterios mencionados [2b]. Así, en un alineamiento múltiple de 209 Lipocalinas, el 90% contenía dos de los motivos (SRC1 y SRC3), mientras que más del 60 % de ellas contenía el tercer motivo (SRC2). Llega a darse el caso en que todos los componentes de ciertos clados de lipocalinas carecen de alguno de estos motivos.

Otras propiedades derivadas de la secuencia primaria de las proteínas permiten identificar a las lipocalinas de los diferentes clados. El punto isoeléctrico (pI), calculado para la secuencia proteica madura, es un factor importante para la solubilidad y plegamiento del polipéptido. Encontramos valores desde pI básico como en la probasina hasta ácido como en RBP o ApoD. También ha sido investigada la capacidad de glicosilación (N-glicosilación y O-glicosilación) de las lipocalinas encontrándose diversidad de estas en los diferentes clados. Así hay clados que presentan los dos tipos de glicosilaciones, bien sólo una de ellas, e incluso clados que presentan escasos lugares de glicosilación. En la tabla 1 se sintetiza la información de los parámetros propios de algunos clados de lipocalinas bien establecidos filogenéticamente.

Nombre	Clado	Longitud(nt)	pI	N-glicosilación	O-glicosilación	S-S
Lipocalinas bacterianas	I	157	8.1	0	0	0-1
Lipocalinas de plantas	I	190	5.1	1	0	0
Lipocalinas de artrópodos	II	189	6.6	1-5	0-1	2
ApoD	II	168	5.9	2	0	2
RBP	III	179	5.9	0	0	3
PGDS	V	169	7.2	2	2-4	1
NGAL	V	177	8.2	1-2	0-3	1
A1mg	VI	182	6.6	1-2	0-1	1
C8GC	VI	181	8.7	1	0	1
RUPs	IX	164	5.1	0-1	0	1
Quimiorrecepción I	X	159	5.9	0-1	0	0-1
ApoM	XII	167	6.1	0-1	0-2	3
Miscelánea de lipocalinas	-	162	6.5	1-2	0-1	0-1

Tabla 1. *Propiedades bioquímicas de lipocalinas pertenecientes a algunos de los clados bien establecidos filogenéticamente (propiedades predichas a partir de su secuencia de proteína). Modificada de referencia [2]*

3.- Semejanzas estructurales en la familia de las Lipocalinas.

La forma común de plegarse de las lipocalinas es una estructura simétrica, toda ella en forma beta.

La estructura predominante es una cadena simple con ocho láminas en beta en forma antiparalela. Dicha estructura está cerrada sobre sí misma formando una estructura de tipo "barril beta" (figura 1), estabilizada mediante puentes de hidrógeno entre láminas beta adyacentes. Las ocho láminas en beta del barril de lipocalinas, normalmente nombradas de la A a la H, están unidas mediante una sucesión de conexiones "+ 1" , en el extremo amino terminal se encuentra una estructura 3-10 y en el carboxilo una alfa hélice (figura 2). La estructura del barril de las lipocalinas se enrolla hacia la derecha y de forma cónica alrededor del eje central, de forma que la lámina A puede unirse a la H mediante puentes de hidrógeno. Un extremo del barril se encuentra abierto, mientras el otro está cerrado, formando los residuos del interior del barril un núcleo muy empaquetado. A continuación de las ocho láminas beta del barril hay una cadena en alfa hélice (figura 3).

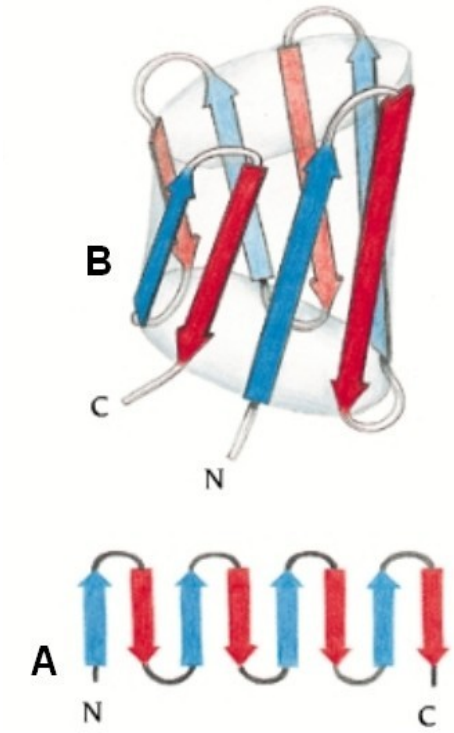


Figura 1. *Estructura típica de "barril beta". A, esquema de la estructura secundaria. B, representación de la estructura tridimensional del barril beta. Flechas: hojas en beta, líneas: lazos de unión entre hojas en beta.*

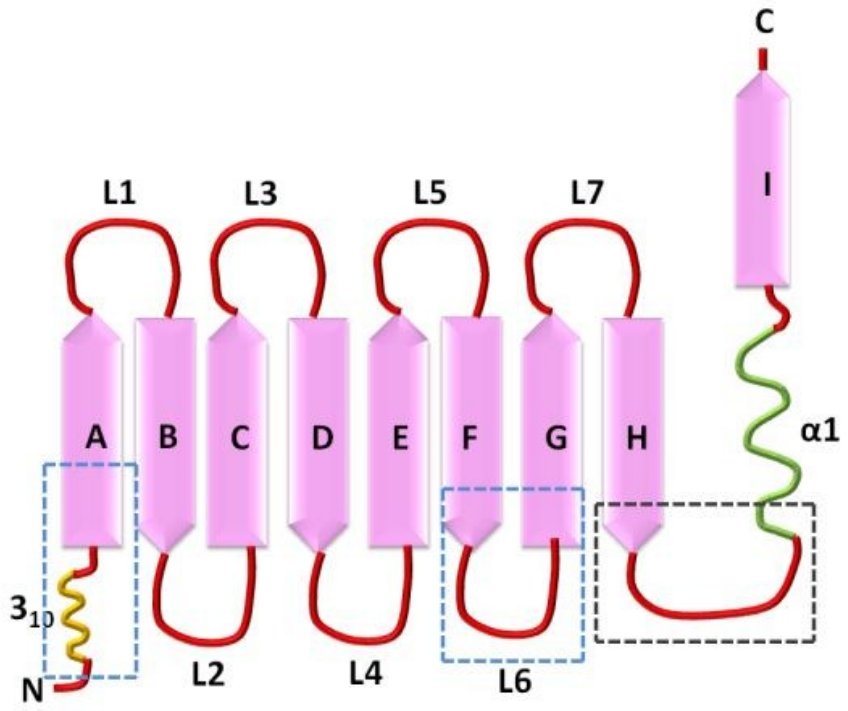


Figura 2. *Esquema de la estructura secundaria típica de las lipocalinas. A a H representan las hojas beta, L1 a L7 los lazos de unión entre hojas beta. En el extremo amino terminal se encuentra una estructura tipo alfa 3-10 y en el carboxilo terminal una alfa hélice (α1). En los recuadros se encuentran las regiones mejor conservadas entre diferentes lipocalinas, azul respecto a estructura y negro en secuencia aminoácidos. Extraída de referencia [4].*

El barril beta da lugar al sitio de unión del ligando formado por una apertura y por una cavidad interna (ver figura 3). La diversidad estructural y de secuencia de este conjunto apertura-cavidad es lo que permite a las diferentes lipocalinas mostrar diversidad en su unión a los ligandos. En contraste con la relativamente bien conservada topología del barril beta, los lazos de unión entre láminas beta difieren bastante entre los miembros de esta familia, tanto en longitud y secuencia de aminoácidos, así como en la conformación que estos adquieren. Estas diferencias son las que dan lugar a la diferente conformación del sitio de unión y a la especificidad de ligando que muestran las diferentes lipocalinas.

Dentro de las características comunes del plegamiento de las lipocalinas destacan tres regiones estructurales bien conservadas (SCRs: SCR1, SCR2 y SCR3), ya comentadas previamente, que se encuentran en regiones concretas de la estructura de las lipocalinas (figura 4). En fechas posteriores han sido identificados otros dos motivos, con un papel relevante en el plegamiento del polipéptido [5]. Uno de ellos localizado próximo al extremo cerrado del barril beta, el otro localizado al comienzo de la alfa-hélice del extremo carboxilo terminal.

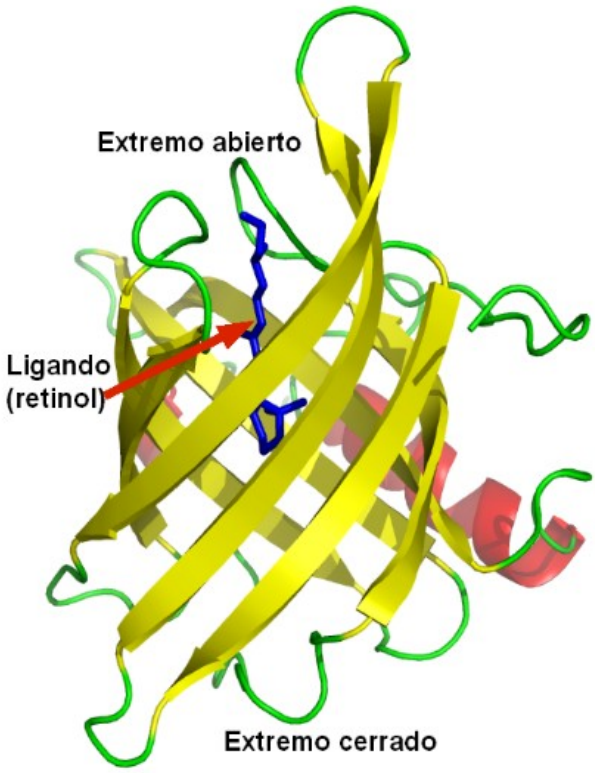


Figura 3. *Estructura tridimensional de la lipocalina RBP unida a retinol.*

Las lipocalinas junto con otras dos familias distintas de proteínas, las FABPs (proteínas de unión a ácidos grasos) y las avidinas (con afinidad por la biotina), forman parte de la superfamilia de las Calicinas [2b]. El barril beta de las FABPs está formado por diez láminas en beta y el de las avidinas aunque esta formado por ocho, como las lipocalinas, es menos elíptico en sección transversal. A pesar de estas diferencias y de la ausencia de similitud global de sus secuencias, los componentes de esta superfamilia adquieren una conformación bastante semejante. Así mismo comparten una semejanza funcional, en cuanto a unión a compuestos hidrofóbicos, o al menos de pequeño tamaño y/o presentan interacciones macromoleculares clave. Con el tiempo la superfamilia de las calicinas ha ido creciendo, gracias a nuevas evidencias de semejanzas estructurales, incluyendo a un variado grupo de proteínas como las inhibidoras de metaloproteasas o las estafostatinas, entre otras [2b].

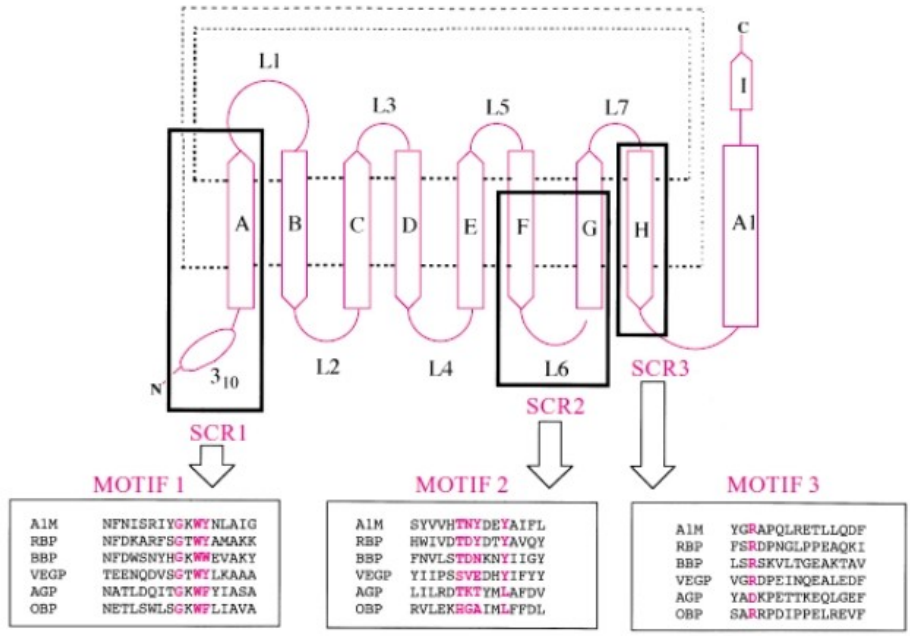


Figura 4. *Posición de las regiones estructurales conservados SCR1, SCR2 y SCR3 en la estructura secundaria de las lipocalinas. En los cuadros inferiores se nuestran los motivos que se corresponden con los SRCs en diversas secuencias de lipocalinas. Imagen extraida de referencia[3].*

4.- Funciones de las Lipocalinas

Si bien las lipocalinas son calificadas como proteínas transportadoras, estas cumplen bien mediante transporte o mediante otros procesos un enorme abanico de funciones. Una enumeración exhaustiva de las funciones estaría fuera de los propósitos de esta tesis, dándose además la circunstancia de que hay lipocalinas cuya función es aún desconocida. A continuación se mencionan ejemplos de las variadas funciones que desempeñan algunos miembros de la familia mejor estudiados.

4.1.- Función transportadora

Las lipocalinas han sido clasificadas generalmente como proteínas transportadoras extracelulares. Dicha función viene tipificada por RBP [2c, 3], proteína transportadora del retinol en el plasma. RBP es producida en el hígado y tras su unión al retinol es secretada al plasma sanguíneo. Ya en la circulación RBP se une a otra proteína la transtiterina (TTR), formando así un complejo

macromolecular. Dicho complejo, debido a su mayor tamaño, previene de la pérdida de RBP por filtración del glomérulo renal. La interacción de RBP con receptores de membrana en células específicas hace que este libere el retinol y pierda su afinidad por TTR, finalmente la RBP resultante es filtrada por el riñón, reabsorbida y degradada.

4.2.- Actividad de feromona

Al menos tres clases de lipocalinas cumplen esta función: el complejo MUP (*Major Urinary Protein*) de ratón, la alfa2u-globulina de la rata y la afrodisina del hamster [2d, 3]. La síntesis de lipocalinas urinarias como UMP y alfa2u-globulina tiene lugar en el hígado y es dependiente de la presencia de testosterona. Dos peculiaridades distinguen a estas dos lipocalinas urinarias de ratón y rata de otras lipocalinas excretadas. En primer lugar actúan como proteínas de unión a odorantes que se producen en el cuerpo y posteriormente en el medio exterior liberan lentamente las sustancias odorantes volátiles. En segundo lugar las proteínas no son secretadas en una forma química definida sino que lo hacen en forma de un complejo de composición variable.

Parece ser necesaria la unión del odorante urinario con la fracción proteica para que la acción de feromona sea efectiva. En la naturaleza los odorantes volátiles atraen a los congéneres y los invita a realizar una prospección química de las proteínas mediante el lamido. Efectos de feromona bien conocidos son los producidos por MUP, como el adelanto del estro en hembras prepuberales (efecto Bandenbergh) y la inducción del estro, tras su supresión (efecto Whitten) [2d].

El receptor de las lipocalinas urinarias es el órgano vomeronasal, que tras ser estimulado enviá la información al hipotálamo donde el ciclo sexual es controlado.

La afrodisina es una lipocalina aislada de las secreciones vaginales del hamster dorado. Es sintetizada *in situ* por las glándulas del cervix del útero. La actividad feromona de la afrodisina se mantiene tras retirar la fracción orgánica unida a ella, pero desaparece con la proteolisis. Esto indica que la fracción proteica es la responsable de dicha acción [2d].

4.3.- Proteínas olfativas y gustativas

Un conjunto de lipocalinas son propias de las secreciones de mucosa nasal, son las llamadas proteínas de unión a odorantes, denominadas OBPs (*odorant binding proteins*) [2 y 3]. Han sido identificadas en diferentes especies de mamíferos y en un único caso de vertebrado no mamífero,

en la rana (BG; proteína de la glándula de Bowman).

Una característica importante de las OBPs es su capacidad de unión reversible a un amplio espectro de moléculas orgánicas de tamaño medio y naturaleza hidrofóbica. Las OBPs son sintetizadas por las glándulas del epitelio nasal, en las regiones respiratorias o vomeronasal, pero no en la olfativa [2]. La función que desempeña no está totalmente clara, pero el sitio donde se expresa, su afinidad por compuestos de tipo feromona y su gran similitud a las proteínas urinarias, ya mencionadas, sugieren que las OBPs juegan un papel en la percepción de las feromonas en el órgano vomeronasal.

Por otra parte hay evidencias que otra lipocalina, la VEGP (*Von Ebner´s-gland protein*), puede colaborar en la sensación del gusto mediante la eliminación del sabor amargo, uniéndose a compuestos que tienen esta propiedad [3]. VEGP es también secretada por la glándula lacrimal en el fluido de la lágrima. Se ha sugerido su colaboración con la lisozima en su acción bactericida y que ella misma podría tener acción bactericida mediante el transporte de compuestos que presentan esta propiedad.

4.4.- <u>Modulación del Sistema Inmune</u>

Las concentraciones en plasma de diversas proteínas varían durante la respuesta en fase aguda, una reacción fisiológica compleja ante el estrés y la inflamación, que juega un importante papel en el progreso de las enfermedades. Entre las proteínas que muestran este aumento se encuentran las lipocalinas: AGP (*alfa1-acid glicoprotein*), NGAL (*neutrofil lipocalin*), PP14 (*pregnancy protein 14*) y A1M (*alfa1-microglobulina*) [3]. Se cree que estas lipocalinas tienen función inmuonsupresora o antiinflamatoria previniendo frente a los daños en tejidos, aunque también parecen desempeñar un importante papel en esta función mediante el transporte de factores. Hemos de citar además a la lipocalina C8γ, que es una de las subunidades de los ocho componentes del complemento (C8), que junto con otros componentes se une a la membrana de organismos patógenos dando lugar a la formación del llamado complejo de ataque de membrana [3].

Se ha comprobado que PP14 tiene efecto supresor de la actividad de los linfocitos T asesinos, así mismo en presencia de PP14 la interleucina-2 pierde su capacidad de incrementar la proliferación de linfocitos T [3]. Se ha demostrado que A1M suprime la proliferación policlonal, inducida por

antígenos, de linfocitos cultivados. Esta misma lipocalina inhibe la migración espontánea de granulocitos neutrófilos *in vitro* e inhibe la atracción quimiotáctica de los granulocitos frente a un gradiente de citoquinas [3]. Respecto a AGP también hay evidencias de su capacidad anti inflamatoria e inmunoreguladora, funciones estas que parecen estar muy relacionadas con el tipo de glicosilaciones que presenta esta lipocalina, que recubren la casi totalidad de su superficie molecular [3].

4.5.- Función Enzimática

La lipocalina PGDS (*prostaglandin D synthase*) es el primer miembro de esta familia en ser reconocido como enzima. Dicha lipocalina cataliza la isomerización del precursor de prostaglandina PGH_2 a PGD_2 , el cual es un potente somnógeno interno, así como un modulador de la nociocepción [2e, 3].

PGDS es una lipocalina muy glicosilada que se expresa principalmente en sistema nervioso central y en órganos genitales de mamíferos. Una vez producida es secretada al líquido cerebro-espinal o al plasma seminal respectivamente. PGDS demuestra además tener afinidad por diversos ligandos de carácter lipófilo (PGD_2, bilirrubina, ácido retinoico, etc) , por ello se sugiere que es una lipocalina multifunción con funciones enzimática y de transporte [2e].

4.6.- Regulación celular

Si hay una lipocalina que parece estar implicada en múltiples procesos, que podríamos calificar de regulación celular, es ApoD (*apolipoprotein D*). En humano esta lipocalina se encuentra en numerosos fluidos, aunque poco expresada en hígado e intestinos, donde por el contrario si son expresadas otras lipoproteínas [2f].

Diversas evidencias sugieren una relación entre la expresión de ApoD y la proliferación celular. Así mismo hay evidencias de la relación de la expresión de ApoD con la fase celular de detención del crecimiento permanente (*growh arrest*). La modulación de la expresión de ApoD puede observarse en diversas patologías: diabetes tipo 2, disfunción renal y daños en tejidos por isquemia, entre otras [2f].

En el hipotálamo ApoD interacciona específicamente con la parte citoplasmática de la forma larga del receptor de leptina (Ob-Rb), implicado en la regulación de la ingesta de alimento y en el peso

corporal. Los indicios sugieren que ApoD estaría implicado en la ruta de señales de transducción que controla la acumulación de grasa corporal [2f].

ApoD está también implicado en la gestación y el desarrollo del feto. En ratón ApoD está selectivamente modulado, desde E9 al nacimiento, en mesenquima y en neuroepitelio. En cerebro de rata, durante el desarrollo y en el periodo neonatal temprano, la inducción de la expresión de ApoD coincide con el periodo de mielinización y con la formación de sinapsis. En el sistema nervioso central ApoD podría estar implicado en el transporte de hormonas esterorideas y por lo tanto participar en los procesos de reinervación. ApoD se incrementa en el fluido cerebroespinal de pacientes con Alzheimer, así como en pacientes de ictus, meningoencefalitis, demencia y otros trastornos. Sin embargo los niveles de ApoD son bajos en el serum de pacientes con esquizofrenia, reforzando este hecho las hipótesis recientes que apuntan a defectos sistémicos en el metabolismo de los lípidos como el origen de esta enfermedad [2f].

La expresión de ApoD se ve incrementada tras daños cerebrales agudos en astrocitos y oligodendrocitos , así como en neuronas. Igualmente ApoD podría ser parte de un sistema de defensa antioxidante y actuar como un eliminador de moléculas relacionadas con hemo, como la bilirubina.

Respecto a cáncer ApoD es sobre-expresado en ciertos tipos de cáncer como de pulmón, ovario y piel entre otros. Sin embargo es sub-expresado en otros o incluso suprimido por metilaciones del ADN, como en el carcinoma esofágico de células escamosas. A pesar de ello la correlación entre la expresión de ApoD y las distintas fases de los tumores permanecen ambiguas. En qué medida la expresión de ApoD es una causa o una consecuencia en estas transformaciones celulares es una cuestión que permanece sin resolver [2f].

La regulación de la expresión de ApoD es compleja y muchos autores han demostrado la influencia de diversos factores biológicos en la modulación de la misma [2f].

5.- Los genes de las lipocalinas y su historia evolutiva

El gen ancestral de las lipocalinas parece haberse originado en un grupo de bacterias y probablemente fue heredado con posterioridad por los eucariotas, debido a una fusión genómica [2g]. Considerando este origen hipotético hemos de esperar que las lipocalinas estén presentes en todos los descendientes del eucariota común ancestral. Actualmente, además de en las bacterias, se

han encontrado lipocalinas genuinas en un protoctista, un hongo, varias plantas, un nematodo, varios artrópodos, un tunicado, un cefalocordado y en muchos cordados [2g].

5.1.- Estructura del gen de las lipocalinas

Bien por secuenciación directa, o por métodos bioinformáticos, se ha podido llegar a conocer la estructura exón-intron de las diversas lipocalinas eucariotas. En la figura 5 se ofrece el panorama de la estructura génica de un determinado número de lipocalinas sin pretensión de ser exhaustivo.

Si nos centramos en animales, cuando se señala la posición de los intrones en los alineamientos de lipocalinas, el patrón que emerge es un modelo de organización con un máximo de cinco exones (e1-e5) y cuatro intrones (A-D) en artrópodos, frente a un máximo de siete exones (e1-e7) y seis intrones (A-F) en cordados (figuras 5 y 6), siendo además el tamaño de los exones y la fase de los intrones semejantes entre diferentes lipocalinas [7, 8]. Estos rasgos han permitido realizar alineamientos de las lipocalinas más atípicas con el resto y así poder realizar análisis filogenéticos más completos de esta familia de proteínas en el reino animal. Más allá de esto, aun no se han identificado suficientes genes de lipocalinas en otros filum, como para extraer de ellos inferencias a partir de su organización exón-intrón.

Además de estas características, existe otro aspecto que puede ser utilizado para la identificación de exones, sin ambigüedad, se trata de los motivos comunes de lipocalinas. Dichos motivos son cortas secuencias que codifican para una secuencia peptídica característica, los llamados motivo GxW en el e2 y el motivo TDY en el e4 (ver figura 6). La presencia de estos motivos es la única herramienta fiable que permite afirmar por ejemplo que los dos únicos exones de las lipocalinas de plantas debe formar parte del conjunto de exones encontrados en animales [2g] (ver figura 5).

En vertebrados, si bien el patrón general de la estructura génica que ya hemos comentado se conserva de forma general, se muestran ciertas desviaciones particulares para algunas lipocalinas. Entre otras podemos citar la pérdida de los intrones D a F en APOD o la incorporación de un exón no codificante 5´adicional en RBP4 y en APOD de mamíferos [2g].

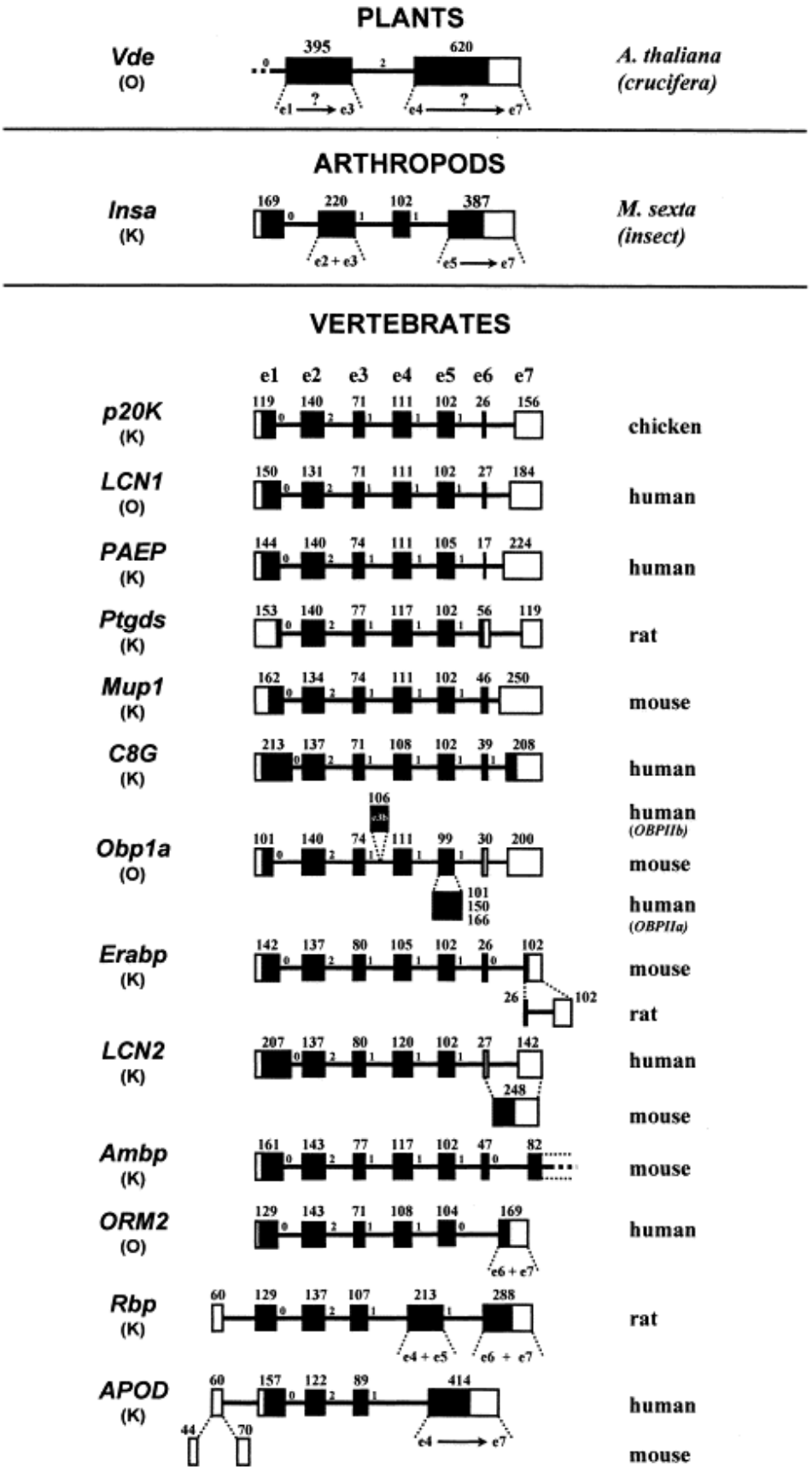


Figura 5. *Organización exón-intrón de los genes de lipocalinas. Debajo del nombre se indica si es una lipocalina principal (K, kernel) o periférica (O, outlier). Los exones representados por bloques; negros ORF, blancos UTRs , sobre ellos un número indica su tamaño (nt). Los intrones representados por líneas continuas no están a escala, sobre ellos se indica la fase del intrón. La correspondencia de ciertos exones, como en* ***Vde****, con los exones arquetípicos (e1-e7) se indica debajo de estos.Los fenómenos de splicing alternativo también aparecen indicados bajo ciertos exones.Extraída de referencia [7]*

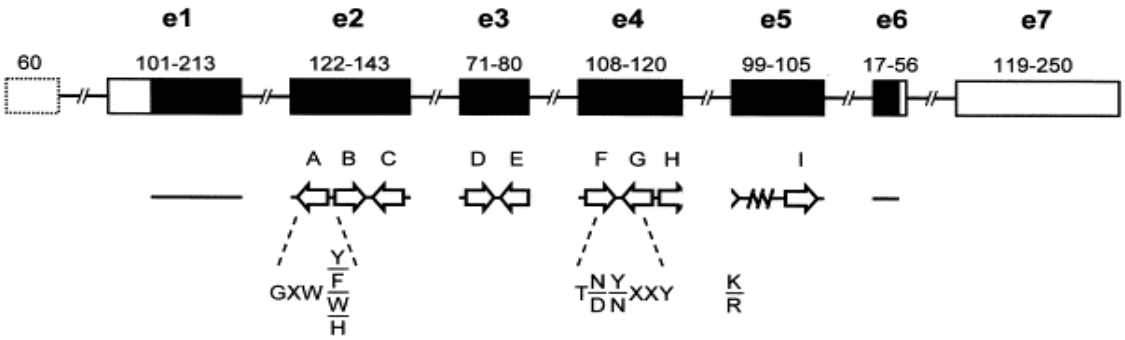


Figura 6. *Organización exón-intrón de lipocalina modelo de vertebrados. Bajo los exones (bloques negros ORF y blancos UTRs) se representa la posición de las hojas beta antiparalelas y otros elementos de su estructura secundaria. Bajo estas se representan los motivos conservados entre lipocalinas. Extraida de referencia [7].*

5.2- Localización cromosómica y clusters de genes

La organización cromosómica de las lipocalinas en vertebrados muestra un curioso patrón. En humano, con la excepción de APOD, APOM y RBP, la mayoría de los genes de lipocalinas están situados en el brazo largo del cromosoma 9 (HSA9q) (Figura 7 y 8). Así mismo sus ortólogos en ratón y rata están agrupados en dos cromosomas diferentes que muestran sintenia con HSA9q [7, 9]. En la gallina un grupo de genes de lipocalinas, que son los equivalentes de HSA9q, se hayan situados en el cromosoma 17 [10]. Así mismo, al igual que en humano, los genes de ApoD y Rbp4 de roedores y gallina se hayan en cromosomas aislados (ver figura 7). En muchas ocasiones , en estos clusters de genes, aparece más de una copia de algunos de ellos. Esto nos indica la tendencia de los genes de lipocalinas a sufrir duplicaciones [2g].

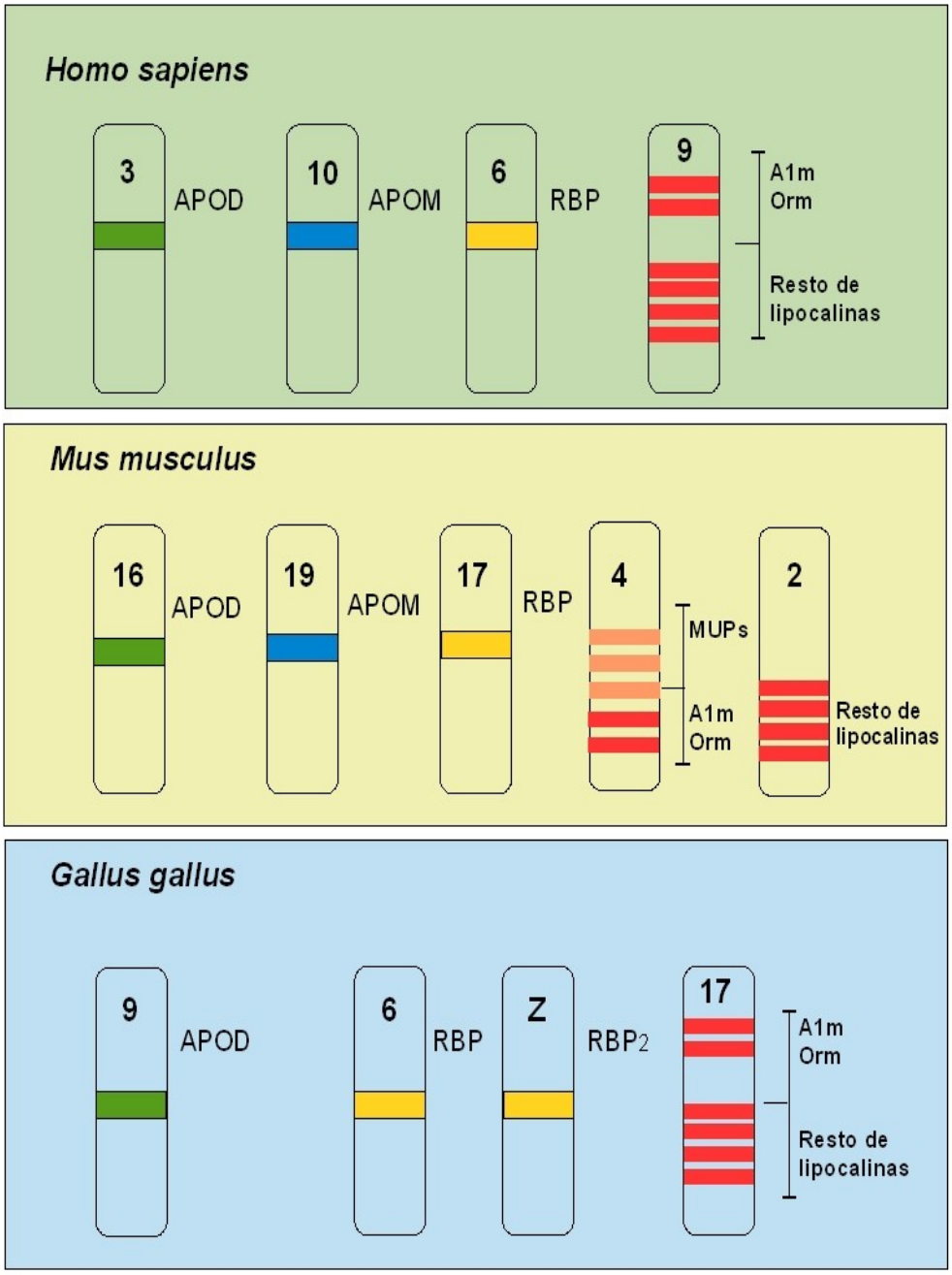


Figura 7. *Localización cromosómica de las lipocalinas en diferentes especies de vertebrados. Las proteínas urinarias (MUPs) del cromosoma 4 de ratón son específicas de roedores. En gallina se ha dado una duplicación de RBP intralinaje.*

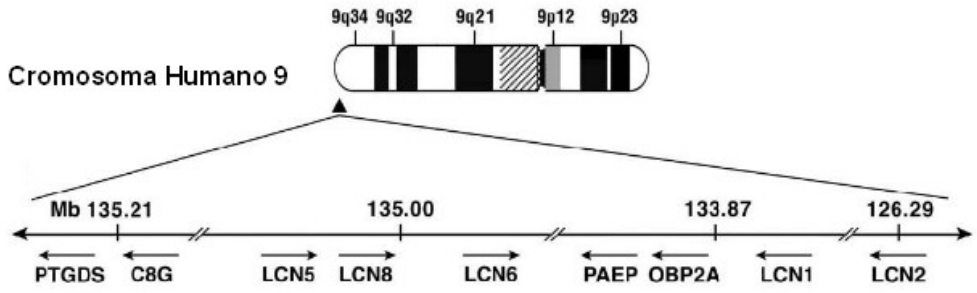


Figura 8. *Organización genómica de algunas lipocalinas del conjunto de las ubicadas en el cromosoma 9 humano. Figura modificada de referencia [14]*

5.3.- Inferencias sobre la evolución de la familia de las lipocalinas

5.3.1.- Evolución de las lipocalinas como resultado del análisis de la estructura génica

Como ya hemos tratado previamente, la estructura génica de las lipocalinas está bien conservada, por lo que podemos asumir que los límites exón-intrón son homólogos y que por lo tanto contienen trazas de la historia evolutiva de estos genes. Para este análisis se desarrolló un método basado en una medida de similitud del límite intrón-exón, límites que son posteriormente mapeados sobre un alineamiento múltiple de secuencias de proteínas de lipocalinas seleccionadas, cuya estructura génica es bien conocida [8]. Tres parámetros son usados para calcular la distancia genética siguiendo este método: el número de intrones presentes en la secuencia codificante (ORF, *open reading frame*), la fase de los intrones y la posición del limite exón-intrón. La restricción del cálculo de esta distancia a la ORF se debe a varios motivos: 1) a las dificultades de alinear las UTRs y por lo tanto a la dificultad de asignar caracteres homólogos de los límites exón-intrón que se encuentren en estas regiones, 2) la menor presión de selección a la que están sometidas estas regiones no codificantes y por lo tanto a su mayor propensión a ganar o perder intrones, y 3) a que la fase de los intrones, presentes en la ORF, es un carácter que aporta una señal filogenética realmente informativa.

Cuando se aplica esta metodología a la familia de las lipocalinas se obtiene un árbol filogenetico

que es congruente con otros obtenidos previamente, basados en el alineamiento de las secuencias de proteínas [11 y 12], dando así soporte y ayudando a interpretar la historia evolutiva de las lipocalinas. Este árbol filogenético obtenido (Figura 9) y , enraizado con la lipocalina del eucariota unicelular Dictyostelium, se muestra acorde con la filogenia de los organismos y clasifica las lipocalinas de artrópodos en tres grupos, dos de ellos con 3 intrones y uno con 4 intrones. La lipocalina de cordados que ocupa una posición más basal en el árbol filogenético es ApoD, el resto de lipocalinas de cordados es monofilético y aparecen separadas en tres grupos, en función de la presencia de 4, 5 o 6 intrones en su ORF.

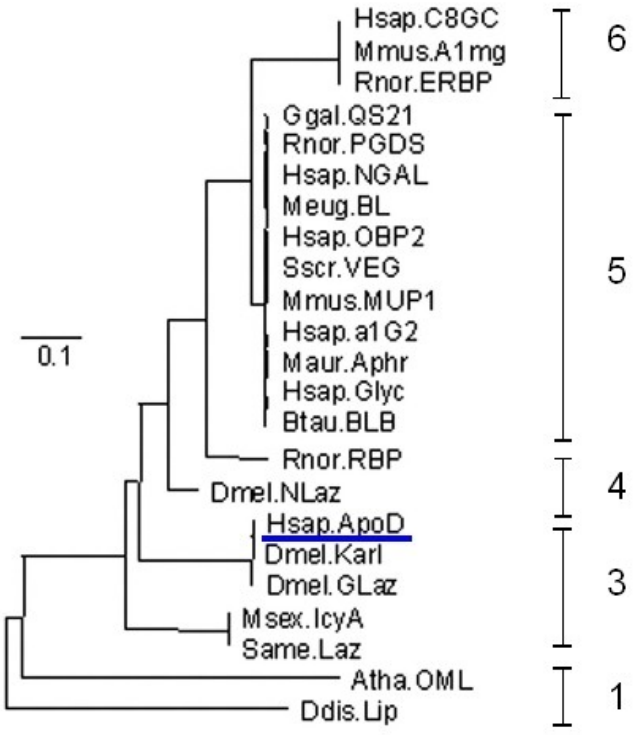


Figura 9. *Árbol filogenético (Neighbor- Joining) de las lipocalinas basado en la organización exón-intrón de los genes seleccionados, enraizado con una lipocalina de protoctista (Ddis.Lip). El número de intrones de la ORF de los genes se indica con un número a la derecha del árbol. Subrayada en azul se muestra la lipocalina de cordados con una posición más basal (Hsap. ApoD), junto a lipocalinas de artrópodos La barra de escala representa la longitud de las ramas (nº de sustituciones de aminoácidos / sitio). Imagen modificada de referencia [8].*

La información más importante que se extrae de este árbol es que las lipocalinas más recientes contienen más intrones en su ORF. Como ya mencionamos anteriormente los intrones E y F no están presentes en las lipocalinas de los no cordados, mientras que los intrones A-D muestran una amplia distribución filogenética, estando los intrones A y C presentes en todos los metazoos

incluidos en este análisis. Por lo tanto la historia evolutiva de las lipocalinas de metazoos puede ser mejor trazada a través de la distribución de los intrones B y D [2g].

5.3.2.- Evolución de las lipocalinas como resultado del análisis de la secuencias de proteínas

El usos del alineamiento múltiple de las secuencias de aminoácidos de las lipocalinas nos permite explorar su historia evolutiva con más detalle. Un gran número de lipocalinas pueden ser utilizadas en estos estudios gracias al gran número de secuencias de proteína y ARNm disponibles en las bases de datos. Este método permite la inclusión de las lipocalinas de procariotas, cosa que no puede hacerse con el método de la estructura génica.

Han podido alinearse las secuencias de 210 lipocalinas, a las que se ha aplicando un filtro de penalización por hueco (*gap penalty mask*), para penalizar los huecos dentro de los elementos de la estructura secundaria y con los mínimas correcciones manuales [11, 12], basadas en el conocimiento de la estructura y función de las lipocalinas. Con este alineamiento, utilizando un método Bayesiano [13], y tomando como raíz las lipocalinas bacterianas se obtiene el árbol filogenético de la figura 10.

Algunos aspectos interesantes se desprenden de este árbol. En primer lugar las lipocalinas de plantas y hongos aparecen relacionadas con las lipocalinas de bacterias, que son las que enraízan el árbol. La lipocalina del protoctista (Dictyostelium) se ubica en uno de los grupos de lipocalinas bacterianas. La lipocalina ApoD se encuentra asociada a un grupo de lipocalinas de artrópodos, que comparten expresión génica en el sistema nervioso, lo que sugiere que ApoD podría ser la lipocalina ancestral de los cordados. Esta hipótesis toma fuerza por el hecho de que la lipocalina del tunicado (Cint.Lip) ocupa una posición basal, junto con el grupo de ApoD y lipocalinas de artrópodos mencionado, y porque su secuencia muestra una similitud máxima cuando se alinea con ApoD [2g].

Podemos observar también que la lipocalina ApoM aparece en una posición basal del sub-árbol de lipocalinas de cordados. Respecto al resto de lipocalinas no hay diferencias a otros árboles filogenéticos obtenidos previamente [11, 12], siendo Rbp4 y el grupo Pgds-Ngal las que se encuentran más relacionadas con la ancestral ApoD.

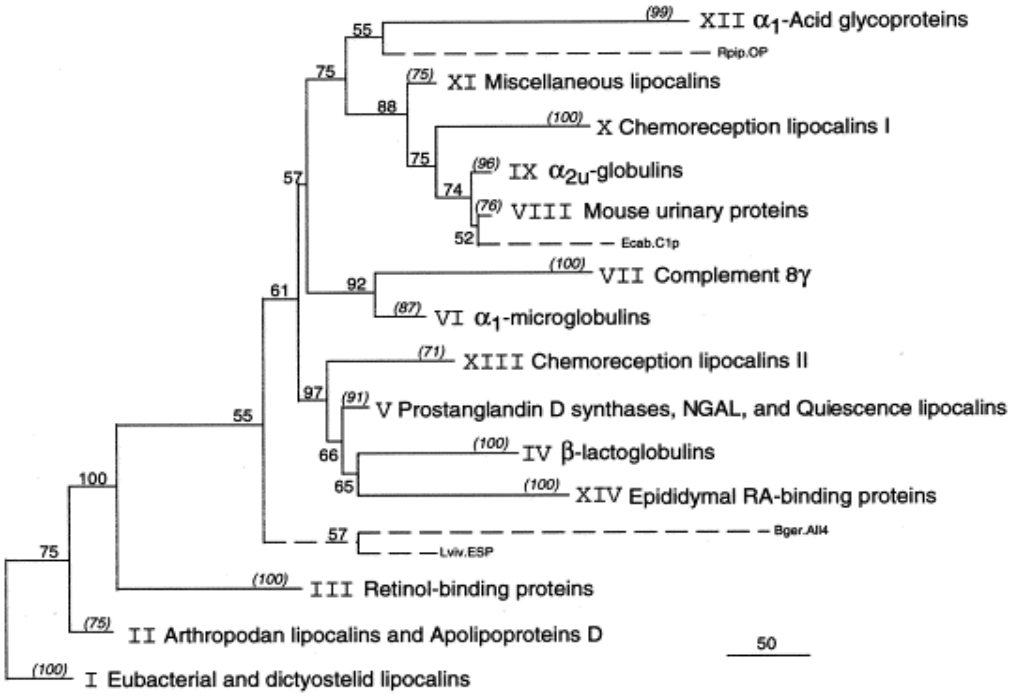


Figura 10. *Árbol filogenético de lipocalinas enraizado con las secuencias de eubacterias y dictisotelido como un grupo externo. Los 14 clados monofiléticos se indican con números romanos. Los valores de LBP (local boopstrap porportion) se indican en cada nodo del análisis ML (Maximun Likelihood). Las lipocalinas no agrupadas en clados aparecen señaladas con líneas discontinuas. La barra de escala representa la longitud de las ramas (sustitución de aminoácidos / 100 residuos).Imagen extraida de referencia [12].*

Esta representación filogenética permite clasificar las lipocalinas en diferentes clados (ver figura 10), que se aplican principalmente a las lipocalinas de cordados, si bien estos clados están sometidos a continua revisión conforme aparezcan nuevos datos. Para el resto de lipocalinas, de otros filum, es necesario la acumulación de más información para que análisis filogenéticos más detallados permitan su agrupamiento en dichos clados.

Este análisis filogenético, junto con el conocimiento extenso de la función y estructura de las lipocalinas, apunta a la siguiente hipótesis [12]. La extraordinaria diversificación de las lipocalinas que ha acompañado a la radiacción evolutiva de los vertebrados aparentemente liberó a las lipocalinas parálogas de las constricciones estructurales y funcionales. Las lipocalinas más modernas parecen haber evolucionado a un mayor ritmo de divergencia en sus secuencias y con una mayor tasa de duplicaciones génicas. Conforme las duplicaciones y esta divergencia han tenido lugar, el bolsillo interno de las lipocalinas más modernas ha evolucionado para unirse a ligandos hidrofóbicos más pequeños y con más eficiencia mayor que las lipocalinas ancestrales.

5.4.- <u>Una hipótesis para la evolución de las lipocalinas</u>

La posición cromosómica de los genes de lipocalinas, que expusimos anteriormente ya nos apunta algo sobre su historia evolutiva. El hecho de que determinadas lipocalinas de mamífero y gallina, pertenecientes a los clados IV-XII, estén agrupadas en un mismo cromosoma nos sugiere que los precursores de estos clados estaban ya presentes en los ancestros comunes de reptiles y aves. Posteriormente han tenido lugar duplicaciones en tandem de este cluster de genes, lo que se confirma por las semejanzas en secuencia y en estructura génica entre ellos y en algunos casos por las similitudes de su expresión génica y función [2g]. El mismo razonamiento puede aplicarse a los genes de APOD y RBP, para los cuales podemos suponer una ubicación en cromosomas separados en los primeros vertebrados terrestres [2g].

En el intento e ofrecer una hipótesis robusta sobre el origen de las lipocalinas hemos de considerar su representación en el árbol de la vida además de las consideraciones filogenéticas mencionadas. Si consideramos el número de lipocalinas legítimas de los distintos taxones, sin considerar las duplicaciones intraespecie, comprobamos que un sólo gen/especie está presente en procariotas, protoctistas y hongos y al menos dos genes para plantas, si bien su peculiar estuctura exón-intrón parece indicar una evolución independiente.

El reino animal heredó el gen ancestral simple de lipocalina de procariotas y tras sufrir duplicaciones daría lugar a dos genes, que evolucionarían hacia diferentes estucturas génicas, bien con 3 o con 4 intrones en su ORF. Basándonos en las similitudes de organización de exón-intron entre los dos filum animales de artrópodos y cordados, podemos sugerir, que estos dos genes de lipocalinas estaban ya presentes en el ancestro común de ambos [2g].

Estos dos genes de lipocalina ancestrales sufrirían una evolución divergente entre estos dos filum ya que estarían expuestos a diferentes paisajes adaptativos. En artrópodos al menos cuatro genes parálogos son comunes en este filum. Respecto a los cordados, por lo ya comentado en los análisis filogenéticos, ApoD es el mejor candidato a sucesor de uno de los dos genes de lipocalinas ancestrales (de 3 intrones). La lipocalina ancestral de 5 exones (4 intrones) fué probalemente una lipocalina tipo RBP, dada la posición basal de RBP en los árboles filogenéticos. Esta hipótesis se ve reforzada por la presencia de RBP en el cefalocordado Branchyostoma [2g].

Posteriormente y coincidiendo con las sucesivas duplicaciones, a escala genómica, ocurridas en la evolución temprana de los cordados, la RBP ancestral sufriría duplicaciones que daría lugar a dos

nuevas lipocalinas situadas en cromosomas separados. Estas dos lipocalinas fueron probablemente los ancestros de las actuales PGDS y APOM, hipótesis reforzada por la filogenia de secuencia y estructura génica y por la presencia de estas dos lipocalinas en peces y de PGDS en Branchyostoma [2g].

Diversos argumentos apuntan a PGDS como la responsable de haber originado, mediante sucesivas duplicaciones durante la evolución de los cordados el cluster de lipocalinas presentes en un sólo cromosoma: 1) una posición basal de PGDS en los diversos árboles filogenéticos, 2) una estructura génica similar a las duplicadas, 3) su presencia en cada cordado estudiado y 4) un lugar de expresión similar a ApoD [2g]. En este proceso de duplicaciones de PGDS, la lipocalina A1m, puede ser propuesta como el primer descendiente de ella, como nos sugiere la presencia de A1m en peces [2g]. Siguientes rondas de duplicaciones en tandem de los genes de PGDS y A1m generarían el resto de componentes del cluster, siguiendo un patrón no del todo conocido aún. Esta hipótesis de proceso evolutivo permanece en continua revisión debido a la información que se incorporará procedente de nuevos genomas y de las nuevas investigaciones sobre función y expresión de las lipocalinas.

5.5.- Conclusiones

El modelo de proceso evolutivo que hemos esbozado pone de manifiesto la presencia de lipocalinas en todos los reinos y su gran expansión en metazoos, en los que se ha mantenido un gran número de genes duplicados parálogos. El camino evolutivo seguido en los dos filum mejor estudiados, artrópodos y cordados, ha sido muy diferente. Un pequeño número de lipocalinas está presente en la mayoría de especies de artrópodos, mientras que duplicaciones intra-linaje han originado lipocalinas específicas a ciertos estilos de vida, en algunas especies de este filum.

En los cordados también se han dado duplicaciones intra-linaje (como ejemplo podemos mencionar las proteínas urinarias de roedores), pero en este caso, si hay un gran número de genes de lipocalinas parálogos comunes al filum, originados por duplicaciones a escala genómica. En general estos parálogos de lipocalinas no mantienen la misma función proteica, como es el caso de otras familias de proteínas (globinas o genes Hox). La divergencia de las secuencias de proteínas de las lipocalinas parálogas, aún manteniendo una estructura semejante, ha abierto nuevos caminos que ha posibilitado nuevas interacciones moleculares y la participación de las lipocalinas en funciones diversas.

6. Bibliografía

[1] Unterman, R.D., Lynch, K.R., Nakhasi, H.L., Dolan, K.P., Hamilton, J.W., Cohn, D.V., Feigelson, P. Cloning and sequence of several alpha 2u-globulin cDNAs. *Proc Natl Acad Sci U S A.* **78,** 3478-82 (1981).

[2a] Akerstrom, B. et al. "Lipocalins 2005: An Introduction". Lipocalins Book Review. *Landes Bioscience* (2006)

[2b] Akerstrom, B. et al. "Lipocalins Protein Family". Lipocalins Book Review. *Landes Bioscience* (2006)

[2c] Akerstrom, B. et al. "RBP". Lipocalins Book Review. *Landes Bioscience* (2006)

[2d] Akerstrom, B. et al. "BLG", MUP & OBP. Lipocalins Book Review. *Landes Bioscience* (2006)

[2e] Akerstrom, B. et al. "PGDS". Lipocalins Book Review. *Landes Bioscience* (2006)

[2f] Akerstrom, B. et al. "Plasma Lipocalins A1agp, ApoD, ApoM, C8GC". Lipocalins Book Review. Landes Bioscience (2006)

[2g] Akerstrom, B. et al. "Lipocalins Genes and their Evolutionary History". Lipocalins Book Review. Landes Bioscience (2006)

[3] Flower, D.R. The lipocalin protein family: structure and function. *Biochem J.* **15,** 1-14 (1996)

[4] Chakraborty, S., Kaur, S., Tong, Z., K. Batra, S. & Guha, S. Neutrophil Gelatinase Associated Lipocalin: Structure, Function and Role in Human Pathogenesis, Acute Phase Proteins - Regulation and Functions of Acute Phase Proteins, Prof. Francisco Veas (Ed.), ISBN: 978-953-307-252-4 (2011).

[5] Greene, L.H., Hamada, D., Eyles, S.J., Brew, K. Conserved signature proposed for folding in the lipocalin superfamily. *FEBS Lett.* **553,** 39-44 (2003).

[6] Frenette Charron, J.-B. *et al.* Identification, Expression, and Evolutionary Analyses of Plant Lipocalins. *Plant Physiology* **139,** 2017–2028 (2005).

[7] Salier, J.-P. Chromosomal location, exon/intron organization and evolution of lipocalin genes. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1482,** 25–34 (2000).

[8] Sánchez, D., Ganfornina, M. D., Gutiérrez, G. & Marín, A. Exon-Intron Structure and Evolution of the Lipocalin Gene Family. *Molecular Biology and Evolution* **20,** 775–783 (2003).

[9] Chan, P., Simon-Chazottes, D., Mattei, M.G., Guenet, J.L., Salier, J.P., Comparative Mapping of Lipocalin Genes in Human and Mouse: The Four Genes for Complement C8 γ Chain, Prostaglandin-d-Synthase, Oncogene-24P3, and Progestagen-Associated Endometrial Protein Map to HSA9 and MMU2, *Genomics* **23,** 145-150 (1994).

[10] Pagano, A. *et al.* Phylogeny and regulation of four lipocalin genes clustered in the chicken genome: evidence of a functional diversification after gene duplication. *Gene* **331,** 95–106 (2004).

[11] Ganfornina, M. D., Gutiérrez, G., Bastiani, M. & S, D. A Phylogenetic Analysis of the Lipocalin Protein Family. *Molecular Biology and Evolution* **17,** 114–126 (2000).

[12] Gutiérrez, G., Ganfornina, M. D. & Sánchez, D. Evolution of the lipocalin family as inferred from a protein sequence phylogeny. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1482,** 35–45 (2000).

[13] Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed

models. *Bioinformatics* **19,** 1572–1574 (2003).
[14] Hamil, G., H. *et al*. Lcn6 a novel human epididymal lipocalin. *Reproductive Biology and Endocrinology*, **1,** 112 (2003)

II

REGULACIÓN DE LA EXPRESIÓN GÉNICA EN EUCARIOTAS

1. - Introducción

A lo largo de la última década se ha hecho evidente que la regulación de la expresión génica en eucariotas superiores es un proceso complejo y finamente regulado, implicando a diversos factores y con diferentes niveles de regulación. Para un determinado gen, son las regiones promotoras, las regiones UTRs (regiones no traducidas) 5´y 3´, junto con los intrones, las principales regiones implicadas en la regulación de su expresión [1]. Las regiones intergénicas, que habían sido calificadas de ADN basura por mucho tiempo, han demostrado tener también un papel relevante en la regulación de la expresión génica [2].

La acumulación de evidencias indica que la complejidad de los organismos superiores, que se correlaciona con un aumento en el tamaño de las regiones no codificantes, proviene de un aumento en el número y complejidad de estas vías de regulación [3] y que además las diferencias fenotípicas entre individuos y entre especies provienen principalmente de las variaciones en estas secuencias no codificantes [4].

Como hemos mencionado la regulación de la expresión génica es ejercida en diferentes niveles, que van desde modificaciones de la cromatina y de las histonas o las metilaciones, hasta la regulación de la transcripción (factores de transcripción, promotores alternativos, etc) y la regulación postranscripcional (maduración del ARNm, estabilidad del ARNm, regulación de la traducción, etc) . En esta revisión nos centraremos en los principales mecanismos del nivel transcripcional y postranscripcional ya que tienen una mayor relación con los contenidos de la tesis, siendo además estos los mecanismos de regulación de expresión génica más relevantes.

2. - Regulación de la transcripción

2.1. - La región promotora

El promotor eucariota es una región reguladora del ADN situada principalmente corriente arriba del gen donde se unen ciertos factores permitiéndose así la coordinación de todos los componentes del complejo de iniciación de la transcripción, incluida la ARN polimerasa II, que comienza la transcripción. El núcleo promotor generalmente se extiende sobre unas 80 pb alrededor

del sitio de inicio de la transcripción (TSS), y en mamíferos pueden diferenciarse dos clases: promotores conservados enriquecidos en la caja TATA (TATA-box), que contienen un solo TSS y promotores variables enriquecidos en CpG, conteniendo múltiples TSS (llamados promotores dispersos) [5]. La segunda clase es abundante en vertebrados, al menos representan el 70 % de los promotores humanos [19] y están relacionados con genes de mantenimiento celular (*housekeeping*) [7]. La expresión mediante estos promotores dispersos implica los efectos combinatorios de un gran número de sitios de unión de factores dentro de la región promotora.

2.1.1.- Promotores con TATA-box

El promotor mínimo de eucariotas se define como la secuencia de ADN mínima requerida para que se inicie una transcripción correcta *in vivo* [16]. Típicamente, consta de 40 pb, situados tanto aguas arriba como aguas abajo del sitio de inicio de la transcripción. Sus elementos componentes son cuatro, ordenados en dirección 5' a 3' (ver figura 1):

- BRE (*TFII*B r*ecognition* e*lement*), secuencia reconocida por el factor de inicio de la transcripción TFIIB.
- Caja TATA , elemento reconocido por la proteína de unión a la caja TATA o TBP (de T*ATA* b*inding* p*rotein*).
- Inr o iniciador, que alberga al punto de inicio de la transcripción y al cual se une el factor TFIID, y, probablemente, la propia ARN polimerasa II,
- DPE (D*ownstream* p*romoter* e*lement*), o elemento aguas abajo del promotor, reconocido por TFIID.

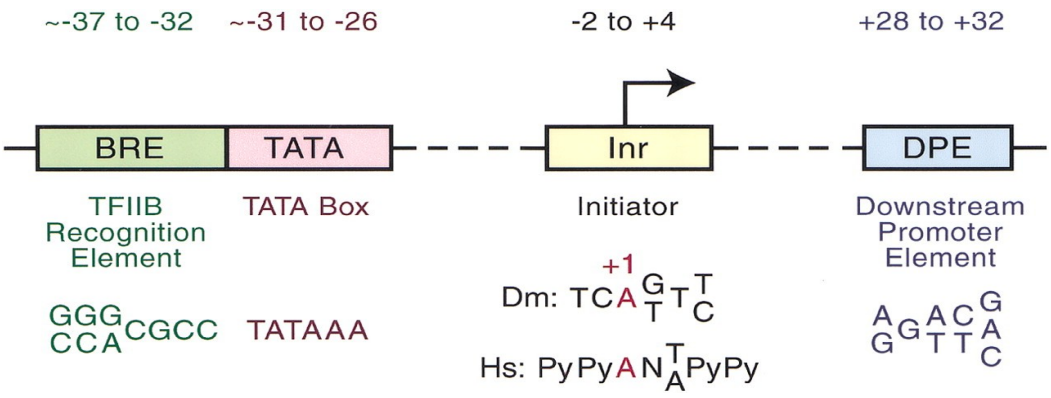


Figura 1. Elementos del *promotor mínimo eucariota*, que se mencionan en el texto. Se indica arriba la posición que ocupan respecto al TSS (indicado con flecha, en posición +1) y abajo la secuencia consenso de los mismos. Imagen extraída de la referencia [16]

No obstante, otras secuencias están implicadas en la regulación del proceso de transcripción. Habitualmente situadas aguas arriba del promotor mínimo, reciben el nombre de secuencias reguladoras. De naturaleza diversa, pueden agruparse en varias categorías, según su estructura y actividad estimuladora o inhibidora.

El mecanismo básico que actuá en una región promotora típica vendría dado por los siguientes pasos [17]:

1) Un activador se uniría a la región diana correspondiente corriente arriba del TSS.

2) Esto facilitaría el reclutamiento del factor de transcripción TFIIA y del complejo TFIID, formado a su vez por TBP (proteína de unión a TATA) y el complejo TAF.

3) Todo este complejo de iniciación reclutaría a la ARN polimerasa II, que procedería al inicio de la transcripción (ver Figura 2).

2.1.2. - Promotores dispersos (asociados a CpG)

Si bien el mecanismo de los promotores con TATA-box es bien conocido, el de los promotores dispersos (carentes de TATA-box) ha mostrado ser mucho más complejo y consecuentemente está peor caracterizado. Estos promotores son generalmente ricos en GC y contienen múltiples sitios de unión de factores de transcripción Sp1 (las llamadas cajas GC) [6]. La presencia de múltiples sitios de unión es un buen ejemplo de como los promotores dispersos funcionan de una manera compleja, dando lugar a que sean utilizados distintos TSS. Múltiples proteínas Sp1 pueden unirse en varios sitios y a la misma vez. Se han identificado diferentes isoformas de Sp1 y ciertas modificaciones postraduccionales pueden hacer que se comporten de forma diferente, bien potenciando o inhibiendo la transcripción. La conformación del complejo TFIID (ver figura 2) puede ser diferente al unirse a diferentes núcleos promotores, estableciendo interacciones con diferentes tipos de activadores transcripcionales [6].

Otros elementos intervienen también en la regulación de la transcripción de estos promotores,

principalmente potenciadores, represores y aisladores, todos ellos actuando de manera selectiva con el resto de elementos para contribuir a la acción promotora [7]. Estudios del efecto causado por delecciones han demostrado que hay regiones que afectan positivamente a la transcripción 300 pb corriente arriba del TSS y regiones que afectan negativamente hasta 1000 pb corriente arriba del mismo [8]. Estos hechos ponen en evidencia que incluso regiones regiones muy distantes del TSS pueden afectar a la actividad promotora.

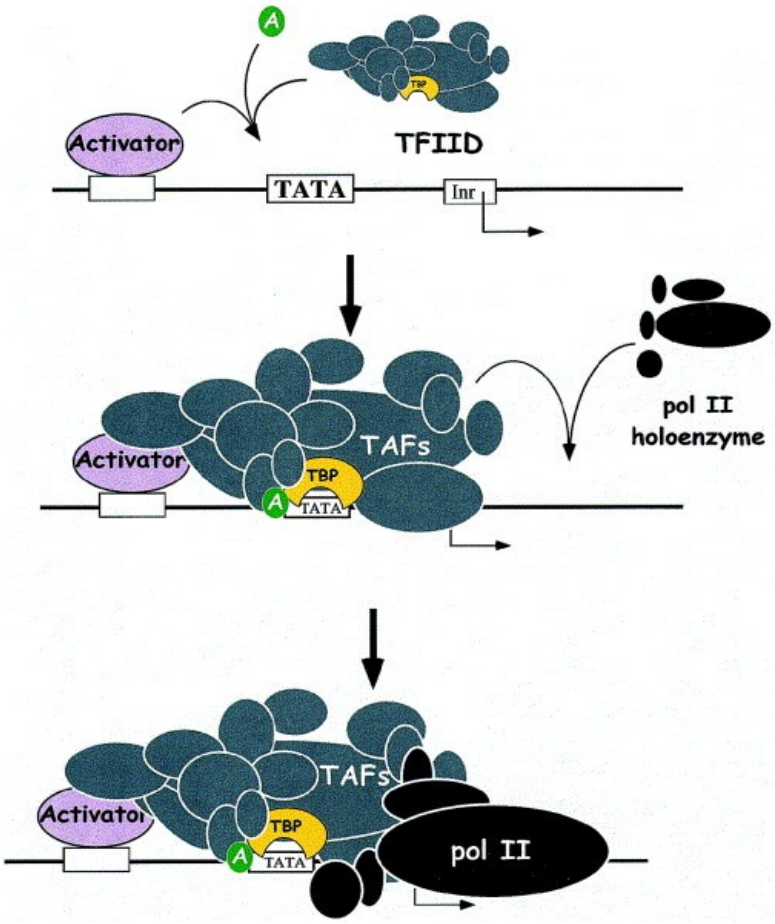


Figura 2. Esquema simplificado de los principales pasos de formación del complejo de inicio de la transcripción. El elemento "A" representa a FTIIA que su uniría a una secuencia tipo BRE. El resto de elementos son los mencionados para el *promotor mínimo eucariota* en el texto. Figura extraída de la referencia [17].

2.2. - <u>Promotores alternativos</u>

Los promotores alternativos son frecuentes en mamíferos y especialmente en humanos y pueden tener un gran impacto sobre la expresión génica [8]. Los promotores alternativos suponen un mecanismo importante en la expresión de genes en células específicas o en estados de desarrollo

específicos [9]. Diversos análisis a escala genómica revelan que entre un 30 y un 50% de los genes humanos presentan promotores alternativos y que pueden llegar a extenderse por cientos de Kilobases [8 y 10].

Un ejemplo del uso de promotores alternativos lo tenemos en el gen de la hemoglobina γ A (HBG1) humana. Uno de los promotores es deficiente en TATA-box y el otro si la contiene, de manera que son utilizados, de forma diferenciada, durante y después el desarrollo embrionario respectivamente [11]. Demostrando así que el aparato de transcripción puede ser reclutado hacia diferentes promotores en función del estado de desarrollo. Otro ejemplo pone de manifiesto la complejidad resultante del uso de promotores alternativos, es el caso del factor de regulación de la transcripción MITF I (que interviene en el desarrollo del ojo de vertebrados). Cada uno de los nueve promotores alternativos asociados a este gen produce una isoforma, conteniendo cada una de ellas un primer exón diferente y como consecuencia un sitio de unión de la proteína diferente, estableciéndose así una diferente regulación espacio-temporal de la expresión durante el desarrollo del ojo [12].

El uso de promotores alternativos puede afectar también a la región UTR 5´, la cual a su vez influye en la estabilidad o eficiencia de la traducción de los ARNm alternativos resultantes. Un ejemplo lo tenemos en el gen homeobox de estatura corta SHOX que utiliza dos promotores, dando lugar a dos UTRs 5´diferentes [13], una de las cuales es más larga y estructurada. Esto tiene como consecuencia que el mismo producto proteico sufra una diferente combinación de mecanismos de regulación transcripcionales y traduccionales (ver figura 3). Otros ejemplos de promotores alternativos que originan transcritos con diferentes UTRs 5´, que produciendo la misma proteína, sufren una regulación de su expresión diferente (según tipo de tejido o estado de desarrollo), los encontramos en los genes p18 murino y NOS1 y PPARγ humanos, entre otros [14]. Los efectos regulatorios de las UTRs 5´serán tratados más ampliamente en un apartado específico posterior.

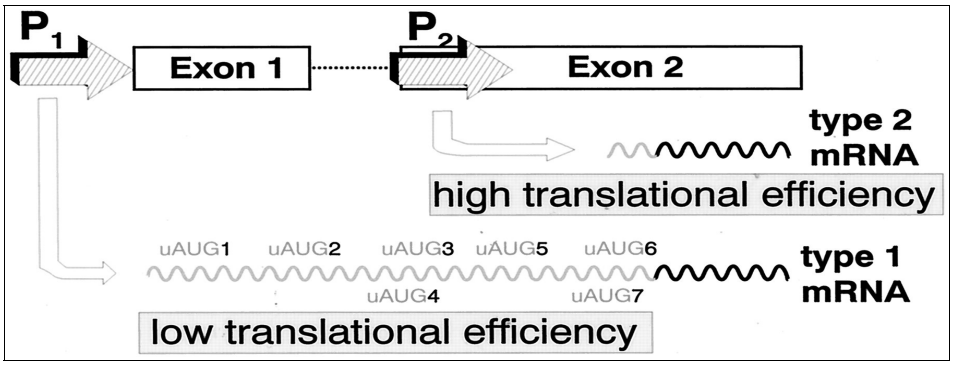


Figura 3. Diferentes transcritos del gen SHOX que difieren en su región UTR 5´(líneas gris claro onduladas)dependiendo del promotor utilizado, P1 o P2. Imagen extraída de referencia [13].

El uso de promotores alternativos tiene efectos que pueden extenderse más allá del efecto directo sobre la transcripción. Existen evidencias de genes (en humano y ratón) en los que el uso de promotores alternativos juega también un papel en la regulación de los mecanismos de splicing, dando como resultado diferencias en el producto proteico resultante [14]. Entre otros podemos citar a los genes humanos NOS1 y CASP2. Por ejemplo, en el gen de la caspasa 2 (CASP2), el uso del segundo de los exones de la región no codificante 5´ (debido a uno de los promotores alternativos) da lugar a la incorporación de un noveno exón codificante variable [15], que tiene como resultado una isoforma más corta (ver figura 4). Varios mecanismos han sido propuestos para explicar esta relación [14]. Por una parte puede estarse dando una interacción diferencial entre los factores que actúan en una región promotora dada y la maquinaria de splicing, modificándose así el comportamiento de esta. Un segundo mecanismo explicaría que ocurriese splicing alternativo como resultado de diferencias en la estructura del ARNm, causadas por la inclusión de una u otra secuencia en el extremo 5´del transcrito. Si bien, estos dos mecanismos no tienen por qué ser mutuamente excluyentes.

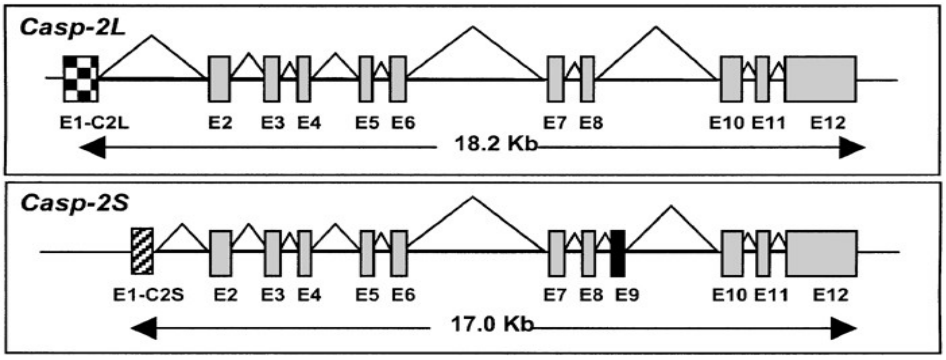


Figura 4. *Isoformas de la caspasa 2 humana. Puede observarse la relación entre el primer exón de la región UTR 5´ y la inclusión alternativa del exón codificante E9. Figura extraida de la referencia [15].*

El abundante uso de promotores alternativos en mamíferos nos lleva a preguntarnos como han llegado estos a aparecer en el proceso evolutivo. Podemos considerar varias vías por las que esto ha podido ocurrir [14]. Una primera podría ser la aparición de mutaciones graduales que en el lugar apropiado diesen lugar a diversos motivos capaces de reclutar a la maquinaria de transcripción, originando con el tiempo una región promotora alternativa. Una segunda posibilidad sería que por recombinación se produjese una duplicación de la región promotora completa y que las

subsiguientes mutaciones alterasen las afinidades o especificidad de tejido de este nuevo promotor. Una tercera posibilidad es la inserción de elementos transponibles (TE) en las proximidades del gen y que este derivase en la formación de un nuevo promotor. Las evidencias genómicas de que en las regiones promotoras de al menos el 25 % de los genes humanos aparecen TEs [18] es una prueba de la importancia que puede tener esta vía en el origen de la complejidad de las vías regulatorias.

2.3. - Conclusión

Es evidente que los promotores eucariotas han evolucionado desde sencillos "interruptores", como los existentes en bacterias, hasta complejas regiones regulatorias multi-factores que encontramos actualmente en vertebrados y especialmente en mamíferos. Los promotores complejos inducen un número variado de respuestas en función de las variaciones ambientales o de las señales celulares, ajustando así el nivel de expresión de los genes a las condiciones requeridas por el tipo de célula o su estado de desarrollo. A esta complejidad contribuye indudablemente el uso de promotores alternativos, característico de mamíferos, aumentando la complejidad de las vías regulatorias de la expresión génica.

2.4. .- Bibliografía

[1]Barrett, L. W. et al., Untranslated Gene Regions and Other Non-coding Elements, *SpringerBriefs in Biochemistry and Molecular Biology*, DOI : 10.1007/978-3-0348-0679-4_1 (2013).

[2] Birney E, et al, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).

[3] Levine, M., Tjian, R. Transcription regulation and animal diversity. *Nature* **424,** 147–151 (2003).

[4] Mattick, J.S. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* **2,** 986–991 (2001).

[5] Carninci P, et al. . Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38,** 626–635 (2006).

[6] Smale ST, Kadonaga JT The RNA polymerase II core promoter. *Annu Rev Biochem* **72**, 449–479 (2003).

[7] Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E Prevalence of the initiator over theTATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389,** 52–65 (2007).

[8] Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res* **16,** 1–10 (2006).

[9] Levine M, Tjian R (2003) Transcription regulation and animal diversity. Nature 424:147–151

[10] Baek, D. et al. (2007) Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. Genome Res. 17, 145–155

[11] Duan ZJ, Fang X, Rohde A, Han H, Stamatoyannopoulos G, Li Q (2002) Developmental specificity of recruitment of TBP to the TATA box of the human gamma-globin gene. Proc Natl Acad Sci USA 99:5509–5514

[12] Bharti K, Liu W, Csermely T, Bertuzzi S, Arnheiter H Alternative promoter use in eye development: the complex role and regulation of the transcription factor MITF. *Development* **135,** 1169–1178 (2008).

[13] Blaschke RJ, Topfer C, Marchini A, Steinbeisser H, Janssen JW, Rappold GA Transcriptional and translational regulation of the Leri-Weill and Turner syndrome homeobox gene SHOX. *J Biol Chem* **278,** 47820–47826 (2003).

[14] Landry, J.-R., Mager, D. L., & Wilhelm, B. T. (s.f.). Complex controls: the role of alternative promoters in mammalian genomes. *Trends in Genetics*, **19**, 640-648.

[15] Logette, E. et al. (2003) The human caspase-2 gene: alternative promoters, pre-mRNA splicing and AUG usage direct isoform-specific expression. Oncogene 22, 935–946

[16] Butler, J. E. F. & Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & Development* **16,** 2583–2592 (2002).

[17] Pugh, B. F. Control of gene expression through regulation of the TATA-binding protein. *Gene* **255,** 1–14 (2000).

[18] Jordan, I.K. et al. (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet. 19, 68–72

[19] Saxonov S, Berg P, Brutlag DL.. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci* **103**, 1412–1417 (2006).

3. - El corte y empalme del ARNm inmaduro o *splicing*

El mecanismo de corte y empalme de exones de eucariotas presentes en un transcrito inmaduro, conocido como *splicing,* tiene un papel fundamental en el origen de la diversidad de proteínas originadas por un organismo y en los mecanismos de regulación génica. Un gran número de genes muestran splicing alternativo, siendo muy frecuentes en mamíferos, con cifras que en humano alcanzan el 70 % de los genes [20].

Mediante el splicing alternativo un mismo gen puede originar diferentes productos proteicos y además de forma selectiva en diferentes tejidos, en respuesta a las diferentes condiciones celulares. Son necesarios factores reguladores del splicing, para que pueda regularse la forma en que este se produce. Un ejemplo lo tenemos en la proteína Nova, exclusiva del cerebro, que regula el splicing de al menos 49 ARNm distintos, produciendo proteínas que no se encuentran en otros tejidos [21].

Las mutaciones que alteran los sitios de splicing causan alteraciones en las proteínas resultantes y esto suele estar relacionado con ciertas enfermedades humanas. Por ello el splicing alternativo es un foco de investigación activo para nuevos diagnósticos o tratamientos terapéuticos.

3.1.- Mecanismo general del *splicing*

Cuando el pre-ARNm se transcribe a partir del ADN, incluye varios intrones y exones. Los exones que van a ser retenidos en el ARNm se determinan durante un proceso de corte y empalme denominado splicing. Nos ocupamos aquí solo de el splicing de los genes nucleares que codifican para proteínas y que requiere la maquinaria del espliceosoma y no del splicing de los intrones tipo I o II (propios de ARNr, ARNt o de ARNm de eucariotas inferiores o propio de orgánulos), que son eliminados mediante autocatálisis.

El típico intrón nuclear eucariota tiene secuencias de consenso que definen regiones importantes para su procesado. Cada intrón tiene GU en su extremo 5 ', se conoce como “sitio donador”. Cerca del extremo 3 ' hay un sitio de ramificación (*branch site*). El nucleótido en el punto de ramificación es siempre una A, aunque el consenso alrededor de esta secuencia varía algo. El sitio de ramificación es seguido por una serie de pirimidinas, y a continuación, por AG en el extremo 3 ',

que se conoce como “sitio aceptor” (ver esquema de sitios de splicing en figura 5).

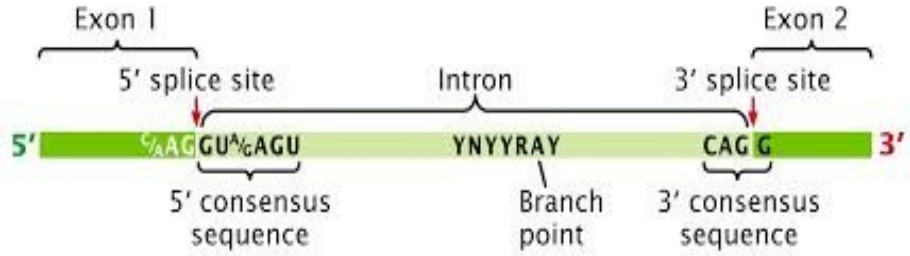


Figura 5. *Sitios necesarios para que se lleve a cabo el splicing del ARNm inmaduro en eucariotas.*

El *splicing* del ARNm inmaduro se lleva a cabo por un complejo de ARN y proteína conocido como espliceosoma, que contiene las snRNPs (*small nuclear ribonucleoproteins*), designadas U1, U2, U4, U5 y U6 (U3 no está involucrado en la unión al ARNm). U1 se une al sitio GU del extremo 5 ' del intrón y U2 se une al sitio de ramificación A con la ayuda de los factores proteicos U2AF. El complejo en esta etapa es conocido como complejo de espliceosoma A. La formación del complejo A es generalmente el paso clave en la determinación de los extremos del intrón que va a ser eliminado, y definiendo los extremos del exón que va a ser retenido.

El complejo U4, U5, U6 se une, y U6 sustituye a U1 (U1 y U4 son liberados). El complejo restante entonces lleva a cabo dos reacciones de transesterificación. En la primera transesterificación, el extremo 5 'del intrón se escinde del exón aguas arriba y se une al sitio de ramificación A por un enlace 2', 5'-fosfodiéster. En la segunda transesterificación, el extremo 3 'del intrón se escinde del exón aguas abajo, y los dos exones se unen mediante un enlace fosfodiéster. El intrón se libera en forma de lazo y es degradado (ver proceso en figura 6).

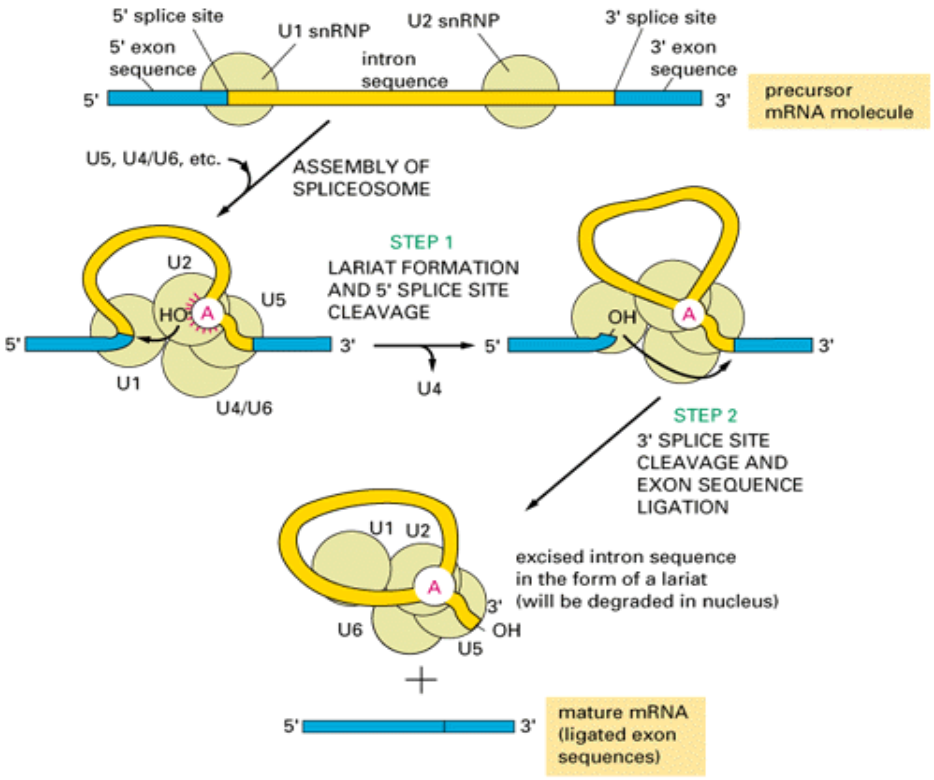


Figura 6. *Mecanismo general de eliminación de intrón en el splicing.*

3.2.- La definición del exón en el splicing de vertebrados

El gen promedio de un vertebrado se compone de varios exones pequeños (de un tamaño medio de 137 nucleótidos) separados por intrones que son considerablemente más grandes. Por lo tanto, la maquinaria de splicing de un vertebrado tiene la difícil tarea de encontrar pequeños exones en medio de intrones mucho más largos. Recordemos que las secuencias de consenso de splicing que guían el reconocimiento de un exón se encuentran en los extremos mismos de los intrones . Además de este gran desafío los sitios de splicing de vertebrados son cortos y mal conservados. De hecho, las secuencias de los sitio de splicing en los mamíferos están menos conservadas que las de sus homólogos en la levadura, a pesar del hecho de que sólo una minoría de genes en Saccharomyces

cerevisiae tienen intrones, y los genes que están divididos por intrones normalmente sólo tienen un único intrón [22]. Por lo tanto, el splicing de vertebrados se enfrenta con un problema más complejo al darse secuencias de splicing más alejadas y menos precisas. Cualquier mecanismo para la orquestación de empalme de genes multiexónicos de vertebrados debe proporcionar una explicación para este rompecabezas.

Los modelos que proponen el reconocimiento de un par de sitios a ambos lados del exón en contraste con un par de sitios, más alejados, a ambos lados del intrón parecen ofrecer una mejor explicación [23]. El modelo de **"definición de exón"** [24] propone que en un pre-ARNm con grandes intrones, la maquinaria de splicing busca un par de sitios de splicing estrechamente espaciados en ambos polos del exón (aceptor, en su extremo 5´ y donador del exón ,en su extremo 3´). Cuando tal par es encontrado, el exón queda definido por la unión de U1, U2, snRNP y los factores asociados de splicing, incluyendo los factores de reconocimiento de splicing 3´ U2AF y SC35 y el factor de reconocimiento de splicing 5´ ASF/SF2. Siguiendo a la definición del exón, los exones vecinos deben ser yuxtapuestos, presumiblemente a través de las interacciones entre los factores que reconocen los exones individuales. Así, desde esta perspectiva, el conjunto del espliceosoma en vertebrados consta de los pasos secuenciales de definición del exón y yuxtaposición de exones (ver figura 7). Posteriormente se eliminarían los intrones según el mecanismo previamente mencionado.

Las perspectivas "orientadas al exón" u "orientadas al intrón" del splicing predicen diferentes fenotipos resultantes de la mutación de sitios de splicing de un exón interno. Un estudio sobre este tipo de mutaciones [25] mostró que el fenotipo más frecuente fue la omisión del exón (*exon skiping*). La omisión de exón es un fenotipo mejor predicho a partir de una perspectiva "orientada al exón", porque la mutación del sitio de splicing en un lado de un exón debe inhibir el emparejamiento de los sitios de splicing a través de los exones y así inhibir el reconocimiento del exón. En caso contrario sería más frecuente la aparición de retención del intrón. Esto nos sugiere que son los exones, y no los extremos de los intrones, las unidades que son reconocidas por el espliceosoma.

A modo de conclusión podemos decir que la definición del exón es el mecanismo predominante en vertebrados, mientras que en otras especies, bien sólo se utiliza el mecanismo de definición del intrón (ver figura 6), como en la levadura *S. pombe* que posee intrones cortos, o bien se alternan ambos tipos al poseer intrones cortos y largos como en *C. elegans* o *D. melanogaster* [26].

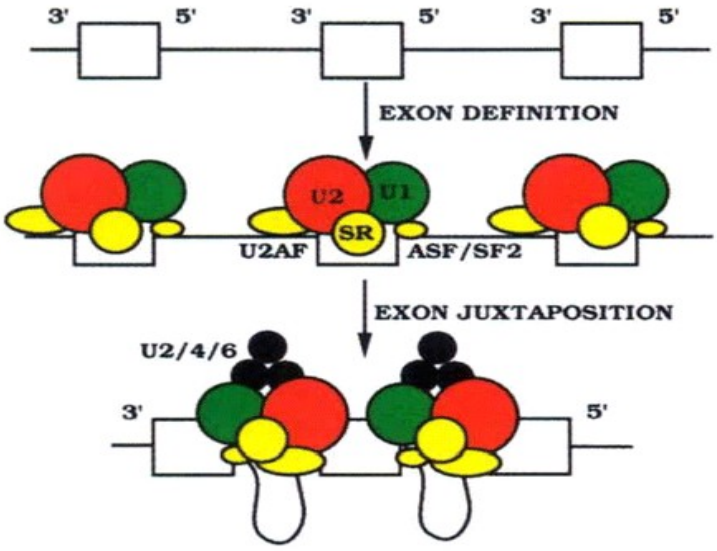


Figura 7. *Mecanismo de splicing basado en el modelo de "definición del exón". Figura extraida de la referencia [23]*

3.3.- La definición del exón en exones terminales

El modelo de "definición del exón" nos sugiere que los exones terminales, primer y último exón del ARNm inmaduro, requieren mecanismos especiales para su reconocimiento. Pensemos que el primer exón termina con un sitio donador de splicing pero carece de señal en su inicio. Existen evidencias [27 y 28], al menos "in vitro", de que la caperuza y las proteínas nucleares que se unen a ella son esenciales para remover el primer intrón del pre-ARNm. Por lo que el primer exón sería reconocido por interacción entre los factores que reconocen la caperuza y los lugares de splicing (sitio donador).

El último exón del pre-ARNm contiene un sitio de splicing aceptor, pero carece de sitio donador , sin embargo termina con una señal de poliadenilación (PAS), por lo que el modelo de definición del exón nos lleva a considerar que existe una interacción entre los factores de unión al sitio de splicing aceptor y los factores implicados en el reconocimiento de PAS. Esta interpretación es respaldada

por el hecho de que es necesaria una correcta poliadenilación para un correcto splicing del último exón. [29].

3.4.- El Splicing alternativo

El splicing alternativo (AS) es un proceso molecular fundamental en la regulación de la expresión génica de los eucariotas. El AS es postulado como el principal mecanismo para aumentar la diversidad de proteínas a partir de un número relativamente limitado de genes (ver figura 8). Según estimaciones recientes entre un 60-75 % de los genes multiexónicos sufren AS [20, 30 y 31].

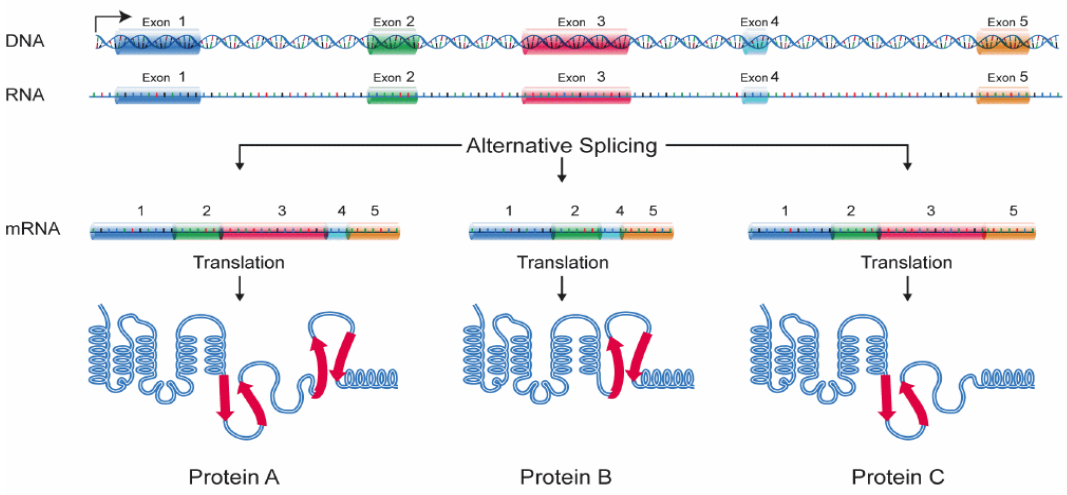


Figura 8. *Diferentes proteínas son producidas a partir de un solo gen mediante splicing alternativo. (Figura extraida de: National Human Genome Research Institute: http://www.genome.gov/Images/EdKit/bio2j_large.gif)*

Tradicionalmente se proponen 5 mecanismos clásicos de AS [32 y 33] (ver figura 9):

a) ***El barajado de exones (exon skipping):*** en este caso uno o varios exones pueden ser incluidos o no en el ARNm maduro. Este es el tipo de splicing más frecuente en mamiferos.

b) ***Exones mutuamente exclusivos:*** uno o dos exones son retenidos en el ARNm maduro o no, pero nunca ambos en un mismo transcrito.
c) ***La retención de intrón:*** una secuencia puede ser eliminada por splicing como un intrón o retenida como parte de un exón. Este es el tipo de splicing más raro en mamíferos.

d) ***Sitios alternativos donadores:*** un exón puede usar sitios donadores alternativos, esto origina exones con diferencias en su extremo 3´.

e) ***Sitios alternativos receptores:*** un exón presenta distintos sitios aceptores, esto origina exones con diferencias en su extremo 5´.

Adicionalmente a estos modos clásicos de splicing existen otros mecanismos que pueden originar diferentes ARNm a partir de un mismos gen. Se trata de los promotores alternativos y los lugares de poliadenilación alternativos. El uso de promotores alternativos es más bien un mecanismo de regulación transcripcional que un mecanismo de splicing, como ya hemos mencionado previamente. Debido al comienzo de la transcripción desde distintos puntos el transcrito generado posee diferentes exones en su extremo 5´. Por otra parte la poliadenilación alternativa da lugar a diferentes extremos 3´del transcrito. Estos dos mecanismos en combinación con los modos de splicing alternativo mencionados permiten que la variedad de ARNm originadas por los genes sea mayor [32].

El estudio a fondo de los transcritos por el proyecto ENCODE nos muestra un escenario mucho más complejo, con toda una variación de los patrones de splicing que implican a múltiples combinaciones de los mecanismos clásicos enumerados [34]. Tal es la complejidad que se ha propuesto un modelo para elaborar una nomenclatura única de todas las posibles formas de splicing [35].

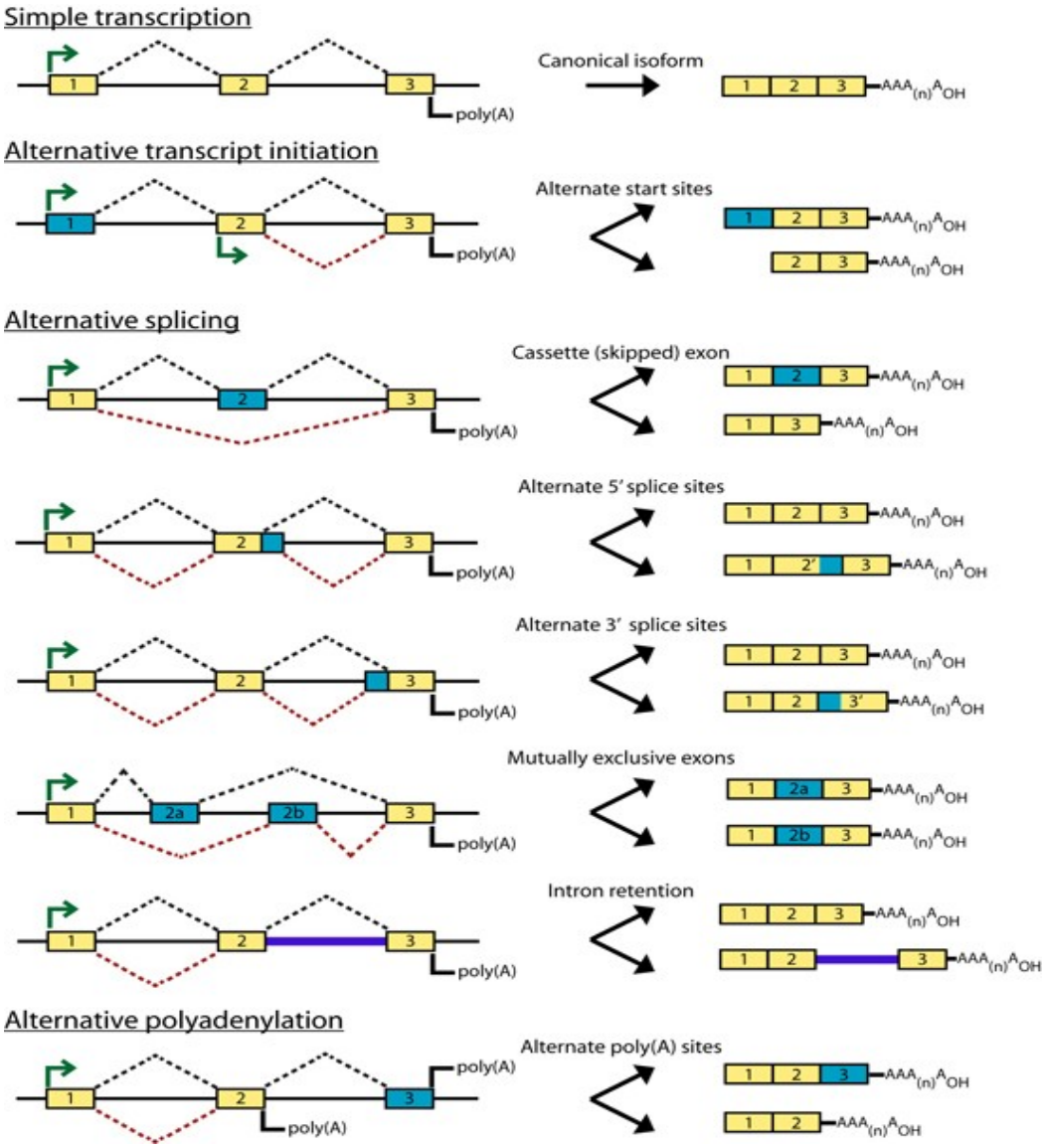


Figura 9. Mecanismos básicos de splicing alternativo.Figura extraída *de web de la plataforma para análisis de micromatrices ALEXA (sección intorducción) (http://www.bcgsc.ca/people/malachig/htdocs/alexa_platform/alexa_arrays/index.htm)*

3.5.- La regulación del splicing

La producción de los ARNm alternativos resultantes del splicing se regula por un conjunto de proteínas que actúan en *trans* que se unen a sitios que actúan en *cis* en el mismo pre-ARNm. Tales

proteínas incluyen activadores de splicing que promueven el uso de un sitio de splicing particular, y represores de empalme que reducen el uso de un sitio en particular.

Hay dos tipos principales de elementos que actúan en *cis* presentes en pre-ARNm y tienen sus correspondientes proteínas de unión a ARN (que actúan en *trans*). Los **silenciadores de splicing** son sitios a los que se unen las proteínas represoras de empalme, lo que reduce la probabilidad de que un sitio cercano sea utilizado como un lugar de splicing (ver figura 10). Estos silenciadores pueden estar situados en el mismo intrón **(***Intronic Splicing Silencers***, ISS)** o en un exón vecino **(***Exonic Splicing Silencers***, ESS)**. Estos elementos varían en secuencia, así mismo existe variación en los tipos de proteínas que se unen a ellos. La mayoría de los represores de splicing son ribonucleoproteínas nucleares heterogéneas (hnRNPs) tales como hnRNPA1 y la proteína de unión al tracto de polipirimidina (PTB).

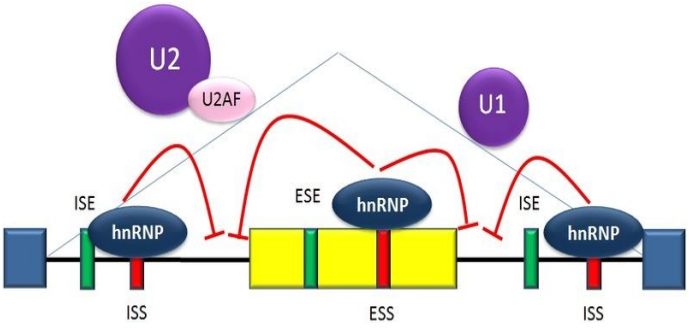


Figura 10. *Sitios silenciadores exónicos (ESS) e intrónicos (ISS) que inhiben a los elementos de la maquinaria de splicing.*

Los **potenciadores de splicing** son sitios a los que se unen proteínas activadoras de splicing, incrementando la probabilidad de que un sitio cercano (donador o aceptor) sea utilizado como lugar de splicing (ver figura 11). Estos sitios pueden encontrarse en el intrón **(***intronic splicing enhancers***, ISE)** o en el exón **(***exonic splicing enhancers***, ESE)**. La mayoría de las proteínas activadoras que se unen a ISEs y ESEs son miembros de la familia de proteínas SR. Estas proteínas contienen motivos de reconocimiento de ARN y dominios ricos en arginina y serina.

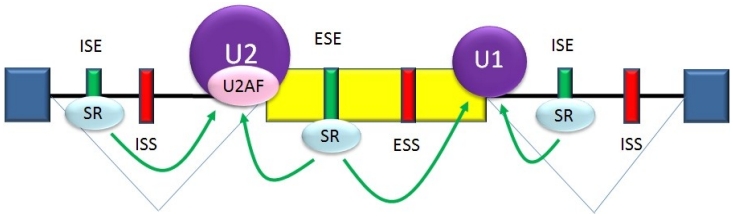


Figura 11. *Sitios potenciadores exónicos (ESE) e intrónicos (ISE) que reclutan a los elementos de la maquinaria de splicing.*

La estructura secundaria del pre-ARNm también puede afectar al mecanismo de splicing mediante su efecto sobre el acceso a sitios de splicing o sobre la afinidad por factores reguladores. Así por ejemplo se ha comprobado que una estructura secundaria en horquilla es la responsable del secuestro el exón 6B del pre-ARNm de la Beta-tropomiosina de gallina, resultando en su exclusión del ARNm maduro [39].

En general los elementos que regulan el splicing trabajan de una manera interdependiente y que además se ve influenciada por del contexto. Así por ejemplo la presencia de un determinado elemento de acción *en cis* en la secuencia de ARN puede aumentar la probabilidad de que un sitio cercano sea sometido a splicing en algunos casos, pero disminuye la probabilidad de que lo sea en otros casos, dependiendo del contexto celular. Existen múltiples evidencias de factores específicos de tipos celulares o de su estado de desarrollo. Por ejemplo el factor PTB (*polypyrimidene tract binding protein*) es expresado en las células progenitoras de células nerviosas, pero su expresión es mucho menor en las neuronas ya diferenciadas, donde sin embargo si se expresa abundantemente nPTB (*neuronal polypyrimidene tract binding protein)* [40].

Las investigaciones esperan dilucidar plenamente los sistemas de regulación implicados en el splicing, de modo que los productos de splicing alternativo de un gen dado, en condiciones particulares, pueda ser predichos por lo que podríamos llamar un "código de splicing" [36]. Se trataría de encontrar el conjunto de reglas que determina el patrón de splicing que puede originar

un determinado transcrito primario. El siguiente paso sería integrar toda esta información para llegar a simular el reconocimiento de exones e intrones. Una primera aproximación a este complejo problema lo tenemos en el algoritmo de simulación ExonScan [37]. Este algortimo utiliza de forma combinada la información de los sitios donadores de splicing, los sitios aceptores, así como los potenciales sitios potenciadores o inhibidores y hace un cálculo combinado determinando los candidatos a exones más favorables (ver figura 12). Pero todavía este tipo de simulaciones tiene que mejorarse incorporando información sobre nuevas señales e interacciones entre factores, provenientes tanto de ensayos experimentales como del estudio de patrones de coevolución de las señales reguladoras [38].

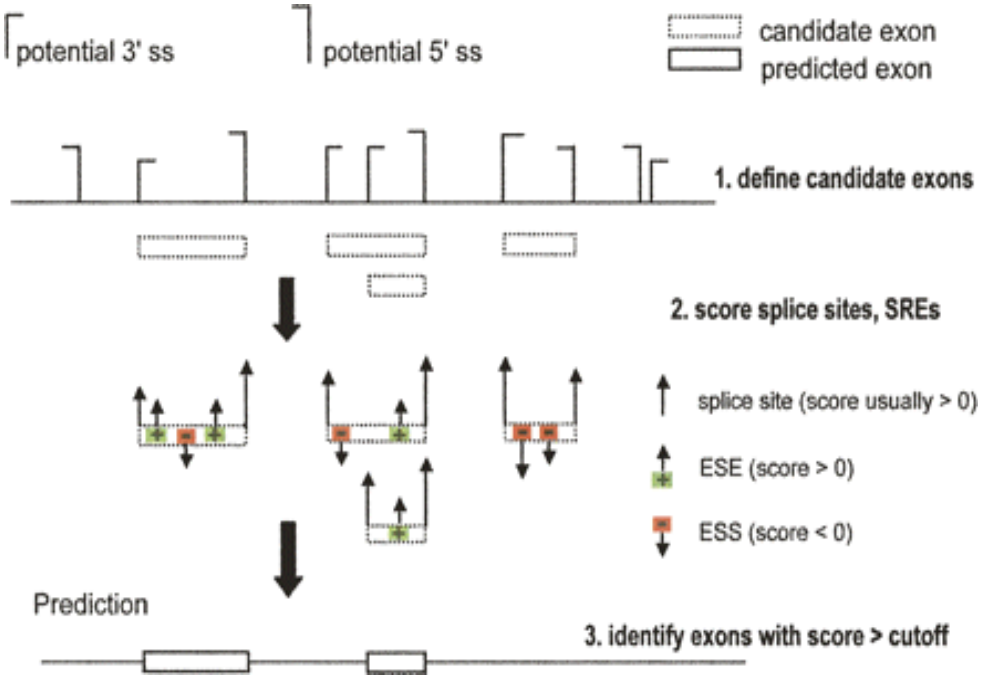


Figura 12. *Esquema que representa los pasos realizados por el algoritmo* ***ExonScan*** *en la predicción de exones en una región genómica dada.Figura extraida de la referencia [37].*

3.6.- Conclusiones

La revisión aquí realizada revela la complejidad de la regulación del splicing alternativo. El splicing alternativo se puede regular en diferentes etapas de organización del espliceosoma por diferentes factores, tanto generales como específicos, y por muchos mecanismos que se basan en elementos

que actúan *en cis*. El splicing alternativo correcto también depende de la estequiometría y las interacciones de las proteínas reguladoras positivas y negativas, incluyendo los CSP. Cada tipo de célula tiene un repertorio único de proteínas hnRNPs y SR, y cambios moderados en su estequiometría relativa puede tener grandes efectos sobre el patrón de splicing alternativo. Es posible que los cambios en la estequiometría de snRNPs perturben la compleja red de factores de splicing y las interacciones entre estos factores y los elementos principales del espliceosoma. Por lo tanto, las redes de regulación de splicing alternativo tienen una arquitectura tan exquisita que la perturbación de un solo paso puede conducir a un splicing alternativo defectuoso.

Las futuras investigaciones sobre regulación de splicing alternativo deben ir encaminadas a la comprensión de como los elementos reguladores controlan los eventos clave del splicing durante el desarrollo y en respuesta a los estímulos ambientales, y cómo la desregulación del splicing alternativo conduce a la enfermedad.

3.7. .- Bibliografía

[20] Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, et al. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302: 2141–2144.

[21] Jensen, K. B., B. K. Dredge, G. Stefani, R. Zhong, R. J. Buckanovich, H. J. Okano, Y. Y. Yang, and R. B. Darnell. 2000. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron 25**:**359-371

[22] Stephanie W. Ruby, John Abelson, Pre-mRNA splicing in yeast, Trends in Genetics, Volume 7, Issue 3, March 1991, Pages 79-85

[23] Berget, S. M. Exon Recognition in Vertebrate Splicing. *Journal of Biological Chemistry* **270,** 2411–2414 (1995).

[24] Robberson, B. L., Cote, G. J. & Berget, S. M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and Cellular Biology* **10,** 84–94 (1990).

[25] Talerico, M. & Berget, S. M. Effect of 5’ splice site mutations on splicing of the preceding intron. *Molecular and Cellular Biology* **10,** 6299–6305 (1990).

[26] Collins L, Penny D (2006) Proceedings of the SMBE Tri-National Young Investigators’ Workshop 2005. Investigating the intron recognition mechanism in eukaryotes. Mol Biol Evol 23: 901–910

[27] Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj IW. A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell. 1994 Aug 26;78(4):657-68

[28] Ohno, M., Sakamoto, H., Shimura, Y. Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. *Proc Natl Acad Sci U S A*. **84**,

5187-5191 (1987).

[29] Niwa M, Berget SM. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. Genes Dev. 1991 Nov;5(11):2086-95

[30] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921

[31] Kim H, Klein R, Majewski J, Ott J (2004) Estimating rates of alternative splicing in mammals and invertebrates. Nat Genet 36: 915–916; author reply 916–917.

[32] Black, Douglas L. (2003). "Mechanisms of alternative pre-messenger RNA splicing". *Annual Reviews of Biochemistry* **72** (1): 291–336.

[33] Pan, Q; Shai O; Lee LJ; Frey BJ; Blencowe BJ (Dec 2008). "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing". *Nature Genetics* **40** (12): 1413–1415.

[34] The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447:799–816.

[35] Michael Sammeth; Sylvain Foissac; Roderic Guigó . Brent, Michael R., ed. "A general definition and nomenclature for alternative splicing events". *PLoS Comput Biol.* **4** (8): e1000147 (2008).

[36] David, C. J.; Manley, J. L. (2008). "The search for alternative splicing regulators: new approaches offer a path to a splicing code". *Genes & Development* **22** (3): 279–85

[37] Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M. and Burge, C. B. (2004). Systematic identification and analysis of exonic splicing silencers. *Cell* 119, 831-845.

[38] Xiao, X., Wang, Z., Jang, M., Burge, C.B.(*2007) Coevolutionary networks of splicing cis-regulatory elements. Proc. Natl. Acad. Sci.104:18583–18588.*

[39] Libri D, Balvay L, Fiszman MY. *In vivo* splicing of the beta tropomyosin pre-mRNA: a role for branch point and donor site competition. Mol Cell Biol. 1992;12:3204–3215.

[40] Boutz PL, et al. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev. 2007;21:1636–1652.

4. - Regulación ejercida por las regiones no traducidas del ARNm (UTRs)

4.1.- Generalidades

El ARNm maduro de los eucariotas presenta una estructura tripartita formada por: una región no traducida en su extremo 5´(UTR 5´), que contiene en su inicio una caperuza de 7-metil-guanosina, seguida de una región codificante formada por tripletes o codones y finalmente una región no codificante en su extremo 3´(UTR 3´), que termina con una cola de poliadenina, tal como se muestra en la figura 1.

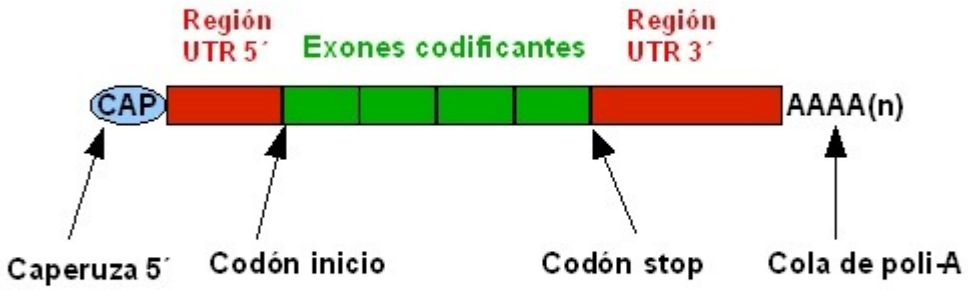


Figura 1. *Estructura típica de un ARNm eucariota maduro.*

Es conocido que las UTRs juegan un papel importante en la regulación postranscripcional de la expresión génica, incluyendo procesos como la eficiencia en la traducción, la regulación de la estabilidad del ARNm o la modulación del transporte del ARNm fuera del núcleo y su posterior localización subcelular [1]. La importancia de las UTRs en la regulación de la expresión génica se pone de manifiesto por la constatación de que las mutaciones que alteran estas regiones dan lugar a graves patologías [3].

La regulación realizada por las UTRs es ejercida mediante diversos mecanismos. En general diversos patrones o motivos de nucleótidos localizados en la UTRs 5´y 3´pueden interaccionar con proteínas de unión a ARN específicas. La actividad reguladora de estos motivos es ejercida normalmente mediante una combinación de elementos de estructura primaria y estructura secundaria (ver figura 2). Así mismo puede darse interacción entre elementos localizados en las

UTRs y diversos ARNs no codificantes (ARNnc), como es el caso de los micro ARN (miARN), que se unen a secuencias diana ubicadas generalmente en las regiones UTRs 3´ (ver figura 2).

La comparación de secuencias, a escala genómica, ha revelado la conservación de algunos aspectos de la estructura de las UTRs. La longitud media de las UTRs 5´es aproximadamente constante entre diversos taxones y oscila entre 100 y 200 nucleótidos, mientras que la longitud media de las UTRs 3 ´es mucho más variable, oscilando entre 200 nucleótidos en plantas y hongos y 1000 nucleótidos en humanos y otros vertebrados [1]. Sin embargo dentro de una especie se observa una gran variación en el tamaño de las UTRs 5´y 3´, pudiendo encontrarse tamaños desde una docena de nucleótidos a varios miles de ellos (ver tabla 1).

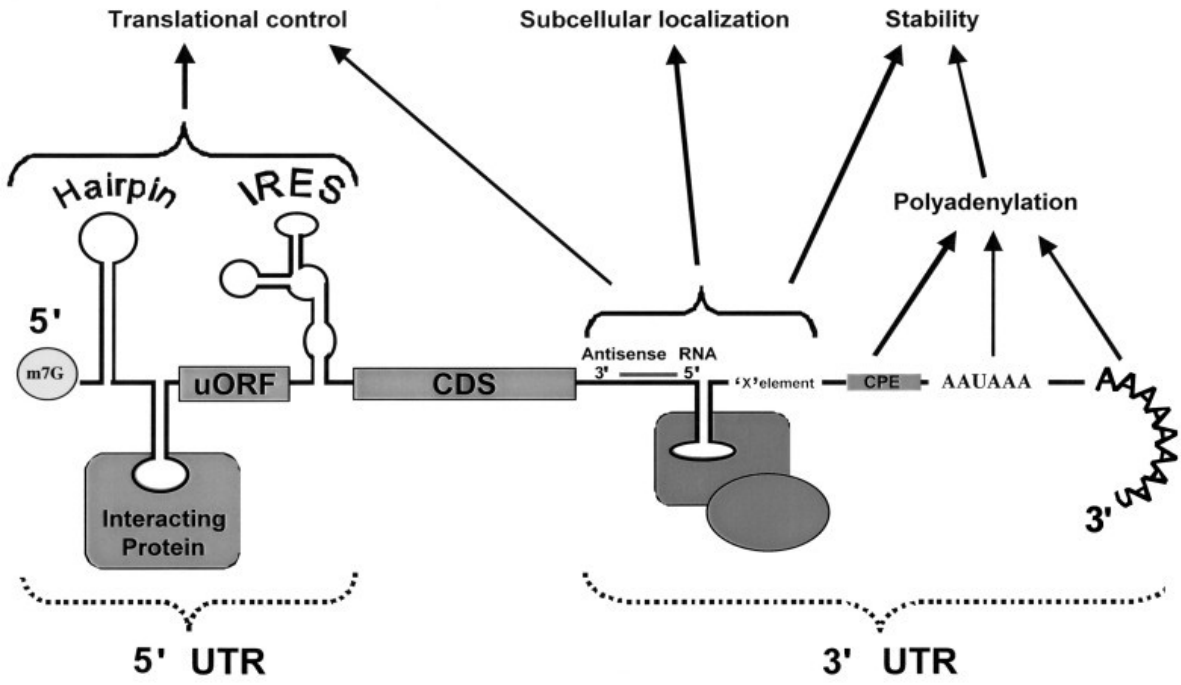


Figura 2. *Estructura general de un ARNm eucariótico que ilustra algunos elementos reguladores postranscripcionales de la expresión génica y su actividad.La regulación mediada por UTR 5´ puede implicar: la caperuza 7-metil-guanosina (m7G); estructuras secundarias similares a horquillas; Interacciones proteína-ARN; marcos de lectura abiertos aguas arriba (uORFs); sitios internos de entrada al ribosoma (IRES). La regulación mediada por UTR 3´ puede involucrar: interacciones de ARN antisentido; Interacciones proteína-ARN, con la participación también de complejos multiproteicos; elementos de poliadenilación citoplasmáticos (CPE); cola poli-A y la variación de su tamaño.*

La composición de bases de las secuencias de las UTRs 5´y 3´muestra diferencias, siendo el contenido en G+C de las UTRs 5´ mayor que el de las UTRs 3´. Esta diferencia está más marcada

en los vertebrados de sangre caliente, cuyo contenido en G+C es de un 60 % en las UTRs 5´y de un 45 % en las UTRs 3´ [2]. Hay por otra parte una interesante correlación entre el G+C de las UTRs y la tercera posición del codón de la correspondiente secuencia codificante, y además hay una correlación inversa significativa entre el G+C de las UTRs y su longitud (ver figura 3) [4]. Esta última seguramente es el reflejo del hecho de que los genes localizados en isocoras ricas en G+C contienen UTRs 5´y 3´más cortas, al igual que ocurre con los intrones y los exones de las secuencias codificantes ubicadas en estas regiones [5].

	UTR 5´			UTR 3´		
	Long. Media	**Long. Máx**	**Long. Min.**	**Long. Media**	**Long. Máx**	**Long. Min.**
Humano	210.2	2803	18	1027.7	8555	31
Roedores	186.3	1786	16	607.3	3354	19
Aves	126.4	620	17	651.9	3990	21
Invertebrados	221.9	4498	14	444.5	9142	15
Vegetal (Liliopsida)	129.8	715	17	273.3	1605	22
Hongos	134.0	1088	16	237.1	1142	25

Tabla 1. *Longitud de las UTRs 5´y 3´de diversos taxones. Simplificada de tabla en referencia [1]*

Hemos de mencionar también que los ARNm eucariotas presentan con frecuencia diferentes tipos de secuencias repetitivas en sus UTRs, incluyendo SINEs (*short interpersed elements*), LINEs (*long interpersed elements*), minisatélites y microsatélites. En los ARNm humanos las secuencias repetitivas se encuentran con más frecuencia en las UTRs 3´ (presentes en alrededor del 36 % de las UTRs 3´, frente al 12 % de las UTRs 5´). Una menor abundancia de repeticiones se encuentran en las UTRs de otros taxones (ver tabla 2), incluidos los mamíferos [1].

Otra diferencia que encontramos en las UTRs es que la presencia de intrones es más frecuente en las UTRs 5´que en las 3´. Alrededor del 30 % de las UTRs 5´de metazoos contienen al menos un intrón, mientras que en las UTRs 3´solo lo contienen entre el 1 % y el 11 % de ellas (ver tabla 3), dependiendo del taxón [1]. Los exones de las UTRs 3´suelen ser los más largos del gen, por el contrario los exones de las UTRs 5´suelen ser bastante más cortos.

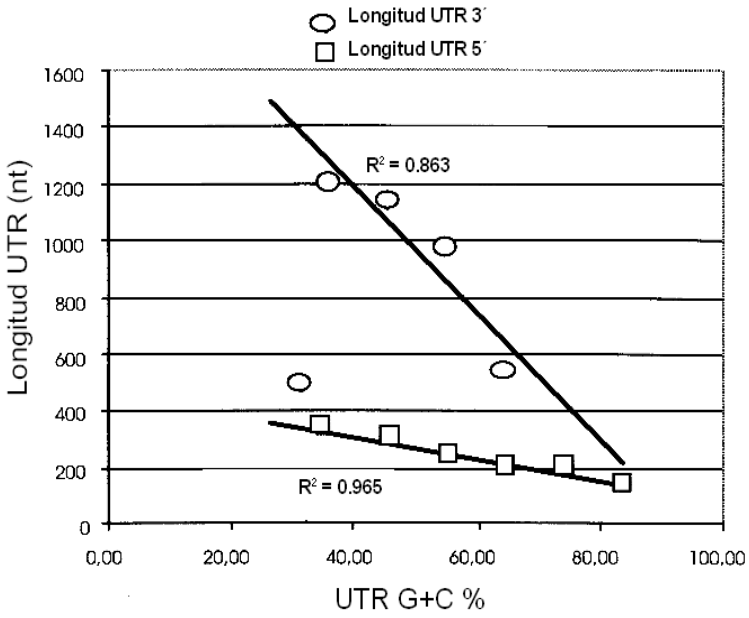


Figura 3. *Relación entre la longitud de las UTRs humanas y su % de G+C. Cada punto representa el valor medio del conjunto de secuencias dentro de un mismo rango de % de G+C. Modificada de figura en referencia [2].*

La formación de transcritos con UTRs alternativos es un fenómeno relativamente frecuente, produciéndose con más frecuencia en las UTRs 5´ que en las 3´. Este hecho se produce por diversos mecanismos: uso de sitios de inicio de transcripción alternativos, la poliadenilación alternativa, uso de donadores/aceptores de splicing alternativos, o a combinaciones de algunos de estos mecanismos [3]

	UTR 5´					**UTR 3´**				
	SR	**LINEs**	**SINEs**	**LTR**	**% del total**	**SR**	**LINEs**	**SINEs**	**LTR**	**% del total**
Humano	140	40	60	21	12.24	1130	213	508	91	36.52
Otros mamíf.	12	11	9	3	9.37	156	36	70	7	19.43
Otros verteb.	9	0	0	0	3.72	79	7	0	0	16.05
Vegetal (Liliopsida)	19	0	0	0	7.34	96	1	0	0	11.74
Hongos	13	0	0	0	2.63	36	0	0	1	6.10

Tabla 2. *Número de UTRs, que contienen los distintos tipos de elementos repetitivos, para distintos grupos taxonómicos.**SR**, short repeats; **LINEs**, long interspersed elements; **SINEs**, short inerspersed elements; **LTRs**, long terminal repeats. Se muestra también el porcentaje de UTRs 5´y 3´respecto al total.(Simplificada de referencia [2]).*

	UTR 5´				UTR 3´			
	1	**2**	**3 o +**	**% del total**	**1**	**2**	**3 o +**	**% del total**
Humano	264	53	21	28.18	69	18	11	7.68
Otros mamíf.	21	7	0	26.76	5	2	0	4.73
Otros verteb.	34	1	2	35.24	1	0	0	0.90
Vegetal (Liliopsida)	203	14	1	12.50	56	5	2	3.94
Hongos	20	2	0	5.67	4	0	0	1.23

Tabla 3. *Número de UTRs 5´y 3´ que contienen 1, 2, 3 o más intrones, de distintos grupos taxonómicos. Se muestra también el porcentaje de UTRs 5´y 3´respecto al total. (Simplificada de referencia [2]).*

4.2. - La región no traducida 5´(UTR 5´)

4.2.1.- La estructura secundaria

La estructura y el contenido en nucleótidos de la UTR 5´parecen jugar un papel fundamental en la regulación de la expresión génica. Estudios a escala genómica han revelado que hay claras diferencias en la estructura y composición de nucleótidos entre genes del mantenimiento celular (*housekeeping genes*) y genes del desarrollo [6]. En general, las UTRs 5´que promueven una traducción eficiente, son cortas, tienen un contenido bajo en G+C, son relativamente desestructuradas y no contienen AUGs aguas arriba (uAUGs) del codón de inicio principal [7]. Por el contrario las UTRs 5´de los genes con poca producción de proteína son en promedio, más largos, con mayor contenido en G+C y poseen un mayor grado de complejidad en las predicciones de sus estructuras secundarias [8].

Entre cierto tipo de genes como factores de crecimiento, protooncogenes y otros, con escasa expresión génica en condiciones normales, se ha encontrado que en el 90 % de los casos sus UTRs 5´contienen estructuras secundarias con una energía libre inferior a -50 Kcal / mol. Dándose además la circunstancia de que en el 60 % de las UTRs 5´ de estos genes las estructuras secundarias se encuentran próximas a la caperuza del extremo 5´ [9]. En un ensayo de células cultivadas, y mediante el uso de un vector con un gen marcador diseñado para manipular la UTR 5´[10], se comprobó que la eficiencia de la traducción disminuía considerablemente cuando la energía libre de

las estructuras en horquilla de la UTR 5´se reducían entre -25 y -35 Kcal / mol, a una distancia de la caperuza comprendida entre 1-46 nucleótidos (ver figura 4) .

En el modelo de escaneo del inicio de la traducción (*scanning model*) se propone que el complejo ribosómico 43s, tras unirse a la caperuza 5´, escanea la UTR 5´hasta encontrar un codón de inicio óptimo, comenzando así la traducción. De forma que la presencia en la UTR 5´de una estructura secundaria estable se interpreta como un indicio de inhibición de la traducción al interferir en el escaneo del ribosoma. Esta barrera puede ser superada si se produce una sobreexpresión de ciertos factores de inicio de traducción, como la helicasa elF4A, en combinación con elF4B. No es de extrañar por ello que la sobreexpresión de factores de inicio de la traducción este asociada a la génesis de tumores, ya que esto supone una forma de saltarse la inhibición de la expresión de los protooncogenes, los cuales presentan UTRs 5´muy estructuradas [11].

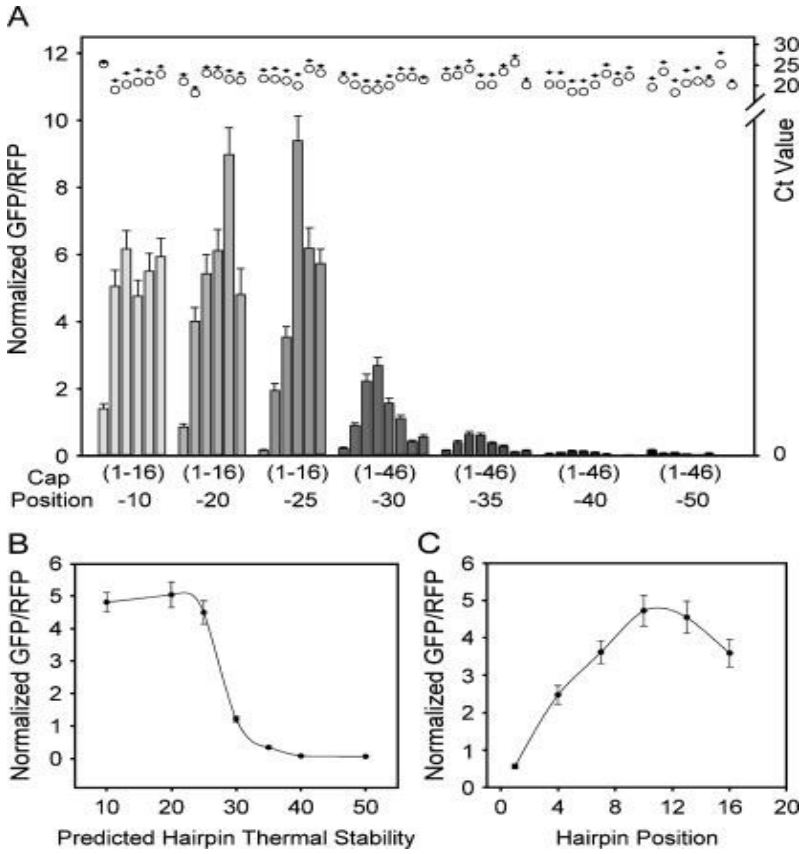


Figura 4. *Influencia de la estabilidad térmica de las estructuras en horquilla de la UTR 5´ y de su distancia a la caperuza sobre la eficiencia de la traducción de un constructo diseñado al efecto.* ***A)****: Influencia en la traducción de cada conjunto de horquillas con diferentes estabilidades térmicas, de -10 a -50 Kcal/mol, para posiciones entre 1-16 o 1-46 respecto a la caperuza.* ***B*** *y* ***C)****: representación gráfica del efecto sobre la traducción de la estabilidad térmica y de la posición de las horquillas respectivamente .Figura extraída de referencia [10].(Para más detalles ver dicha referencia)*

Una estructura bien caracterizada es la estructura cuádruplex de "G" (G4). Dicha estructura se forma a partir de una secuencia rica en guanina que se pliega en una estructura de tetra-hélice no canónica (ver Figura 5), estructura muy estable y que inhibe fuertemente la traducción [12]. La estructura G4 aparece con frecuencia en oncogenes, como el gen TRF2, que está implicado en el control de la función del telómero. La estructura G4 presente en la UTR 5´ de dicho gen puso de manifiesto, en un ensayo realizado con gen marcador, que se produce una inhibición importante de la traducción [13]. El gen TRF2 es sobreexpresado en un buen número de canceres, indicando esto que la actividad de G4 puede ser modulada por una serie de factores [3] (ver figura 5) . Se ha podido comprobar, in vitro, que ciertos ligandos pueden unirse a G4 y modular la expresión de TRF2 [13].

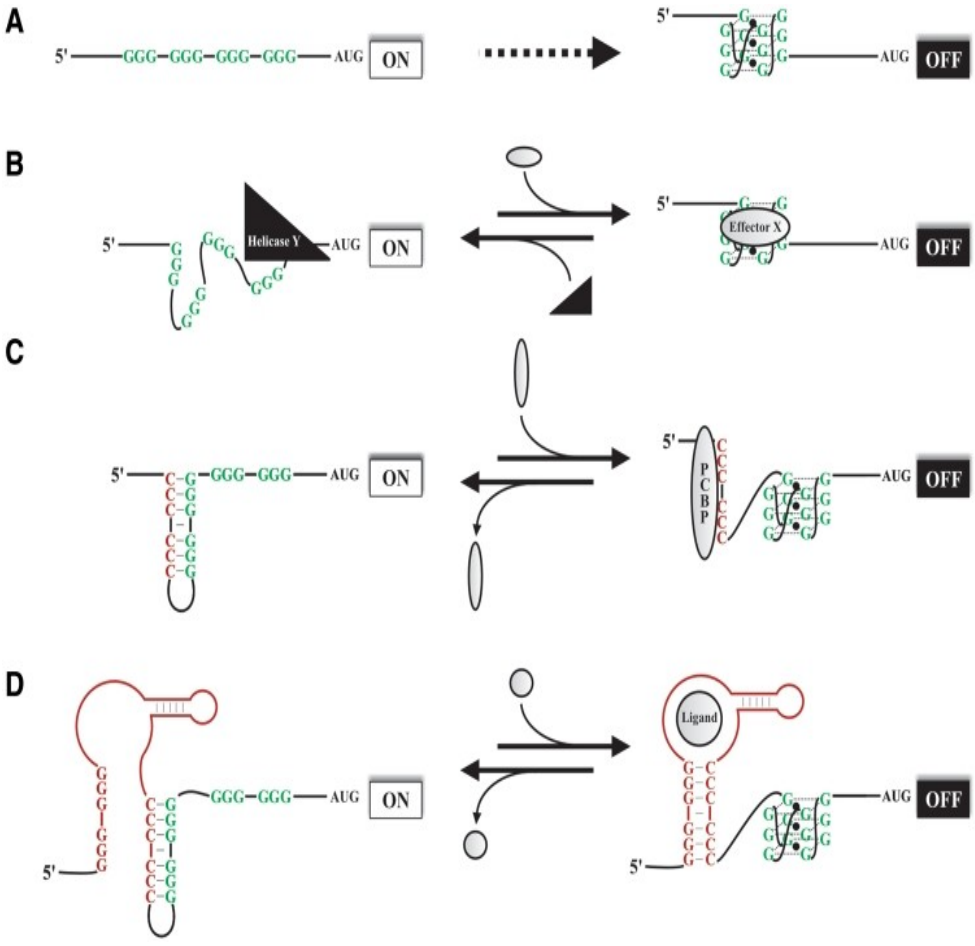


Figura 5. *Formación de estructura cuádruple de G (G4) y modelos propuestos para su regulación .**A)**: formación de la estructura G4 a partir del motivo rico en "G"; **B)**: Formación de G4 gracias a la presencia de concentración suficiente de un efector; **C y D)**: formación de G4 por la presencia de ligandos que desestabilizan una estructura secundaria que impide la formación de G4. (ON traducción activa; OFF traducción inhibida; PCBP proteína de unión a poli-C; los círculos negros en G4 representan cationes necesarios para su formación). Figura extraída de referencia [12].*

El gen TGF-beta1 es otro buen ejemplo de inhibición de la traducción mediada por la estructura secundaria. Un motivo bien conservado en su UTR 5´forma una estructura típica en horquilla muy estable [4]. Sin embargo esta estructura, por si sola, no es suficiente para el bloqueo de la traducción. Es necesario que se produzca la unión de la proteína YB-1, la cual coopera en la formación de dicha horquilla, para que se produzca la inhibición de la traducción [14].

El hecho de que pueda darse una expresión adecuada a pesar de la existencia de estructuras secundarias estables en las UTRs 5´, como en el ejemplo mencionado del efecto de "desliado" ejercido por sobreexpresión de elF4A, así como la necesidad de la presencia de ciertos factores para que la inhibición ejercida por estas estructuras sea efectiva (ejemplo del gen TGF-beta1), nos advierte que, solo basándonos en predicciones *in silico* podemos sacar conclusiones erróneas sobre el nivel de expresión ejercido por una determinada UTR 5´. Si bien la presencia de estas estructuras nos indican indicios de regulación, deben de llevarse a cabo pruebas experimentales adicionales (mediante vectores con genes marcadores, que incluyan a los elementos reguladores) o en su defecto pruebas *in silico* complementarias, como puede ser entre otras, el estudio de conservación evolutiva de estos elementos estructurales. De forma que encontremos un respaldo al hipotético papel regulador de estos elementos.

4.2.2.- Elementos reguladores en la UTR 5´

Mecanismo de inicio de la traducción interna (independiente de la caperuza 5´)

El motivo IRES (*internal ribosome entry site*) es un motivo regulador de ciertos ARNm que facilitan un inicio de la traducción interna, independiente de la caperuza 5´, mediante la unión del ribosoma a un sitio próximo al sitio de inicio de la traducción (ver Figura 6). Este motivo se ha encontrado en diversos ARNm que codifican para proteínas reguladoras, como proto-oncogenes, o factores de crecimiento y sus receptores [1]. El motivo IRES permite el secuestro del ribosoma hacia ARMm con o sin caperuza, en condiciones en las que la traducción dependiente de caperuza es inhibida por estrés, por la fase del ciclo celular o por apoptosis, asegurando que la expresión de genes esenciales se mantiene en la célula [15]. El análisis comparativo de ciertos IRES bien conocidos ha permitido la identificación de un motivo estructural común compartido por muchos de ellos. Se trata de una estructura de horquilla en forma de "Y" justo antes del codón de inicio principal [16]. A pesar de esto se ha descubierto que cortas secuencias complementarias del ARN ribosómico pequeño pueden actuar igualmente como IRES [17].

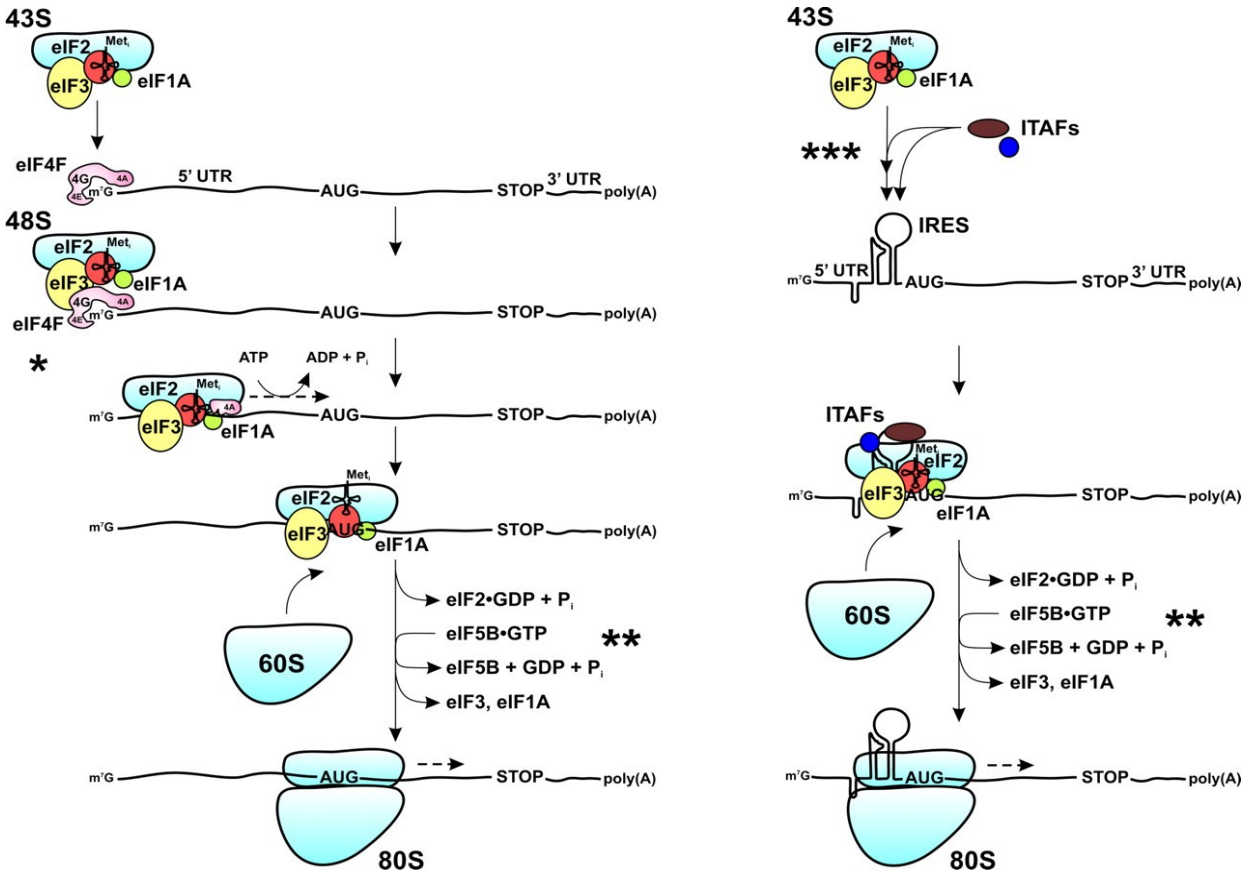


Figura 6. ***Vía de inicio de traducción dependiente de caperuza (izquierda) frente a la vía de iniciación interna (derecha)***. ***Vía dependiente de caperuza***: *En una primera etapa, un complejo de iniciación 43S (que comprende la subunidad 40S, eIF2, GTP Met-tRNAi, y eIF3) se une a la caperuza 5' del ARNm, lo que conduce a la formación del complejo 48 S. En un segundo paso, la subunidad 40S con factores de iniciación asociados y Met-tRNAi explora aguas abajo a lo largo de la UTR-5´ del ARNm en busca del codón de iniciación. En un tercer paso este complejo se desplaza y reconoce el codón de iniciación AUG. Este reconocimiento es seguido por la liberación de los factores de iniciación, y, posteriormente, la subunidad 40S se une a una subunidad 60S para formar un ribosoma 80 S (la unión de esta subunidad es catalizada por el factor de iniciación eIF5B).* ***La vía de iniciación interna*** *postula un solo paso (o en algunos casos dos pasos) del mecanismo por el cual el ribosoma 40S se aproxima al codón de iniciación. Este mecanismo de iniciación de la traducción generalmente es independiente del reconocimiento del extremo 5´del ARNm, e implica el reclutamiento directo de los ribosomas 40 S a la vecindad del codón de iniciación (dirigida por un elemento IRES y la acción del factor ITAFs). Figura extraída de referencia [15].*

El mecanismo del inicio de la traducción interna es aún poco conocido, sin embargo está claro que la eficiencia del motivo IRES está muy influenciada por factores proteicos que actúan en *trans*. Mediante estos factores se consigue una traducción mediada por IRES específica en determinados tipos celulares [8]. La estructura de la UTR 5´ha demostrado tener influencia en la actividad del motivo IRES, que puede ocurrir mediada por la intervención de ciertos factores actuando en *trans* o bien actuando directamente sobre el ribosoma. Un ejemplo de genes donde ocurre esto es en la familia de proto-oncogenes Myc, que están implicados en la proliferación celular [18].

El motivo IRE

El motivo IRE (*iron-responsive element*) se localiza en la UTR 5´de los ARNm que codifican para proteínas implicadas en el metabolismo del hierro (ferritina, 5-aminolevulinato sintetasa y aconitasa). La regulación es ejercida por proteínas de unión a ARN (IRP 1 e IRP 2). Dichas proteínas reconocen una secuencia en un bucle conservado de unos 30 nucleótidos, presentes en el IRE, siendo la señal más importante la secuencia de seis nucleótidos CAGYCX (**Y**= U o A y **X**= U, C o A). Cuando los niveles celulares de hierro son bajos las proteínas IRP 1 y 2 se unen a IRE y bloquean la traducción de la ORF aguas abajo. Cuando los niveles de hierro son altos, un complejo que contiene hierro se une a las IRPs e inhibe su unión al IRE (ver Figura 7), permitiendo así la traducción de la correspondiente proteína[4].

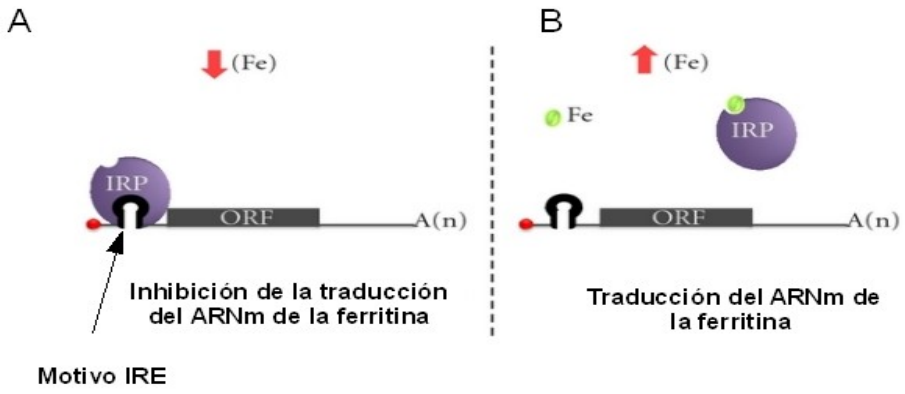


Figura 7. Acción del motivo IRE junto a las proteínas IRP (iron regulatory proteins) según la concentración de hierro sea baja **A)** o alta **B).** Figura extraída de referencia [11].

El motivo TOP

El motivo TOP (terminal oligopyrimidine tract) consiste en una serie de 5 a 15 pirimidinas, entre una citosina y una guanina (C (PY)n G) adyacentes a la caperuza 5´. Se ha relacionado principalmente con los ARNm de proteínas ribosómicas y de factores de elongación de la traducción de vertebrados. Este motivo es utilizado para una represión coordinada de la traducción en ciertas condiciones celulares: durante la detención del crecimiento, la diferenciación, el desarrollo o ciertos tratamientos con fármacos [19]. Más recientemente se han encontrado evidencias de que este motivo no es exclusivo de los tipos de ARNm mencionados más arriba y puede ejercer su función reguladora en un conjunto más amplio de genes, incluyendo a genes relacionados con el lisosoma y con el metabolismo [20].

Codones de inicio y marcos de lectura abiertos "aguas arriba" (uAUGs y uORFs)

Los codones de inicio y marcos de lectura aguas arriba de la secuencia codificante, uAUG y uORF respectivamente, son unos de los principales elementos reguladores de las regiones UTRs 5´. Como su nombre sugiere los uORF son secuencias definidas por un codón de inicio y otro de final, aguas arriba de la secuencia codificante principal. Un gran porcentaje del transcriptoma humano contiene uAUGs o uORFs, con valores que oscilan entre el 44 y 49 %, y valores similares se han encontrado en el transcriptoma de ratón [21 y 22]. Aunque estos números pueden parecer elevados, tanto los uAUGs como los uORFs de las regiones UTR 5´son menos frecuentes de lo esperado, por puro azar, sugiriéndonos que se encuentran sometidos a una fuerte presión selectiva [4].

Un estudio piloto realizado sobre un conjunto de genes de humano y roedores mostró resultados que indican que solo una fracción del conjunto de uAUGs y uORFs se encuentran conservados (38 % de lo uORFs y 24 % de los uAUGs) [23]. Estos escasos valores de conservación y el hecho de que el tamaño medio de los uORFs sea el esperado por azar, ha llevado a proponer que realmente sean escasos los uORFs que se han mantenido con una función reguladora (especialmente los que muestran conservación) [23]. En un estudio diferente, sobre la conservación de uAUGs en las regiones UTRs 5´de genes ortólogos de humanos y roedores, se pone de manifiesto que estos tripletes AUG se conservan en mayor

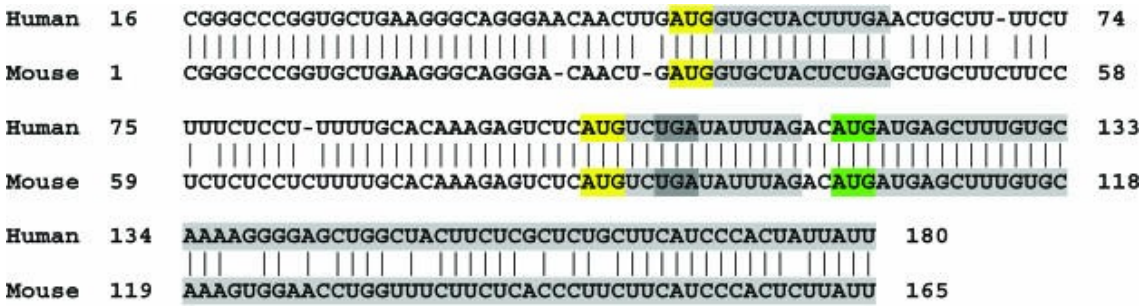


Figura 8. *Alineamiento de las UTRs 5´de las secuencias ortólogas del "colágeno alfa1 tipo III" de humano y ratón. Se observa la conservación de uAUGs (amarillo); en gris se indican los uORF ; en verde el codón de inicio principal. Figura extraída de referencia [24].*

medida que el resto de los 63 posibles tripletes [24] , mientras que en la región codificante o en la región UTR 3´no existe esta diferencia. Este hecho nos indica que dichos codones han sido mantenidos por la selección natural, probablemente con el papel de atenuar la traducción (ver Figura 8). Se ha comprobado además, en este mismo estudio, que los uORFs que comienzan por uAUGs que muestran conservación, son de forma significativa, de menor tamaño que los que lo hacen a partir de uAUGs no conservados (ver Figura 9). Este hecho nos sugiere una posible optimización del nivel de atenuación de la traducción ejercida por estos elementos, evitándose así una reducción completa de la misma [24].

Por otra parte se ha encontrado, que en eucariotas, el número tripletes AUG de las regiones UTRs 5 ´se correlaciona inversamente con la fortaleza del contexto (secuencia de Kozak) del codón de inicio principal [25]. Se sugiere que este resultado anti-intuitivo debe ser interpretado en términos de un papel de los uAUGs en la regulación de la traducción, asegurando una baja expresión basal por la vía de un doble control negativo, creando así oportunidades para potenciales mecanismos de regulación adicionales. Un ejemplo de estos mecanismos, con respaldado experimental, incluiría la posibilidad de eliminar o no la porción de la UTR 5´que contenga a un determinado uAUG, mediante splicing alternativo [25].

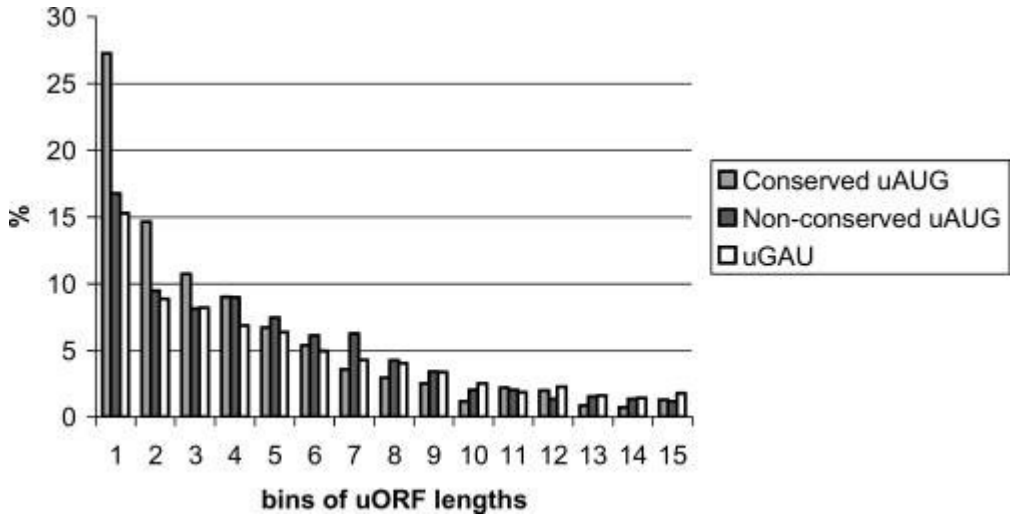


Figura 9. Distribución de la longitud de los uORFs que comienzan con uAUGs conservados, uAUGs no conservados y de pseudo uORFs que comienzan con uGAUs. La longitud de los uORFs se representa mediante bins, donde cada bin incluye 10 codones. *Figura extraída de referencia [24].*

En términos generales la presencia de uORFs en las UTRs 5´de los ARNm de mamíferos puede correlacionarse con una reducción en el nivel de expresión de proteína [21, 26 y 27], que puede ser de alrededor de un 40 % . Los estudios mutacionales ponen de manifiesto el importante papel de los uAUGs y uORFs en la regulación de la expresión génica [11] (ver Figura 10). El gen PAPOLA (poliA-polimerasa-alfa) contiene dos uORFs conservados en su UTR 5´. La mutación del uAUG más próximo al extremo 5´causa un aumento en la eficiencia de la traducción, indicando que el primer uORF ejerce un efecto inhibitorio sobre dicha expresión [28]. Diversas mutaciones que crean o eliminan uORFs han podido relacionarse con diversas enfermedades humanas [4]. Por ejemplo la predisposición al melanoma puede ser causada por una mutación que origina un uORF en la UTR 5´del gen CDKN2A (*cyclin-dependent-kinase-inhibitor-protein*) [29]. Otro caso es el de la trombocitemia hereditaria, causada por una mutación que origina una variante de splicing que elimina un uORF, lo que causa una desregulación de la traducción, provocando un aumento de los niveles de trombopoyetina [30].

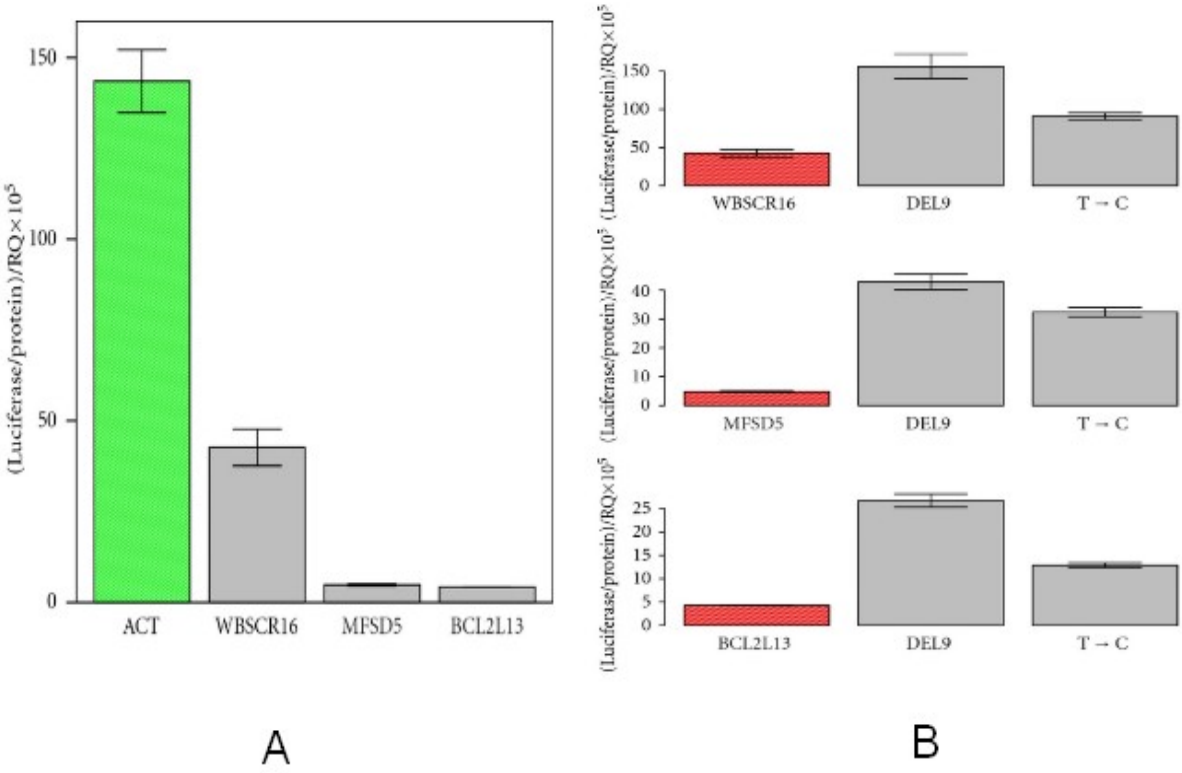


Figura 10. *Impacto de los uAUGs en la regulación de la traducción.* ***A)*** *Comparación de los niveles de luciferasa a partir de constructos que contienen la UTR 5´ de un gen de referencia (ACT, en verde) y otros que contienen la UTR 5´de genes con uAUG en dicha región (WBSCR16, MFSD5 y BCL2L13, en gris).* ***B)*** *Delección o mutación del uAUG de los genes WBSCR16, MFSD5 y BCL2L13, que revierte la inhibición de la traducción como lo demuestra el aumento de luciferasa. Figura extraída de referencia [11].*

Diversas propiedades de los uORFs han demostrado tener efecto positivo sobre su grado de inhibición de la traducción. Entre ellas destacan por su significación: la mayor longitud de la distancia caperuza-uORF, la fortaleza del contexto del uAUG, el número de uORFs presentes en la UTR y el grado de conservación que estos presentan [21].

Es comúnmente aceptado que los uORFs provocan un descenso en la traducción mediante el secuestro del ribosoma y su imposibilidad de emprender una reiniciación, tras la fase de terminación del uORF [3]. Sin embargo las evidencias muestran que la traducción de la secuencia codificante puede producirse, en cierta medida, a pesar de la presencia de uORF en la UTR 5´ y sin que sea necesaria la presencia de IRES [3]. Dos mecanismos alternativos al mecanismo clásico de

desplazamiento del ribosoma, dependiente de la caperuza, han sido propuestos para explicar este hecho (ver Figura 11). Por un lado estaría el "desplazamiento a saltos o defectuoso" (*leaky scanning*), en este caso no todos los ribosomas reconocen el 100 % de los uAUGs, de forma que algunos los saltan e inician en el AUG principal [31]. Por otra parte estaría el "mecanismo de reiniciación", según el cual la subunidad 40 S del ribosoma permanecería unida al ARNm tras la terminación del uORF y continuaría su recorrido hasta encontrar el AUG principal [31]. Este mecanismo es altamente ineficiente, siendo sólo posible cuando los uORFs son cortos, de forma que el escaso tiempo entre inicio y terminación permite la reorganización de la subunidad del ribosoma y los factores de inicio. Esto explicaría también por que la distancia entre dos ORFs aumenta la eficiencia de la reiniciación, al disponer el ribosoma de más tiempo para su reorganización [31].

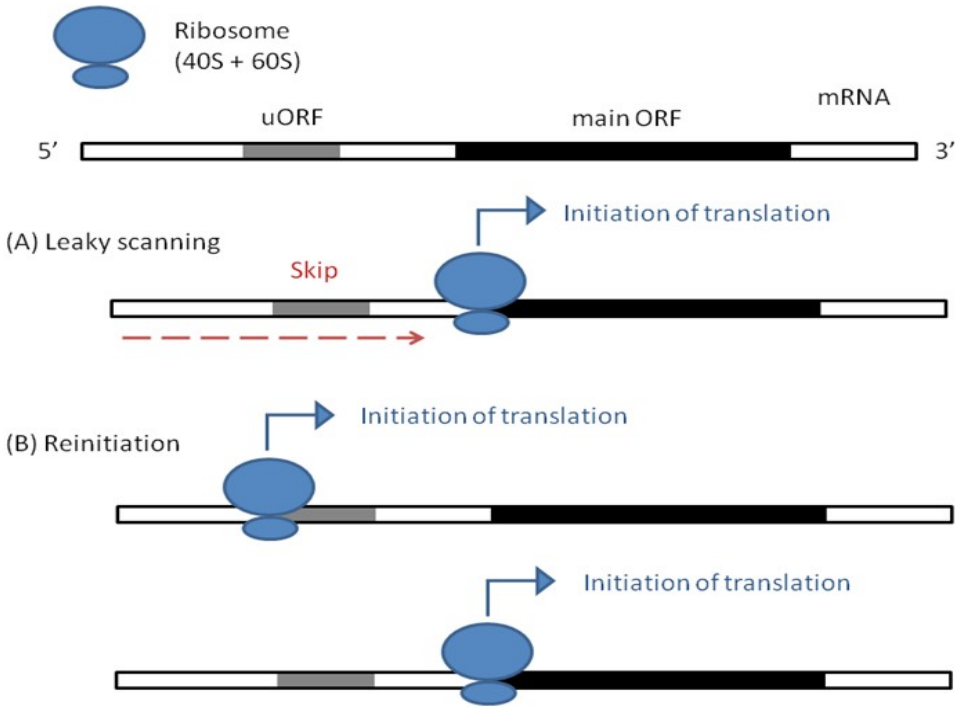


Figura 11 . *Mecanismos de inicio de traducción alternativos al modelo clásico, dependiente de la caperuza. A) Desplazamiento debil o defectuoso (leaky scanning); B) Reiniciación. www.intechopen.com/books/computational-biology-and-applied-bioinformatics/emergence-of-the-diversified-short-orfeome-by-mass-spectrometry-based-proteomics.*

Adicionalmente a la inhibición de la traducción ejercida por los uAUGs/uORFs hemos de considerar el efecto que pueden tener los uORFs sobre la estabilidad de los transcritos. El modelo de "codón de terminación prematura" (PTC) del mecanismo NMD (*Nonsense Mediated Decay*) de los ARNm establece que si a una distancia de más de 55 nt, aguas abajo del PTC, hay un EJC (E*xon Juntion Complex*) depositado por el espliceosoma, se activa la via de NMD (ver Figura 12).

En el caso de que en la UTR 5´ los transcritos contengan algún uORF, el cual presenta efectivamente un codon de terminación prematuro, igualmente pueden promoverse el mecanismo de NMD [32]. Así mismo hay otras situaciones, como intrones en la UTR 3´que pueden promover la degradación del ARNm (ver figura 13). Hay evidencias de uORFs en la UTR 5´ que presentan este efecto de desestabilizar el ARNm, promoviendo el mecanismo NMD [33 y 34], mientras que otros uORFs también demuestran esta propiedad pero mediante una ruta independiente de NMD [34].

Por otra parte, el papel regulador que puedan desempeñar los péptidos resultantes de la traducción de los uORFs es poco conocido, debido a la dificultad de detección de los mismos. Las primeras evidencias de la presencia de estos péptidos se han tenido en células de leucemia mielógena crónica [36], en ellas se han detectado 54 proteínas de menos de 100 aminoácidos que mapean sobre uORFs. Esta evidencias ponen de manifiesto que algunos uORFs son traducidos y los péptidos son mantenidos por un tiempo en la célula, apuntando a que cumplen alguna función, aunque no existe una idea clara de cual es el papel que desempeñan. En un estudio donde se analizaron 200 uORFs bien conservados entre humano y ratón se encontró que la fortaleza del codón de inicio era mayor en estos que en los uORFs que no muestran conservación. Además se han encontrado evidencias de selección purificadora de la secuencia de aminoácidos codificada por los uORFs conservados [32]. Estos resultados suponen un respaldo a la hipótesis de que dichos péptidos cumplan alguna función reguladora.

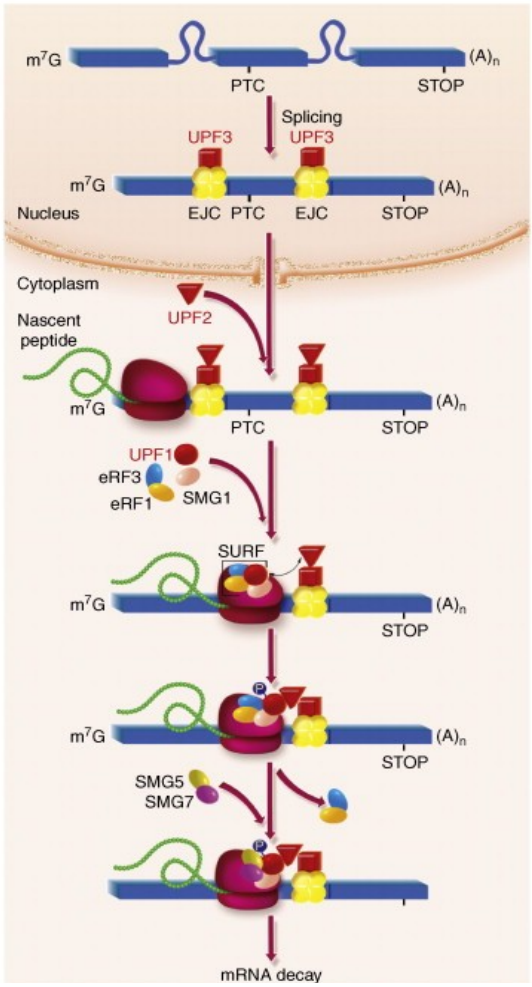
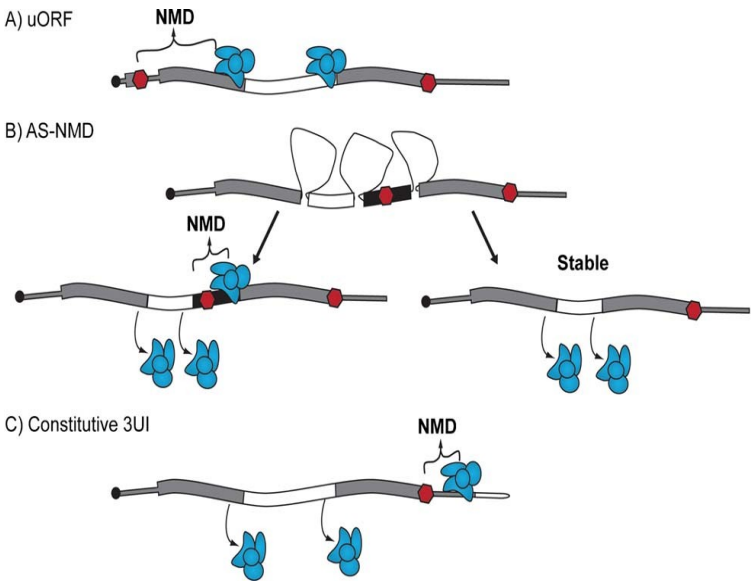


Figura 12. Mecanismo de NMD en mamíferos. Después del splicing del pre-mRNA en el núcleo, los transcritos con EJC que incluyen a UPF3, la proteína principal de NMD, se exportan al citoplasma. En el citoplasma, UPF2 se une a UPF3. Los ribosomas comienzan a traducir el ARNm hasta alcanzar un PTC (*premature termination codon*). En estos ribosomas estancados, cuatro proteínas (SMG1, UPF1, eRF1 y eRF3) generan un complejo conocido como SURF. En los casos en los que hay un EJC aguas abajo, UPF1 (en el sitio de terminación) pueden interactuar con la proteína asociada a EJC, UPF2, lo que conduce a una cascada de reacciones que incluye a factores SMG5 y SMG7, provocando NMD. Código de color: los exones, rectángulos azules; intrones, lazos azules. Figura extraída de referencia [82]

Figura 13. *Diferentes situaciones con PTC que pueden promover NMD. A) uORF en la UTR 5´; B) PTC en un exón alternativo, según sea incluido o no promoverá NMD; C) Presencia de EJC en la UTR 3´, debido a la presencia de un intrón en la misma. En rojo PTC o codón de terminación auténtico, en azul EJC, segmentos gruesos exones codificantes, segmentos finos exones UTR 5´o 3 ´. Figura extraída de referencia [83]*



4.2.3.- UTRs 5´alternativas

Un análisis a gran escala del transcriptoma de mamíferos ha revelado que el uso de UTRs 5´alternativas es un fenómeno frecuente y que prácticamente la mayoría de genes pueden manifestar esta expresión diferencial [37]. Dichas UTRs alternativas puede ser generadas por el uso de promotores alternativos, pero también por el uso de sitios de inicio alternativos de un mismo promotor o por fenómenos de splicing alternativo [3 y 37] (ver Figura 14). La diversidad en la UTR 5´de los genes permite variaciones en su expresión, en función de los elementos reguladores que contengan las UTRs 5´alternativas.

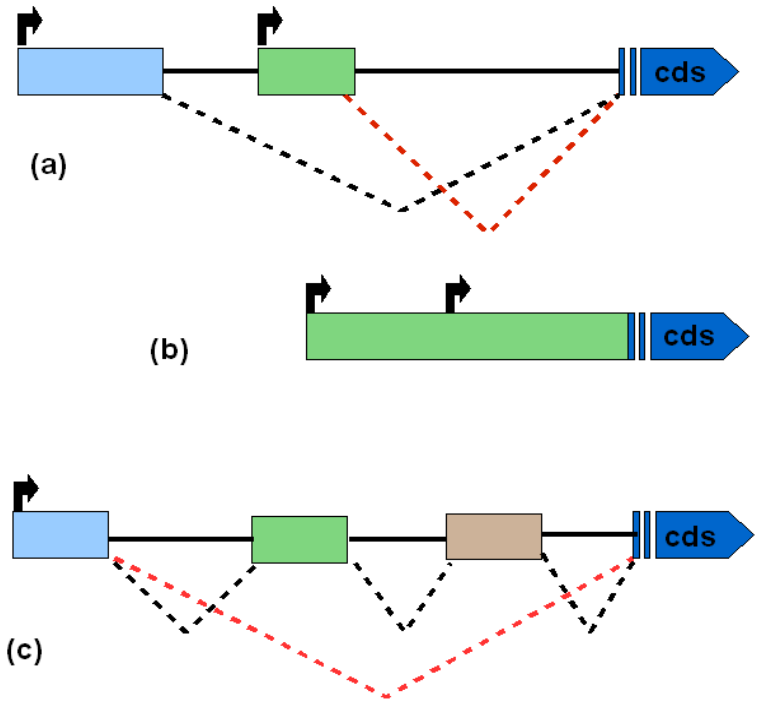


Figura 14. *Mecanismos responsables del origen de UTRs 5´alternativas.* ***a)*** *Uso de promotores alternativos;* ***b)*** *Un solo promotor con sitios de inicio alternativos;* ***c)*** *Un solo promotor y splicing alternativo de los exones de la UTR 5´.(Verde, azul claro y marrón: exones de UTR 5´; azul oscuro: secuencia codificante (cds); líneas discontinuas negras y rojas: eventos de combinación alternativos)*

Las UTRs pueden determinar una expresión específica de un tejido o en unas condiciones fisiológicas, cuando ciertos motivos reguladores están incluidos en algunas variantes de ARNm y no en otras, y dichas variantes muestran un patrón específico en los diferentes tejidos o se expresan de

forma diferente en diferentes condiciones celulares [37]. El gen AXIN2 (*axis inhibition protein-2*) tiene tres promotores, que permiten una expresión específica de tejido de tres UTRs 5´alternativas. Cada UTR contiene diferente estructura secundaria y uORFs confiriendo una estabilidad y eficiencia de la traducción diferente para cada transcrito. La cantidad de axina-2 presente en cada tejido depende del nivel de expresión global de ARNm de axina-2 y de la proporción de los diferentes transcritos con diferentes UTRs 5´ [38, 39]. Diferentes UTRs regulan también la expresión de FGF1 (*fibroblast grow factor 1*) durante el estrés celular producido por hipoxia o apoptosis. Bajo estas condiciones la traducción dependiente de la caperuza es inhibida y esta se lleva a cabo mediante IRES. FGF1 contiene cuatro promotores, con expresión específica de cuatro UTRs 5´ en distintos tejidos, pero sólo dos contienen IRES (transcritos A y C), cuya actividad además es muy diferente (ver Figura 15). Por lo tanto la proporción de los diferentes transcritos alternativos, junto con las condiciones celulares, va a determinar en cada tipo de tejido el nivel de expresión de FGF1 [40].

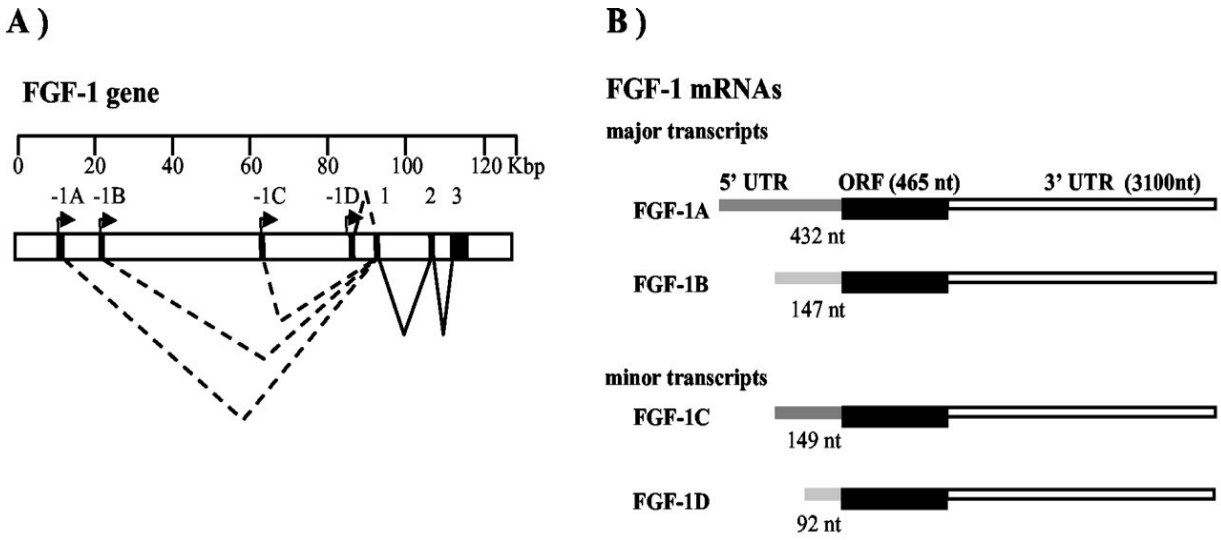


Figura 15. *Gen FGF-1 humano y sus transcritos.* ***A)*** *Organización del gen FGF-1: las flechas indican los diferentes promotores; 1A a 1D representan exones de la UTR 5´.* ***B)*** *Diferentes transcritos que se originan según los promotores utilizados. Figura extraída de referencia [40].*

El gen PPARγ presenta hasta siete transcritos alternativos, con diferentes UTRs 5´, resultantes de splicing alternativo de cinco exones en su UTR 5´. En los ensayos realizados se encontró que la eficiencia de la traducción está inversamente correlacionada con: la estabilidad de la estructura secundaria del ARNm, la presencia de bases apareadas en la secuencia consenso de Kozak, el

número de uAUGs y la longitud de las diferentes UTRs 5´ [41].

4.2.4.- Conclusión sobre UTR 5´

Las UTRs 5´pueden ejercer una regulación de la expresión génica mediante diversos elementos de estructura primaria y secundaria, que a veces actúan de forma combinada. Así mismo, en algunos casos, determinados factores proteicos deben unirse a ciertos elementos estructurales de las UTRs 5 ´para que su regulación sea efectiva (IRES, IRE).

La complejidad de la estructura de la UTR 5´, su estabilidad, así como la existencia de uAUGs y uORFs, se revelan como elementos claves en la regulación negativa de la expresión génica que ejercen estas regiones. La existencia de mutaciones que afectan a los IRES, uAUGs o uORFs y que causan diversas patologías ponen en evidencia la importancia de estos elementos.

La existencia de UTRs 5´alternativas, en las que están presentes o ausentes diferentes elementos o motivos reguladores, ofrece oportunidades para que pueda llevarse a cabo una regulación compleja de la expresión génica, utilizándose las diferentes variantes según las necesidades de diferentes tejidos o diferentes condiciones fisiológicas de la célula.

4.3.- La región no traducida 3´(UTR 3´)

Las regiones UTR 3´, situadas aguas abajo de la secuencia codificante, están implicadas en numerosos procesos de regulación, entre los principales se encuentran: determinación del fin de la transcripción, el grado de poliadenilación y estabilidad del ARNm, la localización del ARNm e incluso efectos sobre la traducción del mismo [3].

En comparación con las regiones UTRs 5´, que contienen secuencias responsables del inicio de la traducción, las restricciones de las secuencias de las UTRs 3´son menores, resultando esto en un gran potencial para la evolución de elementos reguladores [3]. A pesar de ello, las UTRs 3 ´presentan ciertas regiones de elevada conservación, dándose el caso de que en ellas encontramos algunos de los elementos más conservados del genoma de mamíferos [42] (ver Figura 16). Un análisis *in sílico*, a escala genómica, ha revelado que, al contrario que los motivos de la región

promotora, los motivos en la región UTR 3´ se conservan principalmente en una de las dos cadenas del ADN, lo que es consistente con un papel regulador de la UTR 3´ en la fase post-transcripcional [43]. La región UTR 3´sirve como sitio de unión de diversas proteínas reguladoras, así como microARNs (miARNs) [3], aspectos que trataremos en los siguientes subapartados.

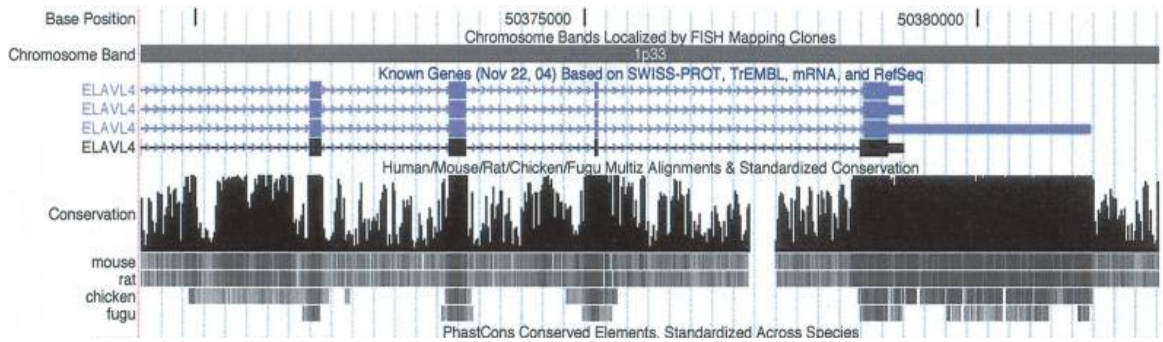


Figura 16. *Conservación extrema de la UTR 3´ (derecha de la imagen)del gen ELAVL4 (HuD). Esta región es el quinto elemento mejor conservado del genoma humano. Figura extraída de referencia [42].*

4.3.1.- MicroARNs y UTR 3´

Los microARNs (miARNs) son cadenas sencillas de ARN que son producidos a partir de genes endógenos por la ARN polimerasa II, en forma de precursores de una longitud entre 70-100 nucleótidos, que adquieren estructura en horquilla. Dicha estructura es escindida en el núcleo por la ARNasa Drosha para generar un pre-miARN, el cual es posteriormente transportado al citoplasma mediante complejos de poros nucleares. Ya en el citoplasma este pre-miARN madura mediante la ARNasa Dicer, que genera una doble cadena de aproximadamente 22 nt y ya está listo para actuar. Tras deshacerse la doble hélice del miARN maduro, este se une a un complejo ribonucleoproteico (RISC), el cual retiene a una de las cadenas que será el miARN funcional y la otra es eliminada [44] (ver Figura 17).

La cadena simple del miARN maduro se aparea con las bases complementarias de una secuencia diana, generalmente ubicada en la UTR 3´, el apareamiento se produce especialmente en el extremo 5´del miARN, en los nucleótidos 2 a 7, a esto se le llama "región semilla" (*seed region*). Dicho apareamiento del miARN con la *seed region* es esencial para una regulación efectiva por parte del miARN [45] (ver Figura 18).

Los miARN generalmente ejercen un efecto negativo sobre la expresión de los genes, bien inhibiendo la traducción o mediante degradación del ARNm, aunque también hay algunas evidencias de que pueden potenciar la expresión de los genes [46]. Si bien en los últimos años se han hecho grandes progresos en el conocimiento de la biogénesis y función de los miARN, todavía existen algunas controversias sobre los mecanismos que emplean los miARNs para la regulación de la expresión génica, si bien algunos de ellos son bien conocidos. En animales la represión de la traducción parece ser ejercida de cuatro formas distintas: inhibición del inicio de la traducción, inhibición de la elongación de la misma, degradación proteica co-traduccional y terminación prematura de la traducción. Si bien el primer mecanismo parece ser el más relevante [47, 48].

Un hecho que es relevante para nuestra discusión es que los ARNm son competentes para la traducción si poseen una estructura de caperuza-5' y una de cola de poliA-3´ . Los factores que se asocian con la caperuza y la cola poli-A interaccionan, así la proteína de unión a poli-A (PABP) interactúa con el factor de iniciación de la traducción 4G (eIF4G), que se asocia con la estructura de la caperuza-5´ a través de la interacción con la proteína de unión a la caperuza (eIF4E) (ver Figura 19). Esta interacción da lugar a ARNm circulares que se protegen de la degradación y son traducidos de manera eficiente [49]. Cada vez hay más pruebas que sugieren que los miARNs animales interfieren con la función del complejo eIF4F (que comprende: eIF4E, eIF4G y eIF4A) y PABPC durante la traducción y / o estabilización del ARNm [46].

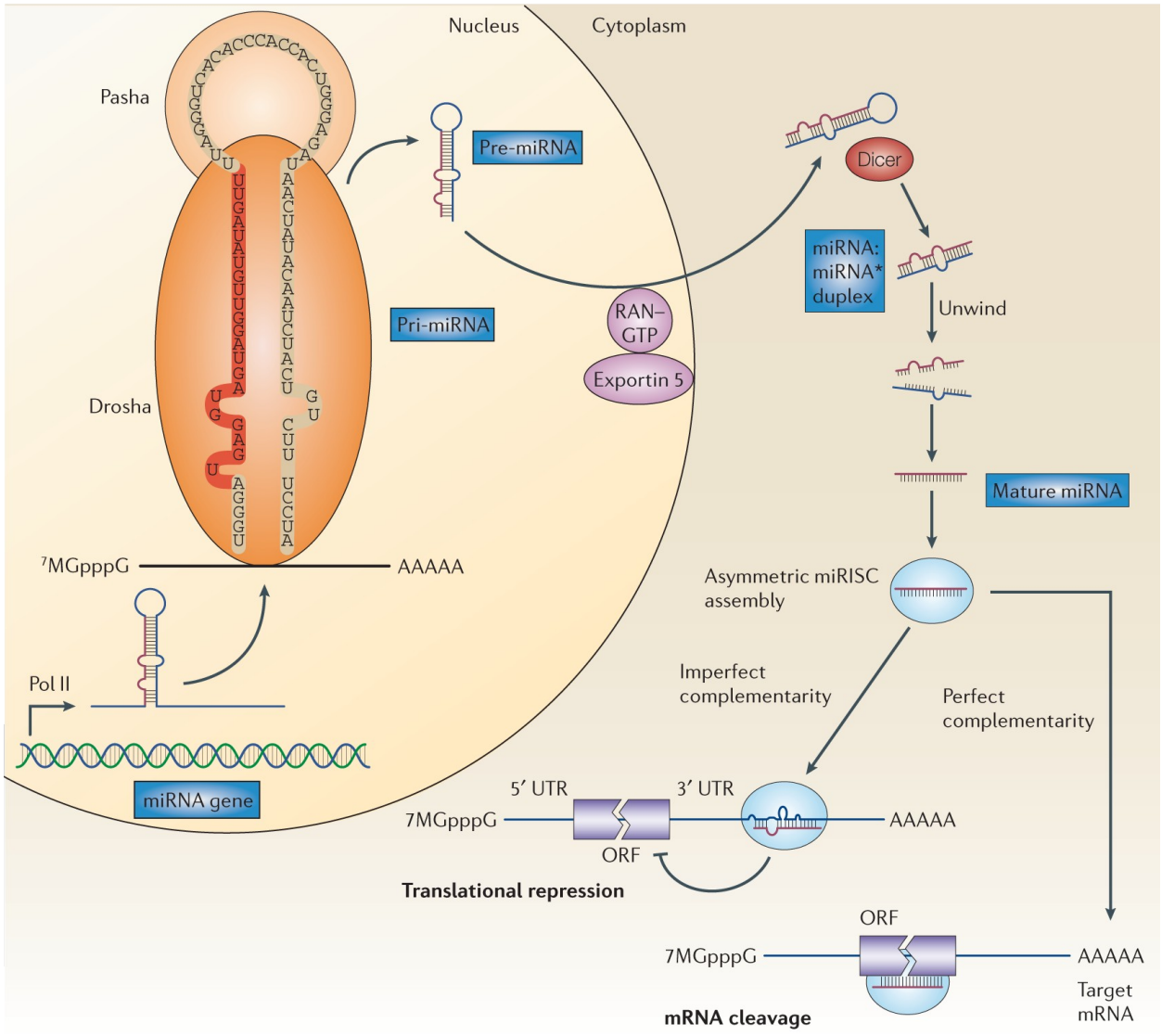


Figura 17 . *Biogénesis y vías de acción de un miARN en eucariotas. Se representan los pasos de la biogénesis tal como se cita en el texto. Se muestran también dos de las posibles vías de acción de los miARN sobre los genes: represión de la traducción y degradación por endonucleosis del ARNm. http://www.nature.com/nrc/journal/v6/n4/fig_tab/nrc1840_F1.html*

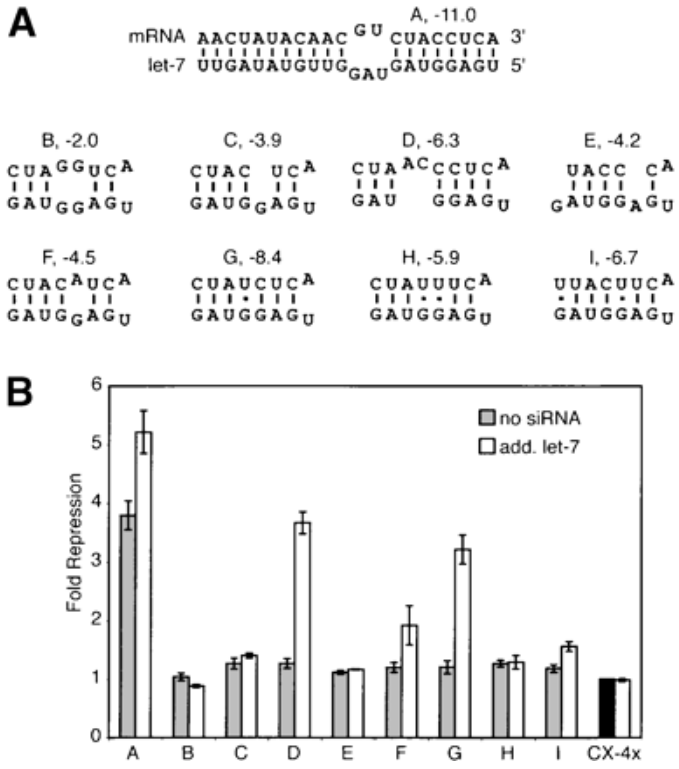


Figura 18. *Confirmación de la importancia de la región 5´ del miARN let-7a, mediante uso de constructos en células HeLa. (A) Representación esquemática predicha del sitio de unión de la UTR 3´, y su interacción con let-7a, junto con ocho mutantes (B-I) para la región 5 ' de let-7a (los números indican el valor de ΔG). (B) Grado de la represión de los diversos constructos que se muestran en A. En gris se muestra el grado de represión ejercido por los let-7a endógenos y en blanco la represión tras la adición de let-7a adicionales. Los valores de expresión se normalizaron respecto a la expresión de luciferasa de luciérnaga, a través de muestras de constructos de control CXCR4-4x, que se muestra en negro. Los valores son promedios de tres experimentos independientes ± desviación estándar. Figura extraída de referencia [45]*

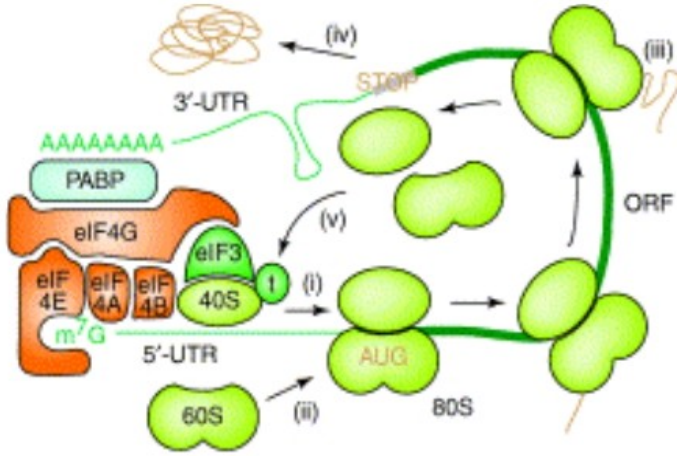


Figura 19. *Circularización del ARNm tal como se describe en el texto. Figura extraída de referencia [70].*

Se da la circunstancia de que un mismo miARN puede regular a un gran número de genes, ya que solo es necesaria la complementariedad de la *seed region.* Adicionalmente un mismo ARNm puede ser regulado por varios miARNs, aumentando así su repertorio de expresión en un momento determinado o en un tipo celular particular [3]. Los ARNm con múltiples dianas presentan una expresión mutuamente exclusiva respecto al miARN, indicando esto que los miARN ejercen un ajuste muy fino de la expresión génica [54].

Muchos miARN se han conservado a lo largo de la evolución, y debido a que no se requieren grandes regiones de apareamiento nuevos miARN pueden aparecer relativamente fácil, aportando nuevas herramientas para la evolución [3]. Un estudio comparativo entre diferentes especies de mamíferos, a escala genómica, encontró que los motivos conservados en la UTR 3´ son como media de 8 nt de longitud y que alrededor de la mitad de ellos parecen estar relacionados con miARNs [43].

Aunque los miARNs se unen preferentemente a las UTRs 3´, algunas dianas han sido identificadas en las UTRs 5´y en las regiones intrónicas de los genes [3].

4.3.2.- Elementos de estabilización

La modificación de la estabilidad de un transcrito permite una rápido control de la expresión sin necesidad de alterar la tasa de traducción. Se ha encontrado que este mecanismo está implicado de forma crítica en diversos procesos vitales como el crecimiento celular, la diferenciación o la adaptación a estímulos externos [55,56]. Los elementos de estabilización mejor estudiados son los elementos ricos en "AU" (AREs) que se encuentran en la UTR 3´de diversos genes. Estos elementos tienen un tamaño que oscila entre 50 y 150 nt y generalmente contienen múltiples copias del pentanucleótido AUUA [57]. Los AREs pueden presentar variaciones en su secuencia de forma que han sido definidas tres clases principales. La clase I y II presentan diferente número de repeticiones del pentanucleótido, mientras que la clase III carece del mismo [58].

Los AREs se unen a proteínas (ARE-BPs) que generalmente promueven el decaimiento del ARNm en respuesta a una variedad de señales intra y extracelulares. Muchas ARE-BPs son expresadas de forma específica en ciertos tejidos o tipos celulares, siendo además la estructura secundaria del ARE

un factor importante en la acción de la ARE-BP [59]. Diferentes ARE-BPs pueden competir por un mismo sitio de unión y dependiendo de la localización celular o factores ambientales la regulación ejercida por el ARE puede tener diferentes resultados para un transcrito [56] (ver Figura 20) . Un ejemplo del efecto de factores ambientales lo tenemos en la expresión de la proteína anti-apoptótica Bcl-X_L, que se ve incrementada tras radiacción UVA, siendo este un proceso implicado en cáncer de piel entre otros cánceres. El examen de las ARE-BPs asociadas con un ARE en la UTR 3´de Bcl-K_L, permitió identificar a la nucleotina como una proteína estabilizadora clave, y los autores sugieren que la radiacción UVA aumenta la capacidad de la nucleotina para unirse a ARE y esta facilita la protección del ARNm frente a la degradación [60]. Este y otros ejemplos ponen en evidencia la versatilidad de los AREs en la regulación de la estabilidad y la traducción, pudiendo producir distintos resultados en función de las señales recibidas [3].

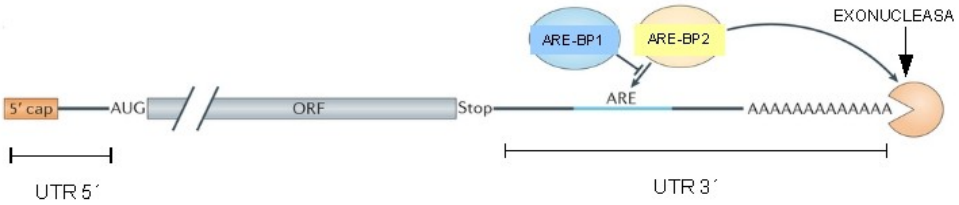


Figura 20. *Diferentes AREs-BP compiten por el motivo ARE de la UTR 3´para activar o inhibir la degradación del ARNm mediante exonucleasa*

El elemento rico en "GU" GRE es otro elemento de estabilidad descubierto posteriormente, el cual interacciona con la proteína CUGBP1, lo que promueve el decaimiento del ARNm asociado [61].

4.3.3.- Cola de poli-A

La cola de poli-A es el resultado de la adición de una serie de adenosinas en el extremo 3´del ARNm. Esto dota al ARNm de un lugar de unión para una clase de factores reguladores llamados proteínas de unión a poli-A (PABPs), que desempeñan un papel regulador de la expresión génica, incluida la estabilidad y el decaimiento del ARNm, así como la exportación del mismo [62]. Estos mecanismos de regulación juegan un papel relevante durante el desarrollo de los vertebrados [63].

En humano se han identificado hasta cinco PABPs diferentes (una nuclear y cuatro citoplasmáticas), todas ellas con diferentes papeles funcionales [63]. La PABPs parecen funcionar como andamios donde pueden unirse otros numerosos factores y así regular de una forma indirecta la expresión de los genes en diferentes circunstancias (recordar por ejemplo su papel en la circularización del ARNm) [3].

La cola de poli-A es sintetizada con una longitud determinada (sobre unos 250 pb en las células humanas) y puede luego ser acortada en el citoplasma para promover la represión de la traducción si las condiciones lo requieren [64]. Es bien conocido que la poliadenilación viene determinada por señales de poliadenilación (PAS) de la región UTR 3´. La señal de poliadenilación más frecuente (señal canónica) es "AAUAAA", localizada a 10-30 nucleótidos al 5´del sitio de corte y posteriormente se encuentra una secuencia rica en G/U a unos 20-30 nucleótidos al 3´del lugar de corte [79] (ver Figura 21). Los dos complejos multiméricos, el de poliadenilación (que se une a la PAS) y el de corte (que se une a la señal rica en G/U) son necesarios simultáneamente para que ambos procesos ocurran adecuadamente.

Alternativamente a la PAS canónica "AAUAAA" y en menor frecuencia se da la señal "AUUAAA" [80]. Se ha determinado, más recientemente, que pueden existir al menos otras 10 posibles variaciones de la señal de poliadenilación, todas ellas de menor frecuencia que la señal alternativa "AUUAAA" y que originan una menor eficiencia de la poliadenilación [81].

La poliadenilación alternativa (APA) puede ocurrir como resultado de la producción de diferentes isoformas de ARNm que difieren en su UTR 3´. La APA puede darse bien por que

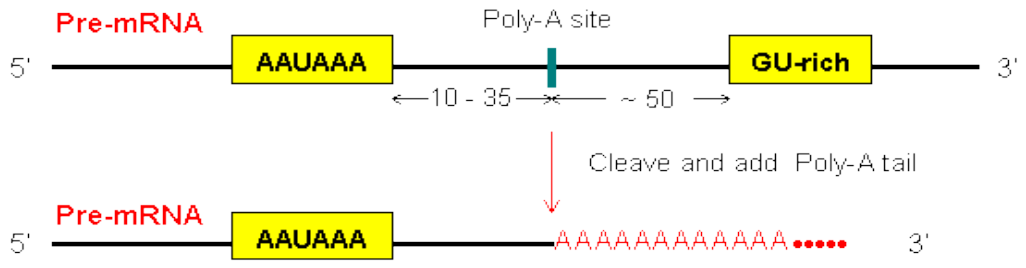


Figura 21. *Señales para el corte y poliadenilación del pre-ARNm en mamíferos. (http://www.web-books.com/MoBio/Free/Ch5A.htm).*

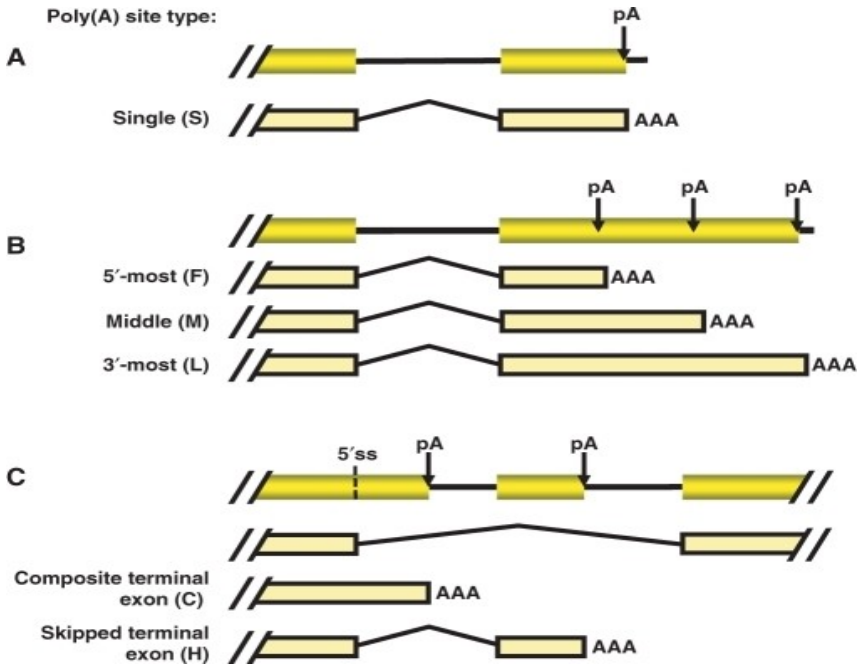


Figura 22. *Diferentes tipos de poliadenilación. (A) Un solo lugar de poliadenilación; (B) Poliadenilación alternativa mediante señales de poliadenilación (pA) alternativas en mismo exón terminal; (C) Poliadenilación alternativa mediante pA alternativas aguas arriba del exón terminal y fenómenos de splicing alternativo. (http://openi.nlm.nih.gov/detailedresult.php?img=2553571_gkn540f1&req=4).*

existan sitios de poliadenilación (PAS) alternativos en una misma UTR 3´o porque se expresen exones terminales mutuamente exclusivos (ver Figura 22). La APA es utilizada en aproximadamente el 50 % de los genes humanos [65]. El uso de UTRs 3´alternativas, con diferente poliadenilación, y corte asociado a ella, es un aspecto importante de la expresión génica durante el desarrollo o la expresión específica de tejidos [37, 66], así mismo cambios en la APA han sido asociados con diferentes tipos de cáncer [67].

4.3.4.- Longitud y estructura secundaria

La necesidad de interacción entre el extremo 5´y 3´, conocida como circularización del ARNm, y que ya mencionamos anteriormente (Figura 19) al tratar los miARN, se ha revelado como un mecanismo clave para una traducción eficiente [49]. En esta interacción 5´-3´ tienen implicaciones la longitud y estructura secundaria de la UTR 3´. Existen evidencias del efecto de UTRs 3´ largas

sobre la expresión, así por ejemplo se ha observado que el aumento de la UTR 3´de 19 a 156 nt disminuye la expresión hasta en 45 veces [68].

Adicionalmente al efecto que la longitud puede tener sobre la interacción 5´-3´, puede afectar también al número de dianas de miARN, ya que a mayor longitud de la UTR 3´existen mayores probabilidades de presencia de las mismas. Un estudio donde se compara la longitud y la presencia de dianas de miARN en genes ribosómicos y en genes de neurogénesis encontró que los genes ribosómicos contenían UTRs 3´más cortas y eran escasos en dianas de miARN, al comparar con secuencias aleatorias de control [54]. Mientras que los genes de neurogénesis presentaban UTRs 3 ´más largas y especialmente enriquecidas en sitios potenciales de unión. Igualmente existen evidencias de genes que presentan UTRs 3´alternativas para el control de la expresión, presentando la forma más larga dianas conservadas de ciertos miARNs que se expresan en ciertas condiciones celulares, reprimiéndose así la traducción [69]. En general las UTRs 3´más largas están relacionadas con menores niveles de expresión, como lo demuestran experimentos que comparan la expresión de isoformas que difieren solamente en la longitud de la UTR 3´[69] (ver Figura 23).

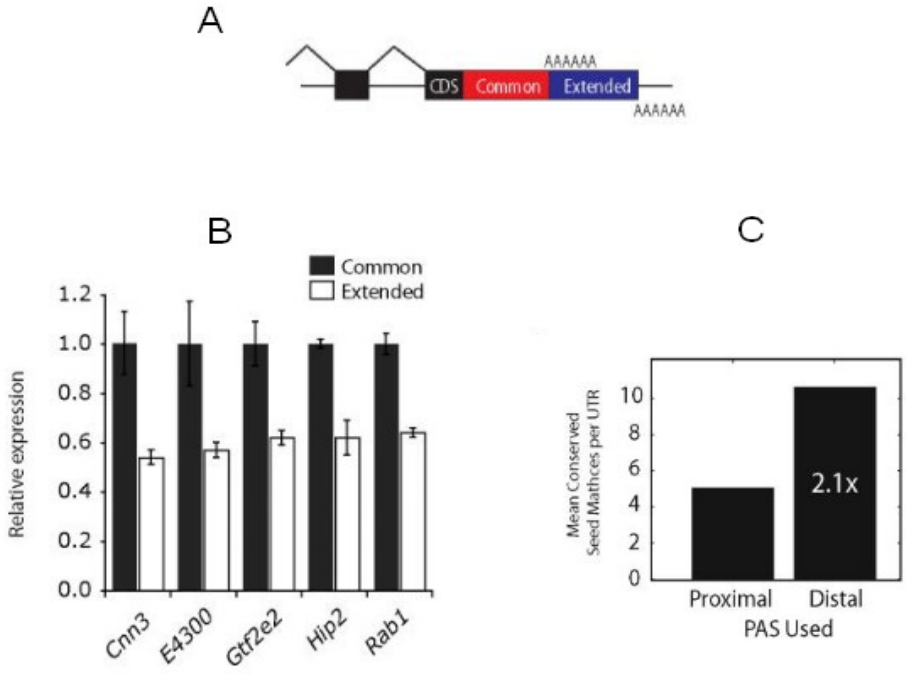


Figura 23. *Efecto de la longitud de la UTR 3´utilizada sobre la expresión génica. (A) Esquema que representa las isoformas alternativas: común (UTR 3´corta) y extendida (UTR 3´larga), con sus diferentes sitios de poliadenilación. (B) Actividad luciferasa en linfocitos T murinos transfectados con constructos que llevan la isoforma de la UTR 3´ común o extendida de los genes indicados.(C) Número medio de dianas de miARN conservadas en la UTR 3´de isoformas de genes que usan la PAS proximal o distal. Imagen modificada de referencia [69].*

Ya habíamos mencionado que las UTRs 3´son considerablemente más largas que las correspondientes UTRs 5´(como media casi cuatro veces mayores en humanos [2]). Esta diferencia ofrece un gran potencial para la regulación de la traducción en dichas regiones, bien sea mediante la ausencia-presencia de dianas de miARNs o mediante las diferencias en otros elementos que puedan afectar a la unión de diversos factores, que afecten a la recircularización o a la estabilidad del ARNm [3, 70]. Adicionalmente parece existir una tendencia a un aumento de la UTR 3´en el proceso evolutivo que va desde los hongos a los invertebrados y de estos a los vertebrados (Figura 24), siendo la longitud media de la UTR 3´humana más del doble de la longitud de los otros mamíferos estudiados [3]. Este incremento de la longitud puede interpretarse como un mayor número de posibilidades de regulación de la expresión, hecho que posiblemente ha tenido relevancia en el aumento de la complejidad de los organismos, incluidos los humanos.

Si bien la longitud de la UTR 3´es un hecho relevante para la eficiencia de la traducción, no lo es menos la estructura secundaria de la misma, como lo ponen de manifiesto el efecto que tienen mutaciones que afectan a dicha estructura. Este hecho queda patente en un estudio sobre 83 enfermedades asociadas a variaciones en la UTR 3´ de ARNm humanos, donde se encuentra una correlación entre la funcionalidad de las variaciones y cambios en la estructura secundaria predicha [71].

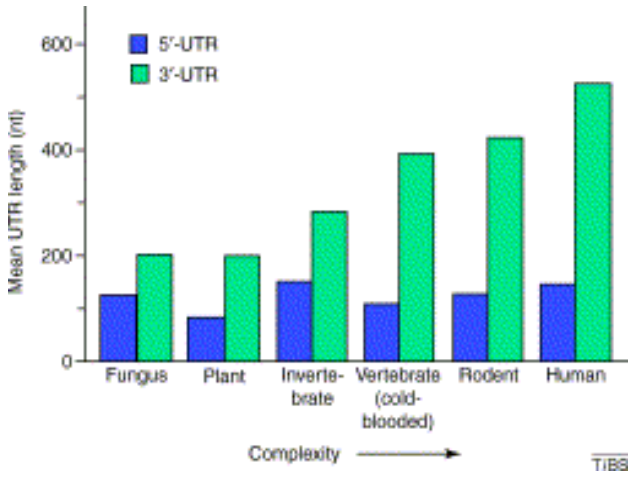


Figura 24. *Longitud media de las UTRs 5´y 3´ en diferentes especies. Imagen extraída de referencia [70].*

Generalmente la localización de una mutación de falta de sentido, anexa al límite exón-exón determina la eficiencia del mecanismo NMD (*nonsense mediated decay*), sin embargo la UTR 3 ´puede tener influencia en este proceso. El mecanismo de terminación de la traducción en un codón de terminación prematura (PTC) se ve afectado por la distancia entre el codón de terminación y PABPCl (*poly-A binding protein*). Se ha demostrado que la extensión de esta región es crucial para determinar la presencia de PTC y acabar dando lugar a NMD, así mismo su acortamiento produce la supresión de NMD. Igualmente las reordenaciones estructurales de la UTR 3´ pueden modular la ruta NMD y suponen un mecanismo adicional de la regulación de la expresión génica [72].

La estructura en forma de horquilla es la forma más común que puede afectar a la expresión génica y esto suele ocurrir en la UTR 3´ mediado por la unión a dichas estructuras de factores proteicos. Diferentes ejemplos demuestran la existencia de estructuras tipo horquilla en la UTR 3´que modulan su afinidad por ciertos factores, afectando así a la estabilidad del ARNm y regulando la expresión génica [73, 74].

4.3.5.- Control de la localización subcelular del ARNm

La localización subcelular de ARNm se ha revelado como un mecanismo clave a través del cual las células se polarizan. La localización de los transcritos es una forma muy eficiente para orientar productos génicos a compartimentos subcelulares individuales o a regiones específicas de una célula o un embrión, por lo que es un nivel importante de la regulación postranscripcional de genes. La localización del ARNm es un fenómeno generalizado que se produce en los organismos unicelulares, en los tejidos de animales y plantas, y en el desarrollo de embriones de una gran variedad de filos animales.

Después de la transcripción en el núcleo, la mayoría de los ARNm salen a través de los poros nucleares hacia el citoplasma, donde son traducidos. Ciertas clases de ARNm, sin embargo, tienen diferentes destinos y están dirigidos a regiones específicas dentro de la célula o embrión y en muchos casos no se traducen hasta que llegan a su destino final. Los pasos iniciales en el proceso de localización están determinados por los elementos que actúan *en cis* dentro del ARNm. Esta serie de señales por lo general, pero no exclusivamente, están ubicadas dentro de la UTR 3´, a las que se unen factores que actúan *en trans* [75].

Hasta ahora, todos los factores que actúan *en trans* se han identificado como proteínas; sin embargo, es muy posible que otras clases de moléculas tales como pequeños ARNs reguladores puedan desempeñar un papel en este proceso. La unión de los factores, que probablemente influye en el plegamiento del ARNm en una configuración espacial específica, facilitan la asociación de una serie de otras proteínas auxiliares y esto produce una gran partícula de ribonucleoproteína (RNP) de transporte. Es probable que a través de los factores proteicos que el ARNm reconoce y asociados con la vía adecuada o estructura subcelular se dirigirá a su destino correcto. En el caso de las vías que puede utilizar la maquinaria del citoesqueleto, probablemente es un mecanismo de transporte activo que implica un motor molecular que, junto con proteínas adaptadoras, impulsará el ARNm. Una vez en su destino, se ancla a través de un anclaje molecular, lo que podría ser o bien proteína o, en algunos casos, otra clase de ARN [75].

Los estudios en diversos sistemas tales como ovocitos, embriones, y células somáticas han demostrado la existencia de varios mecanismos potenciales por los cuales los ARNm pueden estar distribuidos localmente. Estos incluyen el transporte direccional activo de ARNm por los elementos del citoesqueleto, la degradación general y la estabilidad del ARNm localizada y la difusión aleatoria citoplasmática con captura del ARNm (Figura 25). Una combinación de mecanismos puede ser utilizada para localizar diferentes ARNm; sin embargo, las evidencias más convincentes son de los mecanismos de transporte direccional activo por los elementos del citoesqueleto y la degradación combinada con estabilidad localizada [75].

Un claro ejemplo de este control espacial de la expresión, mediante control de transporte activo lo tenemos en MBP, el principal componente estructural de la mielina de la membrana de células del sistema nervioso central, producido por oligodendrocitos. Técnicas de mapeo por delección y experimentos de microinyección con ARNs quiméricos han delineado una secuencia de "ARN de tráfico" de 21-nucleótidos (RTS) en la UTR 3´ del ARNm de MBP, que es necesaria y suficiente para el transporte del ARN en los oligodendrocitos. Secuencias semejantes a RTS también se encuentran en una variedad de ARNm que se localizan en otros tipos de células, lo que sugiere que la RTS es una señal de tráfico de ARN de uso general. Experimentos de unión *in vitro* indican que la RTS se une con especificidad y gran afinidad a hnRNP A2, proteína de unión a ARN expresada de forma ubicua, y que se cree que juega un papel en el tráfico intracelular de ARNm [76].

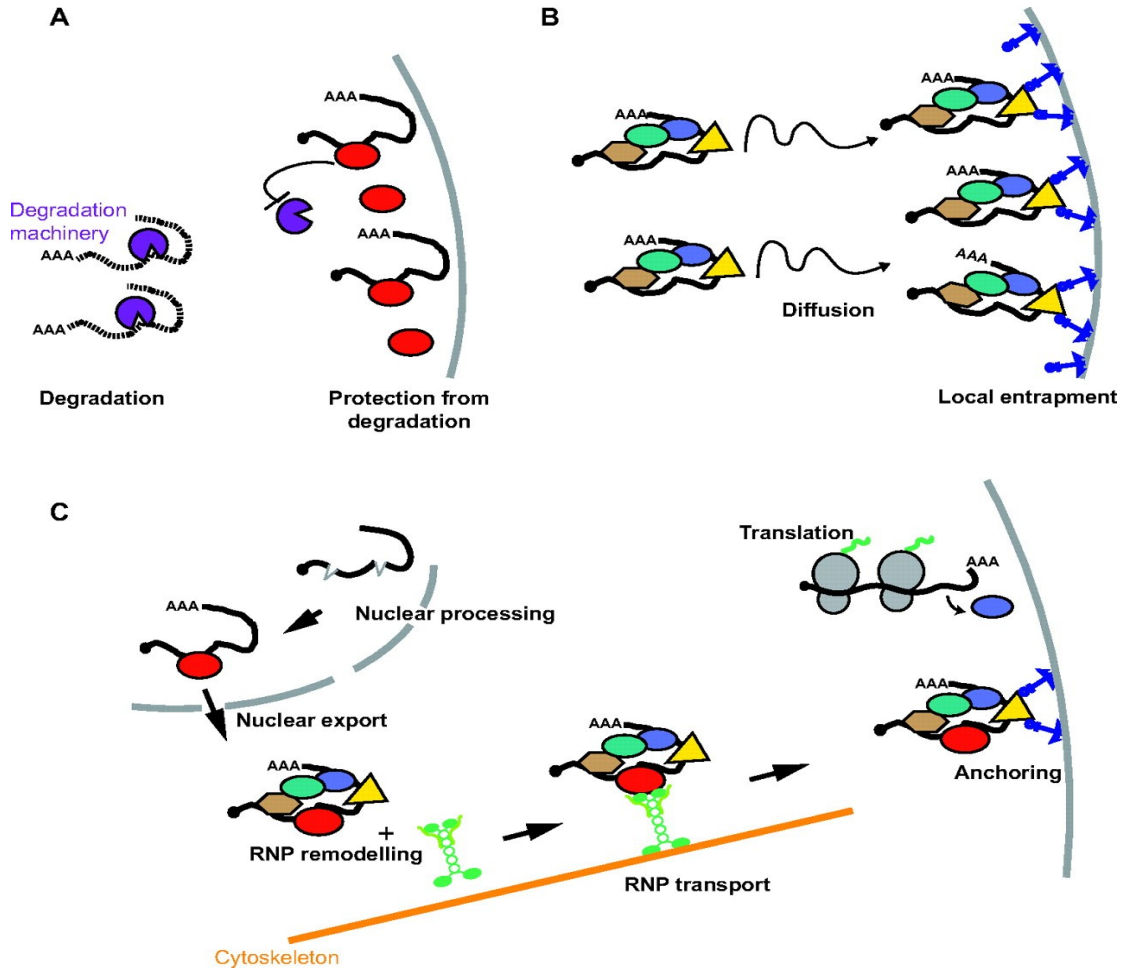


Figura 25. *Mecanismos de transporte del ARNm en la célula. (A) Estabilización local que protege frente a la degradación; (B) Difusión del ARNm y captura local y (C) Transporte activo del ARNm mediante el citoesqueleto. Figura extraída de referencia [84].*

La expresión de una UTR 3´alternativa, más corta o más larga, puede tener influencia en el transporte del ARNm, como en el caso de CaMKIIα (*alfa subunit* Ca^{2+}*/calmodulin-dependent protein kinase II*). Una secuencia requerida para el destino dendrítico de su ARNm se ha mapeado aguas abajo del primer sitio de poliadenilación, lo que sugiere que la isoforma más corta del ARNm se limita al cuerpo celular, mientras que el ARNm más largo se transporta a las dendritas [78].

Distintos ejemplos de localización por estabilización local provienen de Drosophila, durante el desarrollo embrionario temprano. Así los transcritos de la proteína de unión a ARN Nanos o la

proteína de choque térmico Hsp83 son degradados en todas las partes del embrión excepto en el polo posterior. Distintos elementos actuando *en cis,* localizados en la UTR 3´de estos transcritos, median en la degradación global en el embrión y en la estabilización en el citado polo [77].

4.3.6.- Conclusión sobre UTR 3´

La UTR 3´es una región muy versátil y es rica en elementos reguladores, por lo que desempeña un papel crucial en la correcta expresión espacial y temporal de los genes.

La longitud y estructura secundaria de la UTR 3´son aspectos clave, que tienen una gran influencia en el papel regulador que dicha región ejerce. En general las UTRs 3´ de gran longitud tienen un efecto negativo sobre la traducción. Las estructuras típicas en horquilla que se dan en la UTR 3 ´tienen afinidad por diversos factores que modulan así la acción reguladora de estos elementos a nivel de la eficacia en la traducción o en la localización celular del ARNm.

La circularización del ARNm mediante la interacción de la proteína PABP (que se une a la cola de poli-A) con el complejo de iniciación de la transcripción del extremo 5´ parece ser necesaria para una adecuada traducción. Diversos factores (factores proteicos, miARNs u otros ARNs) que tienen afinidad por la UTR 3´ pueden afectar a la interacción necesaria para la circularización, dando lugar a diversos mecanismos reguladores de la expresión génica.

La poliadenilación, fundamental para la estabilidad del ARNm, es un fenómeno que puede ser motivo de regulación mediante el uso de señales de poliadenilación alternativas existentes en la UTR 3´, que dan lugar a una cola de poli-A de diferente longitud. En este hecho tiene un papel importante el uso de UTRs 3´alternativas más largas o más cortas que contienen una o varias señales de poliadenilación. La mayor o menor longitud de la UTR 3´viene dado por el uso de diferentes puntos de corte del ARNm, que se encuentran aguas abajo de la señal de poliadenilación reconocida.

El uso de UTRs 3´alternativas, bien por sitios de corte alternativo, bien por el uso de exones alternativos existentes en esta región, permite también la presencia o ausencia de elementos reguladores, lo que da lugar a una regulación de la expresión génica diferenciada en función del tejido o estado de la célula.

4.4.- Bibliografía

[1] Mignone, F. , Gissi, C. , Liuni, S. and Pesole, G. Unstranslated regions of mRNAs. *Genome Biology* **3,** reviews 0004.1- 0004.10. (2002).

[2] Pesole, G. *et al.* Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* **276,** 73–81 (2001).

[3] Barrett, L. W. & Fletcher, S. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* **69,** 3613–3634 (2012).

[4] Pesole, G., Bernardi, G. & Saccone, C. Isochore specificity of AUG initiator context of human genes. *FEBS Letters* **464,** 60–62 (1999)

[5] Duret L, Mouchiroud D, Gautier C.Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J Mol Evol. 1995 Mar;40(3):308-17.

[6] Ganapathi, M. *et al.* Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes. *BMC Bioinformatics* **6,** 126 (2005).

[7] Kapp, L. D. & Lorsch, J. R. The Molecular Mechanics of Eukaryotic Translation. *Annual Review of Biochemistry* **73,** 657–704 (2004).

[8] Pickering, B. M. & Willis, A. E. The implications of structured 5′ untranslated regions on translation and disease. *Seminars in Cell & Developmental Biology* **16,** 39–47 (2005).

[9] Davuluri, R. V., Suzuki, Y., Sugano, S., & Zhang, M. Q. (2000). CART Classification of Human 5′ UTR Sequences. *Genome Research*, *10*(11), 1807–1816.

[10] Babendure, J. R., Babendure, J. L., Ding, J.-H., & Tsien, R. Y. (2006). Control of mammalian translation by mRNA structure near caps. *RNA*, *12*(5), 851–861. doi:10.1261/rna.2309906

[11] Araujo, P. R. *et al.* Before It Gets Started: Regulating Translation at the 5′ UTR. *International Journal of Genomics* **2012,** e475731 (2012).

[12] Beaudoin, J.-D. & Perreault, J.-P. 5′-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res* **38,** 7022–7036 (2010).

[13}Gomez, D. *et al.* A G-quadruplex structure within the 5′-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res* **38,** 7187–7198 (2010).

[14] Fraser, D. J. *et al.* Y-box protein-1 controls transforming growth factor-β1 translation in proximal tubular cells. *Kidney Int* **73,** 724–732 (2007).

[15] Komar, A. A. & Hatzoglou, M. Internal Ribosome Entry Sites in Cellular mRNAs: Mystery of Their Existence. *J. Biol. Chem.* **280,** 23425–23428 (2005).

[16] Le, S. Y. & Maizel, J. V. A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs. *Nucleic Acids Res* **25,** 362–369 (1997).

[17] Chappell, S. A., Edelman, G. M. & Mauro, V. P. A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. *Proc Natl Acad Sci U S A* **97,** 1536–1541 (2000).

[18] Cobbold, L. C. *et al.* Identification of Internal Ribosome Entry Segment (IRES)-trans-Acting Factors for the Myc Family of IRESs. *Mol Cell Biol* **28,** 40–49 (2008).

[19] Avni, D., Biberman, Y., and Meyuhas, O.. The 5 terminal oligopyrimidine tract conferstranslational control on TOP mRNAs in a cell type- and sequence context-dependent

manner. *Nucleic Acids Res*. **25,** 995–1001 (1997).

[20] Riu Yamashita, et al. Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. Nucleic Acids Research, 2008, Vol. 36, No. 11 3707–3715.

[21] Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *PNAS* **106,** 7507–7512 (2009).

[22] Iacono, M., Mignone, F. & Pesole, G. uAUG and uORFs in human and rodent 5′untranslated mRNAs. *Gene* **349,** 97–105 (2005).

[23] Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *PNAS* **106,** 7507–7512 (2009).

[24] Churbanov, A., Rogozin, I. B., Babenko, V. N., Ali, H. & Koonin, E. V. Evolutionary conservation suggests a regulatory function of AUG triplets in 5′-UTRs of eukaryotic genes. *Nucleic Acids Res* **33,** 5512–5520 (2005).

[25] Rogozin, I. B., Kochetov, A. V., Kondrashov, F. A., Koonin, E. V. & Milanesi, L. Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics* **17,** 890–900 (2001).

[26] Vogel, C. *et al.* Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol* **6,** 400 (2010).

[27] Matsui, M., Yachie, N., Okada, Y., Saito, R. & Tomita, M. Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse. *FEBS Letters* **581,** 4184–4188 (2007).

[28] Rapti, A. *et al.* The structure of the 5′-untranslated region of mammalian poly(A) polymerase-α mRNA suggests a mechanism of translational regulation. *Mol Cell Biochem* **340,** 91–96 (2010).

[29] Liu, L. *et al.* Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat Genet* **21,** 128–132 (1999).

[30] Wiestner, A., Schlemper, R. J., van der Maas, A. P. C. & Skoda, R. C. An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nat Genet* **18,** 49–52 (1998).

[31] Meijer, H. A. & Thomas, A. A. M. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem J* **367,** 1–11 (2002).

[32] Crowe, M. L., Wang, X.-Q. & Rothnagel, J. A. Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides. *BMC Genomics* **7,** 16 (2006).

[33] Lee, M.-H. & Schedl, T. Translation repression by GLD-1 protects its mRNA targets from nonsense-mediated mRNA decay in C. elegans. *Genes Dev* **18,** 1047–1059 (2004).

[34] Gaba, A., Jacobson, A. & Sachs, M. S. Ribosome Occupancy of the Yeast CPA1 Upstream Open Reading Frame Termination Codon Modulates Nonsense-Mediated mRNA Decay. *Molecular Cell* **20,** 449–460 (2005).

[35] Vilela, C., Ramirez, C. V., Linz, B., Rodrigues-Pousada, C. & McCarthy, J. E. Post-termination ribosome interactions with the 5'UTR modulate yeast mRNA stability. *EMBO J* **18,** 3139–3152 (1999).

[36] Oyama, M. *et al.* Analysis of Small Human Proteins Reveals the Translation of Upstream Open Reading Frames of mRNAs. *Genome Res* **14,** 2048–2052 (2004).

[37] Hughes, T. A. Regulation of gene expression by alternative untranslated regions. *Trends in Genetics* **22,** 119–122 (2006).

[38] Hughes, T. A. & Brady, H. J. M. E2F1 up-regulates the expression of the tumour suppressor axin2 both by activation of transcription and by mRNA stabilisation. *Biochemical and Biophysical Research Communications* **329,** 1267–1274 (2005).

[39] Hughes, T. A. & Brady, H. J. M. Expression of axin2 Is Regulated by the Alternative 5′-Untranslated Regions of Its mRNA. *J. Biol. Chem.* **280,** 8581–8588 (2005).

[40] Martineau, Y. *et al.* Internal Ribosome Entry Site Structural Motifs Conserved among Mammalian Fibroblast Growth Factor 1 Alternatively Spliced mRNAs. *Mol. Cell. Biol.* **24,** 7622–7635 (2004).

[41] McClelland, S., Shrivastava, R. & Medh, J. D. Regulation of Translational Efficiency by Disparate 5′ UTRs of PPARγ Splice Variants. *PPAR Res* **2009,** (2009).

[42] Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15,** 1034–1050 (2005).

[43] Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434,** 338–345 (2005).

[44] Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116,** 281–297 (2004).

[45] Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18,** 504–511 (2004).

[46] Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* **12,** 99–110 (2011).

[47] Carthew, R. W. & Sontheimer, E. J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).

[48] Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature Rev. Genet. 9, 102–114 (2008).

[49] Derry, M. C., Yanagiya, A., Martineau, Y. & Sonenberg, N. Regulation of poly(A)-binding protein through PABP-interacting proteins. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 537–543 (2006).

[50] Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594–596 (2004).

[51] Wu, L., Fan, J. & Belasco, J. G. MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl Acad. Sci. USA* **103**, 4034–4039 (2006).

[52] Eulalio, A. *et al*. Deadenylation is a widespread effect of miRNA regulation. *RNA* **15**, 21–32 (2009).

[53] Zhang, R. & Su, B. Small but influential: the role of microRNAs on gene regulatory network and 3′UTR evolution. *Journal of Genetics and Genomics* **36,** 1–6 (2009).

[54] Stark, A., Brennecke, J., Bushati, N., Russell, R. B. & Cohen, S. M. Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3′UTR Evolution. *Cell* **123,** 1133–1146 (2005).

[55] Elkon, R., Zlotorynski, E., Zeller, K. I. & Agami, R. Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics* **11,** 259 (2010).

[56] 1Eberhardt, W., Doller, A., Akool, E.-S. & Pfeilschifter, J. Modulation of mRNA stability as a

novel therapeutic approach. *Pharmacology & Therapeutics* **114,** 56–73 (2007).

[57] Chen, C.-Y. A. & Shyu, A.-B. AU-rich elements: characterization and importance in mRNA degradation. *Trends in Biochemical Sciences* **20,** 465–470 (1995).

[58] Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mRNAs. *Genome Biol* **3,** reviews0004.1–reviews0004.10 (2002).

[59] Meisner, N.-C. *et al.* mRNA openers and closers: modulating AU-rich element-controlled mRNA stability by a molecular switch in mRNA secondary structure. *Chembiochem* **5,** 1432–1447 (2004).

[60] Zhang, J., Tsaprailis, G. & Bowden, G. T. Nucleolin Stabilizes Bcl-XL Messenger RNA in Response to UVA Irradiation. *Cancer Res* **68,** 1046–1054 (2008).

[61] Vlasova, I. A. *et al.* Conserved GU-Rich Elements Mediate mRNA Decay by Binding to CUG-Binding Protein 1. *Mol Cell* **29,** 263–270 (2008).

[62] Gorgoni, B. & Gray, N. K. The roles of cytoplasmic poly(A)-binding proteins in regulating gene expression: a developmental perspective. *Brief Funct Genomic Proteomic* **3,** 125–141 (2004).

[63] Gorgoni, B. *et al.* Poly(A)-binding proteins are functionally distinct and have essential roles during vertebrate development. *Proc Natl Acad Sci U S A* **108,** 7844–7849 (2011).

[64] Kühn, U. *et al.* Poly(A) Tail Length Is Controlled by the Nuclear Poly(A)-binding Protein Regulating the Interaction between Poly(A) Polymerase and the Cleavage and Polyadenylation Specificity Factor. *J Biol Chem* **284,** 22803–22814 (2009).

[65] Dickson, A. M. & Wilusz, J. Polyadenylation: alternative lifestyles of the A-rich (and famous?). *EMBO J* **29,** 1473–1474 (2010).

[66] Wang, E. T. *et al.* Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* **456,** 470–476 (2008).

[67] Mayr, C. & Bartel, D. P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138,** 673 (2009).

[68] Tanguay, R. L. & Gallie, D. R. Translational efficiency is regulated by the length of the 3' untranslated region. *Mol Cell Biol* **16,** 146–156 (1996).

[69] Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3′ UTRs and fewer microRNA target sites. *Science* **320,** 1643–1647 (2008).

[70] Mazumder, B., Seshadri, V. & Fox, P. L. Translational control by the 3′-UTR: the ends specify the means. *Trends in Biochemical Sciences* **28,** 91–98 (2003).

[71] Chen, J.-M., Férec, C. & Cooper, D. N. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. *Hum. Genet.* **120,** 301–333 (2006).

[72] Eberle, A. B., Stalder, L., Mathys, H., Orozco, R. Z. & Mühlemann, O. Posttranscriptional Gene Regulation by Spatial Rearrangement of the 3′ Untranslated Region. *PLoS Biol* **6,** (2008).

[73] Fukuchi, M. & Tsuda, M. Involvement of the 3'-untranslated region of the brain-derived neurotrophic factor gene in activity-dependent mRNA stabilization. *J. Neurochem.* **115,** 1222–1233 (2010).

[74] Fialcowitz, E. J., Brewer, B. Y., Keenan, B. P. & Wilson, G. M. A hairpin-like structure within an au-rich mrna-destabilizing element regulates trans-factor binding selectivity and mrna decay kinetics. *J Biol Chem* **280,** 22406–22417 (2005).

[75] Kloc, M., Zearfoss, N. R. & Etkin, L. D. Mechanisms of Subcellular mRNA Localization. *Cell* **108,** 533–544 (2002).

[76] Kwon, S., Barbarese, E. & Carson, J. H. The Cis-Acting RNA Trafficking Signal from Myelin Basic Protein mRNA and Its Cognate Trans-Acting Ligand Hnrnp A2 Enhance CaP-Dependent Translation. *J Cell Biol* **147,** 247–256 (1999).

[77]1Bashirullah, A., Cooperstock, R. L. & Lipshitz, H. D. Spatial and temporal control of RNA stability. *Proc Natl Acad Sci U S A* **98,** 7025–7028 (2001).

[78] Di Liegro, C. M., Schiera, G. & Di Liegro, I. Regulation of mRNA transport, localization and translation in the nervous system of mammals (Review). *Int J Mol Med* **33,** 747–762 (2014).

[79] Colgan, D.F. and J.L. Manley. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes. Dev.* **11:** 2755–2766.

[80] Wahle, E. and W. Keller. The biochemistry of polyadenylation.
TIBS 21: 247–250.1996.Wahle, E. and W. Keller. The biochemistry of polyadenylation.
[81] Beaudoing, E., et al. Patterns of Variant Polyadenylation Signal Usage in Human Genes. Genome Research 10: 1001-1010. 2000

[82] Linde, L. & Kerem, B. Introducing sense into nonsense in treatments of human genetic diseases. *Trends in Genetics* **24,** 552–563 (2008).

[83] Bicknell, A. A., Cenik, C., Chua, H. N., Roth, F. P. & Moore, M. J. Introns in UTRs: Why we should stop ignoring them. *Bioessays* **34,** 1025–1034 (2012).

[84] Medioni, C., Mowry, K. & Besse, F. Principles and roles of mRNA localization in animal. *Development* **139,** 3263–3276 (2012).

CAPÍTULOS DE RESULTADOS

(I)

IDENTIFICACIÓN Y DESCRIPCIÓN DE LAS UTRs 5´ Y 3´ DE LAS LIPOCALINAS DE MAMÍFEROS

1.- Objetivos

Los objetivos de este capítulo son varios:

- En primer lugar identificar y caracterizar las UTRs 5´y 3´de lipocalinas de mamíferos mediante la determinación y análisis de parámetros básicos de las mismas, como son la longitud, su contenido en G+C y la presencia en ellas de secuencias repetitivas.

- En segundo lugar establecer en qué medida las lipocalinas presentan UTRs alternativas y si las hay, determinar el grado de variabilidad de las mismas.

- Una vez identificadas las lipocalinas que presentan variabilidad en las UTRs, determinar la organización genómica de estas regiones, así mismo tratar de dilucidar los mecanismos responsables del origen de los transcritos con UTRs alternativos.

2.- Métodos

2.1.- Obtención y selección de las secuencias de UTRs 5´y 3´ de lipocalinas

Se obtuvieron las secuencias de las UTRs de 11 lipocalinas de mamíferos (previamente seleccionadas tal como se detalla en los objetivos de la tesis) a partir de bases de datos que dispusieran de información sobre transcritos alternativos. Se decidió utilizar, como fuente principal, la base de datos de secuencias de ARNm **AceView** (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html)[1], debido a la rigurosidad de su metodología, por su útil entorno gráfico, por la completa información de las anotaciones, así como por la facilidad de acceso y descarga de las secuencias. De forma complementaria se utilizó la base de datos **ASPicDB** (http://srv00.ibbe.cnr.it/ASPicDB/index.php)[2].

AceView proporciona una representación depurada y no redundante de todas las secuencias de ARNm públicas (ARNm de GenBank o RefSeq y secuencias de ADNc a partir de dbEST y Trace). La metodología de AceView consiste en que estas secuencias de ADNc experimentales son coalineadas con el genoma, siendo seleccionadas solo si tienen una alta fiabilidad (un alto

porcentaje de identidad a lo largo de toda la longitud de su secuencia). A continuación se agrupan en un número mínimo de variantes (transcritos alternativos) agrupados en los correspondientes genes.

ASPicDB se basa en un método algo diferente y por ello complementario del de Aceview. Dicho método está basado en un algoritmo de alineación múltiple genoma-EST, para la detección de sitios de corte y empalme. Este enfoque realiza una alineación múltiple de los datos de transcripción con la secuencia genómica, basado en el análisis combinado de todos los datos disponibles. Los algoritmos de ASPicDB abordan el problema de predecir splicing constitutivo o alternativo como un problema de optimización, en donde el alineamiento múltiple óptimo de los transcritos minimiza el número de exones y por lo tanto los posibles sitios de corte y empalme.

Se procedió a buscar en la base de datos Aceview los transcritos de las 11 lipocalinas seleccionadas para humano y ratón, que son las especies de mamíferos disponibles en AceView. De todos los transcritos alternativos que ofrece AceView, para cada lipocalina, solo se tomaron las secuencias de los transcritos cuya secuencia codificante se correspondiese con la de la proteína canónica predicha o anotada en RefSeq (NCBI), que es la lipocalina contrastada y anotada como extracelular. En la tabla 1 se muestran dichas secuencias canónicas para humano y ratón. En caso de que el transcrito estuviese incompleto en su extremo 5´o 3´se tomó la UTR del extremo disponible (5´o 3´), solo si los exones codificantes de dicho extremo coincidían exactamente con los de la secuencia canónica.
Los transcritos así seleccionados en Aceview fueron posteriormente contrastados con los datos disponibles en ASPicDB, confirmando los primeros e incluso permitiendo en algunos casos, incluir algunos transcritos alternativos adicionales. En las tablas 2 a 12 del apartado de resultados se muestran los transcritos seleccionados para cada lipocalina. Dichos transcritos o variantes aparecen identificados, en dichas tablas, por las letras con que están identificados en Aceview. Cuando la variante solo se detectó en ASPicDB, se identificó con la letra de la variante de Aceview más parecida a esta, seguida de un número entre paréntesis.

Las secuencias de las UTRs 5´y 3´de todos los transcritos seleccionados fueron obtenidas y almacenadas en formato Fasta. Así mismo se obtuvieron y almacenaron en el mismo formato las secuencias de los diferentes exones, en caso de haberlos, que constituyen las regiones UTRs 5´y 3´.

Lipocalina	Especie	NCBI RefSeq	Long ARNm (pb)	Referencia genómica NCBI
APOD	ratón	NM_001301353.1	2070	Chr: 16(-) NC_000082.6 31296192..31314799
	humano	NM_001647.3	1148	Chr: 3(-) NC_000003.12 195568702..195584205
PTGDS	ratón	NM_008963.2	806	Chr: 2(-) NC_000068.7 25466709..25470113
	humano	NM_207510.3	1851	Chr 9: NC_000009.12 136977504..136981742
RBP4	ratón	NM_001159487.1	1233	Chr: 19(-) NC_000085.6 38116620..38125321
	humano	NM_006744.3	941	Chr: 10(-) NC_000010.11 93591836..93601344
APOM	ratón	NM_018816.1	731	Chr: 17(-) NC_000083.6 35128997..35131752
	humano	NM_001256169.1	1183	Chr 6: NC_000006.12 31652410..31658210
LCN1 (VEGP1)	rata	NM_022945.1	753	Chr: 3 NC_005102.4 4233111..4236960
	humano	NM_001252617.1	781	Chr: 9 NC_000009.12 135521438..135526540
LCN2	ratón	NM_008491.1	853	Chr: 2(-) NC_000068.7 32384637..32387739
	humano	NM_005564.3	840	Chr: 9 NC_000009.12 128149453..128153455

Tabla 1. *Secuencias de ARNm de las lipocalinas estudiadas que contienen la secuencia codificante que se corresponde con la secuencia canónica, obtenidas de RefSeq (NCBI).*

Lipocalina	Especie	NCBI RefSeq	Long ARNm (pb)	Referencia genómica NCBI
LCN8	ratón	NM_033145.1	653	Chr: 2 NC_000068.7 25653118..25656217
	humano	NM_178469.3	918	Chr: 9(-) NC_000009.12 109107983..109111935
OBP2A	ratón	NM_153558.1	738	Chr: 2 NC_000068.7 25697538..25703332
	humano	NM_001293189.1	754	Chr: 9 NC_000009.12 135546139..135549969
C8G	ratón	NM_027062.2	1044	Chr: 2(-) NC_000068.7 25498650..25501719
	humano	NM_000606.2	888	Chr: 9 NC_000009.12 136944885..136946974
LCN12	ratón	NM_029958.1	745	Chr: 2(-) NC_000068.7 25490845..25495883
	humano	NM_178536.3	694	Chr: 9 NC_000009.12 136949580..136955497
ORM2	ratón	NM_011016.2	774	Chr: 4 NC_000070.6 63362449..63365877)
	humano	NM_000608.2	844	Chr: 9 NC_000009.12 114329789..114333256)

Tabla 1 (continuación). *Secuencias de ARNm de las lipocalinas estudiadas que contienen la secuencia codificante que se corresponde con la secuencia canónica, obtenidas de RefSeq (NCBI).*

2.2.- Análisis de parámetros básicos de las secuencias de UTRs 5´y 3´de lipocalinas

Las secuencias UTRs de las lipocalinas fueron analizadas mediante **“** infoseq **”** de EMBOSS (http://emboss.bioinformatics.nl/) [3], para determinar su longitud y contenido en G+C. Dos muestras de UTRs de humano y ratón, tomadas aleatoriamente de la base de datos UTRdb (con 1000 secuencias de cada especie), fueron utilizadas para contrastar con las UTRs de lipocalinas. Estas muestras fueron igualmente analizadas con **“infoseq”** para obtener su longitud y contenido en G+C.

Para determinar la presencia de secuencias repetitivas en las UTRs se utilizó Repeatmasker (A.F.A. Smit, R. Hubley & P. Green RepeatMasker at http://repeatmasker.org). Los análisis estadísticos y representaciones gráficas se realizaron mediante la hoja de cálculo de OpenOffice y el programa estadístico PAST [4].

2.3.- Determinación de la organización genómica y de los procesos responsables del origen de UTRs 5´ alternativos

Solo se ha abordado esta cuestión para las regiones UTRs 5´de las lipocalinas, dado que son estas regiones las que muestran una variabilidad considerable y con una mayor complejidad a la hora de ser interpretada. Mientras que las regiones UTRs 3´muestran escasa variabilidad y cuando la presentan, esta es de fácil interpretación, ya que esta esta asociada a sitios de corte alternativo y no a diferentes exones que sufran splicing alternativo.

La organización genómica de la región UTR 5´de las lipocalinas se ha deducido a partir de la composición de exones de las UTRs 5´alternativas que fueron seleccionadas de las bases de datos, según los criterios establecidos en el punto 2.1. Para ello se recurrió a las anotaciones de Aceview y además de forma complementaria se utilizo “ESIM4” de EMBOSS [3] que permite alinear los transcritos de los UTRs 5´alternativos frente a la región genómica correspondiente. De esta forma pudo obtenerse la estrcutura exón-intrón de dichas UTRs 5´.

Los mecanismos más frecuentes que originan UTRs 5´alternativas son: orígenes de transcripción alternativos, fenómenos de splicing alternativo o una combinación de ambos fenómenos. Para las UTRs 5´de lipocalinas estos mecanismos se han tratado de establecer buscando la explicación más lógica, dentro de las diferentes posibilidades y tratando de encontrar respaldo en una serie de

predicciones que se llevaron a cabo.

2.3.1.- Predicción de regiones promotoras

Se utilizó "Neural Network Promoter Prediction" (NNPP, http://www.fruitfly.org/seq_tools/promoter.html) [5]. Se analizaron las regiones genómicas de las UTRs 5´previamente determinadas más 2000 nucleótidos corriente arriba, para poder detectar posibles promotores alternativos.

2.3.2. -Predicción de exones

Se analizaron las regiones genómicas de las UTRs 5´de lipocalinas con el programa "ExonScan" (http://genes.mit.edu/exonscan/) [6]. Dicho programa hace una predicción de exones combinando la información de diferentes elementos: sitios donadores-aceptores de splicing, los posibles sitios potenciadores (ESE) o silenciadores exónicos (ESS), así como de secuencias GGG intrónicas. Esta predicción de tipo holístico puede darnos una idea más aproximada de la auténtica organización genómica de las regiones UTRs 5´.

2.3.3. -Predicción de sitios donadores-aceptores de splicing

Para predecir estos sitios se recurrió a la herramienta bioinformática "ASSP" (http://wangcomputing.com/assp/index.html) [7], la cual permite hacer predicciones de sitios de splicing, ofreciendo en la predicción la probabilidad de que dichos sitios sean constitutivos, alternativos o sitios crípticos. Este tipo de predicciones es útil a la hora de confirmar el carácter constitutivo o alternativo de los exones que componen las UTRs 5´.

3. - Resultados

3.1.- Transcritos seleccionados

A continuación se presentan los transcritos seleccionados, como se detalla en métodos, para las distintas lipocalinas estudiadas de humano y ratón. Se indica la longitud de los ARNm maduros y de las UTRs 5´y 3´correspondientes.

Tabla 2.APOD				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	a	931	140	221
	b	1016	224	222
	c	2077	358	1149
	d	1074	281	223
	e	1094	214	221
Humano (2007)	a	1129	361	198
	b	680	232	198
	c	553	135	-
	d (2010)	1210	190	-

Tabla 3. PTGDS				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	c	785	80	159
	d	1058	329	159
	e	1326	-	614
Humano (2007)	c	961	72	214
	g	1212	458	178
	j	2495	1283	639

Tabla 4. RBP4				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	a	1233	385	252
	c	765	61	128*
	d	947	89	252
Humano (2010)	b	1316	322	388
	d	563	72	186*
	d(2)	522*	113	186*

Tabla 5. APOM				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	a	1471	781	117
Humano (2010)	d	1183	496	120
	d(2)	760	73	120

Tabla 6. LCN1				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Rata	a	754	55	165
Humano (2010)	b	881	157	185
	h	760	49	185

Tabla 7. LCN2				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	b	894	54	237
Humano (2010)	b	822	72	153
	b(2)	903	72	334

Tabla 8. LCN8				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	a	658	23	107
Humano (2010)	e	822	348	112

Tabla 9. OBP2A				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	b	749	53	165
Humano (2010)	a	689	42	134

Tabla 10. C8G				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	c	1044	275	160
Humano (2010)	a	877	75	193

Tabla 11. LCN12				
	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
Ratón	a	715	55	78
Humano (2010)	c	930	248	103
	c(2)	710	28	103

Tabla 12. ORM2				
Ratón	Variantes(Aceview)	Long ARNm (pb)	Long UTR 5´(pb)	Long UTR 3´(pb)
	a	778	41	113
Humano (2010)	b	917	189	122

Tablas 2 a 12. *Transcritos seleccionados de las diferentes lipocalinas, extraidas de Aceview / ASPicDB. Las variantes se denominan mediante una letra, tal como aparecen identificadas en Aceview. Dado que Aceview ofrece, para humano, dos versiones (2007 o 2010) de la base de datos, se indica en las tablas la versión a la que corresponde la variante seleccionada.*

3.2.- Longitud y composición de las UTRs de lipocalinas

En la tabla 13 se representa la longitud media, máxima y mínima, así como el % G+C medio de las UTRs 5´ de las lipocalinas humanas y de ratón analizadas. A efectos de comparación se han añadido los mismos parámetros pertenecientes al conjunto de UTRs 5´ de la base de datos "UTRdb"[8], para ambas especies. Observamos en las UTRs 5´de lipocalinas (tabla 13) un buen ajuste a los valores medios de longitud y % G+C esperados, aunque en la longitud máxima, se observan valores algo inferiores, tanto en humano como en ratón. Este último resultado es esperable ya que por el gran tamaño de la muestra de la base de datos (UTRdb) de UTRs 5´de humano y roedor, es lógico que aparezcan valores extremos que son de baja frecuencia.

UTR 5´				
	Long media	**Long máx**	**Long mín**	**%GC medio**
Lipoca_hum	232.18	1283	28	63.53
UTRdb_hum	*210.20*	*2803*	*18*	*60*
Lipoc_ratón	186.89	786	23	57.60
UTRdb_roed	*186.30*	*936*	*20*	*60*

Tabla13: *Longitud media, máxima y % G+C medio de las UTR 5´de lipocalinas de humano y ratón y comparación con los valores medios obtenidos de la base de datos UTRdb para las UTR 5´ de dichas especies*

En la tabla 14 se representan los mismos datos que en la tabla 1 pero para las UTR 3´. En este caso

observamos que la longitud media de las UTR 3´de lipocalinas es sensiblemente menor que el valor promedio del conjunto de las UTRs 3´, tanto para humano como para ratón. Comprobamos además que los valores de %G+C en las UTRs 3´de lipocalinas son, por el contrario, sensiblemente mayores de lo que cabría esperar para las dos especies, especialmente en humano.

	UTR 3´			
	Long media	**Long máx**	**Long mín**	**%GC medio**
Lipoca_hum	203.78	639	103	57.24
UTRdb_hum	*1027.70*	*8555*	*21*	*45*
Lipoc_ratón	251.29	1149	78	50.12
UTRdb_roed	*607.30*	*3354*	*19*	*45*

Tabla 14: *Longitud media, máxima y % G+C medio de las UTR 3´de lipocalinas de humano y ratón y comparación con los valores medios obtenidos de la base de datos UTRdb para las UTR 3´ de dichas especies*

Para analizar con más detalle esta aparente anomalía se procedió a comparar la distribución de los datos de longitud y % G+C de las UTRs 3´ de lipocalinas de humano y de ratón frente a los de una muestra representativa de secuencias de UTRs 3´de estas mismas especies extraída de UTRdb. En la figura 1 y 2 se observan los gráficos Box-plot de la longitud y del % G+C de las UTRs 3´para las dos especies.

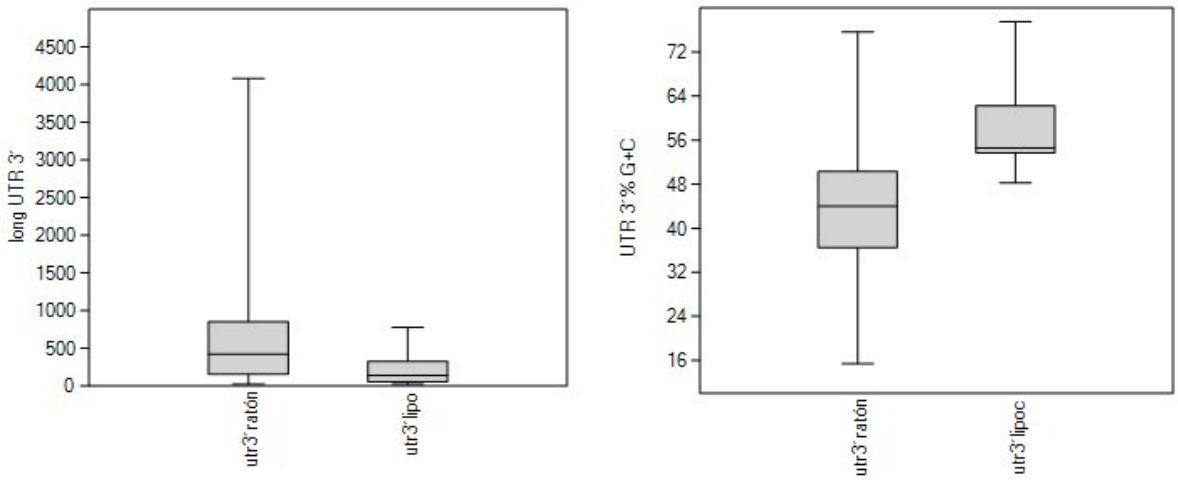


Figura 1. *Representaciones Box-plot de la longitud y el % G+C de las UTRs 3´de lipocalinas de ratón frente a una muestra de UTRs 3´ de la misma especie de la base de datos UTRdb. Las cajas representan los cuartiles (25% a 75%), la línea horizontal representa la media y los segmentos superior e inferior los valores máximos y mínimo.*

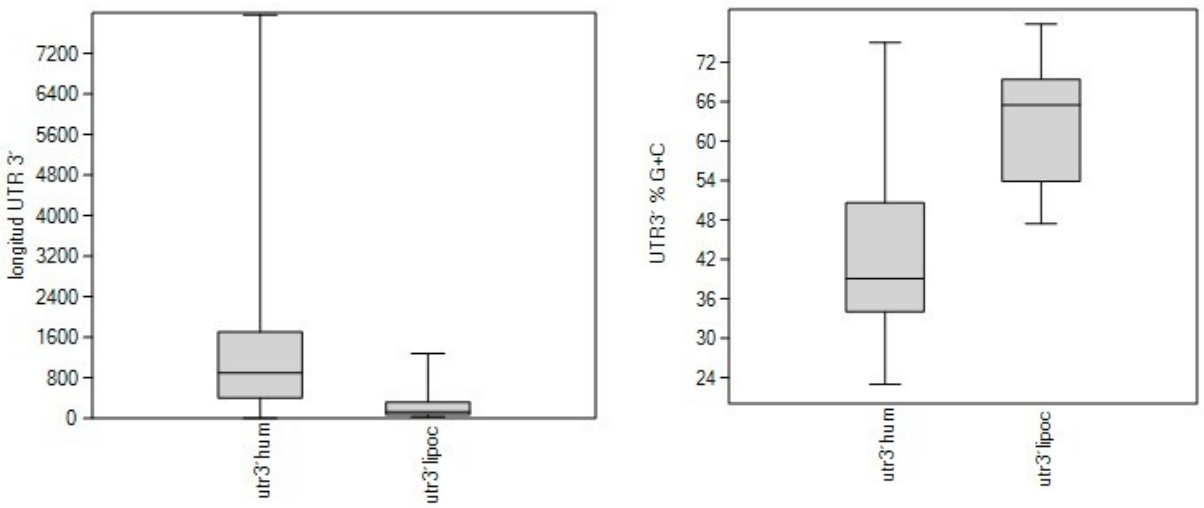


Figura 2. *Representaciones Box-plot de la longitud y el % G+C de las UTRs 3´de lipocalinas de humano frente a una muestra de UTRs 3´ de la misma especie de la base de datos UTRdb. Las cajas representan los cuartiles (25% a 75%), la línea horizontal representa la media y los segmentos superior e inferior los valores máximos y mínimo.*

Observamos en las figuras 1 y 2 que, si bien la distribución de longitudes de las UTRs 3´de lipocalinas de humano y ratón se encuentra en un intervalo inferior del rango de variación del conjunto de las UTRs 3´para cada especie, es respecto al % G+C de las lipocalinas donde se observan de forma más evidente las diferencias.

La explicación de estas diferencias entre las UTRs 3´de las lipocalinas y el conjunto de UTRs 3´ podría encontrarse en una adaptación de estas a la región genómica donde se encuentran, ya que es conocido que los genes situados en isocoras ricas en G+C tienden a ser más cortos que los que se encuentran en isocoras pobres en G+C [9] y por lo tanto también tendrán UTRs más cortas. Si esto fuese así debería afectar igualmente a la región UTR 5´, cosa que no ocurre. Este hecho será tratado de forma más detallada en la discusión de este capítulo.

Por otra parte es conocido que las longitudes de las UTRs de humano son en promedio mayores que las de otros mamíferos [10]. En las gráficas 3 y 4 se han representado de forma conjunta las longitudes de las UTRs de lipocalinas humanas y de ratón, 5´y 3´ respectivamente. La observación de estas gráficas nos muestra que la longitud de la UTRs 5´de humano es claramente superior que la correspondiente en ratón en 5 casos, siendo sólo mayor la UTR 5´de ratón en 2 casos y de longitud semejante entre las dos especies en 4 casos. Mientras que para la UTR 3´no se observa la tendencia

a que dicha región sea mayor en humano, ya que salvo en un caso (ApoD) el resto de lipocalinas muestran longitudes semejantes para las dos especies.

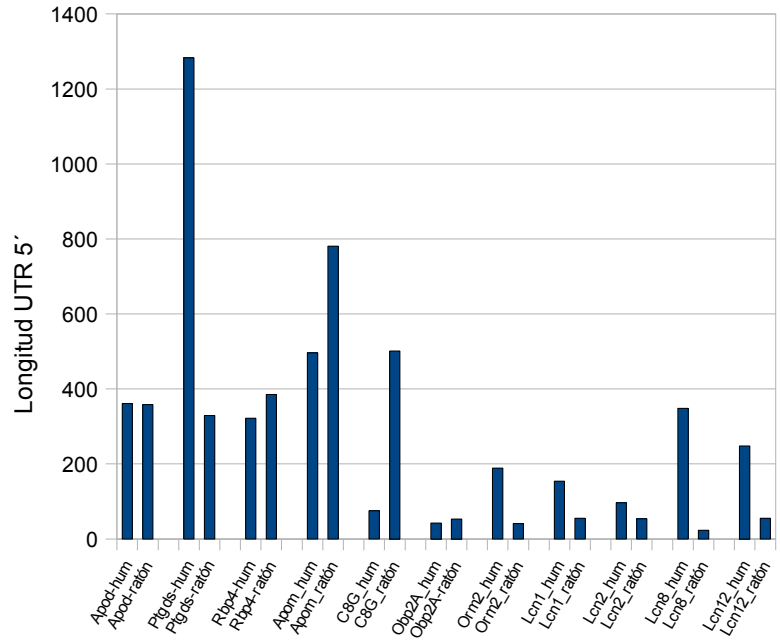


Figura 3. *Longitud de las UTRs 5´ de las diferentes lipocalinas humanas y de ratón. En el caso existir UTRs alternativas se ha incluido la UTR 5´de mayor longitud*

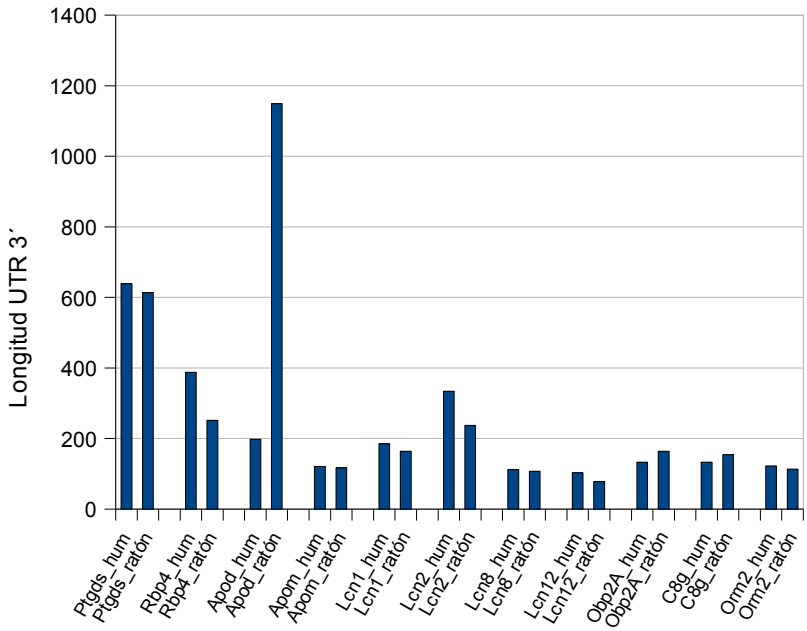


Figura 4. *Longitud de las UTRs 3´ de lipocalinas humanas y de ratón que no presentan variación o que la muestran sólo en una de las dos especies, en este caso se muestra la longitud máxima de las alternativas posibles.*

3.3.- Elementos repetitivos en las UTRs de lipocalinas

Un aspecto interesante que puede estudiarse en estas regiones genómicas es la presencia de elementos repetitivos. Diversos elementos de esta clase (STR, LINE, SINE y LTR) se encuentran con cierta frecuencia en las UTRs de los ARNm de eucariotas y además han mostrado desempeñar un papel funcional en determinados casos [10]. Análisis realizados sobre el conjunto de UTRs de mamíferos han puesto de manifiesto que diversos elementos repetitivos son más frecuentes en humanos que en roedores y que en otras especies de mamíferos [10]. Así mismo se ha encontrado que estos elementos son más abundantes en la región UTR 3´que en la 5´, posiblemente debido a la mayor longitud que suele mostrar la primera [10].

En las tablas 15 y 16 se muestran las secuencias repetitivas que pudieron identificarse mediante Repeatmasker en las UTRs 5´y 3´de lipocalinas humanas y de ratón respectivamente. Los elementos más frecuentes son "SINE/ALU" y "STR", que son también los elementos repetitivos más frecuentes encontrados en las UTRs de mamíferos [10].

La observación de los datos de las tablas 15 y 16 nos indican que estos elementos parecen ser ligeramente más abundantes en las lipocalinas humanas que en las de ratón, para el caso de las UTRs 5´, no obrservándose lo mismo para las UTRs 3´. Por último dichos elementos no muestran mayor abundancia en las UTRs 3´que en las UTRs 5´, como sería de esperar. El hecho de que las longitudes de las UTRs 3´de lipocalinas sean por lo general menores que la longitud promedio del conjunto de UTRs 3´de mamíferos, podría ser la causa de esta diferencia.

Humano	UTR 5´	UTR 3´
APO-D	- exon 1: STR (CA)n, 9 a 43 - exon 3: SINE/ALU, 1 a 73 (*)	--
LCN12	- exón 2: SINE/ALU, 1 a 122 (*)	-exón 2: STR (CA)n, 138 a 300
OMR2	- exón 1: LTR/ERVL-MaLR, 1 a 50	--

Tabla 15. *Elementos repetitivos identificados mediante Repeatmasker en las UTRs de lipocalinas humanas. Se indica el exón de la UTR donde aparece dicho elemento, así como la posición que ocupa en el mismo. El asterisco (*) indica que la repetición ocupa la longitud total del exón.*

Ratón	UTR 5´	UTR 3´
APO-D	--	-exón 1: SINE/ALU, 708 a 736 -exón 1: STR (TTTG)n , 708 a 736 -exón 1: LC (AT_rich), 1123 a 1146
APO-M	-exón 1: SINE/ALU, 80 a 205 -exón 1: STR(GC)n , 206 a 228 -exón 1: STR (A)n, 340 a 366	--

Tabla 16. *Elementos repetitivos identificados mediante Repeatmasker en las UTRs de lipocalinas de ratón. Se indica el exón de la UTR donde aparece así como la posición que ocupan en el mismos.*

3.4.- Existencia de UTRs alternativas en lipocalinas

El análisis de los datos de las UTRs de lipocalinas muestra que existe cierto grado de variabilidad en la expresión de las mismas. Observamos, en la figura 5, que siete de las once lipocalinas estudiadas presentan UTRs 5´alternativas. Como puede comprobarse en dicha figura, el número de UTRs 5´ alternativos es mayor en las lipocalinas evolutivamente más antiguas (Apod-D, Ptgds, Rbp4), presentando la más ancestral (Apo-D) el mayor nivel de variabilidad en su UTR 5´.

Respecto a las UTRs 3´(figura 6) hay un menor número de lipocalinas (4 de las 11) que presentan formas alternativas, mostrando un número de variantes por lipocalina, por lo general menor, que en las correspondientes UTRs 5´. Así mismo comprobamos que se repite el patrón en el que las lipocalinas más ancestrales son las que muestran mayor variabilidad en esta región UTR 3´.

Por último podemos añadir que la observación de las gráficas 5 y 6 nos muestra una tendencia, que nos permite comentar que el número de UTRs alternativas es algo mayor en humano que en ratón.

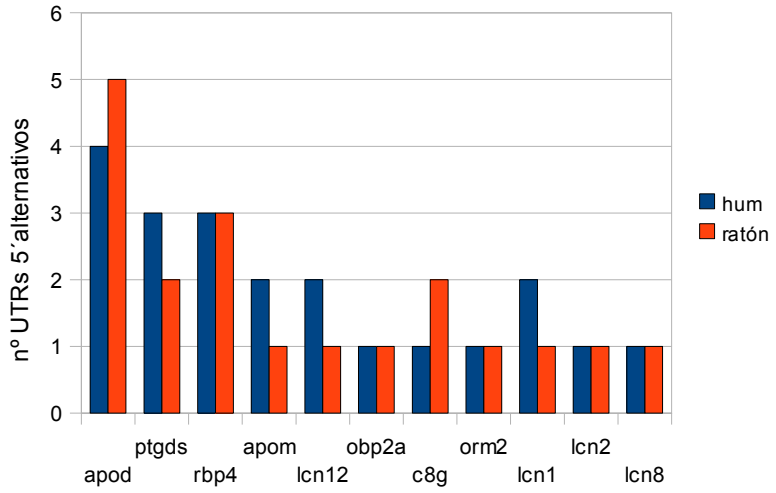


Figura 5: . *Número de UTRs 5´alternativos para las diferentes lipocalinas. Se muestran los datos de humano (hum) junto a los de ratón para cada caso*

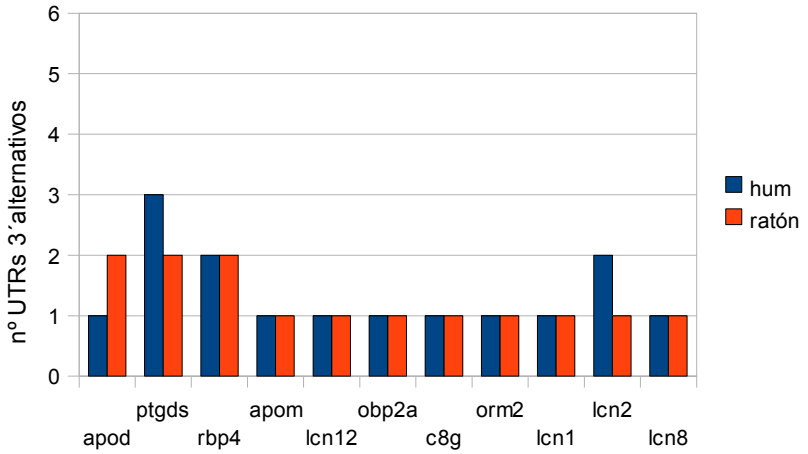


Figura 6: . *Número de UTRs 3´alternativos para las diferentes lipocalinas. Se muestran los datos de humano (hum) junto a los de ratón para cada caso*

En las figuras 7 y 8 se han representado las longitudes de las UTRs 5´ y 3´de lipocalinas ortólogas de humano y ratón, que muestran variabilidad simultáneamente en las UTRs de ambas especies. En estas gráficas sólo se ha representado la longitud máxima y mínima de dichas UTRs alternativas.

La única diferencia importante observable entre las dos especies es en el valor máximo de la UTR 5´de Ptgds.

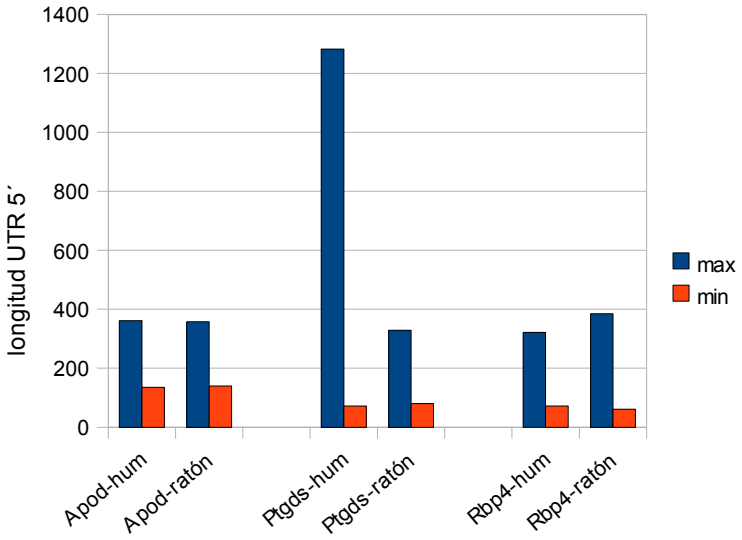


Figura 7. *Longitud de las UTRs 5´de lipocalinas ortólogas entre humano y ratón, que presentan variantes en las dos especies. Se muestra la longitud máxima y mínima de las alternativas existentes para cada especie.*

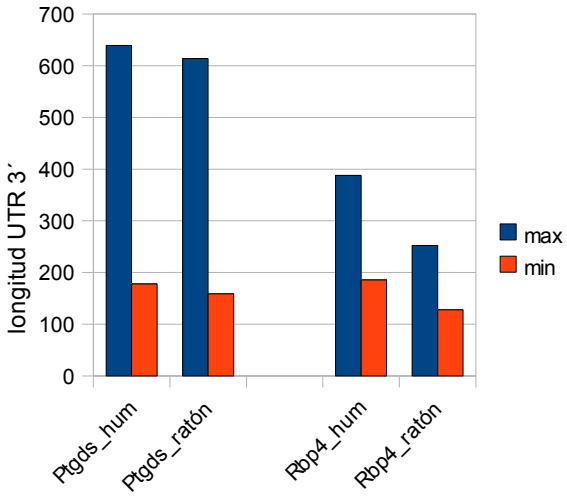


Figura 8. *Longitud de las UTRs 3´de lipocalinas ortólogas entre humano y ratón, que presentan variantes en las dos especies. Se muestra la longitud máxima y mínima de las alternativas existentes para cada especie.*

El número de exones que constituyen las regiones UTRs 5´y 3´ de las lipocalinas aparece representado en las figuras 9 y 10 respectivamente. Observamos que para las UTRs 5´ el número de exones, como es lógico, es mayor en las lipocalinas con mayor número de UTRs 5´alternativas. Algunas lipocalinas como Apom y Lcn1 solo presentan un exón en su UTR 5´ y sin embargo muestran variabilidad en dicha UTR. Hemos de suponer como explicación la existencia de sitios de inicio de la transcripción alternativos. Para el resto de las lipocalinas que muestran diversidad en las UTRs 5´, su origen podría explicarse por una combinación de sitios de inicio alternativos y mecanismos de splicing alternativo. Esta cuestión se trata ampliamente en el siguiente apartado de esta tesis.

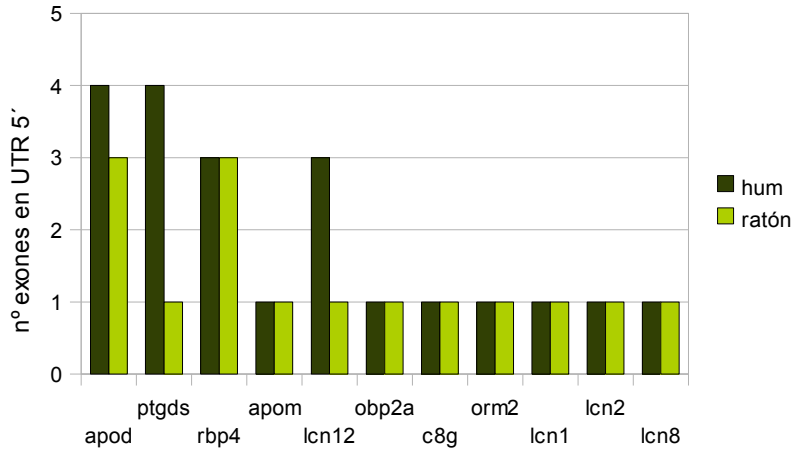


Figura 9: *Número de exones presentes en las UTRs 5´ de las diferentes lipocalinas. Se muestran datos de humano junto a los de ratón para cada caso.*

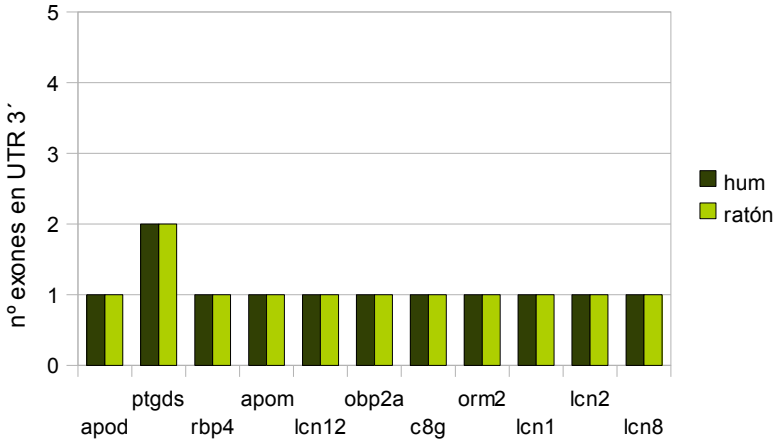


Figura 10: *Número de exones presentes en las UTR 3´de las diferentes lipocalinas. Se muestran datos de humano junto a los de ratón para cada caso.*

Respecto al origen de la diversidad de UTRs 3´, salvo para la lipocalina Ptgds que presenta 2 exones en esta región (ver figura 10) y esto permite la posibilidad de splicing alternativo, para el resto de lipocalinas que presentan diversidad en esta región el mecanismo de corte alternativo (debido a la existencia de señales de poliadenilación alternativas) es la causa más probable.

3. 5.- Organización genómica de las UTRs 5´ de lipocalinas que presentan formas alternativas y procesos responsables de su formación

3.5. 1.- ApoD

UTR 5´ de ApoD en ratón

A partir de la información que aportan los distintos transcritos alternativos de ApoD, deducimos que la organización genómica de la UTR-5´ de este gen quedaría como se observa en la figura 11. Existen 3 exones principales, que se extienden en una región de unos 3700 pb, estando el último "exón" del UTR-5´ transcrito en un exón común junto con el primer exón de la secuencia codificante.

Para poder explicar el origen de las UTRs 5´alternativas de esta lipocalina (A, B, C, D, E, ver en figura 11) es necesario apelar a la existencia de al menos tres promotores alternativos. Así habría tres inicios de transcripción (flechas anguladas en figura11); uno que daría lugar al transcrito E, otro que daría lugar a los transcritos A y B y un tercero que daría lugar a los transcritos C y D.

Con el objeto de determinar la presencia de estos hipotéticos promotores se analizó la región genómica completa de 3700 nt del UTR-5´ más 2000 nt adicionales corriente arriba mediante NNPP ("Neural Network Promoter Prediction"). La salida del programa se muestra en la tabla 17.

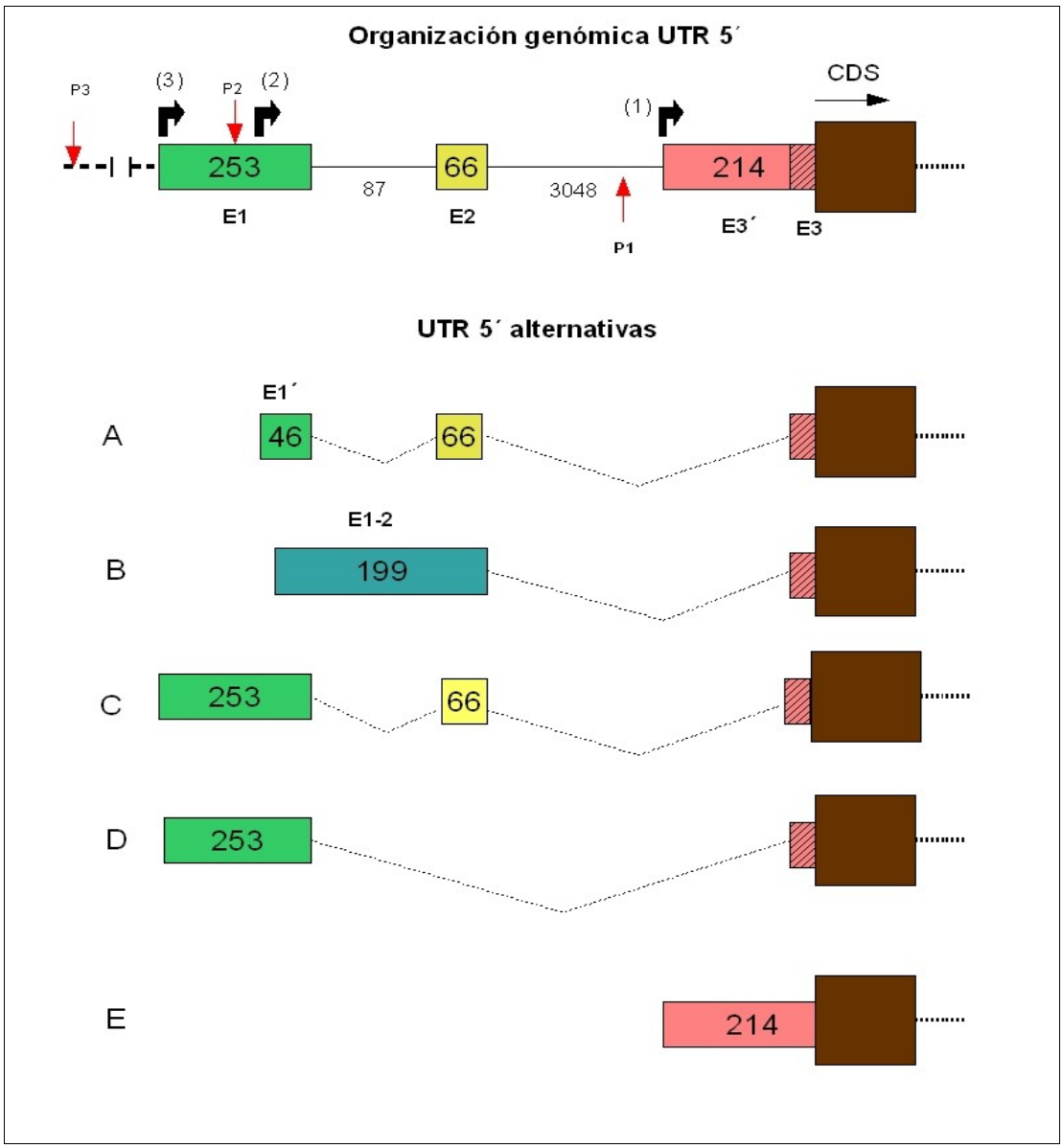


Figura 11. *Organización genómica de la UTR 5´de Apo-D de ratón (arriba) y diferentes UTR 5´alternativos expresados (abajo). P1, P2 y P3 (flechas rojas) indican la posición de los promotores alternativos. Las flechas negras y anguladas indican los inicios de transcripción alternativos. Se indica con números el tamaño de exones e intrones de esta región UTR 5´. En marrón el primer exón de la secuencia codificante, resto de colores exones de la UTR 5´.*

Neural Network Promoter Prediction NNPP version 2.2

Start	End	Score	Promoter Sequence
180	230	0.96	cacactgaactttaaagggtgggtggcatagagtatatgcattctactgc....**P3**
2169	2219	0.96	tgggagcctataaagtgacttgggagaagccacacacctcacttggagga....**P2**
4318	4368	0.93	ggcaaggaggcacaaaaggggaacagagggaaggcaagtcagggagaaag....**P1**?
5169	5219	0.85	aatggcaaggataaaatgtgaggctctggggagccttccctgatggacac....**P1**?

Tabla17. *Promotores predichos por NNPP v 2.2 para la región genómica UTR 5´(+ 2000 nt corriente arriba) de Apo-D de ratón.*

En esta predicción aparece una posible región promotora que se encuentra dentro del E1 (indicado como P2 en la tabla 1), cuyo inicio de transcripción coincide con el inicio del primer exón de los transcritos A y B. Respecto a los transcritos C y D podrían originarse a partir del inicio de transcripción dependiente de una región promotora alternativa corriente arriba, tal como la predicha por NNPP (indicada como P3). Finalmente para explicar la formación del transcrito E sería necesaria la presencia de una región promotora adicional. Como observamos en la predicción de NNPP hay dos potenciales promotores (P1), dentro del segundo intrón de la región UTR-5´, que podrían explicar el origen de transcripción del exón E3´ de este transcrito.

Además de las evidencias de inicios de transcripción alternativos hemos de encontrar evidencias de los mecanismos de splicing necesarios para explicar el origen de las diversas variantes de la UTR 5 ´. Así en la formación del transcrito B intervendría la retención del primer intrón, que daría lugar al exón que llamamos "E1-2". Respecto a la formación del transcrito D, este se produciría por la omisión (exon skipping) del E2.

Los resultados de la predicción de exones de esta región genómica con ExoScan pueden verse en la tabla 18. Observamos que el programa predice la expresión de los exones E1, E2 y E3 pero no la del exón E1-2, exón que resultaría de la retención del primer intrón. Esto apoya la idea de que el E1-2 sería un exón alternativo que se expresaría en unas condiciones específicas, en las que el espliceosoma no eliminaría el correspondiente intrón. El hecho de que el intrón a retener sea corto (tan solo 87 nucleótidos) es un factor favorable a que pueda sufrir retención por el espliceosoma [11]. Además ExonScan predice la presencia de un exón adicional desconocido (indicado con "?" en tabla 2) que estaría situado entre los exones E2 y E3.

ExonScan results page

Predicted exons:

Begin	-	End	(3'ss	5'ss	ESE	ESS	GGG	total)	Exón de 5´UTR
27	-	253	(106	102	9	-21	9	205)	.....**Exón1**
337	-	406	(88	110	11	0	18	227)	.....**Exón2**
3253	-	3446	(77	63	47	-2	9	194)	..... **?**
3640	-	3790	(97	91	18	6	3	215)	.....**Exón 3 + CDS**

Tabla 18. *Resultados de ExonScan para la región genómica de la UTR 5´ de Apo-D de ratón*

Para explicar la omisión (exon skipping) del E2 del transcrito D, dado que dicho exón se expresaría

según la predicción de ExonScan, hemos de suponer la presencia de señales inhibidoras desconocidas o la presencia de los adecuados factores de la maquinaria del espliceosoma que omitiría dicho exón de la UTR 5´de Apo-D de ratón. Las predicciones de ASSP sobre los sitios de splicing se muestran en la tabla 19. Esperaríamos que el sitio donador del primer exón (E1) fuese constitutivo, ya que se expresa en tres de los cinco transcritos, sin embargo aparece como inclasificable, quizás esto esté relacionado con los mecanismos que permiten la omisión del intrón contiguo (intrón 1) en el transcrito B. El sitio aceptor del tercer exón (E3) si aparece como constitutivo lo cual es de esperar en un exón que podemos clasificar de constitutivo ya que aparece en cuatro de los cinco transcritos. Respecto al exón E2 que podemos considerar más alternativo, si muestra al menos su sitio aceptor como alternativo o críptico. Por último comprobamos que los sitios de splicing del exón adicional desconocido, predicho por ExonScan, son identificados por ASSP como sitios alternativos/crípticos.

Exón	Posición	Sitio splicing	Confidencia
E1	253	Donador sin clasificar	--
E2	341	Aceptor alternativo/criptico	0.127
E2	406	Donador constitutivo	0.605
?	3253	Aceptor alternativo/criptico	0.911
?	3446	Donador alternativo/criptico	0.414
E3	1463	Aceptor constitutivo	0.598

Tabla 19. *Predicciones de ASSP para los sitios de splicing de los exones de la UTR 5´de Apo-D de ratón*

UTR 5´ de ApoD en humano

A partir de la información de los distintos transcritos encontrados para ApoD humana, la organización genómica de la UTR 5´ de este gen quedaría como se observa en la figura 12. Existen 4 exones, que se extienden en una región de 4740 pb, estando el último "exón" de la UTR 5´ incluido en un exón conjunto con el primer exón de la secuencia codificante.

Podemos observar (Fig.12) que el primer exón (E1) en los diversos transcritos posee diferentes tamaños. Serían necesarios pues tres inicios de transcripción alternativos, por lo tanto tres promotores, para explicar estas diferencias. Además hemos de apelar a diferentes mecanismos de splicing alternativo para explicar la presencia de los exones E2 y E3 en los transcritos A y B respectivamente.

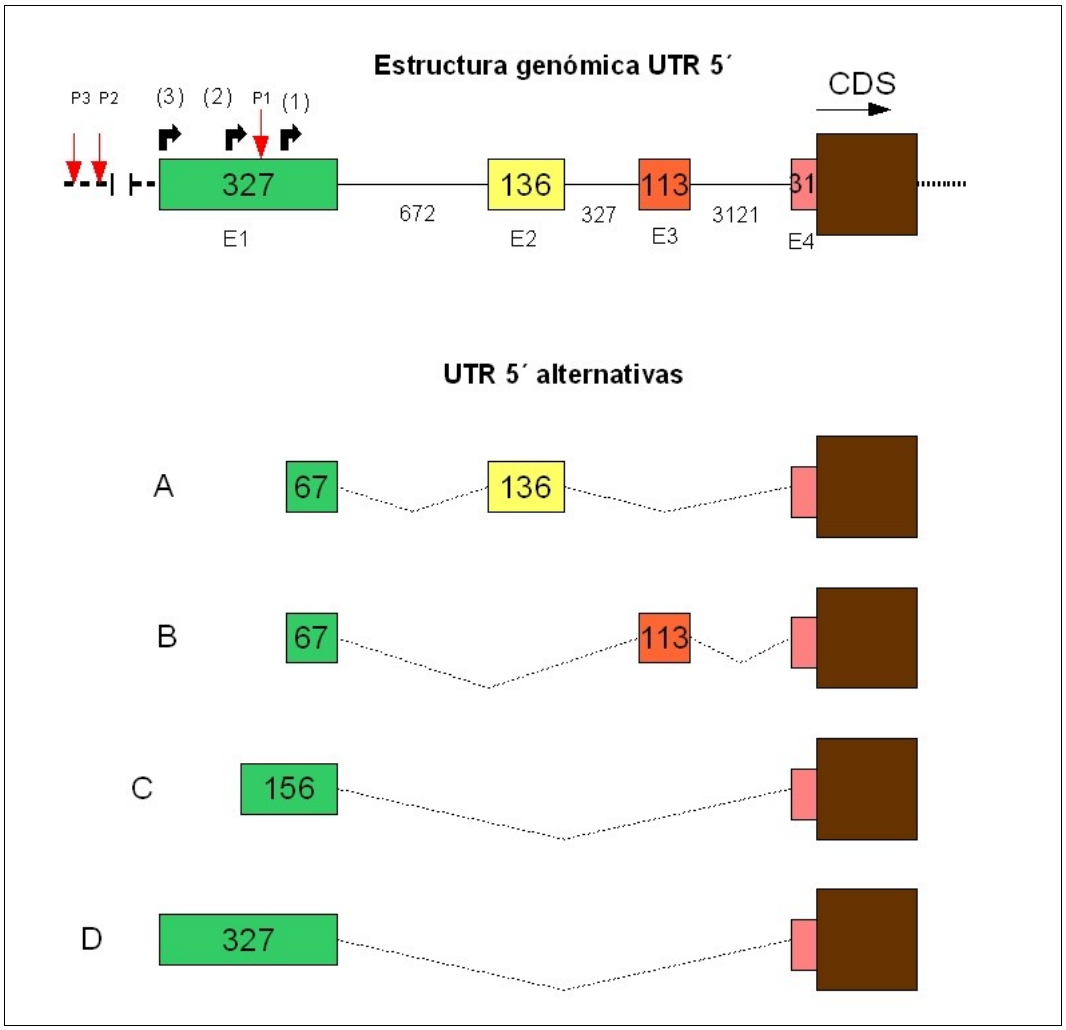


Figura 12. *Organización genómica de la UTR 5´de Apo-D humana (arriba) y diferentes UTR 5´alternativos expresados (abajo). P1, P2 y P3 (flechas rojas) indican la posición de los promotores alternativos. Las flechas negras y anguladas indican los inicios de transcripción alternativos. Se indica con números el tamaño de exones e intrones de esta región UTR 5´. En marrón el primer exón de la secuencia codificante, resto de colores exones de la UTR 5´.*

Se tomó la región genómica del UTR-5´más 2000 nt adicionales corriente arriba. Se analizó dicha región con NNPP y el programa encontró tres potenciales promotores en las regiones indicadas en la tabla 20.

Neural Network Promoter Prediction NNPP version 2.2

Start	End	Score	Promoter Sequence
937	987	0.98	caggataatg**tataaa**taaaggagagaatttcaggtaaagatataatgaa: P3
1529	1579	0.95	ctgcaagggt**cataaa**agggacagagaacagagcgacagaagatgtcttc: P2
2228	2278	0.98	aagaagct**tataaa**atagcttgggagaggccagtcaccaagacaggcat ......P1 *

Tabla 20. *Promotores predichos por NNPP v 2.2 para la región genómica UTR 5´(+ 2000 nt corriente arriba) de Apo-D humana.*

De manera que de ser correcta esta predicción, podríamos explicar la aparición de transcritos con un exón E1 de diferentes tamaños (A, B y C), al utilizarse diferentes orígenes de transcripción (ver Fig. 2).

Se da la circunstancia de que en el mismo exón E1 se encuentra la región promotora, que podemos considerar región promotora canónica para ApoD humana, la cual es correctamente predicha por NNPP (promotor 1, P1 en figura 2).

Por otra parte cada transcrito alternativo sufre un splicing diferente, de forma que los transcritos A y B incluyen un exón intermedio, E2 o E3, de forma excluyente y finalmente empalman con E4. Los otros transcritos C y D sufren un splicing que empalma directamente su primer exón con el E4. De esta situación podemos deducir que los exones E1 y E4 son constitutivos, ya que se expresan en todas las variantes, mientras que E2 y E3 son alternativos.

Se sometió la región genómica de UTR-5´de Apo-D humana al análisis con ExonScan, al igual que con ratón. Los resultados de su predicción pueden verse en la tabla 21. Observamos que predice los exones E1, E2 y E4, pero no E3. Además se predice la presencia de un posible exón adicional desconocido (indicado con "?" en tabla 21) que estaría situado entre los exones E1 y E2.

ExonScan results page

Predicted exons:

Begin -	End	(3'ss	5'ss	ESE	ESS	GGG	total)		
									Exón de 5´UTR
147 -	327	(82	98	43	-19	12	216)	---------	**Exón 1**
413 -	493	(83	102	22	-5	0	202)	---------	**?**
1003 -	1138	(58	91	36	4	12	201)	---------	**Exón 2**
4710 -	4789	(101	65	14	3	9	192)	---------	**Exón 4 + CDS**

Tabla 21. *Resultados de ExonScan para la región genómica de la UTR 5´ de Apo-D humana.*

Dado que la predicción no muestra a E3 como un exon probable hemos de suponer que dicho exón debe poseer sitios de splicing alternativos, de forma que solo en ciertas condiciones, con la presencia de los factores adecuados, tenga lugar su expresión en la UTR 5´. Respecto al E2, puesto que si parece poseer sitios de splicing fácilmente reconocibles, su falta de expresión es probablemente causada por señales inhibidoras, como los elementos silenciadores intrónicos (ISS, no considerados por ExonScan). Su presencia en la variante B sería posible en presencia de cierto factor que bloquee esta inhibición del spliceosoma.

Tras someter a la región genómica de la UTR 5´de Apo-D humana al análisis de predicción de sitios de splicing con ASSP se obtuvieron los resultados de la tabla 22.

Exón	Posición	Sitio splicing	Confidencia
E1	328	Donador constitutivo	0.847
E2	1000	Aceptor alternativo/criptico	0.884
E2	1136	Donador alternativo	0.329
E3	1463	Aceptor alternativo/criptico	0.883
E3	1580	Donador alternativo	0.542
E4	4700	Aceptor alternativo/criptico	0.804

Tabla 22. *Predicciones de ASSP para los sitios de splicing de los exones de la UTR 5´de Apo-D humana*

Los resultados muestran el sitio donador del primer exón, como es de esperar, es clasificado como constitutivo. Observamos por otra parte que los sitios aceptores/donadores de los exones E2 y E3 son clasificados de alternativos. Esto estaría en consonancia con que su inclusión o no pueda estar sometida a regulación en el splicing alternativo, como ya hemos comentado previamente.

3.5.2.- PTGDS

UTR 5´ de Ptgds en humano

La organización genómica de la UTR 5´de Ptgds de humano, tal como se deduce de las diferentes alternativas, es la que se muestra en la figura 3. Para poder explicar el origen de las diferentes alternativas (ver figura 13) son necesarios dos orígenes de transcripción. Uno daría lugar a la alternativa "C" y el otro a las alternativas "G" y "J".

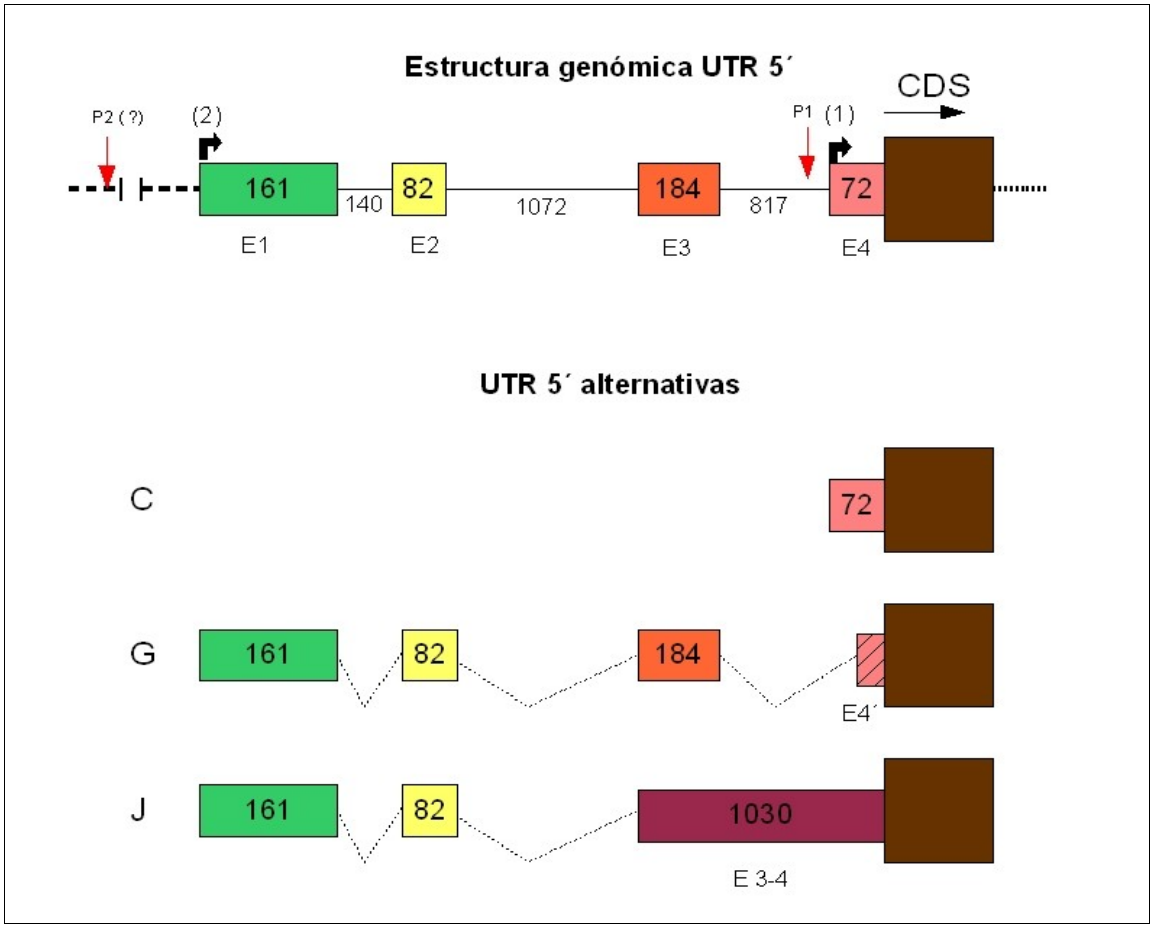


Figura 13. *Organización genómica de la UTR 5´de Ptgds humana (arriba) y diferentes UTR 5´alternativos expresados (abajo). P1, P2 (?) (flechas rojas) indican la posición de los promotores alternativos. Las flechas negras y anguladas indican los inicios de transcripción alternativos. Se indica con números el tamaño de exones e intrones de esta región UTR 5´. En marrón el primer exón de la secuencia codificante, resto de colores exones de la UTR 5´.*

La predicción de regiones promotoras con NNPP ofrece los resultados de la tabla 23.

Neural Network Promoter Prediction NNPP version 2.2

Start	End	Score	Promoter Sequence
2640	2690	0.89	ctgccccgggccacccgccaccacaccccagagcttgctcaccaccggga ….....P ?
4111	4161	0.91	cagcccctgccctatcagggccgctgggttggtgggccctgccagcagga…….....P2 ?
4257	4307	0.96	caggcctcttgcataagcgccctgtgggggacgaggctgcagctgtgcct…….....P2 ?
4371	4421	0.98	aagcctggcccataaataggggtctcctcagtgccctccgctcctcctgc…….....P1

Tabla 23. *Promotores predichos por NNPP v 2.2 para la región genómica UTR 5´(+ 2000 nt corriente arriba) de Ptgds humana.*

Se observa que hay varios posibles promotores. El promotor P1 (ver en tabla 23 y en figura 13) es el candidato más favorable para dar lugar al inicio de transcripción que forma la variante "C", ya que dicho inicio de transcripción coincide con el extremo 5´del exón E3 de la UTR 5´. El resto de promotores (indicados con "P ?", ver tabla 23) serían de función desconocida o falsos positivos. Los resultados muestran que en la región estudiada no aparece un promotor alternativo que pueda explicar el inicio de transcricpción que origina las variantes "G" y "J". Puede que este promotor se encuentre en una región corriente arriba del extremo 5´de la UTR 5´ mayor de 2000 nt, lo que no es descartable, ya que se han encontrado promotores alternativos en genes humanos a unos 100 Kb corriente arriba de la región codificante [12].

La predicción realizada por Exonscan, sobre la región UTR 5´de Ptgds humana, ofrece los resultados de la tabla 24 Como puede observarse solo predice el exón E2.

ExonScan results page

Predicted exons:

Begin - End (3'ss 5'ss ESE ESS GGG total)	
----------	**Exón de UTR 5´**
302 - 382 (63 66 30 6 27 192)---------	-------**E2**

Tabla 24. *Resultados de ExonScan para la región genómica de la UTR 5´ de Ptgds humana.*

Se repitió el análisis de esta región UTR 5´ con ExonScan, pero excluyendo la detección de elementos ESE y ESS y el resultado fué negativo, no se predijo ningún exón. De este resultado deducimos que los pares aceptores/donadores de splicing deben de ser débiles para los exones presentes esta región UTR 5´.

Se sometió al análisis con ASSP esta región genómica y los resultados se muestran en la tabla 25.

Exón	Posición	Sitio splicing	Confidencia
E1	162	Donador constitutivo	0.528
E2	301	No clasificado	0.000
E2	385	Donador alternativo	0.183
E3	1454	Aceptor alternativo	0.877
E4´	2455	-	-

Tabla 25. *Predicciones de ASSP para los sitios de splicing de los exones de la UTR 5´de Ptgds humana*

Comprobamos que, a excepción del sitio donador del primer exón, los sitios de splicing del resto de los exones son propuestos como sitios alternativos, no clasificados o incluso uno de estos sitios no es detectado (sitio aceptor de E4´). Estos resultados nos sugieren la existencia de señales reguladoras de splicing, que en presencia de los oportunos factores reguladores, podrían modificar la afinidad del spliceosoma por estos sitios y producir la combinación necesaria de exones que aparecen en las formas alternativas "G y J" de esta UTR 5´. En la variante "J" estaría teniendo lugar la retención del intrón número 3.

3.5.3.- RBP4

UTR 5´ de RBP4 en humano

La organización genómica de la UTR 5´de Rbp4 de humano se muestra en la figura 14. Para que puedan originarse estas tres variantes es necesario la existencia de tres promotores alternativos.La predicción de promotores de NNPP puede observarse en la tabla 26. Comprobamos que existen suficientes promotores en las posiciones esperadas para que se originen cada una de las variantes.

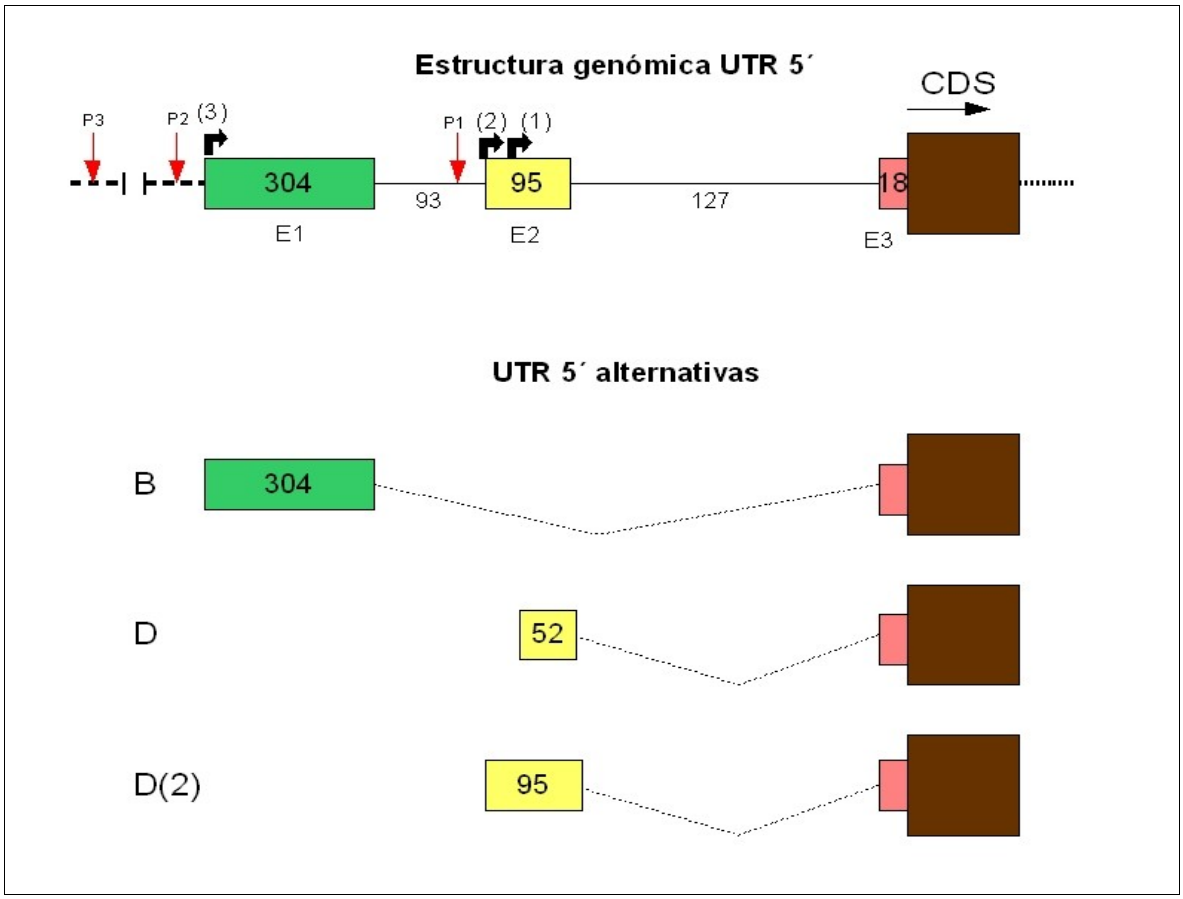


Figura 14. *Organización genómica de la UTR 5´de Rbp4 humana (arriba) y diferentes UTR 5´alternativos expresados (abajo). P1, P2 y P3 (flechas rojas) indican la posición de los promotores alternativos. Las flechas negras y anguladas indican los inicios de transcripción alternativos. Se indica con números el tamaño de exones e intrones de esta región UTR 5´. En marrón el primer exón de la secuencia codificante, resto de colores exones de la UTR 5´.*

En la tabla 26 se muestra, los que, por su localización, serían los candidatos más probables para el origen de las variantes (P1: variante D, P2 variante D(2) y P3 variante B).

Neural Network Promoter Prediction NNPP version 2.2

Start	End	Score	Promoter Sequence
893	943	0.87	ccagggtgcatagatatataccccatagggtcctgcaggagacgatctga...**P3?**
973	1023	0.99	cacagtcttctataaaactggcccaatcagaagatttcctagtcagcttg...**P3?**
1811	1861	0.83	tttctggagaatatttaacagggagggttttaacgcttttaaagatgttg...**P2**
2384	2434	0.97	acccccctccccccggcgctataaagcagcggggcggccgcggcgcgctcg...**P1?**
2393	2443	1.00	ccccggcgctataaagcagcggggcggccgcggcgcgctcgcctccctcg...**P1?**

Tabla 26. *Promotores predichos por NNPP v 2.2 para la región genómica UTR 5´(+ 2000 nt corriente arriba) de Apo-D de ratón.*

La predicción de exones de ExonScan se muestra en la tabla 27. Son predichos el primer y último exon de la UTR 5´de esta lipocalina. Dado que el exón E2 se expresaría como un primer exón, por la utilización de uno de los promotores alternativos de esta región, es lógico que ExonScan no pueda detectarlo ya que no dependería de la existencia de un sitio aceptor. De ser así esperaríamos que las predicciones de ASSP no detectasen ningún sitio aceptor en la posición requerida y este es el resultado que se obtiene al aplicarlas a la región genómica de la UTR 5´ (ver tabla 28). Por el contrario si es detectado el sitio donador de este exón, que es propuesto por ASSP como "donador constitutivo". Así mismo el sitio aceptor de E3 también es detectado y clasificado como aceptor constitutivo.

ExonScan results page

Predicted exons:

```
  Begin -   End (3'ss 5'ss ESE ESS GGG total)
-------------------------------------------- Exón de UTR 5´
   97  -    306  ( 83     109   0    -11  12     193 )..................E1

  620  -    752   ( 70      55  61      3  42     231)..................E3 + CDS
```

Tabla 27. *Resultados de ExonScan para la región genómica de la UTR 5´ de Rbp4 humana.*

Exón	Posición	Sitio splicing	Confidencia
E1	304	Donador constitutivo	0.846
E2	493	Donador constitutivo	0.287
E3	620	Aceptor constitutivo	0.919

Tabla 28. *Predicciones de ASSP para los sitios de splicing de los exones de la UTR 5´de Rbp4 humana*

Dado que el sitio aceptor del exón E2 parece ser débil, esto explicaría porqué en la variante B dicho exón es omitido por la maquinaria de splicing.

UTR 5´ de RBP4 en ratón

La región genómica de la UTR 5´ de esta lipocalina contiene al menos tres exones (ver figura 15). Para explicar el origen de las tres variantes (A, C y D) es necesario recurrir a la presencia de al menos tres promotores. Un primer promotor (ver figura 15) originaría la variante "C", un segundo, más corriente arriba, originaría la variante "D" y un tercero más corriente arriba aún daría lugar a la variante "A".

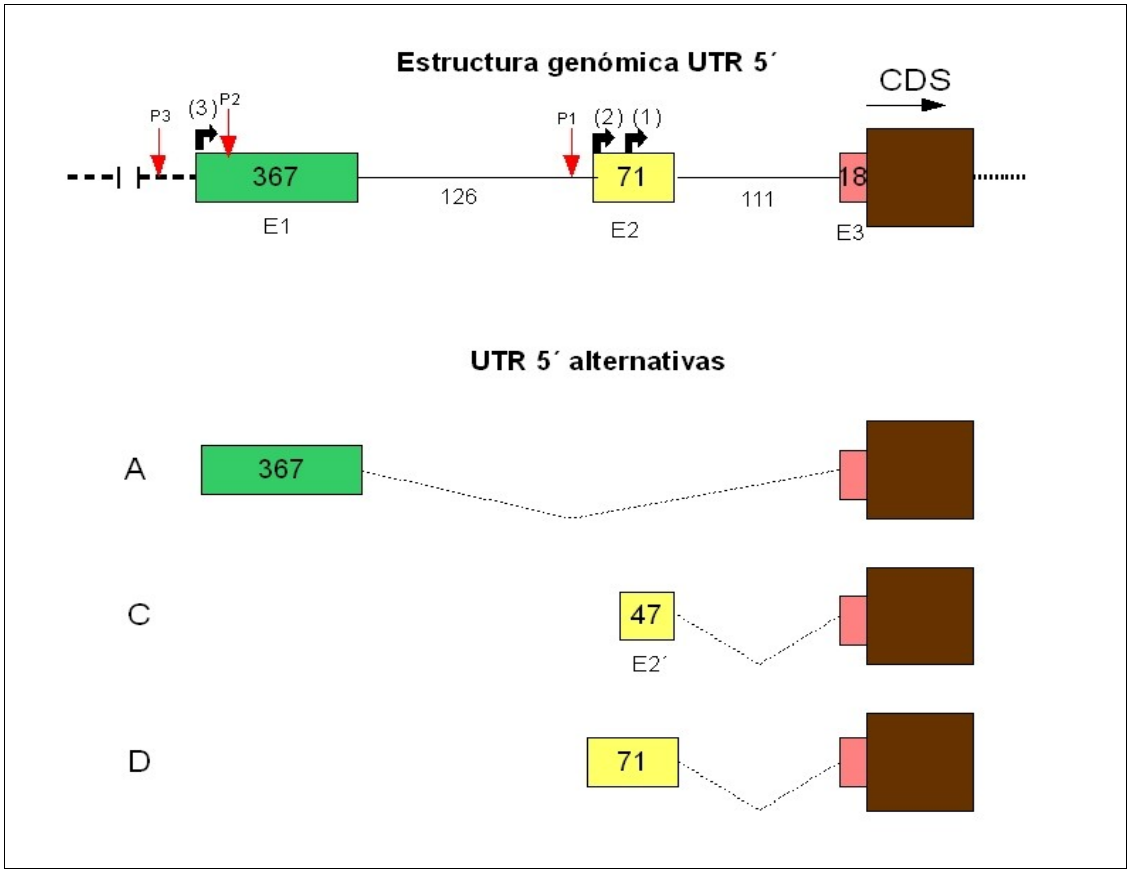


Figura 15. *Organización genómica de la UTR 5´de Rbp4 de ratón (arriba) y diferentes UTR 5´alternativos expresados (abajo). P1, P2 y P3 (flechas rojas) indican la posición de los promotores alternativos. Las flechas negras y anguladas indican los inicios de transcripción alternativos. Se indica con números el tamaño de exones e intrones de esta región UTR 5´. En marrón el primer exón de la secuencia codificante, resto de colores exones de la UTR 5´.*

La predicción de NNPP sobre esta región genómica más 2000 nt corriente arriba ofrece los resultados de la tabla 29. En dicha tabla se indican, según su localización, a que promotores darían lugar cada uno de ellos con mayor probabilidad (ver los correspondientes inicios de transcripción en figura 15)

Neural Network Promoter Prediction NNPP version 2.2

Start	End	Score	Promoter Sequence
1752	1802	0.80	ggcttagaaataaaaatgcatggtaaacacttggcaattatgtttttcag....**P3?**
1894	1944	0.83	tttctagagaatatttaacagggagcggtttagtccttctaaagatgatg....**P3?**
1945	1995	0.89	aatgaaagaataaatattgacccaaacagcaccacaactcatcaaagagt....**P3?**
2011	2061	0.84	caaaggggaaaaaaaaacagccaaaatatgccaaaaagcttctcacaac....**P2**
2487	2537	1.00	cccccggagctataaaggaccgacggccgctcggctccgtcgctccacgc....**P1**

Tabla 29. *Promotores predichos por NNPP v 2.2 para la región genómica UTR 5´(+ 2000 nt corriente arriba) de Rbp4 de ratón.*

La predicción de exones de esta región UTR 5´con ExonScan (ver tabla 30) solo puede detectar el exón E3 + CDS. La predicción de sitios de splicing de ASSP ofrece resultados semejantes a los de RBP4 humana (ver tabla 31). De tal forma que al ser el sitio aceptor de E2 débil esto explicaría su omisión en la variante A, tal como ocurre en humano.

ExonScan results page

Predicted exons:

Begin - End (3'ss 5'ss ESE ESS GGG total)

Exón de UTR 5´

675 - 761 (110 77 1 -5 36 219) ----------**E3+CDS**

Tabla 30. *Resultados de ExonScan para la región genómica de la UTR 5´ de Rbp4 de ratón*

Exón	Posición	Sitio splicing	Confidencia
E1	162	Donador alternativo	0,416
E2	564	Donador constitutivo	0,816
E3	674	Aceptor constitutivo	0,845

Tabla 31. *Predicciones de ASSP para los sitios de splicing de los exones de la UTR 5´de Rbp4 de ratón*

3.5.4.- LCN12

UTR 5´de Lcn12 Humana

La organización genómica de la UTR 5´de Lcn12 humana se muestra en la figura 16. Para explicar el origen de las variantes de esta UTR 5´es necesario que existan dos promotores. La predicción de NNPP (ver tabla 32) indica que hay tres posibles promotores en esta región genómica (UTR 5´+ 2000 nt corriente arriba). Dos de ellos están en las posiciones adecuadas para originar las dos UTR 5´alternativas (P1 daria lugar a la variante "B" y el P2 a la "C").

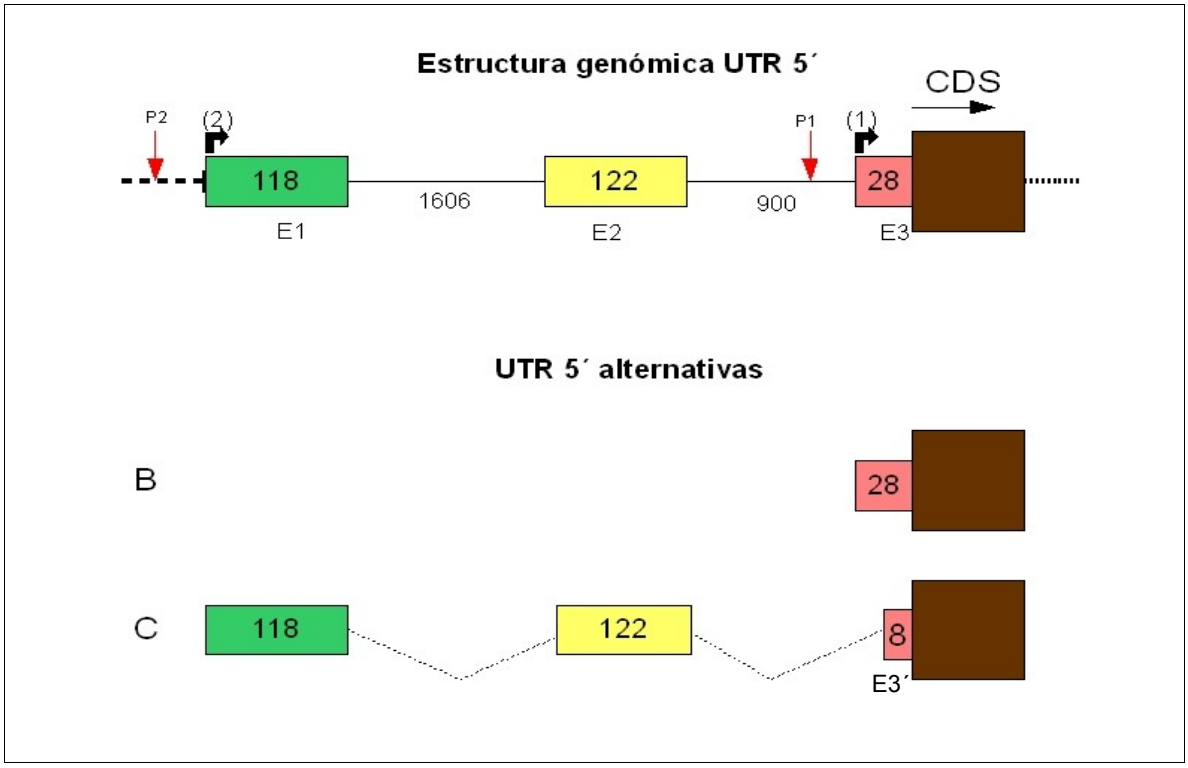


Figura 16. *Organización genómica de la UTR 5´de Lcn12 humana (arriba) y diferentes UTR 5´alternativos expresados (abajo). P1 y P2 (flechas rojas) indican la posición de los promotores alternativos. Las flechas negras y anguladas indican los inicios de transcripción alternativos. Se indica con números el tamaño de exones e intrones de esta región UTR 5´. En marrón el primer exón de la secuencia codificante, resto de colores exones de la UTR 5´.*

Neural Network Promoter Prediction NNPP version 2.2

Start	End	Score	Promoter Sequence
930	980	0.90	ggcttacgcctataatcccagcactttgggaggctgaggcgggtggatca...P2
3010	3060	0.81	agtgccctgcgataagacgggctccgggaggggtgcctgctgcgctgaga...P?
4091	4141	1.00	gagactgcgtttaaaatgccccagctgtcctcaccaggttgcctgggtgc...P1

Tabla 32. *Promotores predichos por NNPP v 2.2 para la región genómica UTR 5´(+ 2000 nt corriente arriba) de Lcn12 humana.*

La predicción de exones con ExonScan (ver tabla 33) no es capaz de detectar los exones de la UTR 5´esperados. El resultado del análisis revela sin embargo la posible presencia de exones alternativos desconocidos, en las posiciones que se indican en la tabla 33.

ExonScan results page

Predicted exons:

Begin	-	End	(3'ss	5'ss	ESE	ESS	GGG	total)	
317	-	448	(89	76	21	2	33	221)	.....exón al 3´de E1(en intrón 1 de UTR 5´)
2376	-	2574	(52	85	50	2	42	231)	.....exón al 5´de E3 (en intrón 2 de UTR 5´)

Exón de UTR 5´

Tabla 33. *Resultados de ExonScan para la región genómica de la UTR 5´ de Lcn12 humana*

El análisis con ASSP de los sitios de splicing para la UTR 5´ se muestra en la tabla 34. Observamos que el sitio donador del primer exón es detectado, pero no puede ser clasificado. Respecto a los sitios aceptores y donadores de los otros exones no son detectados. Sin embargo si son detectados por ASSP los sitios de splicing de los exones desconocidos, propuestos por ExonScan.

Exón	Posición	Sitio splicing	Confidencia
E1	118	Donador sin clasificar	0.000
E? (1)	317	Aceptor constitutivo	0.301
E? (1)	448	Donador alternativo	0.285
E? (2)	2376	Aceptor constitutivo	0.536
E? (2)	2574	Donador constitutivo	0.924

Tabla 34. *Predicciones de ASSP para los sitios de splicing de los exones de la UTR 5´de Lcn12 humana*

Estos resultados probablemente nos estén indicando que la región UTR 5´de Lcn12 humana quizas sea más compleja aún de lo que conocemos. Por una parte los exones expresados en la UTR 5 ´quizás tengan sitios de splicing extraños (crípticos), indetectables para la herramienta de predicción usada, por otra puede que existan exones alternativos adicionales, como los detectados por ExonScan.

3.5.5.- Ausencia de exones alternativos en las UTRs 5´de lipocalinas que muestran nula o escasa variabilidad

Las lipocalinas que no presentan formas alternativas en su UTR 5´ o aquellas que las presentan, pero que solo dependen de la existencia de promotores alternativos y no de exones alternativos, fueron analizadas con ExonScan para comprobar la posible presencia de posibles exones desconocidos en la UTR 5´. Dado que la región UTR 5´de estas lipocalinas suele ser corta se añadiéron 2000 nucleótidos corriente arriba de la misma.

El resultado de los análisis de predicción con ExonScan de todas estas lipocalinas ofreció resultados negativos, no pudo detectarse ningún exón en ninguna de ellas. Estos resultados son los esperables por las evidencias que disponemos de la ausencia o escasez de diversidad en las UTRs 5´ de estas lipocalinas.

4. - Discusión

Los resultados obtenidos muestran que las UTRs 5´de las lipocalinas muestran valores de longitud y composición en G+C que se encuentran en consonancia con los valores medios de la globalidad de las UTRs 5´de mamíferos. Sin embargo las UTRs 3´de esta familia de proteínas muestran tener una longitud menor y, por otra parte un contenido en G+C claramente superior, que la media de la globalidad de UTRs 3´de mamíferos.

Si la causa del pequeño tamaño de las UTRs 3´ y de su elevado G+C en lipocalinas de mamíferos fuese la restricción genómica, como ya se ha comentado previamente, debería de haber una buena correlación entre el G+C de las UTRs y el $G+C_3$ de la secuencia codificante correspondiente (ya que este está relacionado con el G+C de la región genómica donde se ubica). Se determinó el $G+C_3$ de las secuencias codificantes de cada una de las lipocalinas humanas en estudio y se obtuvo la correlación de este con el G+C de sus correspondientes UTRs 5´y 3´. Como se observa en las gráficas (ver figura 17) se obtiene una buena correlación entre el $G+C_3$ y el G+C de las UTR 5´ (R= 0.832; P= 0.001) y baja y no significativa con el G+C de las UTR 3´ (R= 0.159; P= 0.225).

Esta falta de correlación para las UTRs 3´, demuestra que estas no reflejan el contenido G+C de la región genómica donde se ubican. Esto podría ser consecuencia de algún mecanismo de adaptación relacionado con las necesidades de regulación de la expresión génica propia de lipocalinas. Es conocido que los micro ARN (miARN) actúan principalmente en las regiones UTRs 3´ y ejercen generalmente una acción de inhibición de la expresión génica. Las características de las UTRs 3´ de lipocalinas podrían entonces ser una respuesta de adaptación para evitar la acción de estos miARN, ya que hay evidencia experimental de que a más cortas UTR 3´ menor probabilidad de existencia de dianas de miARN [13]. Así mismo hay evidencias de que un mayor G+C en la UTR3´ da como resultado una menor accesibilidad a esas dianas de miRNA, al formarse estructuras secundarias más estables en dicha región [14]. El estudio de las dianas de miARN en estas regiones UTRs 3´ se aborda en un apartado posterior de esta tesis.

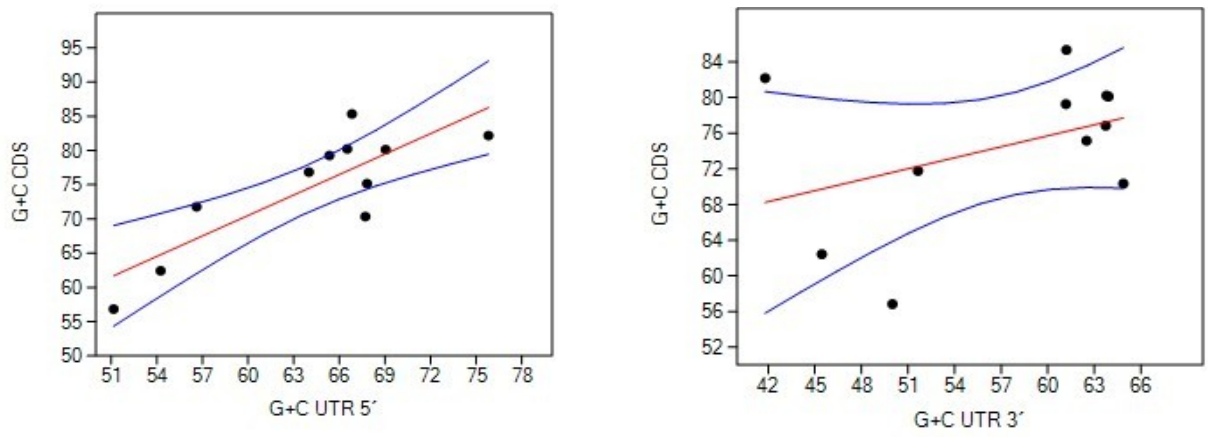


Figura 17: *Relación entre el GC_3 de la zona codificante y el contenido en GC de la UTR 5´ y de la UTR 3´de las lipocalinas humanas. Línea roja regresión, líneas azules intervalo de confianza del 95%.*

Respecto a la variabilidad de las regiones UTRs, los resultados obtenidos muestran que las lipocalinas de mamíferos poseen un mayor número de UTRs alternativos en su extremo 5´ que en el 3´ y que esto debe estar relacionado con el mayor número de exones presentes en estas regiones. Estos resultados obtenidos para las UTRs de lipocalinas de mamíferos están en concordancia con los resultados de diferentes estudios. En uno de estos estudios se ha encontrado que el porcentaje de UTRs 5´que presentan dos exones o más es del 28% para humanos y 26% para roedores, mientras que para las UTR 3´este valor ronda solo el 8% para ambas especies [10]. Por lo tanto hay una mayor oportunidad para originar variabilidad en las UTRs 5´que en la UTRs 3´.

El uso de exones alternativos en la región UTR 3´ parece ser algo poco frecuente y suele estar relacionada con el mecanismo de NMD (*nonsense mediated degradation*)[15]. Esto es lo que observamos en las UTRs 3´de lipocalinas de mamíferos, ya que salvo en el caso de Ptgds que presenta 2 exones, para las demás lipocalinas que muestran variaciones en dicha región, el mecanismo de corte alternativo es el único origen posible. Dicho mecanismo está por lo general relacionado con sitios de poliadenilación alternativos, la presencia de los cuales será abordado en un apartado posterior de esta tesis.

Los datos obtenidos además parecen indicar que existe una mayor variabilidad en las UTRs (analizando conjuntamente las regiones 5´y 3´) de lipocalinas humanas que en las de ratón, hecho que está en concordancia con resultados previamente obtenidos para la globalidad de las UTRs de mamíferos [10]. Esta circunstancia nos indica una mayor necesidad de regulación de la expresión de determinadas proteínas en humano, que sus correspondientes en roedores, entre las cuales estarían

incluidas las lipocalinas.

La variabilidad en las UTRs, tanto en las 5´como en las 3´, es mayor en las lipocalinas que desde un punto de vista evolutivo son más antiguas (esto se constata para las dos especies: humano y ratón). Este resultado es esperable si consideramos que dichas lipocalinas ancestrales tienen una función más diversa frente a las más recientes, con funciones más específicas, necesitando las primeras de una mayor regulación de su expresión.

El estudio de elementos repetitivos en las UTRs de lipocalinas revela que estos se presentan en algunas de ellas, especialmente en sus UTRs 5´, estando presentes en algunos de sus exones alternativos, pero no en los otros (ver tablas 15 y 16). En el caso de la UTR 5´de Apo-D encontramos elementos repetitivos en dos de sus exones alternativos, pero son de distinta naturaleza (ver tabla 15). Estos resultados están en consonancia con un papel regulador de los elementos en cuestión ya que se presentarán en las UTRs alternativas que expresan los citados exones, pero no en las otras.

Dentro de los diferentes elementos repetitivos, las STRs, se han mostrado especialmente relevantes en la regulación de la transcripción o de la traducción y en concreto las STRs de excepcional longitud, que parecen ser elementos reguladores específicos de primates [20]. Una de estas largas STR se encuentra en la UTR 5´de Apo-D (ver tabla 15), dicha STR ha sido detectada, con diferente metodología a la utilizada en esta tesis, en un estudio realizado a escala genómica. Se trata de una STR específica de primates y, debido a la implicación de Apo-D en la regeneración neuronal, es propuesta como un elemento que ha contribuido a la divergencia de primates y no primates [20]. De los resultados de esta tesis se desprende además que solo la UTR 5´alternativa “d” expresaría el exón 1 completo (ver figura 12), que es el que posee en su extremo 5´ la STR mencionada, pudiendo ejercerse así una regulación específica por la citada UTR 5´alternativa.

Otro aspecto a destacar de los elementos repetitivos es la presencia de elementos SINE/ALU en las UTRs 5´de Apo-D y Lcn12 humanas (ver tabla 15), que constituyen exones completos de estas regiones, por lo que se pone de manifiesto su papel como fuente de variación en dichas regiones genómicas.

Del análisis realizado sobre la organización genómica y los mecanismos de splicing de las regiones UTRs 5´ de las lipocalinas, que muestran formas alternativas, podemos comentar que las predicciones realizadas ofrecen en gran medida un respaldo a las hipótesis propuestas. La predicción de promotores ofrece resultados que están, en todos los casos, en consonancia con los promotores esperados para poder explicar el origen de transcripción de las diferentes UTR 5

´alternativas. La predicción de exones de la UTRs 5´ por ExonScan, si bien se ajusta a lo esperado en las UTRs 5´de Apo-D y Rbp4, en otros casos, como Ptgds y Lcn12 humanas solo ofrece resultados incompletos. A pesar de que ExonScan realiza una predicción global incluyendo, además de los sitios de splicing otros elementos reguladores [6], la regulación del splicing alternativo es un fenómeno complejo en la que intervienen un conjunto de señales y factores, que comienzan a dilucidarse [16, 17 y 18], pero aún no del todo conocidos. Esta puede ser la causa de que esta herramienta predictiva sea incapaz de predecir toda la estructura exón/intrón de las diferentes UTR 5´.

Hemos de comentar, en relación a la capacidad predictiva de ExonScan, que este algoritmo reconoce el primer exón de la UTR 5´en varios de los casos estudiados (Apo-D humana y de ratón y Rbp4 humana), pero si bien predice correctamente el extremo 3´del exón (sitio donador) no acierta a predecir correctamente el extremo 5´del mismo. Esto es debido a que el reconocimiento del sitio de inicio del primer exón no está determinado por un lugar de splicing (sitio aceptor), como ocurre en los exones interiores, sino que depende de otros factores de la maquinaria de inicio de transcripción [19], no tenidos en cuenta por esta herramienta de predicción. El hecho de que ExonScan prediga correctamente el extremo 3´del primer exón y que en la mayoría de estos casos ASSP identifique estos mismos sitios como donadores constitutivos nos permite asegurar la veracidad de estos primeros exones de las UTRs 5´ con un mayor nivel de confianza.

Un último aspecto a discutir en relación con las predicciones de ExonScan es la veracidad de algunos de los exones desconocidos predichos por dicha herramienta en las regiones UTRs 5´de las lipocalinas Apo-D (humana y de ratón) y en Lcn12 humana. El hecho de que ExonScan prediga de forma correcta, una fracción considerable de los exones en las UTRs 5´de lipocalinas que muestran variabilidad en dicha región y sin embargo no detecte ninguno en las UTRs 5´de las lipocalinas que no muestran dicha variabilidad, es un argumento a favor para considerar que al menos algunos de los nuevos exones alternativos detectados por ExonScan podrían ser correctos.

Las predicciones de sitios de splicing realizadas con ASSP ofrecen, en la mayoría de los casos aquí estudiados, y más allá de la cuestión del primer exón ya comentada, resultados coherentes con el tipo de regulación del splicing que parecen estar sufriendo estas regiones genómicas. De forma que la mayoría de los exones que son expresados de forma alternativa, muestran sitios de splicing (donadores, aceptores o ambos) que son identificados por ASSP como alternativos o crípticos y en menor número de casos no son detectados.

Podemos concluir diciendo que la organización genómica de la región UTR 5´de las lipocalinas que

muestran formas alternativas es de cierta complejidad. Los mecanismos que originan las diferentes formas de UTR 5´ resultan de una combinación de promotores alternativos junto a splicing alternativo (barajado de exones, omisión de exón y retención de intrón, entre otros). En estos mecanismos parece existir una fina regulación del splicing debida a la existencia de ciertos elementos reguladores que modifican la afinidad del espliceosoma, bien aumentando esta afinidad por sitios de splicing alternativos (débiles) o reduciéndola para sitios de splicing constitutivos (fuertes). La presencia o ausencia de los oportunos factores reguladores de los promotores y del splicing, en ciertos tipos celulares o condiciones fisiológicas dadas, estaría dando lugar a la expresión de diferentes UTRs 5´, según las necesidades.

5. - Bibliografía

[1] Thierry-Mieg, D. & Thierry-Mieg, J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biology* **7**, 1-14 (2006).

[2] Castrignano, T., D'Antonio, M., Anselmo, A., Carrabino, D., D'Onorio De Meo A, D'Erchia AM, Licciulli F, Mangiulli M, Mignone F, Pavesi G, Picardi E, Riva A, Rizzi R, Bonizzoni P, Pesole G. ASPicDB: a database resource for alternative splicing analysis. *Bioinformatics* **15**, 1300-4 (2008).

[3] Rice, P., Longden, I.. and Bleasby, A. EMBOSS The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276-277 (2000).

[4] Hammer, Ø., Harper, D.A.T., and Ryan, P. D. 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* **4**, 1-9 (2001).

[5] Reese, M.G. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput Chem* **26**, 51-6 (2001).

[6] Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., Burge, C.B. Systematic identification and analysis of exonic splicing silencer. *Cell* **119,** 831–845 (2004).

[7] Wang, M. and Marin, A. Characterization and Prediction of Alternative Splice Sites. *Gene* **366,** 219-227 (2006).

[8] Grillo, G., et al. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Research* **38**, D75–D80 (2010).

[9] Duret, L., Mouchiroud, D., Gautier, C.. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC rich isochores. *J. Mol. Evol* **40**, 308-317 (1995)

[10] Pesole, G., et al. Structural and funtional features of eukaryotic mRNA untranslated regions. *Gene* **276**, 73-81 (2001).

[11] Sakabe, N. J. & de Souza, S. J. Sequence features responsible for intron retention in human, BMC *Genomics*, **8**:59 (2007).

[12] Kamat, A. et al. A 500-bp region,,40 kb upstream of the human CYP19 (aromatase) gene, mediates placenta-specific expression in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A*. **96**, 4575–4580 (1999).

[13] Stark, A., et al. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* **123,** 1133–1146 (2005).

[14] Michael, K., et al. The role of site accesibility in miRNA target recognition. *Nature Genetics* **39,** 1278-1284 (2007).

[15] Mignone, F. , Gissi, C. , Liuni, S. & Pesole, G. Unstranslated regions of mRNAs. *Genome Biology* **3** (3), reviews 0004.1- 0004.10 (2002).

[16] Carstens, R., W. McKeehan, & Garcia-Blanco, M. An intronic sequence element mediates both activation and repression of rat fibroblast growth factor receptor 2 pre-mRNA splicing. *Mol. Cell. Biol.* **18,** 2205–2217 (1998).

[17] Modafferi, E.& Black, D. Combinatorial control of a neuronspecific exon. *RNA* **5**, 687–706. (1999).

[18] Southby, J., Gooding, C., Smith, C.W. J. Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of a-actinin mutually exclusive exons. *Mol. Cell. Biol.* 19, 2699–2711 (1999).

[19] Berget, S.M. Exon recognition in vertebrate splicing. *J. Biol. Chem,* **10**, 2411-4 (1995).

[20] Namdar-Aligoodarzi, P., et al., Exceptionally long 5′ UTR short tandem repeats specifically linked to primates. *Gene* **10**, 88-94 (2015).

II

ENSAYOS EXPERIMENTALES CON APO-D DE RATÓN

1. - Objetivo del capítulo

Con el objeto de confirmar la expresión en las células de las UTRs 5´alternativas de lipocalinas consideradas en esta tesis se procedió a realizar algunos ensayos experimentales. Se utilizó como modelo de lipocalina Apo-D de ratón, ya que es la lipocalina que muestra la mayor variabilidad en su UTR 5´ y dado que se disponía de muestras de tejidos de esta especie. Se realizaron ensayos de PCR con transcriptasa inversa (RT-PCR) sobre muestras de diferentes tejidos, con la idea de detectar la presencia de estos transcritos y comprobar si existen patrones de expresión diferentes entre los diferentes tejidos. De forma complementaria se realizaron algunos ensayos de RT-PCR en tiempo real o cuantitativa (Q-RT-PCR).

2. - Material y Métodos

2.1. - Preparación de muestras de ADNc y ADNg

Los tejidos de ratón elegidos para este análisis de PCR son los que se muestran en la tabla 1. Para obtener el ADNc se extrajo el ARN, de los diferentes tejidos homogeneizados, con TRIzol (Invitrogen), siguiendo el protocolo de la casa comercial, y el ARN total (1ug) fué retrotranscrito con Prime-Script^{Tm} (Takara) y tratado con DnaseI.

ABREVIATURA	TEJIDO
Adipose	Tejido adiposo
Heart	Tejido cardiaco
Colon	Colon
Lung	Pulmón
Cb-P10	Cerebelo día 10 posnatal
Cb-6M	Cerebelo de 6 meses posnatal
Cb-6M-PQ	" " " tratado con paraquat
E13.5 Head	Cerebro de día 13,5 de gestación

Tabla 1. Tejidos sobre los que se llevó a cabo el ensayo de RT-PCR

Para obtener el ADN genómico (ADNg) se utilizó también TRIzol. Se partió de la fase de

separación del homogeneizado de tejidos, paso común a la extracción del ARN, y se aplicó el protocolo de la casa comercial para aislar ADN.

2.2. - Selección de primers

Los primers fueron diseñados con la aplicación Primer-BLAST (NCBI) [1]. Se utilizaron los parámetros por defecto, salvo leves modificaciones para adaptarse a las necesidades específicas de nuestra amplificación:

- PCR product size: min: 90 - max: 110
- Organism: Mus musculus
- Primer GC content: min: 40,0 - max: 80,0

Los primers elegidos se muestran en la tabla 2. Diferentes combinaciones de estos primers permiten detectar las diferentes UTRs 5´alternativas de Apo-D de ratón. Así mismo se eligieron primers para detectar ciertas regiones codificantes de dicha proteína (CDS-R, EAK-F y PET-R, VSE-R, ENG-R y ApoD-R). Las diferentes combinaciones de primers utilizadas en los diferentes ensayos se detallan en el apartado de resultados.

Primer	Secuencia	Tm	GC
Región UTR 5´			
Ex1/2-F	5´-GGAGGATTCTGGGTGGAAACTTCAG-3´	57.04°C	52.00%
Ex3-F	5´-CCTCGGTGCTGAGGAGAATTCCA-3´	58.14°C	56.52%
Ex1-F	5´-AGGGGACAGACACAGCATCCCA-3´	59.25°C	59.09%
Ex2-F	5´-AGTTGGAGCTTGCACTTGGGGT-3´	58.61°C	54.55%
Ex2-R	5´-AGCCTTCAGTTGGTGCTCACTGT-3´	58.44°C	52.17%
Región codificante			
CDS-R	5´-CGTGGCCAGGAACATCAGCATG-3´	58.48°C	59.09%
EAK-F	5´-GAAGCCAAACAGAGCAACG- 3´	59.60°C	52.60%
PET-R	5´-TGTTTCTGGAGGGAGATAAGGA- 3´	60.10°C	45.60%
VSE-R	5´-AGCTTGGCTGGCTCTGAGACG- 3´	58.40°C	62.00%
ENG-R	5´-AGCACTTCGATGTTTCCGTTCTCC- 3´	59.50°C	50.00%
ApoD-R	5´-CGGGCAGTTCGCTTGATCTGT- 3´	61.80°C	57.15%

Tabla 2. *Primers seleccionados para la detección de las diferentes UTRs 5´alternativas de Apo-D de ratón.*

2.3. - Ensayos de RT-PCR

Para realizar la PCR se utilizó la solución lista para usar de Gotaq ® Colorless Master Mix. Se trabajó con un volumen de PCR de 15 µl por muestra, constituido por la adición de los volúmenes ajustados de las disoluciones de los primers, la del ADNc y la del Gotaq Mix. Cada muestra de tejido fue preparada con la combinación de primers que permitiese detectar las oportunas variantes de la UTR 5´.

La PCR se llevó a cabo en un termociclador con un programa que incluía una desnaturalización inicial, posteriormente los ciclos de desnaturalización, alineamiento y elongación y por último una extensión final. Las condiciones de este programa fueron: 95 ºC 2´/(95ªC 30´´ - 58ºC 30´´ - 72ºC 30´´) x 35 / 72ºC 5´.

Las muestras ya amplificadas fueron sometidas posteriormente a electroforesis en gel de agarosa al 2%, añadiendo previamente bromuro de etidio y finalmente las placas obtenidas fueron fotografiadas bajo luz ultravioleta. La imagen digital de las bandas de electroforesis fue sometida a densitometría óptica (niveles de grises en la imagen de 8 bits) para obtener una medida aproximada de los niveles de expresión relativa de las diferentes UTRs 5´ alternativas.

2.4. - Ensayos de Q-RT-PCR

El ADNc de los diferentes tejidos de ratón, obtenidos como se ha indicado anteriormente, fue sometido a amplificación de PCR cuantitativa mediante SYBR Green I (Takara). Se utilizaron los pares de primers necesarios para amplificar las oportunas alternativas de la UTR 5´de Apo-D (ver tabla 1) y el par de primers del gen Rpl18 de ratón, que fue utilizado como gen de referencia. Los primers para este gen se indican en la tabla 3.

Primer	Secuencia	Tm	GC
MrPL18-F	5´ -TTCCGTCTTTCCGGACCT- 3´	60.6	55.50%
MrPL18-R	5´-TCGGCTCATGAACAACCTCT- 3´	60.8	50.00%

Tabla 3. *Primers seleccionados para el gen de referencia de la Q-RT-PCR (gen Rpl18 de ratón).*

La amplificación fue realizada por cuadruplicado en un termociclador (ABI Prism 7900HT) con el mismo programa de ciclos que se utilizó para la RT-PCR, previamente detallado. Las diferencias en los niveles de transcripción fueron determinados por el método $2^{-\Delta\Delta CT}$ [2].

3. - Resultados

3.1. - RT-PCR

Se seleccionaron las parejas de primers (ver material y método) con el objeto de poder detectar de forma específica las distintas variantes de la UTR 5´ de Apo-D de ratón. En la tabla 4 se muestran las combinaciones de los mismos. En la Figura 1 se muestra la posición que ocupan los distintos primers en los diferentes exones expresados en las variantes de la UTR 5´.

El primer ensayo realizado fue uno de control (+) de la expresión de ApoD en las diferentes muestras de tejidos. Se realizó RT-PCR utilizando unos primers que amplifican una región de la secuencia codificante (primers EAK-F y PET-R, ver en material y métodos). Como control negativo (-) se utilizó una muestra que carecía de retrotranscriptasa. Los resultados, que se observan en Fig.2, demuestran la expresión de ApoD en todos los tejidos y parece observarse, por la intensidad de las bandas, un mayor grado de expresión en los tejidos del sistema nervioso.

Posteriormente se procedió a comprobar la amplificación sobre muestras de ADN genómico de ratón con diferentes combinaciones de los primers para la región UTR 5´. Los resultados obtenidos

mostraron un patrón de bandas que se corresponde con los tamaños esperados de los fragmentos amplificados (ver figura 3), indicándonos que la amplificación ocurría de forma correcta.

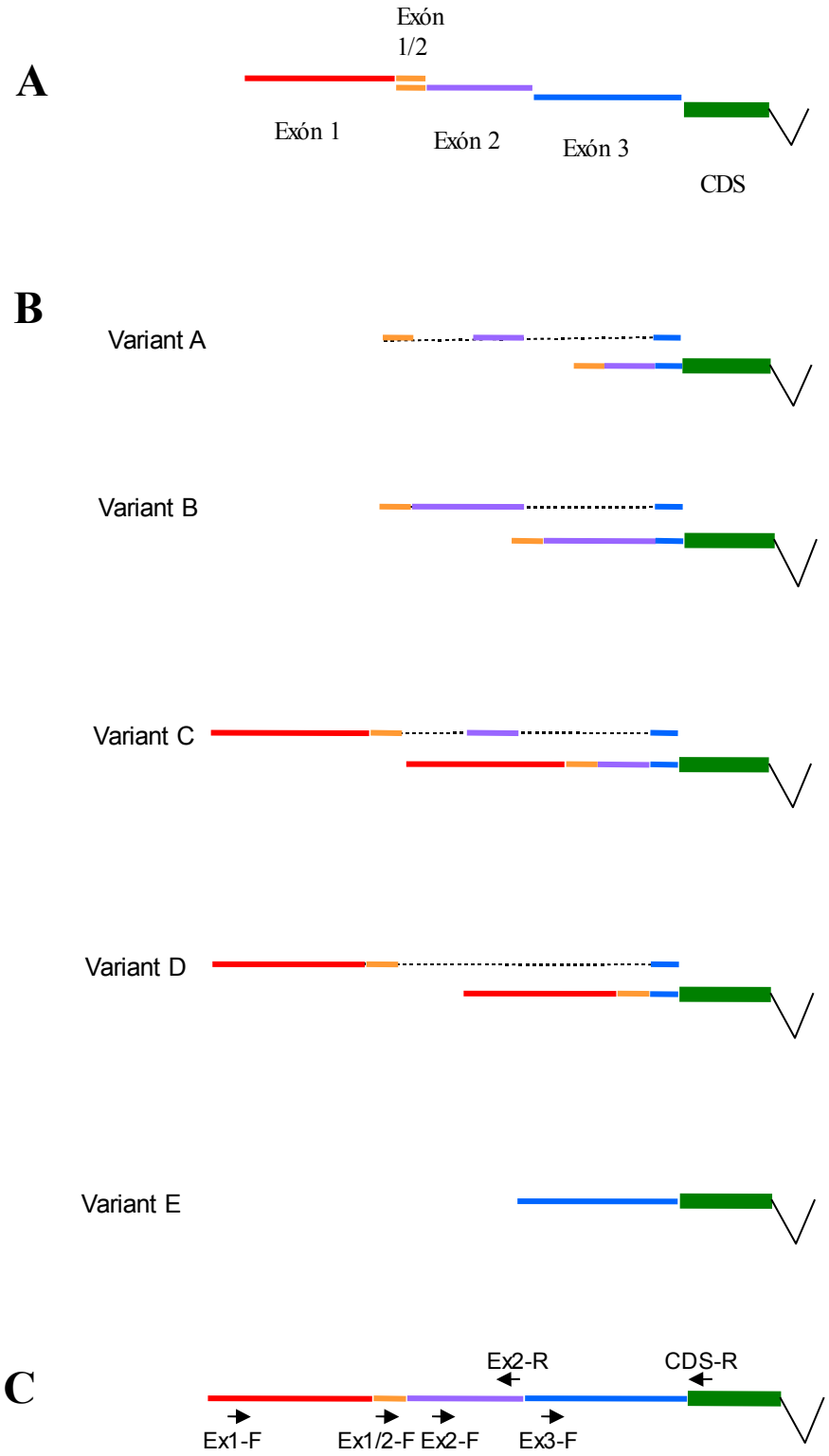


Figura 1. **A**: *Diferentes exones que conforman la UTR 5´de Apo-D de ratón.* ***B****: Lista de UTR 5´alternativas con la combinación de exones correspondiente.* ***C****: posición de los primers en los diferentes exones de la UTR 5´*

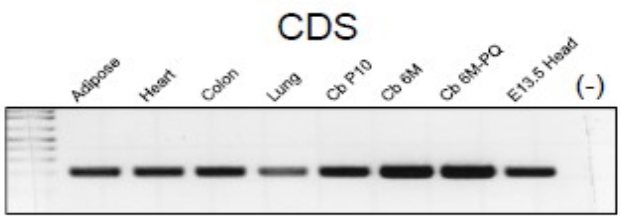


Figura 2. *Resultado del test de RT-PCR para comprobar la expresión de Apo-D en diferentes tejidos*

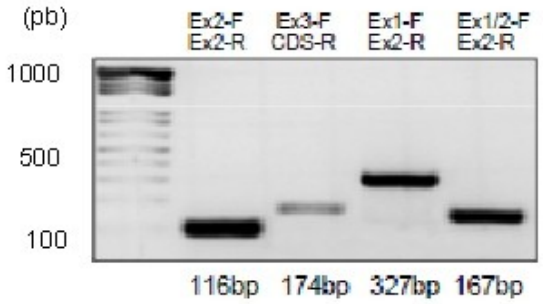


Figura 3. *Resultado del test de eficacia de RT-PCR para los primers de la región UTR 5´de Apo-D*

Posteriormente se procedió a amplificar las diferentes variantes de la UTR 5´ a partir de las muestras de ADNc de los diferentes tejidos. Las combinaciones de primers disponibles nos permitieron resolver, de forma inequívoca, las variantes B, C y E de la UTR 5´de ApoD de ratón. Para las variantes A y D, la combinación de primers no diferencian totalmente entre A/B y C/D (ver tabla).

Primers	Variante amplificada
E1-F / E2-R	C
E1-F / CDS-R	C y D
E1/2-F / E2-R	A y B
E2-F / CDS-R	B
E3-F / CDS-R	E

Tabla 4. *Combinaciones de primers y variantes de la UTR 5´que amplifica cada una de ellas.*

Los resultados de los productos amplificados y corridos en geles se muestran en la figura 4. Pueden observarse claramente los patrones de expresión de las variantes B, C y E (que pudieron amplificarse de forma individual). Respecto a los resultados de las otras variantes A y D, no separables en su amplificación de B y C respectivamente, si pueden diferenciarse tras correr los productos en los geles, debido a los diferentes tamaños de la amplificación.

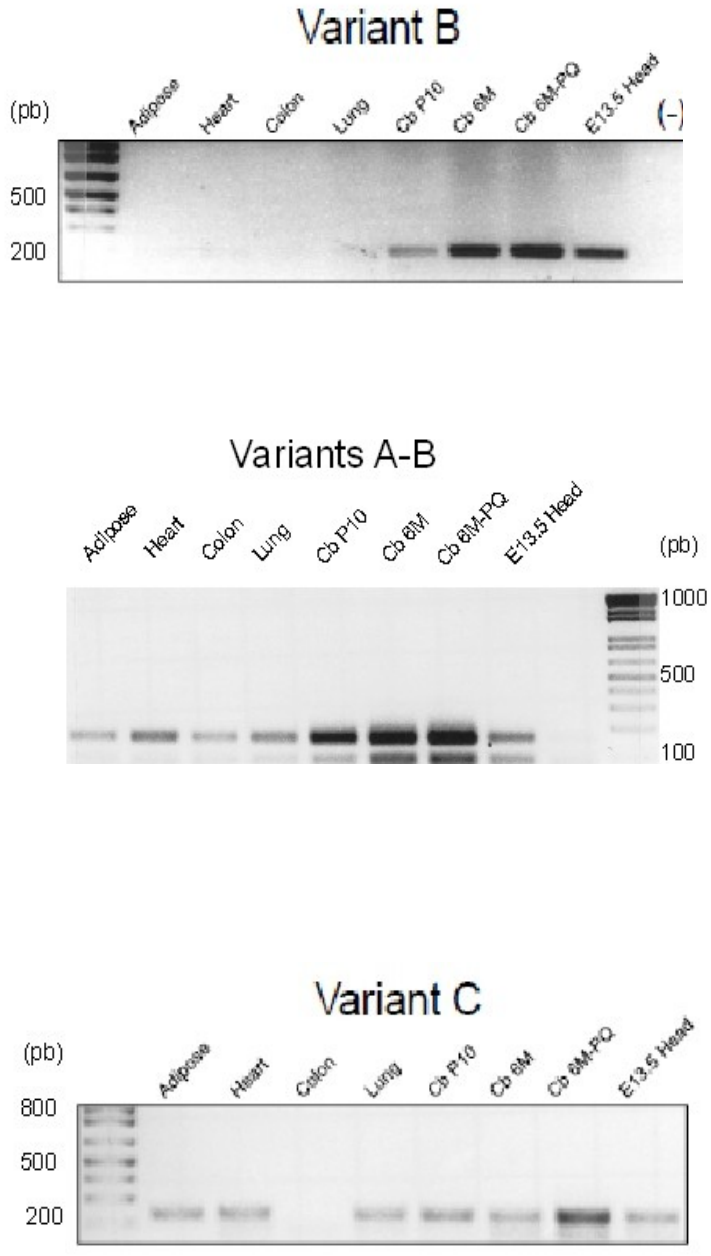


Figura 4. *Geles obtenidos tras la amplificación por RT-PCR de las variantes de la UTR 5´ de Apo-D en los diferentes tejidos de ratón.*

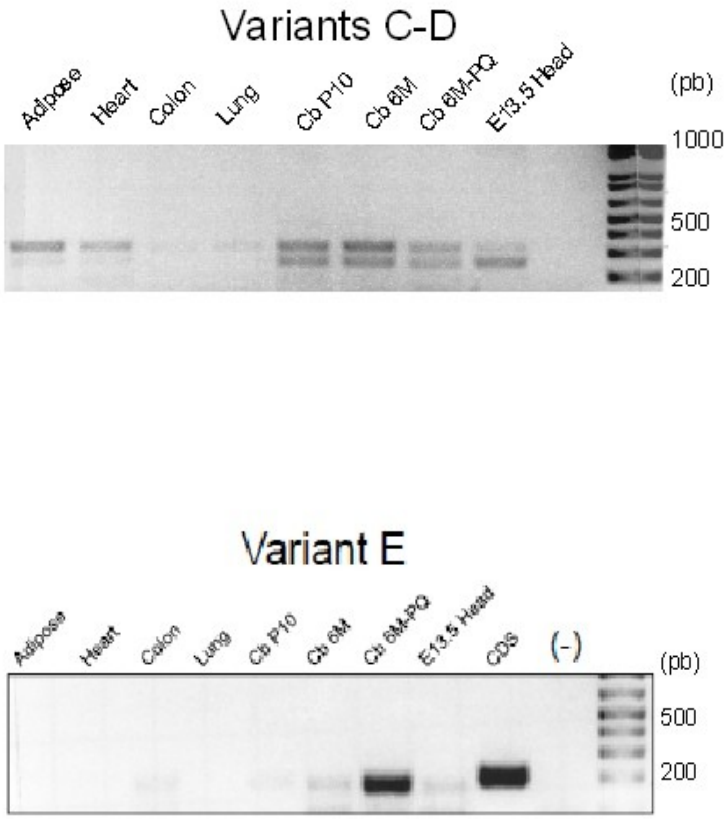


Figura 4 (continuación). *Geles obtenidos tras la amplificación por RT-PCR de las variantes de la UTR 5´ de Apo-D en los diferentes tejidos de ratón.*

Para obtener una imagen más clara de las diferencias de expresión entre las diferentes variantes se sometieron las imágenes obtenidas de las placas de electroforesis a una cuantificación mediante densitometría óptica. Los resultados numéricos (en forma de % relativo) se representaron en las gráficas de las figuras 5, 6 y 7. Con estos datos pudo analizarse de forma más clara las preferencias de expresión de las diferentes variantes. Así las variantes A, B y D son variantes enriquecidas en sistema nervioso, la variante C es de expresión más generalizada o distribuida y la variante E es una variante sensible al tratamiento con PQ en sistema nervioso.

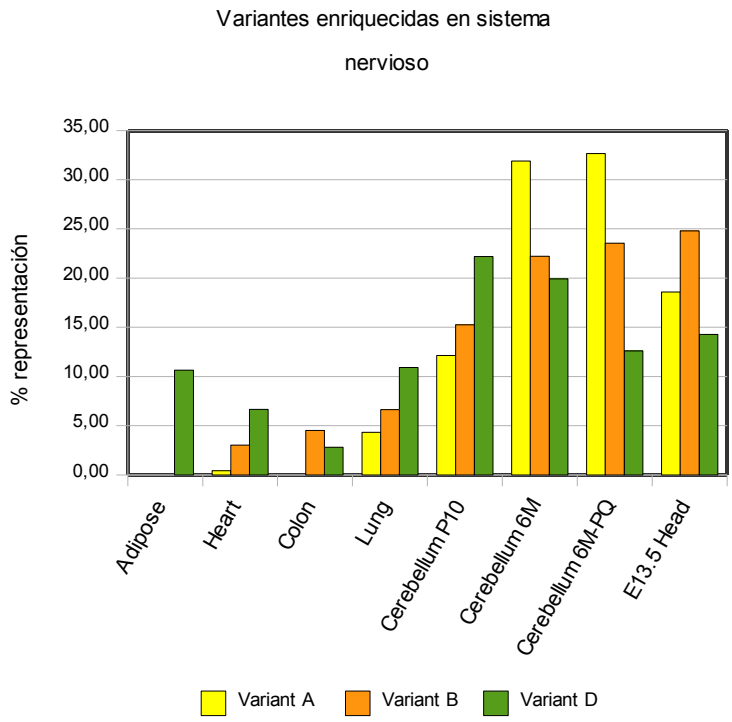


Figura 5. *Perfil de expresión de las variantes de la UTR 5´ que muestran preferencia en sistema nervioso. Datos obtenidos por densitometría óptica de imágenes de geles resultantes de amplificación.*

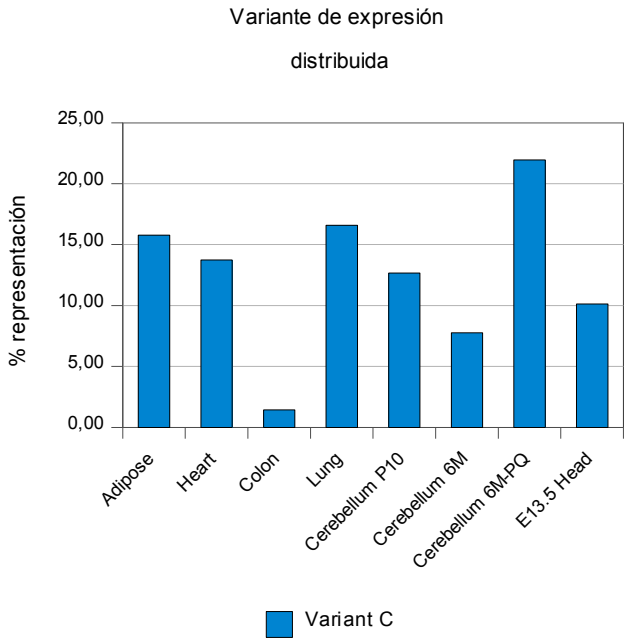


Figura 6. *Perfil de expresión de la variante de expresión distribuida Datos obtenidos por densitometría óptica de imágenes de geles resultantes de amplificación.*

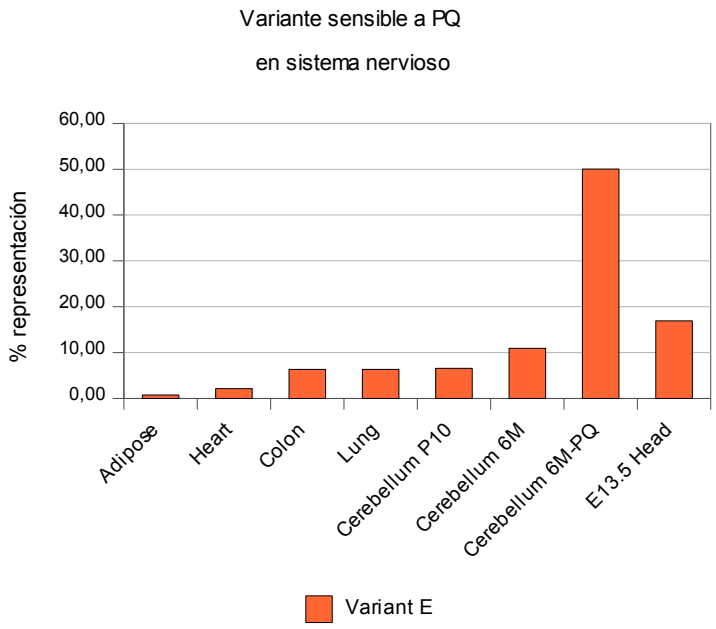


Figura 7. *Perfil de expresión de la variante sensible a paraquat en sistema nervioso. Datos obtenidos por densitometría óptica de imágenes de geles resultantes de amplificación.*

Dado el interés que presenta la variante "E" de la UTR 5´, por su especificidad en tejido de cerebelo y su sensibilidad al estrés oxidativo, se realizaron ensayos adicionales de RT-PCR para confirmar la transcripción del ARNm completo portador de dicha UTRs 5´. Para ello se utilizaron los primers "Ex3-F" (de la región UTR 5´de esta variante) y otros elegidos en diferentes puntos de la región codificante (ver material y métodos). La ubicación de todos estos primers puede verse en la figura 8. Para estos ensayos se utilizó la muestra de ADNc de tejido de cerebelo tratada con paraquat (Cb-PQ).

Las diferentes combinaciones del primer "forward" de la región UTR 5´ con los primers "reverse" de los exones codificantes fueron sometidas a amplificación y posteriormente corridas en geles de agarosa. El resultado de estos ensayos, como se observa en los geles de la figura 9, fue positivo, a pesar de que para detectar el producto amplificado de mayor longitud (resultado del par: E3-F/ApoD-R) hubo que ajustar las temperatura de los ciclos de amplificación. De estos resultados concluimos que se produce la transcripción completa del ARNm de esta variante

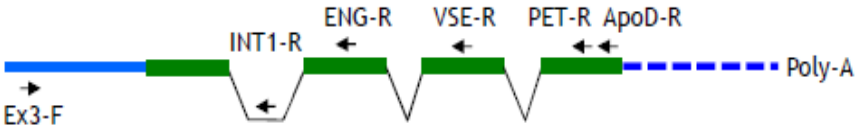


Figura 8. *Ubicación de los primers utilizados para la verificación de la transcripción completa del ARNm de la variante E de ApoD de ratón. En azul la región UTR y en verde los exones de la región codificante.*

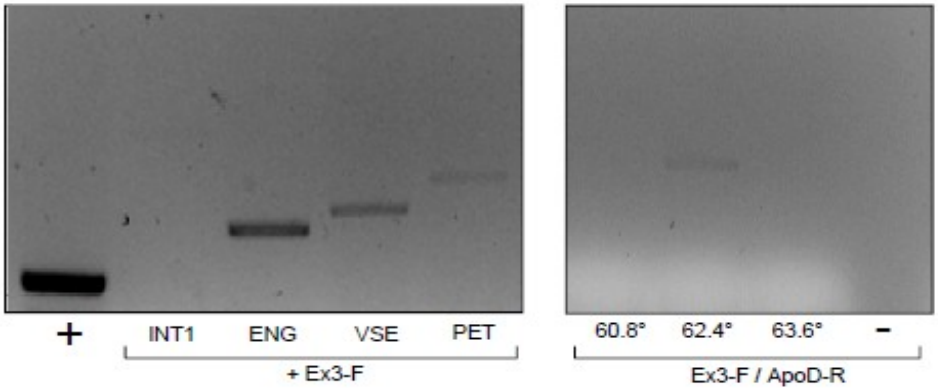


Figura 9. *Geles obtenidos tras la amplificación de los pares de primers seleccionados para verificar la transcripción completa del ARNm de la variante E de ApoD de ratón.*

3.2. - Q-RT-PCR

Para obtener una idea más precisa de los niveles de expresión que las basadas en la densitometría óptica se realizaron algunos ensayos de PCR en tiempo real o cuantitativa (Q-RT PCR). Estos se realizaron sobre las variantes B y E, cuyos perfiles de expresión muestran claras diferencias entre diferentes tejidos.

Para la variante B se realizó el ensayo de su expresión en cerebelo (Cb) frente a corazón (Heart) y para la E de cerebelo (Cb) frente a cerebelo tratado con paraquat (Cp-PQ). Los primers utilizados en cada caso fueron E2-F/CDS-R para variante B y E3-F/CDS-R para la E. Los resultados se observan en la figura 8.

En el caso de la variante B se obtiene que la expresión en tejido nervioso es del orden de 20 veces superior que en otros tejidos, como corazón. Respecto a la variante E observamos que su expresión en cerebelo con paraquat es del orden de 3 veces superior que en cerebelo sin tratar, diferencia importante pero de magnitud inferior a la obtenida por la densitometría.

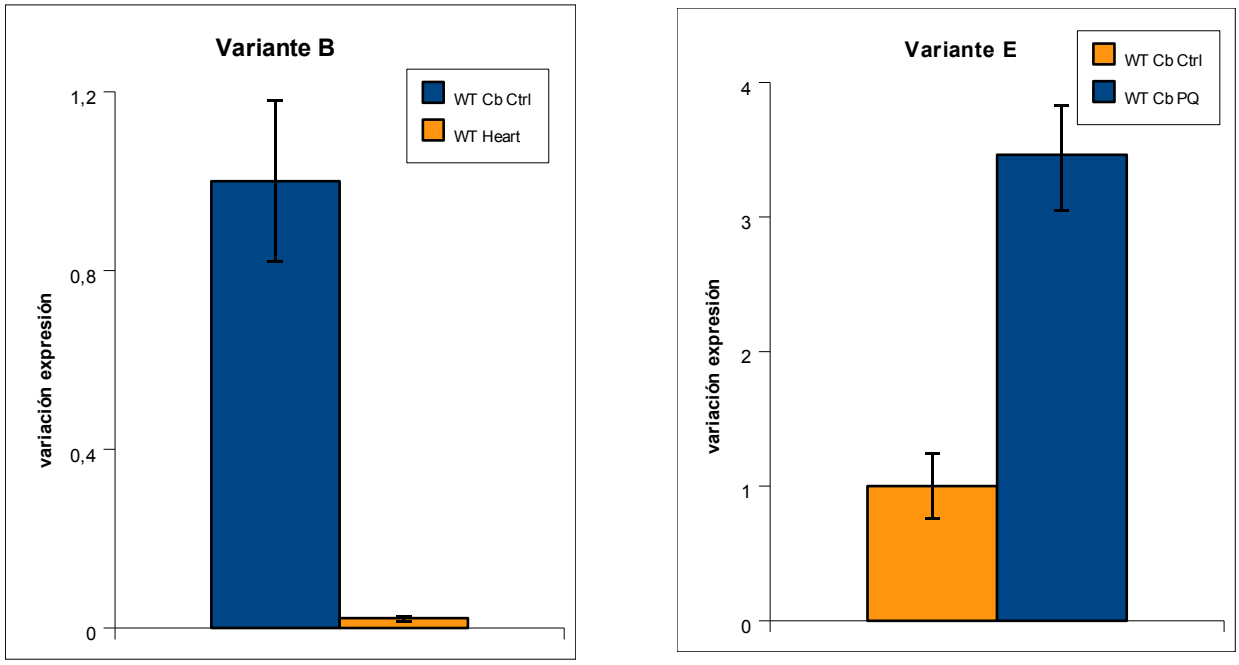


Figura 8. *Resultados de Q-RT-PCR para variantes B y E de ApoD de ratón. Para "B" se comparó la expresión entre cerebelo (Cb) y corazón (Heart). Para "E" se comparó cerebelo (Cb Ctrl) y cerebelo tratado con paraquat (Cb PQ)*

4. - Discusión

Los resultados de estos ensayos de PCR en ApoD de ratón aportan sólidas evidencias de la expresión de todas las UTRs 5´ alternativas seleccionadas de las bases de datos, respaldando así su realidad biológica. Dichos resultados confirman la validez de los criterios y la metodología empleada en la selección de los transcritos de lipocalinas a partir de las bases de datos (ver métodos de capítulo I).

Se han obtenido evidencias claras de que las diferentes variantes se expresan de forma diferenciada

en los distintos tejidos. Los perfiles de expresión muestran una clara preferencia de expresión en tejido nervioso, excepto para la variante C, de expresión más distribuida. La variante E sería una variante expresada en condiciones más especiales en las que el tejido nervioso es sometido a condiciones de estrés oxidativo.

Los ensayos de RT-PCR cuantitativa ponen de manifiesto que hay claras diferencias en la representación de estas variantes de la UTR 5´ entre diferentes tejidos (caso de variante B) o en diferentes condiciones fisiológicas de los mismos (caso de variante E).

La diferente composición de exones de las diferentes UTRs 5´(ver figura 1) estaría relacionada con el diferente perfil de expresión entre las mismas. La variante E, con el perfil de expresión más específico, esta constituida por un exón alternativo, no expresado en ninguna de las otras variantes, lo que daría cuenta de su comportamiento. Sin embargo las variantes C y D tienen una composición de exones muy semejantes, siendo su perfil de expresión diferente a pesar de ello.

El hecho de que tres de las variantes (A, B y D) muestren preferencias de expresión en tejido nervioso nos da una idea del nivel fino de regulación que la expresión de Apo-D debe estar teniendo, ya que hemos de suponer que estas tres variantes, de diferente composición exónica (especialmente A y B respecto a D) no deben están ejerciendo una acción reguladora equivalente. El diferente papel regulador que podría estar ejerciendo cada una de las UTRs 5´alternativas será abordado extensamente en capítulos posteriores.

5. - Bibliografía

[1] Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L.. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* **13**:134 (2012) .

[2] Livak, K.J., Schmittgen, T.D. Analysis of relative gene expresion data using real time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402-408 (2001).

III

CONSERVACIÓN DE LAS REGIONES UTRS DE LIPOCALINAS DE MAMÍFEROS.

1. - Objetivos

El objetivo de este capítulo es conocer si las UTRs 5´y 3´de las lipocalinas de mamíferos se han conservado a lo largo de la evolución, para ello se ha procedido a:

- Determinar en qué medida los exones que forman estas regiones están presentes en diferentes especies de mamíferos y si lo están en que grado de conservación.

- Obtener evidencias de que la conservación que se detecte a nivel del genoma es expresada en las diferentes especies de mamíferos.

- Realizar alineamientos múltiples de los transcritos de diferentes especies de mamíferos para los que se obtengan evidencias de secuencias UTRs ortólogas, de forma que obtengamos una mayor certeza de la conservación de estas regiones y una visión más detallada de la evolución que han seguido las mismas.

2. - Métodos

2.1. - Evidencias genómicas de conservación de exones de UTRs

Para obtener evidencias genómicas de conservación en mamíferos de los diferentes exones de las UTRs 5´y 3´ de lipocalinas humanas y de ratón, que previamente habían sido identificados, se utilizó BLAT (UCSC; https://genome.ucsc.edu/cgi-bin/hgBlat) [1]. Se aplicó esta herramienta sobre el genoma de especies de mamíferos de los siguientes órdenes: primates, roedores, artiodáctilos y carnívoros. Complementariamente se utilizó la herramienta BLAST (NCBI, nucleotide Blast; http://blast.ncbi.nlm.nih.gov/Blast.cgi) [2], seleccionando la opción de buscar en bases de datos genómicas. Las secuencias obtenidas que mostraron un porcentaje de identidad (PI) mayor del 60% y que se localizaron en las posiciones genómicas correctas, fueron tomadas como exones ortólogos de la UTR (5´o 3´) respecto a humano, ratón o ambos, según el caso.

2.2. - Evidencias de expresión de UTRs 5´y 3´ortólogas

Una vez confirmado en el genoma de los diferentes mamíferos la existencia de exones ortólogos en las UTRs de lipocalinas, se procedió a comprobar que estos son expresados en las correspondientes UTRs. Se utilizó BLAST (NCBI, nucleotide Blast; http://blast.ncbi.nlm.nih.gov/Blast.cgi) [2], seleccionando la opción de buscar en bases de datos de expresión. Se utilizaron como secuencias de partida las correspondientes UTRs 5´y 3´ de humano y ratón, obteniéndose así las secuencias de los correspondientes transcritos ortólogos para las otras especies.

2.3. - Obtención de alineamientos múltiples de UTRs ortólogas

Los transcritos de las diferentes lipocalinas, obtenidos como se indica en el apartado anterior, fueron alineados junto a los de humano, ratón o ambos, según el caso, mediante "emma" (EMBOSS; http://emboss.bioinformatics.nl/) [3], una interfaz al programa ClustalW [4], y los alineamientos múltiples obtenidos fueron visualizados mediante "prettyplot" (EMBOSS) [3].

2.4. - Comparación del grado de conservación entre las secuencias UTRs ortólogas y sus correspondientes secuencias ortólogas codificantes.

Para conocer el grado de semejanza entre los pares de secuencias ortólogas de las UTRs se recurrió a los alineamientos múltiples (obtenidos como se cita en apartado anterior). Dichos alineamientos fueron analizados con "dismat" (EMBOSS) [3], obteniendo así la matriz de distancias correspondiente, y a partir de aquí pudo calcularse el porcentaje de identidad (PI) entre pares de secuencias.

Para las secuencias codificantes correspondientes se procedió a alinearlas mediante "tranalign" (EMBOSS) [3], que realiza un alineamiento múltiple de las secuencias de nucleótidos, guiado por un alineamiento de las secuencias de proteínas correspondiente. Posteriormente este alineamiento de las secuencias codificantes fue analizado con "distmat" (EMBOSS) [3] y se obtuvo, mediante la selección oportuna en las opciones, la matriz de distancia para la primera, segunda y tercera posición de cada codón. Finalmente se obtuvieron a partir de estos datos los porcentajes de identidad para cada una de estas posiciones de los pares de secuencias codificantes.

3. - Resultados

3.1. - Conservación en la región UTR 5´

3.1.1. - Evidencias genómicas

La simple comparación visual de la organización genómica de las UTR 5´ de las lipocalinas ortólogas de humano y ratón (ver figuras del capítulo I) nos permite observar ciertas semejanzas, especialmente al comparar las regiones UTRs 5´ de Apo-D y Rbp4. Comprobamos que la extensión de estas regiones es semejante, igualmente observamos semejanzas respecto al número y tamaño de los exones e intrones que constituyen estas regiones genómicas en ambas especies.

Una primera aproximación al grado de conservación que existe en estas regiones podemos obtenerla realizando un "dotplot" (dotmatcher, EMBOSS) entre las regiones genómicas completas de las UTRs 5´ ortólogas. En las figuras 1 y 2 se observa el dotplot para las regiones genómicas de UTRs 5´de Apo-D y de Rbp4 de humano frente a ratón. Comprobamos (ver diagonales principales) que hay diferentes zonas, a lo largo de toda la región UTR 5´ que muestran conservación.

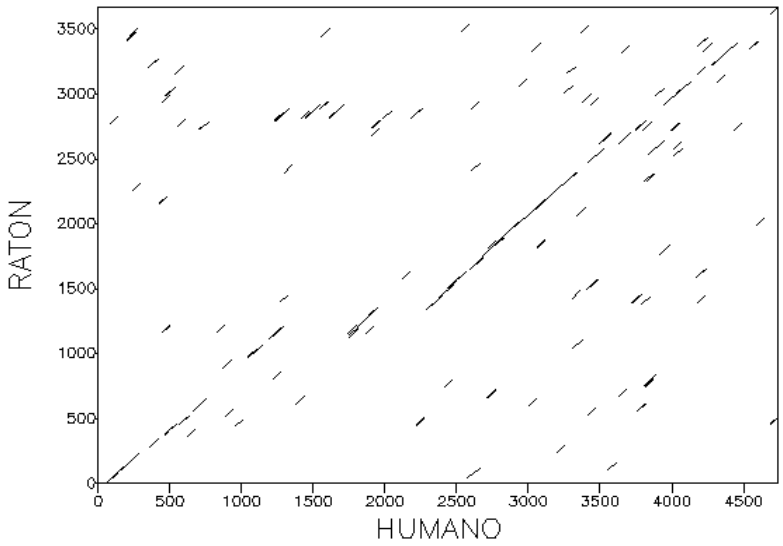


Figura 1. Dotplot de UTR 5´ de Apo-D humana frente a la de ratón (window=50 nt)

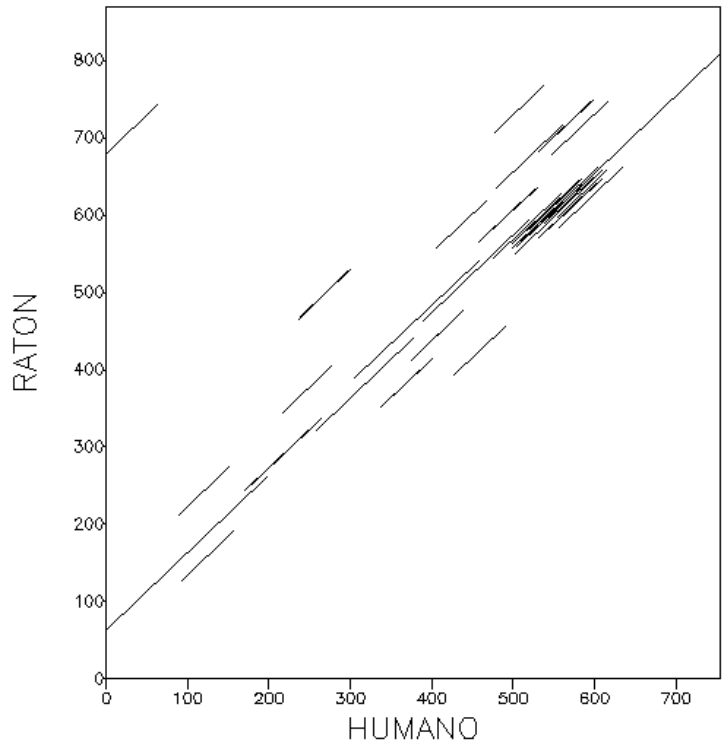


Figura 2. Dotplot de UTR 5´ de Rbp4 humana frente a la de ratón. (window=50nt)

Lo que nos interesa, para los propósitos de esta tesis, es conocer si los exones de las UTRs 5´se han conservado en mamíferos, ya que dichos exones son los responsables de que se formen UTRs 5 ´alternativas. Para ello se procedió a estudiar, en que medida, los exones que componen dicha región en humano y ratón son ortólogos entre sí y si existen exones ortólogos en otros órdenes de mamíferos diferentes.

Con este propósito se llevó a cabo un análisis de los exones, previamente identificados, en las UTRs 5´ de las lipocalinas de humano y ratón frente a los genomas de otros mamíferos. Se analizaron otras especies de primates y roedores y especies de mamíferos de órdenes diferentes como carnívoros y artiodáctilos (ver métodos).

Los resultados obtenidos revelan que los exones de las UTRs 5´ humanos de las lipocalinas se encuentran todos muy bien conservados (95% a 100% de Identidad) entre los primates. Ocurre lo mismo con los exones de ratón, al menos cuando comparamos con el genoma mejor estudiado de otro roedor como la rata (*R. norvegicus*). La visualización en Genome Browser (USCS) permite

comprobar como se conserva la estructura genómica de la región UTR 5´ (tamaño y orden de exones y tamaño de intrones) dentro de cada uno de estos órdenes de mamíferos. En la figura 3 puede verse este hecho para la UTR 5´de Apo-D en primates.

Cuando se analizó la conservación de estos exones entre especies de mamíferos, de órdenes diferentes, se encontraron distintos resultados. Para un primer conjunto de lipocalinas (Apo-D, Ptgds, Rbp4 y Apo-M), que son las que presentan una mayor diversidad en su UTR 5´, se constató que hay conservación entre los diferentes órdenes de mamíferos de algunos de los exones alternativos de dicha región. El porcentaje de identidad que muestran dichos exones ortólogos oscila alrededor del 80 % (ver tablas 1 a 4). Mientras que para otros exones alternativos, de este mismo grupo de lipocalinas, bien de humano o de ratón, no existen evidencias genómicas de exones ortólogos en los otros órdenes de mamíferos (ver tablas 1 a 4).

Si tomamos como ejemplo el caso de Apo-D (ver tabla 1) comprobamos que dos de los cuatro exones que poseen la especie humana y el ratón en su UTR 5´pueden considerarse ortólogos (pares humano/ratón: E1/e1 y E4/e3). En los otros órdenes de mamíferos estudiados también se encuentran exones ortólogos a estos mismos. Los otros dos exones que presentan, tanto humano como ratón, en su UTR 5´ no pueden considerarse ortólogos entre sí y tampoco hay evidencias de que existan ortólogos a ellos en otros órdenes de mamíferos, por lo tanto podemos considerar a estos exones como específicos de primates y roedores respectivamente.

Para el resto de lipocalinas, de este primer grupo (tablas 1 a 4), se encuentran resultados semejantes, algunos exones ortólogos entre especies de los diferentes órdenes de mamíferos y otros que parecen ser específicos, bien de primates o de roedores.

No siempre el tamaño de los exones ortólogos es coincidente, hay en ocasiones disparidad en el tamaño de los mismos. En las tablas (1 a 4) se han querido destacar, señalándose con un asterisco, los exones que poseen un tamaño inferior al 50% respecto a su pareja ortóloga de mayor tamaño.

Para el resto de lipocalinas estudiadas: Lcn1, Lcn2, Obp2a, C8g, Lcn8, Lcn12 y Orm2 (que solo presentan un exón en su UTR 5´, con la excepción de LCN12 que presenta 3), no se encontró ninguna evidencia genómica de la existencia de exones ortólogos entre las especies de mamíferos de órdenes diferentes.

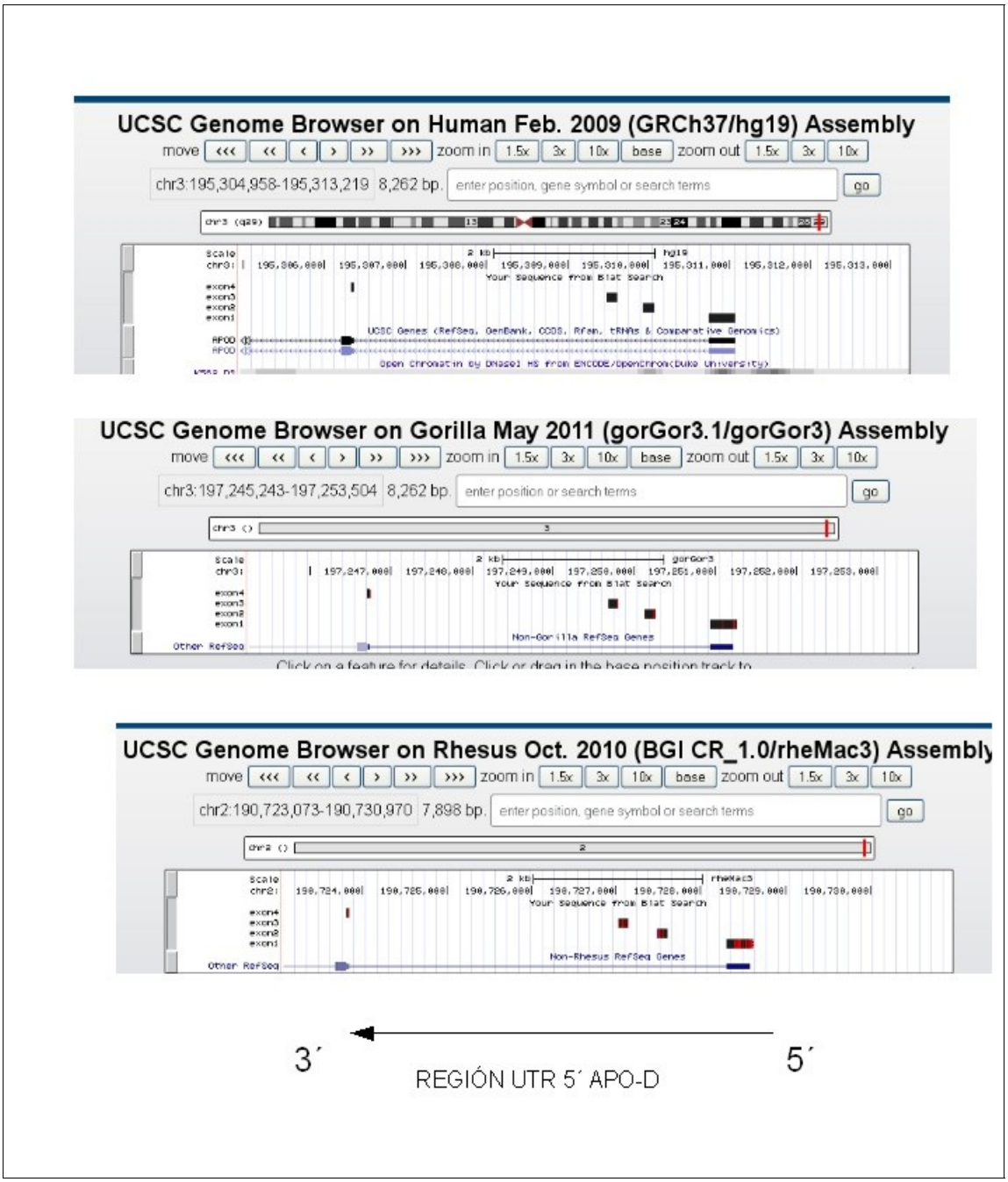


Figura 3. *Resultados gráficos en Genome Browser (UCSC), tras realizar búsqueda en BLAT con exones de la UTR 5´ de Apo-D humana (arriba), sobre genoma de los primates Gorila y Resus (debajo).*

Lipocalina	Exones UTR 5´	Long(pb)	% Ident Hum/ratón	% Ident Hum/artiod	% Ident Hum/carniv	% Ident Ratón/artiod	% Ident Ratón/carniv
APOD Hum							
	E1	327	75.44	85.55	88.50		
	E2	136	–	–	–		
	E3	113	–	–	–		
	E4	31	82	89.25	87.60		
APOD Ratón							
	E1	260	75.44			76.24	76.74(*)
	E1-2	199	–			–	–
	E2	66	–			–	–
	E3´	214	82			85	85

Tabla 1. *Evidencias genómicas de exones ortólogos a exones en UTR 5´ de Apo-D de humano y ratón*

Lipocalina	Exones UTR 5´	Long(pb)	% Ident Hum/ratón	% Ident Hum/artiod	% Ident Hum/carniv	% Ident Ratón/artiod	% Ident Ratón/carniv
PTGDS Hum							
	E1	162	–	–	–		
	E2	92	–	–	–		
	E3	1030	63.40	100 (*)	79.30 (*)		
PTGDS Ratón							
	E1	329	63.40 (*)			78.70 (*)	74.70

Tabla 2. *Evidencias genómicas de exones ortólogos a exones en UTR 5´ de Ptgds de humano y ratón*

Lipocalina	Exones UTR 5´	Long(pb)	% Ident Hum/ratón	% Ident Hum/artiod	% Ident Hum/carniv	% Ident Ratón/artiod	% Ident Ratón/carniv
RBP4 Hum							
	E1	304	72	76	72		
	E2	95	89	83	80		
	E3	18	?	?	?		
RBP4 Ratón							
	E1	367	72			69	–
	E2	71	89			81	81
	E3	18	?			?	?

Tabla 3. *Evidencias genómicas de exones ortólogos a exones en UTR 5´ de Rbp4 de humano y ratón*

Lipocalina	Exones UTR 5´	Long(pb)	% Ident Hum/ratón	% Ident Hum/artiod	% Ident Hum/carniv	% Ident Ratón/artiod	% Ident Ratón/carniv
APOM Hum							
	E1	496	83 (*)	78	82		
APOM Ratón							
	E1	781	83			81 (*)	81 (*)

Tabla 4. *Evidencias genómicas de exones ortólogos a exones en UTR 5´ de Apom de humano y ratón*

3.1.2. - Evidencias de la expresión de UTRs 5´ortólogas

La mera presencia de estos exones de la UTR 5´ conservados en el genoma de los diferentes órdenes de mamíferos no es garantía de que se estén expresando en la UTR 5´de los transcritos de las correspondientes lipocalinas, o que si se expresan lo estén haciendo con la misma combinación de exones que lo hacen en el caso de humano o ratón. Para abordar esta cuestión se realizó un análisis, mediante BLAST sobre bases de datos de expresión de mamíferos, utilizando como punto de partida los diferentes UTR 5´ que sabemos están expresándose en humano y ratón (ver métodos).

Apo-D Humana				
Variante UTR 5´	Combinación Exones	Evid. Expres. Primates	Evid. Expres. Otros Mamif.	
a	E1, E4	si	si	vaca, perro, oso y ratón
b	E1´, E2, E4	no	-	–
c	E1´, E3, E4	si	-	–
d	E1´, E4	si	si	cerdo y perro
Apo-D Ratón				
Variante UTR 5´	Combinación Exones	Evid. Expres. Roedores	Evid. Expres. Otros Mamif.	
a	E1´, E2	si	–	–
b	E1-2, E3	si	–	–
c	E1, E2, E3	si	–	–
d	E1, E3	si	si	vaca, perro y humano
e	E3´	no	-	–

Tabla 5. *Evidencias de expresión de las UTR 5´de Apo-D de humano y ratón en otros mamíferos*

Ptgds Humana				
Variante UTR 5´	Combinación Exones	Evid. Expres. Primates	Evid. Expres. Otros Mamif.	
c	E3´´	si	si	vaca, oso, rata y topo
g	E1, E2, E3´, E3´´	si	-	–
j	E1, E2, E3	si	-	–
Ptgds Ratón				
Variante UTR 5´	Combinación Exones	Evid. Expres. Roedores	Evid. Expres. Otros Mamif.	
c	E1´	si	si	oso, vaca y cabra
d	E1	si	si	mono gibón, papión y vaca

Tabla 6. *Evidencias de expresión de las UTR 5´de Ptgds de humano y ratón en otros mamíferos*

RBP4 Humana				
Variante UTR 5´	Combinación Exones	Evid. Expres. Primates	Evid. Expres. Otros Mamif.	
b	E1, E3	si	si	vaca, cerdo y ratón
d	E2´, E3	si	si	perro, vaca y ratón
d(2)	E2, E3	si	–	–
RBP4 Ratón				
Variante UTR 5´	Combinación Exones	Evid. Expres. Roedores	Evid. Expres. Otros Mamif.	
a	E1, E3	si	si	cerdo y humano
c	E2´, E3	si	si	macaco
d	E2, E3	si	si	mono tití y humano

Tabla 7. *Evidencias de expresión de las UTR 5´de Rbp4 de humano y ratón en otros mamíferos*

Apom Humana				
Variante UTR 5´	Combinación Exones	Evid. Expres. Primates	Evid. Expres. Otros Mamif.	
d	E1	si	si	vaca, perro y ratón
d(2)	E1´	si	si	oveja
Apom Ratón				
Variante UTR 5´	Combinación Exones	Evid. Expres. Roedores	Evid. Expres. Otros Mamif.	
a	E1	–	si (p)	humano y perro

Tabla 8. *Evidencias de expresión de las UTR 5´de Apom de humano y ratón en otros mamíferos*

Los resultados de este análisis (ver tablas 5 a 8) nos muestran, en primer lugar, que entre las especies de primates así como entre las de roedores se están expresando las UTR5´alternativas encontradas en humanos y ratón respectivamente. En segundo lugar los resultados nos indican que en especies de órdenes diferentes a primates y roedores también existen evidencias de estar expresándose las UTRs 5´, para las cuales, ya habíamos encontrado evidencias genómicas de

exones ortólogos respecto a los de humano o ratón. Si observamos el caso de Apo-D (tabla 5) comprobamos que hay evidencias de expresión de UTR 5´donde están presentes bien los exones E1 y E4 de humano o E1 y E3 de ratón. Observamos resultados similares para el resto de lipocalinas, en las que previamente se había encontrado evidencia de exones ortólogos en la UTR 5´ (ver tablas 6 a 8).

3.1.3. - Alineamientos múltiples de UTRs 5`ortólogas

Para los casos en que se constató la existencia de lipocalinas con UTRs 5´ortólogas se realizaron alineamientos múltiples de las secuencias de diferentes especies de mamíferos. En las figuras 4, 5 y 6 pueden observarse los alineamientos múltiples más relevantes. Se corresponden estos con alineamientos de secuencias ortólogas a los transcritos humanos de Apo-D (variante "a"), Rbp4 (variante "b") y Apo-M (variante "d") . En dichos alineamientos observamos que a pesar de la existencia de inserciones, delecciones y sustituciones, existen bloques (en color azul) que muestran una considerable conservación (con un valor del PI que oscila entre el 70- 85% de identidad según los pares de secuencias ortólogas elegidos).

A partir del alineamiento múltiple obtenido para la UTR 5´ de Apo-D ha podido encontrarse evidencia de un hipotético exón específico para úrsidos. En la figura 4 se muestra dicho alineamiento. Se indica en esta figura la ubicación de los dos exones (E1 y E4) que constituyen esta variante de UTR 5´ de Apo-D. En este alineamiento podemos observar que no hay buena conservación en un fragmento del extremo 5´ de la región UTR 5´, excepto para oso y panda, para los que si hay buena conservación en este fragmento de unos 80 nt (en color verde en figura 4). Alineando este fragmento de 80 nt frente a su región genómica correspondiente, en oso y panda, encontramos que alinea en una región que está corriente arriba del exón E1, separado de él por un intrón de 874 nucleótidos, o sea, es un exón adicional de la UTR-5´ de Apo-D. Este exón se indica "E0" en la figura 4 . No hay evidencia de presencia genómica de dicho exón E0 en las otras especies de mamíferos, lo que es un hecho a favor de considerarlo un exón específico de úrsidos.

Cuando se compara el porcentaje de identidad (PI) entre los pares de regiones UTRs 5´ ortólogas (obtenidos a partir de los alineamientos múltiples) con el PI existente entre los correspondientes pares de secuencias codificantes (CDS) ortólogas , se observa que el PI de las primeras (que presentan un valor medio de 75,5 %) muestra valores semejantes al de las segundas en la 3ª

posición del codón (que presentan un valor medio de 72,5 %). Para el caso de Apo-D (UTR 5 ´variante "a"), sin embargo, los valores de PI entre los pares de UTR s 5´ortólogas si muestran valores claramente superiores a los de la tercera posición del codón. En la tabla 11 se muestran estos valores para algunos de los pares de especies de mamíferos estudiadas. Destaca el valor del PI en la pareja humano-vaca que es equiparable al PI de la 1ª y 2ª posición del codón, prueba del efecto que la selección natural está ejerciendo sobre esta región UTR 5´.

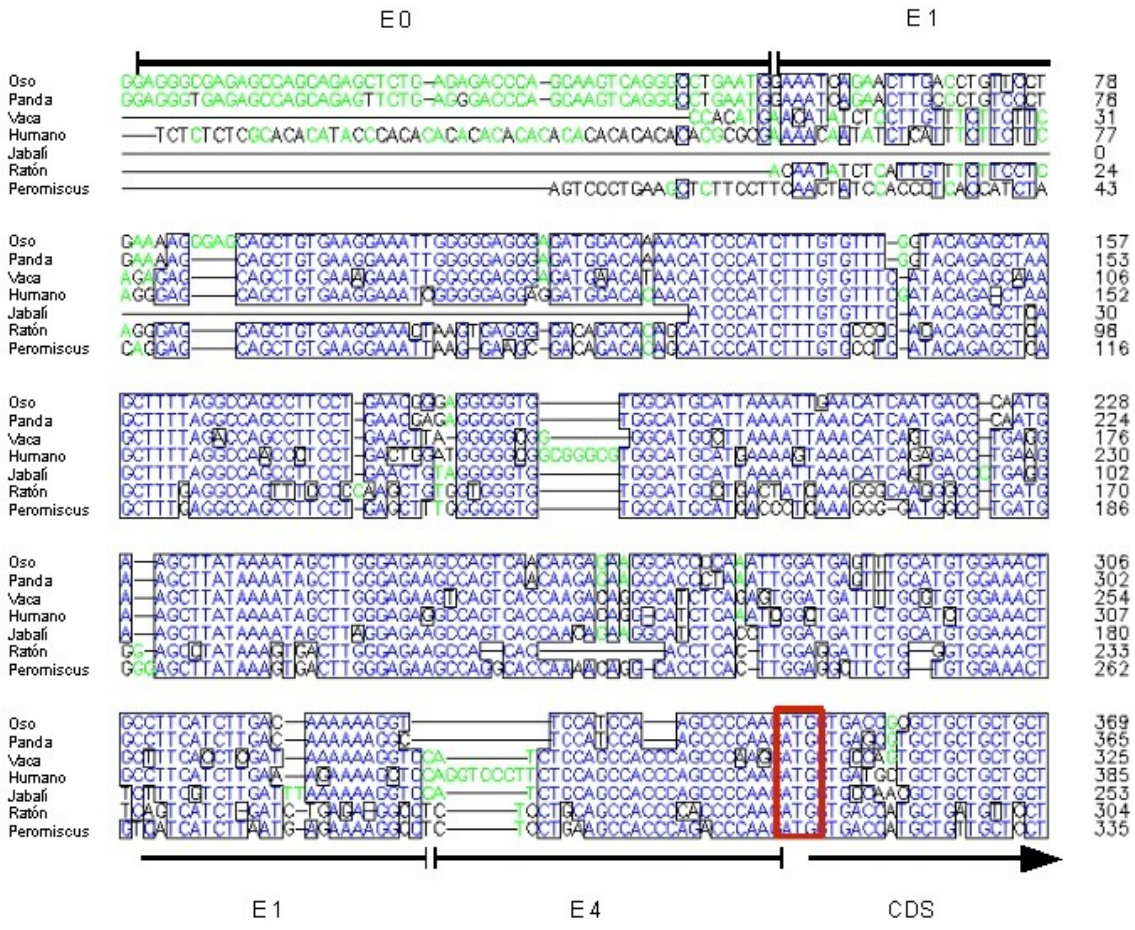


Figura 4. *Alineamiento múltiple de región UTR 5´ de transcritos ortólogos al de Apo-D humana, variante "a" . El recuadro rojo indica el codón de inicio. E1 y E4 indican los exones que constituyen esta variante. E0 indica un posible exón adicional de esta UTR 5´, específico de úrsidos.*

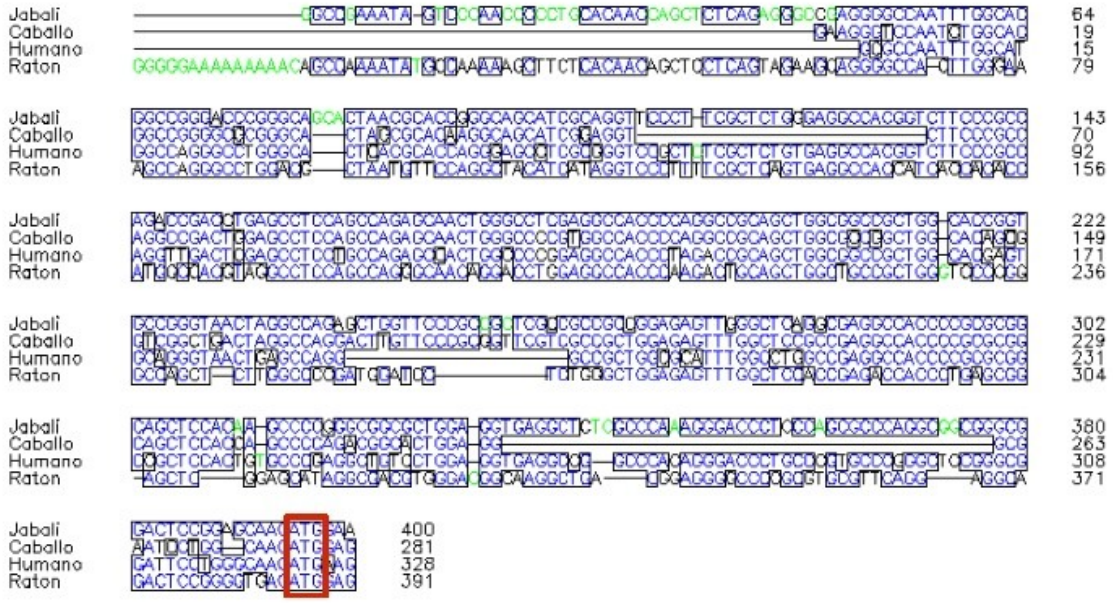


Figura 5. *Alineamiento múltiple de región UTR 5´ de transcritos ortólogos al de Rbp4 Humano variante "b". El recuadro rojo indica el codón de inicio.*

Figura 6. *Alineamiento múltiple de región UTR 5´ de transcritos ortólogos al de Apo-M Humano variante "d". El recuadro rojo indica el codón de inicio.*

Pares de especies	PI (%) CDS 1ª y 2ª posiciones de Apo-D	PI(%) CDS 3ª posición de Apo-D	PI(%) UTR 5´ de Apo-D (variante "a" humana)
Humano-Vaca	88,90	79,90	89,36
Humano-ratón	83,33	67,20	71,97
Vaca-ratón	86,77	65,61	73,28

Tabla 11. *Comparación de PI de UTRs 5´ortólogas respecto al PI de la 1ª, 2ª y 3ª posición de la región codificante correspondiente.*

3.2. - Conservación en la región UTR 3´

3.2.1. - Evidencias genómicas

Se procedió de la misma forma que con las UTRs 5´. Las regiones UTRs 3´ poseen una organización genómica más sencilla, por lo general no contienen intrones, excepto Ptgds y Lcn2, que presentan uno.

Los resultados obtenidos revelan que las regiones UTRs 3´ de humanos se encuentran muy bien conservadas (90% a 99% de Identidad) entre los primates. Ocurre lo mismo con las regiones UTRs 3´ de ratón cuando comparamos con otras especies de roedores (75 a 85% de identidad). Tras estudiar el grado de conservación con mamíferos de los otros órdenes (artiodáctilos y carnívoros), se pudo comprobar que existen, para algunas lipocalinas, regiones UTRs 3´ que pueden considerarse ortólogas (ver Tabla 10).

Al igual que ocurre en las regiones UTRs 5´ de las lipocalinas , hay con más frecuencia, conservación de las regiones UTRs 3´ entre diferentes especies de mamíferos en las lipocalinas más ancestrales. Si bien en el caso de las UTRs 3´existe una diferencia importante respecto de las regiones UTRs 5´. Cuando se encuentran regiones UTRs 3´ortólogas a las de primates en carnívoros

y artiodáctilos, no se encuentran, salvo la excepción de ApoM, regiones ortólogas en roedores (ver tabla 10).

Lipocalina	Exones UTR 3´	Long(pb)	% Ident Hum/ratón	% Ident Hum/artiod	% Ident Hum/carniv	% Ident Ratón/artiod	% Ident Ratón/carniv
APOD Hum							
	Único	198	-	91	89		
RBP4 Hum							
	Único	388	-	70	72		
APOM Hum							
	Único	121	70	78	-		
APOM Ratón							
	Único	117	70			68	-
LCN2 Hum							
	E1	153	-	70	75		
	E1´	334	-	-	-		

Tabla 10. *Evidencias genómicas de UTRs 3´ ortólogas entre diferentes especies de mamíferos*

3.2.2. - Evidencias de la expresión de UTRs 3´ ortólogas

El análisis mediante BLAST, sobre bases de datos de expresión de mamíferos, demuestra que las UTRs 3´de humano y de ratón, de las diferentes lipocalinas, son también expresadas en otras especies de primates y roedores respectivamente. Para los otros órdenes de mamíferos (artiodáctilos y carnívoros) solo se encontró evidencia clara de la expresión de UTRs 3´ para la UTR 3´de ApoD humana, en vaca, oveja y turón (O. aries y B. taurus y M. putorius).

3.2.3. - Alineamientos múltiples de UTRs 3´ortólogas

Se procedió a realizar los alineamientos múltiples de las secuencias UTRs 3´para las que se encontraron ortólogas entre diferentes órdenes de mamíferos , al igual que con las UTRs 5´. En este caso dado que solo hay evidencias claras de expresión para los ortólogos de la UTR 3´de ApoD humana (variante “a”) solo se muestra dicho alineamiento múltiple (ver figura 7) .

Al igual que en el caso de las UTRs 5´, a lo largo de esta UTR 3´, existen bloques con una elevada conservación entre especies. En este caso se obtienen valores de PI que oscilan entre el 80 y el 95 % según los pares de especies seleccionados, valores algo superiores a los obtenidos a partir de los alineamientos múltiples obtenidos para las UTRs 5´.

En el alineamiento múltiple de la UTR 3´de Apo-D puede observarse como se ha conservado el motivo (AATAAA) que muy probablemente sea la señal de poliadenilación.

Figura 5. *Alineamiento múltiple de región UTR 3´ de transcritos ortólogos al de ApoD Humano. En recuadro rojo se observa la conservación de la señal de poliadenilación.*

4. - Discusión

El grado de conservación encontrado (alrededor del 80% de identidad) entre las UTRs 5´ y 3´de lipocalinas ortólogas, de los diferentes grupos de mamíferos, puede considerarse elevado y se encuentra en consonancia con los mejores valores del porcentaje de identidad encontrado entre las UTRs para otros genes ortólogos estudiados [5 y 7]. El hecho de que se obtengan además alineamientos múltiples, igualmente con elevada identidad, es una prueba sólida de que existe conservación de las UTRs de algunas lipocalinas entre los diferentes linajes de los mamíferos. Hemos de interpretar la existencia de esta conservación como una prueba de la funcionalidad de

estas UTRs, las cuales deben desempeñar un importante papel en la regulación de la expresión génica.

El hecho de que, para las lipocalinas con varios exones en su UTR 5´, algunos se hayan conservado entre diferentes órdenes y otros no, por lo que estos serían exones específicos dentro de un determinado orden (primates, roedores, etc), es un fenómeno que ha podido observarse en otros estudios realizados sobre la región UTR 5´de diferentes genes que muestran diversidad en dicha región [5 y 6].

Al igual que tenemos la evidencia de la presencia de exones específicos en las UTRs 5´ de humano o ratón, como se desprende del estudio realizado, es muy probable que existan otros exones específicos desconocidos en la UTRs 5´de estas lipocalinas, propios de los otros órdenes como artiodáctilos y carnívoros. Esto ha podido ponerse de manifiesto al encontrar, a partir del análisis del alineamiento múltiple de Apo-D, evidencias de uno de estos exones específicos en especies de úrsidos.

El escenario que nos muestran los resultados encontrados para la conservación de las UTRs 5´ es que existe conservación de parte de la arquitectura de la UTR 5´ entre las lipocalinas ortólogas que muestran una UTR 5´más compleja (que son las evolutivamente más antiguas) y nula conservación en las lipocalinas con una UTR 5´ simple (que son las evolutivamente más recientes). Si bien incluso en las lipocalinas que muestran cierta conservación parece haberse dado cierta divergencia entre los diferentes linajes de mamíferos, hemos de suponer que en función de las diferentes necesidades de regulación de la expresión de dichas lipocalinas.

Respecto a las UTRs 3´, también los resultados muestran que se da una mayor conservación de estas en las lipocalinas más ancestrales. Se da la peculiaridad de que, con la excepción de ApoM, los roedores no muestran secuencias UTRs 3´ortólogas a las de los otros grupos (primates, artiodáctilos y carnívoros) (ver tabla 10). Este hecho nos sugiere que en los roedores, las regiones UTRs 3´ de las lipocalinas han sufrido un camino evolutivo diferente derivando en una mayor divergencia respecto a los otros órdenes.

5. - Bibliografía

[1] Kent, W.J. . BLAT -- The BLAST-Like Alignment Tool. *Genome Research* **4**: 656-664 (2002)

[2] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. . Basic local alignment search tool. *J. Mol. Biol.* **215,** 403-410 (1990)

[3] Rice, P., Longden, I.. and Bleasby, A. . EMBOSS The European Molecular Biology Open Software Suite. *Trends in Genetics* **16***,* 276-277 (2000).

[4] Thompson, J.D., Higgins, D.G. and Gibson, T.J. . CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680 (1994)

[5] Martin, M. S., et al. Characterization of 5′ untranslated regions of the voltage-gated sodium channels SCN1A, SCN2A, and SCN3A and identification of cis-conserved noncoding sequences. *Genomics* **90,** 225-235 (2007).

[6] Tan, J.S., Mohandas, N., Conboy, J.G. . Evolutionary conserved coupling of transcription and alternative splicing in the EPB41 and EPB41L3 genes. *Genomics*. **86,** 701-707 (2005).

[7] Shabalina, S. A., et al. Comparative analysis of orthologous eukaryotic mRNAs: potential hidden funtional signals. *Nucleid Acids Research* **32,** 1774-1782 (2004).

IV

PAPEL REGULADOR DE LAS UTRs 5´ DE LIPOCALINAS DE MAMÍFEROS

1. - Objetivos

El objetivo de este capítulo es obtener información sobre cuales son los posibles elementos reguladores presentes en las UTRs 5´de las lipocalinas de mamíferos, para conocer en que forma estas regiones ejercen su regulación sobre la expresión génica. Dado que se ha encontrado que cierto número de lipocalinas presentan formas alternativas en sus UTRs 5´ es de especial interés conocer las diferencias de elementos reguladores existentes en ellas para dilucidar las distintas formas de regulación que estas podrían estar ejerciendo.

Para abordar este análisis este se ha recurrido a diversas herramientas bioinformáticas de probada eficacia. Estas herramientas buscan regularidades o patrones confirmados experimentalmente o procedentes de estudios a escala genómica. Siempre que haya sido posible, se ha comprobado si existe conservación evolutiva, dado que dicha conservación implicaría una necesidad funcional.

2. - Métodos

2.1.- Clasificación de las UTRs 5´

Para una primera aproximación a la clasificación de las UTRs 5´de lipocalinas se utilizó el modelo CART [29]. Este modelo de regresión, basado en el análisis de 2312 UTR 5´ humanas, permite clasificar las UTR 5´en tres categorías: la **Clase I** se corresponde con ARNm que son escasamente traducidos, la **Clase II** con ARNm que son regulados en relación con etapas del desarrollo (TOP regulated RNAm) y la **Clase III** se corresponde con ARNm que sufren escasa regulación y dan lugar a una alta traducción.

Dichas categorías son establecidas en función de los valores de parámetros de las UTRs 5´ como la longitud, el contenido en G+C, la energía libre (ΔG) de plegamiento mínima (MFE), la presencia de uAUGs y la presencia del motivo TOP (terminal oligopyrimidine tract) [30]. El modelo ofrece un *árbol de decisión,* basado en los parámetros citados, que puede aplicarse para clasificar las UTRs 5´. Dicho árbol de decisión se muestra en el apartado de resultados.

La longitud, el % de G+C y la presencia de uAUGs de las UTRs 5´fueron determinados con los

recursos de EMBOSS [28]. La MFE se calculó con RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) [6]. La presencia del motivo TOP se determinó con UTRscan de UTRsite (http://utrsite.ba.itb.cnr.it/) [1]. Todos estos parámetros fueron introducidos en una hoja de cálculo para finalmente aplicar los criterios del árbol de decisión de CART y establecer la categoría de cada UTR 5´.

Dado que el motivo TOP es característico de cierto tipo de genes (ARNm de proteínas ribosómicas o factores de elongación), en las UTRs 5´alternativas en las que las predicciones identifican dicho motivo y el arbol de decisión nos lleva a considerarlas de clase II, se ha tomado con cierta reserva el que este motivo sea funcional en las UTR 5´de las lipocalinas. Por ello se ha añadido en las tablas de resultados 1 y 2 la clase opcional a la que pertenece la correspondiente UTR 5´, si elegimos en el arbol de decisión la no presencia de TOP.

Aunque el modelo CART ha sido elaborado para UTRs 5´ humanas, se utilizó también para ratón ya que las UTRs muestran propiedades comunes en mamíferos.

2.2.- Búsqueda de motivos validados, en las UTRs 5´

Se utilizó para este propósito la herramienta UTRscan. Dicha herramienta pertenece a la base de datos UTRsite (http://utrsite.ba.itb.cnr.it/) [1], la cual contiene una colección de patrones de secuencias funcionales de las regiones UTR 5´y 3´. UTRscan analiza las secuencias objetivo aplicando patrones de secuencia y/o estructura de motivos validados y ofrece los resultados de las coincidencias encontradas.

Para contrastar las predicciones de motivos estructurales predichos por UTRscan en las UTRs 5´se utilizaron las herramientas bioinformáticas para ARNs estructurales: RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) [6] , RNAshape (http://bibiserv2.cebitec.uni-bielefeld.de/rnashapes?id=rnashapes_view_submission) [7] y RNAlocomotif (http://bibiserv.techfak.uni-bielefeld.de/locomotif/submission.html) [8]. La forma en que se utilizaron estas herramientas se detalla en los resultados.

2.3.- Determinación de oligonucleótidos sobrerrepresentados

Se procedió de la siguiente forma: Se tomaron todas las secuencias UTRs 5´de lipocalinas de humano y ratón y se eliminaron las variantes que suponían alguna redundancia (debido a la

semejanza en la composición de exones alternativos de su UTR 5´). Posteriormente se sometieron estas secuencias (humanos por un lado y ratón por otro) al análisis de óligos mediante el algoritmo "oligo-analysis" de RSA TOOLS (http://rsat.ulb.ac.be/) [9].

Esta herramienta permite realizar múltiples selecciones y ajustes en el análisis de oligonucleótidos. Se seleccionaron las siguientes opciones: 1) Realizar una purga de las secuencias (se eliminan así las regiones duplicadas de más de 40 nucleótidos) que pueden provocar una desviación en las estimaciones de frecuencia de diferentes "palabras", 2) Prevenir patrones solapados, se evita así la desviación provocada por la tendencia que tienen los patrones periódicos a aparecer agrupados, 3) Se eligió la opción de oligos de 6, 7 y 8 nucleótidos, 4) El modelo de referencia elegido para la validación de los oligonucleótidos fue la composición de nucleótidos corriente arriba de los genes de humano y ratón respectivamente.

2.4. - Identificación de dianas de miARN

Con el objeto de predecir la existencia de dianas de miARN en las regiones UTRs 5´de las lipocalinas, se sometieron dichas regiones a un análisis mediante el algoritmo PITA de Segal Lab (http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html) [13]. Dicho algoritmo considera, además del apareamiento de bases entre el miARN y su diana, aspectos energéticos sobre la accesibilidad a dicha diana, considerando para estos cálculos además ambos flancos de la diana. La importancia de considerar estos aspectos energéticos ha sido puesta de manifiesto al comprobar que se mejoran así las predicciones sobre el nivel de regulación ejercido por ciertos miARNs, determinados experimentalmente [13].

Este algoritmo calcula un parámetro "ddG" que es el resultado de la diferencia entre la variación de energía libre del apareamiento entre el miARN y su diana (dGduplex) y la variación de energía libre de deshacer la estructura secundaria (2D) que contiene a la citada diana (dGopen). Los valores negativos de "ddG" (ddG = dGduplex - dGopen) indicarían uniones miARN-diana energéticamente favorables.

Al aplicar el algortimo PITA se eligieron las siguientes opciones: 1) Búsqueda de todos los miARN conocidos, para humano o ratón, 2) Tamaño mínimo de "seed" = 8, permitiendo un solo apareamiento U-G y una única posición desapareada, 3) Se eligió la opción de considerar los flancos a la diana para el cálculo energético de accesibilidad (ddG) a la misma.

Una vez obtenidas las predicciones se ha considerado que una diana es “accesible” si el parámetro de “ddG ” es < -10 Kcal/mol, teniendo en cuenta los valores que se alcanzan para miARNs validados[13]. Si además de este valor la diana posee un “ dGopen “ >= -10 Kcal/mol (que indica que su estructura 2D puede deshacerse con facilidad), podemos considerarla como un candidato a diana muy probable, ya que sería “especialmente accesible” [13].

Cuando en la UTRs 5´se han encontrado varias dianas de un mismo miARN que se solapan, solo se ha contabilizado como una, eligiendo la que presenta un valor de ddG más favorable.

2.5. - Estudio de los uAUGs y uORFs en las UTRs 5´

Se determinaron los uAUGs y la posición que ocupan en las UTRs mediante “dreg” de EMBOSS [28]. La fortaleza del contexto de AUGs y de los uAUGs se determinó teniendo en cuenta si en las posiciones -3 y +4 (respecto a posición +1 que es la “A” del AUG) se encuentran los nucleótidos adecuados para la iniciación de la traducción (“A o G” en posición -3 y “G” en posición +4) [18 y 19]. Si se encuentran los nucleótidos adecuados en las dos posiciones citadas se considera contexto “óptimo”, si sólo se encuentra uno de ellos “adecuado” y si ninguno “inadecuado”. Para la determinación de los uORFs en las UTRs 5´ se utilizó “Getorf” de EMBOSS [28], seleccionando en las opciones: encontrar uORFs entre un AUG y cualquier codón de terminación, en cualquier fase de lectura y con al menos 3 codones de longitud.

Para el estudio de la conservación de estos elementos se procedió de la siguiente forma. Se seleccionaron las UTRs 5´de lipocalinas ortólogas de diferentes mamíferos que tuviesen al menos 200 nt de longitud, que presentasen uAUGs y/o uORFs y con la condición de que las UTRs ortólogas no difiriesen demasiado en su longitud. Para considerar que un uORF está conservado, o sea es ortólogo de otro, se tomó el doble criterio de que la distancia entre el uAUG y el AUG principal y el número de codones contenidos en dichos uORFs fuesen semejantes [22]. Estos criterios fueron utilizados más flexiblemente para los uORFs que solapan con la región codificante, dado su carácter especialmente inhibidor.

Para encontrar indicios de “péptidos bioactivos”, resultantes de los uORFs de las UTRs 5´de lipocalinas, se realizó un análisis sobre los uORFs ortólogos, previamente encontrados en ellas. Se seleccionaron los casos con uORFs ortólogos en al menos tres especies de mamíferos, con al menos 10 codones y que no difiriesen mucho en longitud. Posteriormente se alinearon las secuencias de estos uORFs, para finalmente extraer los que muestran un grado considerable de conservación. Una

vez seleccionadas así, los grupos de secuencias de uORFs ortólogas se sometieron al análisis de sustituciones sinónimas y no sinónimas (ps/pn) mediante el programa SNAP (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html) [25], que se basa en el método de Nei y Gojobori [34]. Los valores de ps/pn mayores que "1" han sido tomados como prueba de selección natural positiva.

3. - Resultados

3.1.- Aproximación a la clasificación de las UTR 5´de lipocalinas

La longitud, el contenido en G+C y la estructura secundaria de la UTR 5´ paracen jugar un importante papel en la regulación de la expresión génica. Existen evidencias de que las UTRs 5´ que permiten una traducción eficiente son cortas, con bajo G+C, son relativamente desestructuradas y no contienen codones de inico corriente arriba uAUG [31]. Mientras que las que inhiben la síntesis de proteínas son largas, con alto G+C y poseen un alto grado de estructura secundaria [32].

Para hacer una primera aproximación a las funciones que cumplen las UTR 5´de las lipocalinas podemos clasificarlas en función de los parámetros citados, así como de la posible presencia del motivo TOP (Terminal Oligopyrimidine Tract). Dicho motivo se encuentra en el extremo 5´ terminal de los ARNm de proteínas ribosómicas y de factores de elongación de vertebrados [30]. Este motivo consiste en una serie de 5 a 15 pirimidinas, entre una citosina y una guanina (C (PY)n G). Mediante la unión de una proteína a dicho motivo y en ciertas condiciones (diferenciación, desarrollo o bajo ciertos tratamientos con fármacos) interviene en la represión de la traducción [3].

Para hacer una clasificación de las UTRs 5´ de forma más objetiva se utilizo el "árbol de decisión" del modelo de regresión CART [9] (ver métodos). Puede verse el árbol de decisión en la Figura 1.

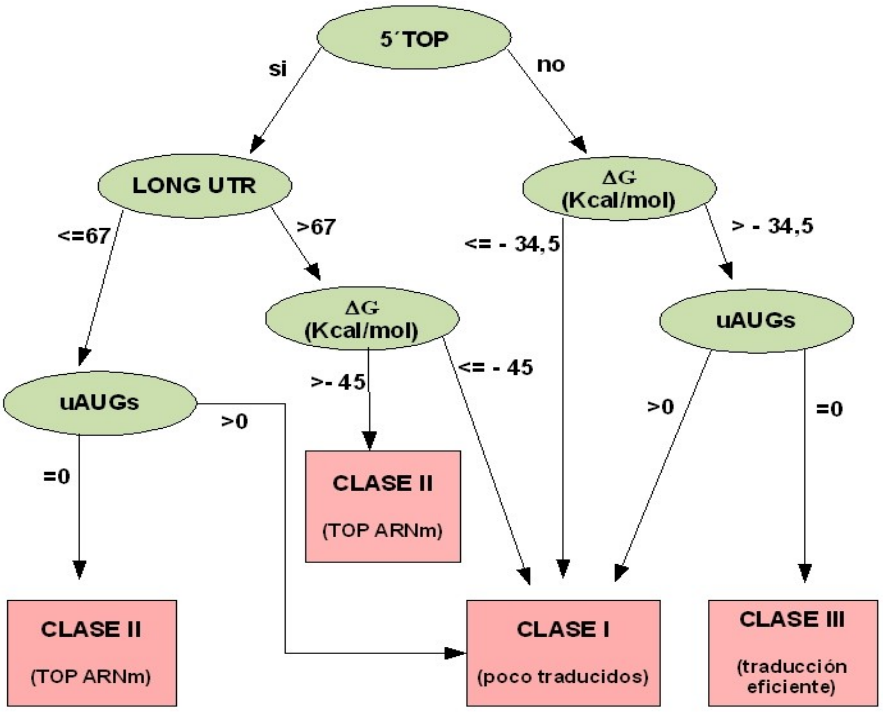


Figura 1: *Árbol de decisión para la clasificación de las UTR 5´ del modelo de regresión CART*

Una vez calculados los parámetros necesarios de las secuencias de las UTRs 5´de las lipocalinas y aplicando el arbol de decisión del modelo CART obtenemos la clasificación de dichas UTRs que se observa en las Tablas 1 y 2.

Observamos que tanto para humano como para ratón son frecuentes las UTRs 5´de "clase I" (potencialmente inhibitorias de la traducción), especialmente entre las lipocalinas de origen evolutivo más antiguo, que son las que presentan mayor número de alternativas (Apo-D, Ptgds, Rbp4). Mientras que las UTRs 5´de tipo III (potencialmente de traducción intensa) son más frecuentes entre las lipocalinas con un origen evolutivo más reciente, que muestran menor número de alternativas. Esta tendencia se observa de forma más evidente en ratón (ver tabla 2).

Lipocalina	long	%GC	MFE	nºuAUG	TOP	CART class
5utr_APOD.a_hum	361	51.25	-102.2	4	s	I
5utr_APOD.b_hum	232	53.88	-87.5	1	n	I
5utr_APOD.c_hum	135	47.41	-35.5	2	n	I
5utr_APOD.d_hum	190	52.11	-60.8	3	n	I
5utr_Ptgds.c_hum	72	69.44	-20.5	0	s	II / III
5utr_Ptgds.g_hum	458	65.5	-203.8	4	n	I
5utr_Ptgds.j_hum	1283	65.55	-644.51	11	n	I
5utr_RBP4.b_hum	322	72.67	-179.5	1	n	I
5utr_RBP4.d_hum	72	77.78	-28.7	0	n	III
5utr_RBP4.d(2)_hum	113	76.99	-46.4	0	n	I
5utr_Apom.d_hum	496	49.6	-161.77	6	n	I
5utr_Apom.e_hum	73	58.9	-17.4	0	n	III
5utr_C8G_a_hum	75	64	-21.5	0	n	III
5utr_OBP2A_b_hum	42	69.05	-15.2	0	n	III
5utr_ORM2_b_hum	189	56.61	-76.5	0	n	I
5utr_ORM2_d_hum	63	52.38	-10.2	2	s	I
5utr_LCN1.b_hum	154	63.64	-63.4	0	n	I
5utr_LCN1.h_hum	49	69.39	-9	0	n	III
5utr_LCN2.b_hum	72	67.71	-13	0	n	III
5utr_LCN2.b(2)_hum	96	67.71	-25.3	0	n	III
5utr_LCN8.e_hum	348	67.82	-166.1	3	n	I
5utr_LCN12.c_hum	248	59.27	-99.8	1	n	I
5utr_LCN12.c(2)_hum	28	71.43	-8.1	0	n	III

Tabla 1: *Clasificación de las UTR 5´de lipocalinas humanas según el modelo CART. Cuando existen UTR 5´alternativos aparecen con una letra distintiva después del nombre de la lipocalina correspondiente.*

Lipocalina	long	%GC	MFE(Kcal/mol)	nº uAUG	TOP	CART class
5utr_Apod.a_mouse	140	50.71	-42	0	n	I
5utr_Apod.b_mouse	224	49.55	-69.1	1	n	I
5utr_Apod.c_mouse	358	51.4	-126.94	2	n	I
5utr_Apod.d_mouse	281	53.74	-103.24	2	n	I
5utr_Apod.e_mouse	214	59.35	-74.4	2	n	I
5utr_Ptgds.c_mouse	80	56.25	-20.2	0	s	II / III
5utr_Ptgds.d_mouse	329	58.36	-149.23	2	s	I
5utr_Rbp4.a_mouse	385	62.08	-148.1	4	n	I
5utr_Rbp4.c_mouse	61	77.05	-18.7	0	n	III
5utr_Rbp4.d_mouse	89	77.53	-37.7	0	n	I
5utr_Apom.a_mouse	781	48.27	-240.6	14	n	I
5utr_C8G_b_mouse	501	53.69	-188.12	6	n	I
5utr_C8G_c_mouse	275	53.82	-104.22	4	n	I
5utr_LCN13_a_mouse	53	62.26	-17.5	0	n	III
5utr_ORM2_a_mouse	41	53.66	-3.4	0	n	III
5utr_Vegp1.a_rat	55	54.55	-18.1	0	n	III
5utr_Lcn2.b_mouse	54	53.7	-7.3	0	n	III
5utr_Lcn8.a_mouse	23	65.22	-3.6	0	n	III
5utr_Lcn12.a_mouse	55	69.09	-30.3	0	n	III

Tabla 2: *Clasificación de las UTR 5´de lipocalinas de ratón según el modelo CART. Cuando existen UTR 5´alternativos aparecen con una letra distintiva después del nombre de la lipocalina correspondiente.*

Aplicando el citado árbol de decisión solo se obtienen UTRs 5´de la clase II (reguladas por el motivo TOP) en una de las UTRs alternativas de la lipocalina Ptgds, tanto en humano como en ratón (ver tablas 1 y 2). La clase opcional a la que pertenecerían estas UTRs 5´, de no ser funcional el motivo TOP, sería la clase III.

3.2.- Búsqueda de motivos validados, en las secuencias de UTRs 5´.

Para la búsqueda de posibles motivos reguladores en la región UTR 5´ de las lipocalinas humanas y de ratón se sometieron dichas regiones a un análisis mediante la herramienta UTRscan [1]. El resultado de este análisis se muestra en la tabla 3. Los motivos predichos son el motivo "TOP" (Terminal Oligopyrimidine Tract), ya mencionado previamente y el motivo "IRES" (Internal Ribosome Entry Site). También son predichos marcos de lectura abierta (uORF) en las UTRs 5´ de diversas lipocalinas, pero no se muestran aquí ya que estos elementos son tratados específicamente en un apartado posterior.

UTRs 5´	MOTIVO	POSICIÓN	SECUENCIA
5utr_APOD.a_hum	TOP	[1-8]	CTCTCTCG
5utr_Ptgds.c_hum	TOP	[1-7]	CCTCCTG
5utr_Ptgds.c_mouse	TOP	[6-14]	CTCCTTCTG
5utr_ORM2_d_hum	TOP	[1-7]	C TCTTTG
5utr_RBP4.b_hum	IRES	[240,322]	CTGTG CC CGAGG CT GTCCT GGAGGTG AGGCC GGCCC ACA GGGAC CCTGC CCGTG CCCGG GC TCCGG GCGGATT CCTGG GCAAG

Tabla 3. *Resultado del análisis de las UTRs 5´de lipocalinas humanas y de ratón mediante UTRscan*

Respecto al motivo TOP, ya hemos mencionado que se presenta en la UTR 5´del ARNm de genes específicos y en ciertas condiciones [30]. Más recientemente se han encontrado evidencias de que el motivo TOP no es exclusivo de los genes mencionadas, sino que puede ejercer también funciones reguladoras en la expresión de otros diferentes [4]. Ante estas evidencias no podemos desechar la posibilidad de que los motivos TOP predichos en las UTRs 5´ de lipocalinas sean funcionales y cumplan, en mayor o menor medida, una función de represión de la tranducción.

El motivo IRES supone un mecanismo alternativo al inicio de traducción convencional y por lo tanto independiente del " 5´cap ". Inicialmente fué descubierto en picornavirus y posteriormente

también ha sido encontrado en ARNm eucariotas. El estudio comparativo de los IRES de diversos ARNm celulares ha permitido identificar, en su región UTR 5´, un motivo estructural común (ver figura 2) formado por una estructura típica en "Y", seguida de estructuras en horquilla justo antes del codón de inicio de la traducción [5].

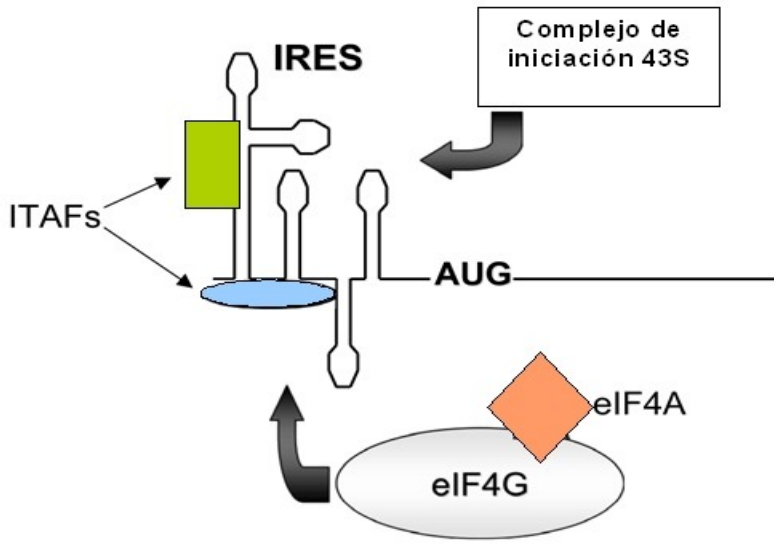


Figura 2. *Representación equemática del inicio de traducción mediada por IRES. Se muestran las estructuras secundarias típicas de este motivo (líneas en forma de horquillas). También se muestran los principales factores (eIF4A y 4G), así como las proteínas específicas ITAFs (IRES trans-acting factors) necesarias para la funcionalidad de este motivo.*

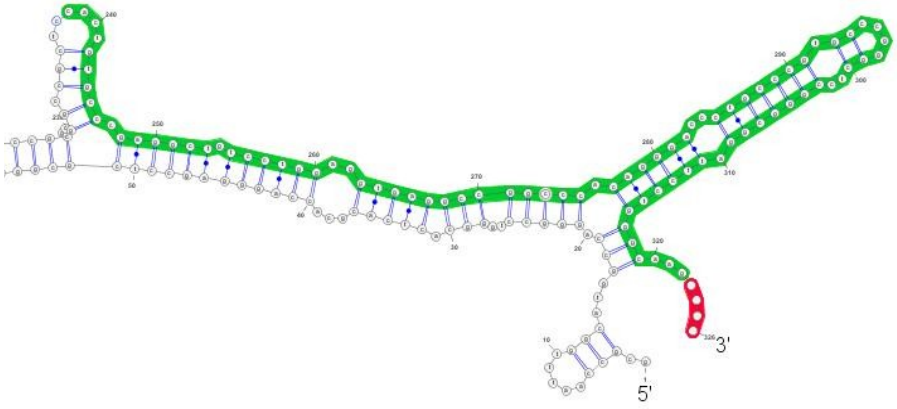


Figura 3. *Estructura 2d de la UTR 5´de Rbp4 humana (variante b) obtenida con RNAfold . Se muestra detalle de la región (en verde) donde UTRscan predice la presencia de un IRE. En rojo la posición que ocuparía el codón de inicio de la secuencia codificante.*

Con objeto de comprobar, en qué medida la predicción del motivo IRES en la UTR 5´de la variante "b" de Rbp4 humana presenta la estructura típica de dicho motivo, se procedió a obtener la estructura secundaria más estable (MFE) de dicha UTR 5´mediante RNAfold [6]. En la figura 3 se muestra un detalle de la estructura 2D de la región donde es predicho el motivo IRES (marcado en verde). Observamos que no muestra la complejidad estructural típica de dicho motivo, destacando la ausencia de la estructura típica en "Y". Dado que la estructura secundaria nativa no tiene por qué ser la estructura con menor energía de plegamiento, sino que puede ser una estructura subóptima próxima en el conjunto de estructuras posibles, se procedió a estudiar dichas estructuras mediante RNAshape (ver introducción) [7]. En ninguna de las estructuras representativas posibles que ofrece esta herramienta, dentro de su repertorio, se encontró en la región adecuada, una estructura secundaria similar a la del motivo IRES. También se procedió de forma inversa, es decir comprobar en qué medida esta UTR 5´puede llegar a formar una estructura semejante al motivo IRES. Se utilizó para ello RNA locomotif (ver introducción) [8] y el resultado obtenido es que la secuencia de la UTR 5´de Rbp4 humana (b) es incapaz de formar un motivo estructural típico del motivo IRES, dentro de sus posibilidades de plegamiento alternativas.

Si bien no encontramos semejanza de las estructuras que puede adquirir esta UTR 5´de Rbp4 humana con el motivo IRES, sí observamos la presencia de una estructura en horquilla, justo antes del codón de inicio. No podemos desechar la idea de que esta estructura tenga algún papel relevante en la regulación de la traducción. El análisis de estas y otras estructuras peculiares de las UTRs 5´de lipocalinas de mamíferos será realizado de manera más amplia y detallada en un capítulo específico de la tesis.

3.3.- Oligonucleótidos sobrerrepresentados

El análisis de los oligonucleótidos sobrerrepresentados es una manera de encontrar posibles motivos implicados en alguna función reguladora. Para detectar oligonucleótidos sobrerrepresentados, en el conjunto de las UTRs 5´de lipocalinas, se utilizó el algoritmo de análisis de oligonucleótidos "oligo analysis", disponible en "RSA tools" [9]. Se analizaron palabras de 6, 7 y 8 nucleótidos en las secuencias de las UTRs 5´de lipocalinas humanas y de ratón (ver métodos).

Los resultados de aplicar dicho algoritmo a las UTRs 5´de las lipocalinas de estas dos especies demuestran que hay un cierto número de oligonucleótidos sobrerrepresentados (con mayor número en humano, ver datos en tablas 4 y 5), que pueden considerarse significativos por su baja probabilidad esperada de ocurrencia, la cual viene dada por un "*indice de significación*" mayor que "0".

El análisis de los datos de las UTRs 5´de lipocalinas humanas revela que algunos oligos (de igual o diferente longitud) solapan entre sí, siendo esta una propiedad que muestran diversos motivos confirmados experimentalmente [9]. Además, la circunstancia de que estos oligos sean los que ocupan las primeras posiciones respecto a su significación, nos permite considerar a estos como los candidatos más probables a representar motivos reguladores.

En la tabla 6 se han representado dichos motivos, indicándose en mayúsculas el oligo que muestra un nivel de significación mayor dentro de cada hipotético motivo. Observamos en dicha tabla que no se produce un aumento de la significación con el aumento del tamaño del óligo, sino que el mayor nivel se encuentra en el tamaño de 6 o 7 nucleótidos. Esto es un indicio de que estos motivos no deben formar parte se secuencias de mayor tamaño, sino que tienen su propia identidad, en un rango de tamaño entre 6 y 8 nucleótidos.

UTR 5´lipocalinas humanas			
oligo 6 nt			
Oligo	**Observado**	**Esperado**	**Ind. Signific**
TGCCAG	16	2.77	3.77
CTGGCA	16	2.77	3.77
GGGTGG	17	4.15	2.13
CCACCC	17	4.15	2.13
CCTGGC	20	5.69	2.02
GCCAGG	20	5.69	2.02
AGGGCC	13	2.61	1.82
GGCCCT	13	2.61	1.82
CCCGGA	9	1.29	1.44
TCCGGG	9	1.29	1.44
GGTCCC	11	2.18	1.13
GGGACC	11	2.18	1.13
GCAGGG	14	3.96	0.55
CCCTGC	14	3.96	0.55
GCCCAG	17	6.03	0.12
CTGGGC	17	6.03	0.12
CCCAGC	20	7.82	0.12
GCTGGG	20	7.82	0.12
GACTCG	5	0.52	0.08
CGAGTC	5	0.52	0.08
GGCCAG	14	4.41	0.08
CTGGCC	14	4.41	0.08
oligo 7 nt			
Oligo	**Observado**	**Esperado**	**Ind. Signific**
GCCACCC	8	0.74	1.71
GGGTGGC	8	0.74	1.71
GGCCCTG	9	1.18	1.17
CAGGGCC	9	1.18	1.17
CTGGCAG	8	0.94	0.98
CTGCCAG	8	0.94	0.98
CCAGCTC	8	1.11	0.44
GAGCTGG	8	1.11	0.44
GCTGTCC	6	0.57	0.32
GGACAGC	6	0.57	0.32
GGTCCCT	6	0.59	0.26
AGGGACC	6	0.59	0.26
GCCCGGA	5	0.36	0.21
TCCGGGC	5	0.36	0.21
CCGCTGG	5	0.36	0.19
CCAGCGG	5	0.36	0.19
oligo 8 nt			
Oligo	**Observado**	**Esperado**	**Ind. Signific**
AGGGACCT	5	0.16	1.27
AGGTCCCT	5	0.16	1.27
CTGGCAGG	6	0.31	1.24
CCTGCCAG	6	0.31	1.24
CCCAGCAA	5	0.24	0.48
TTGCTGGG	5	0.24	0.48
GCCGCTGG	4	0.13	0.15
CCAGCGGC	4	0.13	0.15

Tabla 4 . *Oligonucleótidos sobrerrepresentados, de tamaño 6, 7 y 8 nt, en las UTRs 5´de lipocalinas humanas*

UTR 5´lipocalinas ratón			
oligo 6 nt			
Oligo	**Observado**	**Esperado**	**Ind. Signific**
CTTGGG	12	2.04	2.18
CCCAAG	12	2.04	2.18
GGGTGG	11	2.54	0.54
CCACCC	11	2.54	0.54
oligo 7 nt			
	Observado	**Esperado**	**Ind. Signific**
TGCCCAG	6	0.64	0.04
CTGGGCA	6	0.64	0.04

Tabla 5 . *Oligonucleótidos sobrerrepresentados, de tamaño 6 y 7 nt, en las UTRs 5´de lipocalinas de ratón.*

Motivo	**Longitud oligonucleótido**		
	6	**7**	**8**
CTGGCAgg	<u>3,77</u>	1,71	1,24
ccTGCCAGg	<u>3,77</u>	0,98	1,24
gCAGGGCC	0,55	<u>1,17</u>	-
gCCACCC	<u>2,13</u>	1,71	-

Tabla 6 . *Motivos hipotéticos, formados por solapamiento de oligonucleótidos, en las UTRs 5´de lipocalinas humanas. Para cada motivo se indica el índice de significación del oligonucleótido más representativo. Se muestra subrayado el índice de mayor valor en cada caso.*

En la tabla 7 se muestran las lipocalinas humanas en cuyas UTRs 5´aparecen los oligos más

significativos, referidos a los oligos de la tabla 6. Observamos que dichos oligos se dan con más frecuencia entre las lipocalinas más ancestrales como Apo-D, Apo-M, Ptgds y Rbp4, aunque alguno de los oligos se presentan también en lipocalinas evolutivamente más recientes (Lcn1, Lcn2, Lcn8, Lcn12 y ORM).

En el caso de los oligos sobrerrepresentados de ratón comprobamos que uno de ellos "CCACCC" está tambien sobrerrepresentado en humano y lo hace en las UTRs 5´de algunas lipocalinas comunes a ambas especies (ver tabla 8). Otro de los óligos de ratón "CTGGGCA" es muy semejante al oligo "CTGGCA" humano, e igualmente se presenta en algunas lipocalinas comunes a ambas especies (ver tabla 8).

Oligos	Lipocalinas Humanas
CTGGCA	Ptgds (j); Rbp4 (b) y Orm (b)
TGCCAG	Ptgds (j); Rbp4 (b), Apo-M (d); Lcn2 (b2); Lcn12 (c); Lcn1 (b) y Lcn8 (e)
CAGGGCC	Ptgds (j); Rbp4 (b) y Lcn8 (e)
CCACCC	Apo-D (a, b, c y d); Ptgds (j); Rbp4 (b) y Lcn2 (b2)

Tabla 7. *Oligonucleótidos más representativos en humano y variantes en las que aparecen*

Oligos	Lipocalinas de Ratón
CTGGGCA	Apo-D (b, c ye); Ptgds (d) y Rbp4 (a)
CCACCC	Apo_D (e); Rbp4 (d) y Apo-M (a)

Tabla 8. *Oligonucleótidos más representativos en ratón y variantes en las que aparecen*

Se analizó la posición que ocupan los oligos, mencionados en las tablas 7 y 8, en las UTRs 5´de las diferentes lipocalinas donde se presentan y no pudo encontrarse, en ninguno de ellos, una tendencia clara a ocupar una posición preferente en dichas regiones.

Para el caso concreto de Apo-D, si se constata que el motivo “CCACCC” aparece en el extremo 3´de la UTR 5´, tanto en humano como en ratón. En el alineamiento múltiple de las regiones UTRs 5´ ortólogas de Apo-D de mamíferos puede identificarse dicho motivo dentro de un bloque donde la secuencia esta muy bien conservada (ver figura 4). Al introducir dicho motivo en la base de datos de miARNs TargetScan [14], si bien no se correponde con la diana de ningún miARN conocido, es clasificado como un sitio conservado de un posible miARN en las UTRs 3´ortólogas de 103 genes humanos.

Figura 4. *Alineamiento múltiple de secuencias ortólogas de la UTR 5´ de Apo-D de mamíferos. Sólo se muestra la región donde se encuentra el motivo CCACCC. En verde bloque donde se encuentra el motivo, en rojo codón de inicio de la secuencia codificante.*

Al introducir los otros motivos de la tabla 7 en TargetScan [14] se obtiene para “CTGGCA” y “TGCCAG” la clasificación de sitios conservados de nuevos miARNs en las UTRs 3´ ortólogas de más de 100 genes humanos. Para el motivo “CAGGGCC” se obtiene que es una diana de miR-1291, siendo un sitio poco conservado y una familia de miARN también poco conservada.

Estos resultados, que deben ser tomados con cautela ya que TargetScan está basado en el análisis de la conservación de estas dianas en las UTRs 3´ y no en UTRs 5´, nos sugieren que los oligonucleótidos mencionados, con cierta probabilidad, podrían ser dianas de micro-ARNs desconocidos. La presencia de dianas de miARNs requiere un análisis más específico y detallado que se realiza en el siguiente apartado.

3.4.- Dianas de micro ARN en las UTRs 5´de lipocalinas

Las dianas de miRNAs, aunque más frecuentes en las UTRs 3´, también pueden presentarse en las UTRs 5´ [10, 11]. Con el objeto de predecir la existencia de dianas de miARNs en las regiones UTRs 5´de las lipocalinas, se sometieron dichas regiones a un análisis mediante el algoritmo PITA de Segal Lab [13] . Dicho algoritmo se considera muy eficiente, ya que además del apareamiento de bases entre el miARN y su diana, considera aspectos energéticos de la interacción entre estos dos elementos (ver métodos).

Los resultados obtenidos tras estos análisis realizados sobre las UTRs 5´de lipocalinas, con el citado algoritmo, se muestran en las tablas 9 y 10. Se han representado los datos para humano y ratón de forma conjunta en dichas tablas, siendo el criterio para separarlos en ellas dos, la existencia diversidad en su UTR 5´(tabla 9) frente a las que no la presentan (tabla 10).

Como hecho más destacable puede observarse (ver tabla 9) que en las lipocalinas que muestran más variación en su UTR 5´ (que pueden considerarse las más ancestrales) es donde más frecuente se dan dianas de miARN muy accesibles(ya que cumplen los valores establecidos para ddG y dGopen simultáneamente, ver métodos). Mientras que en las lipocalinas que presentan UTRs 5´ con nula o escasa variabilidad, sólo en unas pocas son predichas dianas de miARN de este tipo y el número de dichas dianas en ellas es significativamente más escaso (ver tabla 10).

Comprobamos también que, como es de esperar, hay diferencias en el número de dianas de miARNs entre las UTRs 5´ alternativas de una misma lipocalina. En estas lipocalinas, cuyas UTRs 5´ presentan diversidad deberíamos de esperar, dado el hipotético diferente papel regulador de las formas alternativas, que hubiese pocas coincidencias entre las dianas presentes en dichas alternativas. Para ello se buscaron, entre las dianas más accesibles, cuales eran compartidas por las diferentes UTRs 5´alternativas de una misma lipocalina. Los resultados, que se muestran en la tabla 11, nos indican que existe un cierto número de coincidencias, aunque a pesar de ello sigue existiendo suficientes diferencias en la dianas que presentan las formas alternativas, a excepción de en la lipocalina Ptgds humana (ver tabla 11).

UTR 5´	Variantes	Nº dianas accesibles (ddG < -10 Kcal/mol)	Nº dianas muy accesibles (ddG < -10 y dGopen > -10 Kcal/mol)
ApoD-Humana	a	11	5
	b	5	2
	c	7	7
	d	9	0
ApoD- Ratón	a	3	0
	b	7	5
	c	5	3
	d	2	2
	e	7	5
Ptgds_Humano	c	8	3
	g	12	0
	j	25	3
Ptgds-Ratón	c	10	7
	d	5	0
Rbp4_Humano	b	0	0
	d(2)	2	0
Rbp4-Ratón	a	0	0
	d	0	0
Apom-Humana	d	5	0
	e	4	2
Apom_Ratón	a	14	3

Tabla 9. *Dianas de miARN predichas por PITA en las UTRs 5´de lipocalinas que presentan diversidad en dicha región.*

UTR 5´	Nº dianas accesibles (ddG < -10 Kcal/mol)	Nº dianas muy accesibles (ddG < -10 y dGopen > -10 Kcal/mol)
C8G-Ratón-c	3	0
Lcn8-Humano-e	3	1
Lcn12-Humano-c	14	0
Lcn2-Humano-b	0	0
ORM2-Humano-b	1	0

Tabla 10. *Dianas de miARN predichas por PITA en las UTRs 5´de lipocalinas que presentan nula o escasa diversidad en dicha región.*

Especie	Dianas de miARN en UTR 5´alternativas	Dianas de miARN comunes entre alternativas
Humano	Apo-D - variante a: 5 - variante b: 2 - variante c: 7	2 entre variantes b y c
	Ptgds - variante c: 3 - variante j: 3	3 entre variante c y j
Ratón	Apo-D -variante b: 5 - variante c: 3 - variante d: 2 - variante e: 5	7 entre diversos pares de las variantes

Tabla 11.*Dianas de miARN (solo dianas más accesibles) comunes entre las UTRs 5´alternativas de ciertas lipocalinas*

3.4.1.- Realidad biológica de las dianas de miARNs en las UTRs 5´de lipocalinas

Dado que la longitud de las UTRs 5´de las lipocalinas es variable, cabe considerar que las diferencias encontradas en el número de dianas de miARN, de ser estas falsos positivos, puedan ser debidas a la mayor o menor longitud de estas y por lo tanto a la mayor o menor probabilidad de su aparición. Utilizando los datos de todas las UTRs 5´de lipocalinas de humano y ratón, conjuntamente, no se encuentra una correlación significativa entre la longitud de la UTR y el número de dianas ($r = 0.154$; $p = 0.472$), (ver figura 5).

Otro factor que podría considerarse que tuviese alguna influencia es la energía mínima de plegamiento (MFE) de la estructura 2D de la UTR 5´. A mayor valor de la MFE global de la UTR 5 ´, podríamos esperar una mayor dificultad de que se deshaga dicha estructura para que se forme el apareamiento entre el miARN y su diana. Utilizando la misma muestra de UTRs 5´ que para la longitud no se observa tal correlación entre la MFE de las UTRs 5´y el número de dianas de miARN presentes ($r = 0.033$; $p = 0.878$), (ver figura 6).

Estos resultados nos permite rechazar la hipótesis de que las diferencias en el número de dianas de miARN encontradas sea debida a diferencias en la longitud o en la MFE de las UTRs 5´, y por lo tanto debe responder a necesidades de regulación de las diferentes lipocalinas.

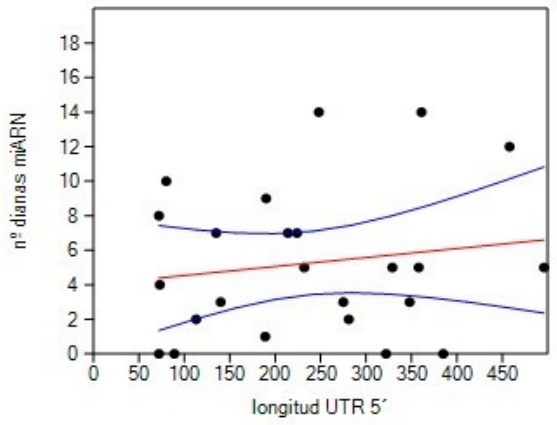


Figura 5. *Relación entre el número de dianas de miARN encontradas y la longitud de la UTR 5´. En rojo línea de regresión, en azul intervalo del 95% confianza*

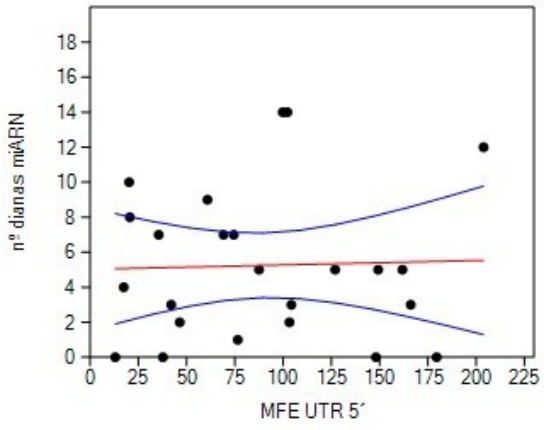


Figura 6. *Relación entre el número de dianas de miARN encontradas y la MFE de la UTR 5´. En rojo línea de regresión, en azul intervalo del 95% confianza Se han representan los valores de MFE en positivo.*

3.5.- uAUGs y uORFs en las UTRs 5´de lipocalinas

La presencia en las UTRs 5´ de codones de inicio corriente arriba (uAUG) y de marcos de lectura abiertos, también corriente arriba (uORF), es interpretada generalmente como un mecanismo de regulación negativa de la traducción [15, 16 y 17]. Por ello es importante el estudio de estos elementos en dichas regiones.

Los resultados del análisis realizado sobre uAUGs y uORFs presentes en las UTRs 5´ de las lipocalinas de humano y ratón (ver métodos) se muestran en las tablas 12 y 13. En humano se observa la existencia de uAUGs en una gran mayoría de las lipocalinas, incluidas sus distintas alternativas. Es de destacar que en muchos casos los uAUGs dan lugar a uORFs. Se observa además que un número considerable de las UTRs 5´ presentan más de un uORF.

En ratón se observa que hay un menor número de lipocalinas que presentan uAUGs (ver tabla 13), pero las que los presentan, en la mayoría de casos, estos dan lugar a uORFs. También observamos, al igual que en humano, que hay un alto porcentaje de UTRs 5´que presentan más de un uORF.

Aunque se ha podido comprobar experimentalmente que uORFs de diferente tamaño y posición en la UTR 5´ pueden ejercer un efecto de inhibición sobre la traducción, hay ciertas propiedades de los uORFs que han demostrado tener un mayor efecto en el grado inhibición de la síntesis de proteínas [20]. Son las siguientes:

- El contexto del uAUG, de forma que en un contexto “fuerte” se produce una mayor inhibición que en uno “débil”.
- La distancia a la caperuza del extremo 5´, de manera que a mayor distancia del uORF respecto a esta se produce una mayor inhibición.
- La presencia de múltiples uORFs, produciendose una mayor inhibición a un mayor número de los mismos.
- La conservación evolutiva de los uORFs, siendo mayor el efecto inhibitorio a mayor conservación de los mismos.

En las tablas 12 y 13 observamos, especialmente en humano, que la mayoría de uORFs presentan un contexto adecuado, e incluso algunos muestran un contexto óptimo. Por otra parte hay un porcentaje elevado de UTRs 5´ de lipocalinas donde existe más de un uORF, concrétamente en humanos representan un 60 %, mientras que el porcentaje en la globalidad de transcritos humanos

que los presentan en sus UTRs 5´ es de un 50 %´[20].

Lipocalina	UTRs 5´alternativas	Nº uAUGs	Nº uORFs	contexto uORFs		
				optimo	adec	inadec
Apod	a	4	3	1	2	0
	b	1	1	0	1	0
	c	2	1	1	0	0
	d	3	2	0	2	0
Ptgds	c	0	0	0	0	0
	g	4	4	0	2	2
	j	11	5	1	3	1
Rbp4	b	1	1	1	0	0
	c	1	1	0	1	0
	d	0	0	0	0	0
Apom	d	6	5	0	2	3
	e	2	2	0	2	0
Lcn12	a	0	0	0	0	0
	b	0	0	0	0	0
	c	1	1	0	1	0
Lcn8	e	3	3	0	1	2
Orm2	b	0	0	0	0	0

Tabla 12. *Número de uAUGs y de los correspondientes uORFs(en caso de haberlos) de las UTRs 5´de lipocalinas humanas estudiadas, incluidas todas sus alternativas. Se indica la fortaleza del contexto de los uAUGs encontrados (en relación a la secuencia consenso de Kozak).*

Lipocalina	UTRs 5´alternativas	Nº uAUG	Nº uORF	contexto uORF		
				optimo	adec	inadec
Apod	b	1	1	0	1	0
	c	2	2	0	2	0
	d	2	2	0	2	0
	e	2	1	1	0	0
Ptgds	c	0	0	0	0	0
	d	2	2	1	0	1
Apom	a	14	8	0	2	6
C8g	b	6	6	1	1	4
	c	4	4	1	1	2

Tabla 13. *Número de uAUGs y de los correspondientes uORFs(en caso de haberlos) de las UTRs 5´de lipocalinas de ratón estudiadas, incluidas todas sus alternativas. Se indica la fortaleza del contexto de los uAUGs encontrados (en relación a la secuencia consenso de Kozak).*

La distancia entre la caperuza y los uORFs presentes en las UTRs de las lipocalinas se muestra en las gráficas de las figuras 6 y 7. En estas se ha representado la posición relativa de los uORFs dentro de la secuencia de las UTRs 5´. Como se observa en las gráficas una gran parte de los uORFs se muestra en posiciones centrales (2º cuarto en humano y 2º y 3º cuarto en ratón) de la UTR 5´, por lo que su posición es suficientemente distante de la caperuza. También se aprecia claramente la escasez de uORFs en las proximidades de la secuencia codificante. Esto puede ser un factor importante a la hora de que el ribosoma disponga de espacio para reiniciar su escaneo a lo largo del ARNm hasta encontrar el AUG principal [21].

Respecto a la longitud de los uORF presentes en las UTRs 5´ de las lipocalinas existe una gran variabilidad, como puede observarse en las figuras 8 y 9. Si bien hay una frecuencia considerable de uORFs de gran longitud, de forma más destacada en humano. El valor medio de longitud de uORF en el conjunto de transcritos, tanto de humano, como de ratón es de 48 nt [20], mientras que en las lipocalinas es de 80.51 nt para humano y de 65.77 nt para ratón, ambos valores ampliamente superiores.

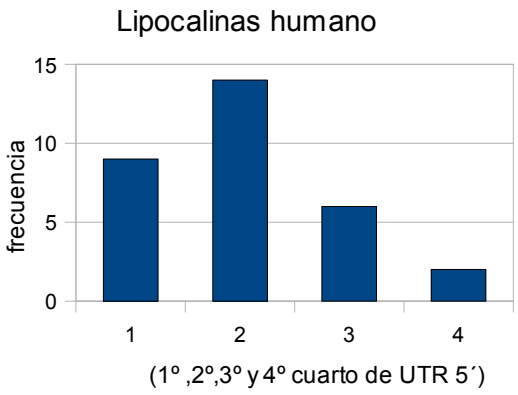


Figura 6. *Frecuencia de las posiciones que ocupan los diferentes uORFs en las UTRs 5´de lipocalinas humanas.*

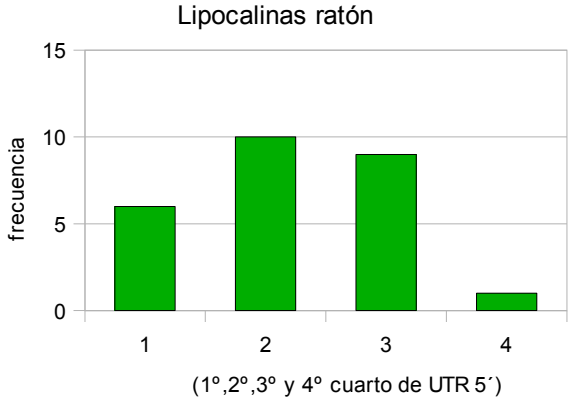


Figura 7. *Frecuencia de las posiciones que ocupan los diferentes uORFs en las UTRs 5´de lipocalinas de ratón.*

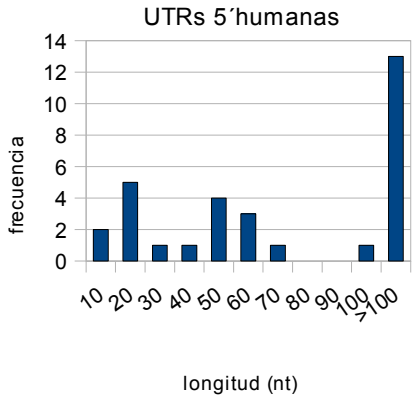


Figura 8. *Frecuencia del tamaño de los diferentes uORFs en las UTRs 5´de lipocalinas humanas*

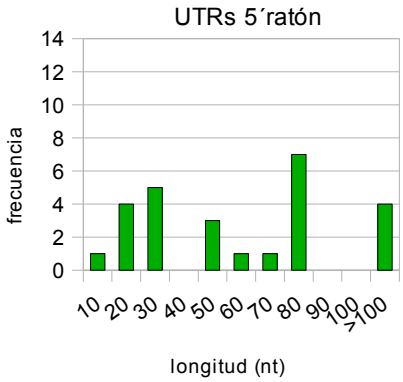


Figura 9. *Frecuencia del tamaño de los diferentes uORFs en las UTRs 5´de lipocalinas de ratón.*

Se observa otra peculiaridad en los uORFs de lipocalinas, además de su mayor longitud media, y es que estos elementos muestran una claro incremento de su longitud en relación a la mayor fortaleza del contexto del AUG de la secuencia codificante (figura 10). El hecho de la mayor longitud media de los uORFs de lipocalinas frente a la globalidad de transcritos y su aparente relación con la fortaleza del contexto del AUG son signos que nos llevan a valorar que la longitud de los uORF podría ejerce un importante el papel regulatorio en las lipocalinas.

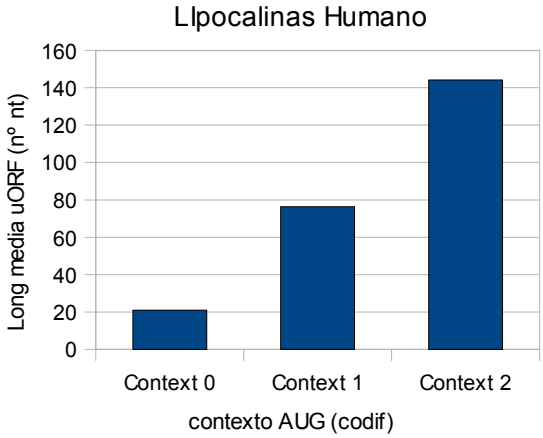


Figura 10. *Relación entre el contexto (0= inadecuado. 1= adecuado, 2= óptimo) del AUG principal y la longitud de los uORFs presentes corriente arriba en la UTR 5´.*

En cuanto a las UTRs 5´alternativas que muestran las lipocalinas, éstas poseen patrones específicos de uORFs, con diferencias claras entre ellas, en alguna o varias de las propiedades ya mencionadas: el número de uORF, la longitud de los mismos, la distancia a la caperuza o el contexto del uAUG. En la tabla 14 y en la figura 11 se representan los datos para las UTR 5´alternativas de Apo-D humana, donde a modo de ejemplo se observan las diferencias mencionadas. Estas diferentes combinaciones de parámetros de los uORFs es de esperar, den lugar regulaciones de la traducción de diferente intensidad, según las necesidades celulares en que son expresados los diferentes transcritos.

uORF	Cap-uORF dist	uAUG context	uORF long(nt)
Apod-a-1	111	óptimo	60
Apod-a-2	198	adecuado	12
Apod-a-3	178	adecuado	294
Apod-b-1	102	adecuado	129
Apod-c-1	63	óptimo	27
Apod-d-1	27	adecuado	12
Apod-d-2	7	adecuado	294

Tabla 14. *Diferente composición de uORFs en las UTRs 5´alternativas de Apo-D humana*

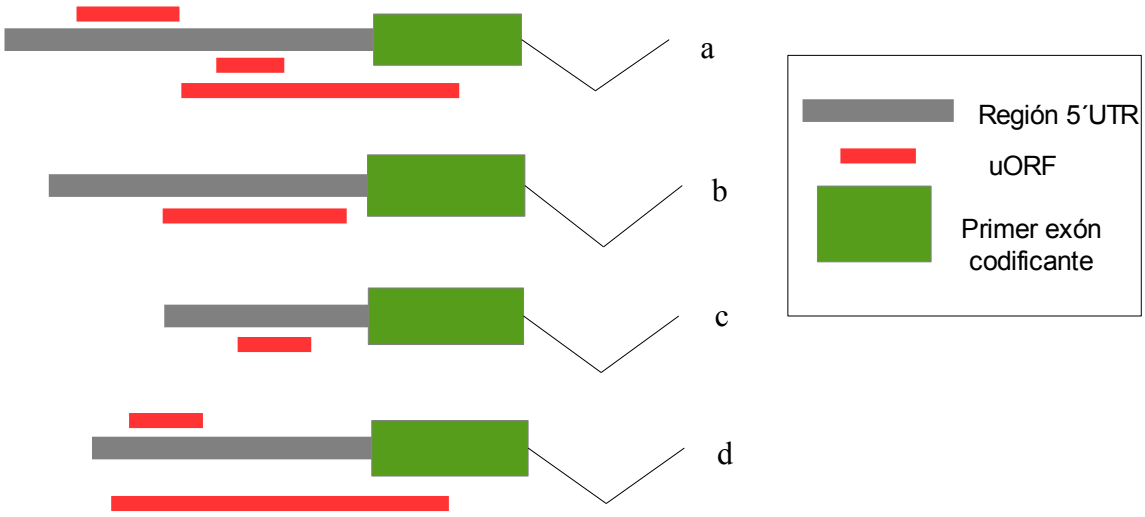


Figura 11. *Representación gráfica de los diferentes uORFs presentes en las diferentes UTRs 5´alternativas de Apo-D humana.*

3.5.1.- Conservación de uAUGs y de uORFs

La conservación de uAUGs y uORFs entre las UTRs 5´ ortólogas de las diferentes lipocalinas de mamíferos supondría un argumento a favor de que dichos elementos tuviesen relevancia biológica.

Aplicando los criterios mencionados en métodos pudo detectarse la presencia de uAUGs y uORFs ortólogos en las UTRs 5´de las lipocalinas Apo-d, Ptgds y Apom. En la tabla 15 se muestra la proporción de los uAUGs y uORFs conservados en las UTRs 5´ortólogas de mamíferos de estas lipocalinas. En las tablas 16 a 18 se muestran los detalles de los uORFs considerados ortólogos para las diferentes lipocalinas.

Como se observa en estos datos (tabla 15), para estas lipocalinas, la porporción de uAUG y uORF conservados entre mamíferos es considerable, siendo claramente mayor para los uORFs humanos que para los de ratón. En las tablas se observa que en Apo-D los uORFs ortólogos se encuentran hasta en cuatro especies de mamíferos, reforzando la idea de la relevancia biológica de dichos elementos.

Lipocalinas (pares ortólog UTR 5´)	uAUGs conserv/total	uORFs conserv/total	Especies con uAUGs y/o uORFs ortólogos
UTR 5´Apod_a_humano	3/4	3/3	A. melanoleuca, U. americanus, B. taurus, S. scrofa
UTR 5´Apod_d_ratón	2/2	1/2	A. melanoleuca, U. americanus, H. sapiens
UTR 5´ Ptgds_g_humano	1/4	1/4	P. paniscus
UTR 5´ Ptgds_j_humano	5/10	4/5	N. leucogenes, P. anubis
UTR 5´ Ptgds_d_ratón	1/2	0/2	N. leucogenes
UTR 5´ Apom_d_humano	5/6	4/5	M. putorius, C. lupus, M. musculus
UTR 5´ Apom_a_ratón	4/8	3/6	M. putorius, C. lupus, H. sapiens

Tabla 15. *uAUGs y uORFs conservados entre diferentes especies de mamíferos, en las lipocalinas: Apo-D, Ptgds y Apom.*

uORF ortólogos	dist uAUG-AUG	nº codones
5utr_APOD.a_hum_1	250	20
A. melanoleuca	229	20
U. americanus	229	20
B. taurus	236	20
5utr_APOD.a_hum_2	163	4
5utr_Apod.d_mouse_1	113	9
A. melanoleuca	128	5
U. americanus	128	5
5utr_APOD.a_hum_3	183	98(s cds)
A. melanoleuca	60	57(s cds)
U. americanus	60	60(s cds)
B. taurus	157	168(s cds)

uORF ortólogos	dist uAUG-AUG	nº codones
5utr_Apom.d_hum_2	432	6
C. lupus	430	5
5utr_Apom.d_hum_3	362	84
5utr_Apom_a_Mus musc_3	352	77
5utr_Apom.d_hum_4	358	17
M. putorius	369	17
C. lupus	369	13
5utr_Apom_a_Mus musc_4	348	26
5utr_Apom.d_hum_5	222	21
M. putorius	215	30
C. lupus	216	14

uORF ortólogos	dist uAUG-AUG	nº codones
>5utr_Ptgds.g_hum_4	114	127(s cds)
P. paniscus	113	81(s cds)
5utr_Ptgds.j_hum_1	1126	3
N. leucogenes	738	3
P. anubis	1126	3
5utr_Ptgds.j_hum_2	1092	18
N. leucogenes	704	17
P. anubis	1091	18
5utr_Ptgds.j_hum_3	1048	68
P. anubis	1047	68
5utr_Ptgds.j_hum_5	458	119
P. anubis	449	112

Tablas 16, 17 y 18. *Datos de los diferentes uORFs considerados ortólogos entre especies de mamíferos, según su longitud y distancia al AUG principal.*

3.5.2.- Posibilidad de la formación de peptidos bioactivos resultantes de la traducción de uORFs

La idea más aceptada sobre el papel de los uORFs, al igual que para los uAUGs, es que tienen un efecto inhibitorio sobre la traducción, al retrasar o entretener la actividad del ribosoma de su unión al AUG principal (según el modelo *leaky scanning* [33]). Igualmente los uORFs puenden estar implicados en una terminación prematura de la traducción del ARNm, mediante su efecto facilitador del mecanismo de NMD [21].

Por otra parte, de forma más alternativa, se han encontrado evidencias de que los péptidos resultantes de la traducción de algunos uORFs podrían tener un papel regulador, actuando ellos mismos sobre la eficiencia del ribosoma o sobre la estabilidad del ARNm [23]. Así mismo se ha constatado la presencia de un número considerable de uORFs conservados entre los genomas de humano y ratón y que además muestran una proporción elevada de sustituciones "sinónimas / no sinónimas" (una proporción mayor que "1" es interpretado como resultado de selección positiva). Hechos que sugieren que los péptidos resultantes pueden tener un papel bioactivo [24].

Para encontrar algún indicio de péptidos que puedan tener este papel en las UTRs 5´de lipocalinas se realizó un análisis sobre los uORFs ortólogos previamente encontrados en ellas. Una vez seleccionados los alineamientos de estas secuencias ortólogas, estas se sometieron al análisis de sustituciones sinónimas y no sinónimas (ps/pn) mediante el programa SNAP [25] (ver métodos).

uORF ortólogos	Nº de aminoácidos del péptido	Proporción sust. sinon. / no sinon.(ps /pn)
5utr_APOD.a_hum_1 A. melanoleucas U. americanus B. taurus	19	1.587
5utr_Ptgds.j_hum_2 N. leucogenes P. anubis	17	0.704
5utr_Apom.d_hum_4 M. putorius C. lupus 5utr_Apom.a_M. musc_4	16	1.309

Tabla 19. *Número de aminoácidos y proporción de sustituciones sonónimas/no sinónimas (ps/pn) de los uORFs ortólogos cuyas secuencias se encuentran conservadas en mamíferos.*

Los resultados obtenidos se muestran en la tabla 19. Hay dos uORFs, uno en la UTR 5´de Apo-D (variante humana "a") y otro en la de Apo_M (variante "d" humana y "a" en ratón) que tienen valores de " ps / pn" superiores a 1. En una investigación, ya mencionada, se encontró que el valor medio de ps/pn obtenido, para un conjunto de uORFs conservados entre humano y ratón, es de 1,65 [24], mientras que se obtuvo un valor de ps/pn de 0.99 en un conjunto de secuencias de contraste generadas artificialmente (con valores que oscilan entre 0.70 y 1.36). A la luz de estos datos podemos concluir que al menos en el caso del uORF-1 de la UTR 5´ de Apo-D(a) humana hay claros signos de selección positiva, siendo así el peptido resultante un candidato a cumplir algún papel en la regulación de la traducción de dicha lipocalina. En la figura 12 se muestra el grado de consenso de este péptido entre las diferentes especies.

Consensus	**DkTSHLCVwYRAKLLGQPS**
Panda	.....
Oso	.....
Vaca	nI.....S.....D...
Humano	.T.....SIqT.....

Figura 12. *Alineamiento múltiple y consenso del péptido uORF(1) de la UTR 5´de Apo-D(a) humana y sus ortólogos en mamíferos.*

4. - Discusión

La aplicación del modelo CART nos ha permitido establecer una primera clasificación de las UTRs 5´de las lipocalinas de mamíferos. Hemos encontrado para las lipocalinas Ptgds y Rbp4, en las dos especies estudiadas, y Apom, Lcn1 y Lcn12 sólo en humanos, que presentan UTRs 5´alternativos que pertenecen, como cabría esperar, a clases diferentes: clase I o III alternativamente (siendo las primeras inhibidoras de traducción y las segundas con expresión elevada). En otras lipocalinas encontramos que esto no ocurre. El caso más llamativo es Apo-D que, tanto en humano como en ratón, presenta el mayor número de variantes en su UTR 5´ y todas pertenecen a la "clase I". La explicación que podemos dar a dicha circunstancia es que en estas UTRs 5´ alternativas

existan diferentes elementos, como uORFs , miARNs u otros, que sean los responsables de un efecto inhibitorio más o menos acusado entre las distintas variantes, o bien que dicho efecto sea o no ejercido en función de la presencia o ausencia de factores que interaccionen con la UTR 5´ en ciertos tipos celulares o condiciones fisiológicas. De ser esto correcto podríamos afirmar que la expresión de estas proteínas necesita de una regulación muy fina.

Respecto a la existencia de dianas de miARN en las UTRs 5´ de lipocalinas el objetivo que se ha perseguido es conocer en qué medida en dichas regiones existen secuencias complementarias accesibles a hipotéticos miARNs conocidos, para así tener una idea de la importancia que este tipo de regulación puede estar teniendo en ellas. Para la predicción de dianas de miARNs en las UTRs 5 ´de lipocalinas hemos utilizando un algoritmo (PITA) que considera aspectos energéticos del apareamiento entre un miARN y su diana, y aún habiendo aplicando un criterio restrictivo en el valor del balance energético de este apareamiento (ddG < - 10 Kcal/mol), son predichas un número considerable de dichas dianas en las UTRs 5´de las lipocalinas. Incluso aplicando un segundo filtro a estos datos obtenidos (ddG< -10 y dGopen > -10 Kcal/mol), que indicaría apareamientos (miARN-diana) altamente accesibles, se siguen obteniendo un número de dianas de miARN considerable. El número de dianas predichas es, por lo general, mayor en las lipocalinas que muestran mayor diversidad en su UTR 5´, necesitadas hipotéticamente de mayor regulación, en comparación con las que muestran escasa o nula variabilidad en dicha región. Así mismo, como cabría esperar, se encuentran diferencias en el número y tipo de dianas de miARN entre las UTR 5 ´alternativas de una misma lipocalina, aunque en algunos casos dichas variantes presenten dianas comunes.

Si a las circunstancias anteriores añadimos el hecho de que el número de dianas de miARN no muestre una correlación significativa con la longitud de las UTRs 5´ni con la MFE de las mismas, encontramos un respaldo a la hipótesis de que estas dianas tengan realmente una función regulatoria de la expresión génica de las lipocalinas donde son predichas. La importancia que puedan tener las dianas de miARN en el papel regulatorio de las UTRs 5´será mejor ponderada cuando sean comparadas con los resultados obtenidos para las UTRs 3´, aspecto que se trata en el siguiente capítulo.

Un número importante de UTRs 5´de lipocalinas presentan uAUGs, que en numerosos casos dan lugar a uORFs, existiendo con frecuencia varios de ellos en una misma UTR 5´. Esto ocurre en un mayor número de casos en lipocalinas humanas al comparar con las de ratón. También claramente esto es más frecuente entre las lipocalinas que presentan mayor diversidad en su UTR 5´.

Si bien cualquier uORF es de forma potencial capaz de inhibir la síntesis de proteína, en conjunto los uORFs presentes en las UTRs 5´de las lipocalinas, presentan unas caracteristicas (entre otras: longitud, el nº de uORFs y su posición en la UTR 5´) compatibles con un efecto inhibidor de la traducción importante. Si a este hecho unimos que hemos podido poner de manifiesto la existencia de uORFs ortólogos, al menos en algunas lipocalinas, podemos concluir que los uORFs existentes en las UTRs 5´ de lipocalinas de mamíferos son buenos candidatos a ejercer un papel regulador manteniendo bajos los niveles de la proteína correspondiente. Las diferencias encontradas entre UTRs 5´alternativas de una misma lipocalina, respecto al número y tipo de uORFs, nos sugiere que está teniendo lugar en ellas un ajuste fino de la traducción.

Se ha encontrado que existe un importante número de uORFs conservados en mamíferos. La existencia de una proporción de uORFs no conservados puede estar indicándonos un papel regulador específico, de los mismos, en diferentes órdenes de mamíferos. Las mayores diferencias encontradas entre ratón y otros mamíferos, respecto a la conservación de estos elementos, podrían indicarnos necesidades de regulación específicas del taxón de roedores. En este mismo sentido observamos como en Ptgds humana (tabla 18) sólo se encuentran indicios de uORFs ortólogos entre primates. Esto reforzaría la idea sugerida de necesidades de regulación comunes a todos los mamíferos, frente a otras específicas, propias de los diferentes órdenes.

En los casos en que los uORFs ortólogos muestran buena conservación entre sus secuencias y que además se muestran evidencias de que ha actuado una selección positiva en ellos, esto ha de interpretarse como una prueba de que los péptidos resultantes pueden tener un papel bioactivo. En nuestro estudio se ha encontrado evidencias de esta situación especialmente en uno de los uORFs existente en la UTR 5´de Apo-D(a) humana. Por ello hemos de considerar la posibilidad de que este péptido tenga un papel bioactivo, que establecería otro mecanismo adicional de regulación de la traducción de dicha lipocalina.

Es conocido que en presencia de codones de terminación prematuros seguidos de intrones tiene lugar el mecanismo de degradación de ARNm conocido como NMD. Dado que, las UTRs 5´de las lipocalinas donde aparecen en mayor frecuencia los uORFs, son el resultado del splicing y transcripción alternativa de varios exones, ofrecen la situación típica para que pueda darse este mecanismo de NMD [21, 26]. Si bien es sabido que no siempre que se dan estas circunstancias se activa dicho mecanismo NMD, debido a la entrada en juego de otros factores [27]. Por ello podemos ver en la diversidad de uORFs (en cuanto a número, tamaño y localización) presentes en

estas lipocalinas una oportunidad para que se establezca cierta regulación mediante la activación o no de este mecanismo de NMD. Respecto a esto podemos hacer especial hincapié nuevamente en las diferencias, respecto a los uORFs y composición de exones, entre las UTRs 5´alternativas de una misma lipocalina.

Podemos concluir esta discusión diciendo que las UTRs 5´de lipocalinas, especialmente las que presentan mayor número de formas alternativas en dichas regiones (que son las lipocalinas más ancestrales), parecen estar fuertemente reguladas mediante la presencia de diversos elementos y que esto respondería a las necesidades de un ajuste fino de su expresión génica en diferentes tejidos o en diferentes condiciones fisiológicas de cada uno de ellos.

5. - Bibliografía

[1] Grillo, G., et al. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Research* **38**, D75–D80 (2010).

[2] Kato, S., Sekine, S., Oh, S.W., Kim, N.S., Umezawa, Y., Abe, N., Yokoyama-Kobayashi, M. and Aoki, T. Construction of a human full-length cDNA bank. *Gene* **150**, 243-50 (1994).

[3] Meyuhas, O., Avni, D. and Shama, S. Translational control of ribosomal protein mRNAs in eukaryotes Translational Control. Cold Spring Harbor, Cold Spring Harbor Laboratory Press 1996. 363-368.

[4] Riu Yamashita, et al. Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. *Nucleic Acids Research* **36**, 3707–3715 (2008).

[5] Le, S.Y., and Maizel, J.V., Jr. A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs. *Nucleic Acids Research* **25**, 362-69 (1997).

[6] Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA Websuite**.** Nucleic Acids Res. 2008. Vol 36: w70-w74.

[7] Giegerich, Robert and Voss, Bjoern and Rehmsmeier, Marc. Abstract Shapes of RNA, Nucleic Acids Research, 2004. Vol 32: 4843-4851

[8] Reeder, Janina, Reeder, Jens and Giegerich, Robert: Locomotif: From Graphical Motif Description to RNA Motif Search in *Bioinformatics*, 2007, 23(13) , Pages:i392-400

[9] van Helden, J., André, B. and Collado-Vides, J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. J Mol Biol, 1998, 4;281(5): 827-42.

[10] Nien-Pei Tsai, Ya-Lun Lin and Li-Na-Wei. MicroRNA mir-346 targets the 5′-untranslated region of receptor-interacting protein 140 (RIP140) mRNA and up-regulates its protein expression. Biochem. J., 2009, **424:** 411–418.

[11] Francesca Moretti, Rolf Thermann, and Matthias W. Hentze. Mechanism of translational regulation by miR-2 from sites in the 5′ untranslated region or the open reading frame. RNA, 2010, 16: 2493-2502.

[12] H.Y. Huang, C.H. Chien, K.H. Jen, and H.D. Huang (2006) "RegRNA: A regulatory RNA motifs and elements finder" Nucleic Acids Research, Vol 34, W429-W434

[13] Michael Kertesz, Nicola Iovino, Ulrich Unnerstall, Ulrike Gaul & Eran Segal. The role of site accessibility in microRNA target recognition. Nature Genetics, 2007, 39, 1278 – 1284

[14] Andrew Grimson, Kyle Kai-How Farh, Wendy K Johnston, Philip Garrett-Engele, Lee P Lim, David P Bartel. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing *Molecular Cell*, 2007, 27:91-105

[15] C. Vogel, R. de Sousa Abreu, D. J. Ko et al., "Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line," *Molecular Systems Biology*, 2010 vol. 6, article 400

[16] S. E. Calvo, D. J. Pagliarini, and V. K. Mootha, "Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans," Proceedings of the National Academy of Sciences of the United States of America, 2009, vol. 106, no. 18, pp. 7507–7512

[17] M. Matsui, N. Yachie, Y. Okada, R. Saito, and M. Tomita, "Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse," *The FEBS Letters*, 2007, vol. 581, no. 22, pp. 4184–4188

[18] Kozak, M. Possible role of flanking nucleotides in recognition of the AUG initiate codon by eukaryotic ribosomes. Nucleic Acids Res. 1981, 9: 5233-5262

[19] Kozak, M. Compilation and analysis of sequences upstream from the translational start site in eukaryiotic mRNA. Nucleic Acids Res. 1984, 12: 857-872

[20] Sarah E. Calvo, David J. Pagliarini, and Vamsi K. Mootha. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A. 2009, May 5;106(18):7507-12.

[21] Hedda A. Meijer and Adri A.M. Thomas. Control of eukaryotic protein synthesis by upstream open reading frames in the 5′-untranslated region of an mRNA. Biochem. J. 2002, **367:** 1–11

[22] L. Karagyozov and F.D. Böhmer. Conservation of the upstream uAUGs and uORFs in the human and mouse 5′ untranslated region of the mRNAs for protein tyrosine phosphatases. XI Anniversary Scientific Conference. 120 years of academic education in biology . Special Edition/on-line Biotechnol. & Biotechnol. eq. 23/2009/se

[23] Inhibition of CHOP translation by a peptide encoded by an open reading frame localized in the chop 5′UTR. Nucleic Acid Res. 2001, Vol. 29, Iss.21, pp: 4341-4351

[24] Mark L Crowe, Xue-Qing Wang and Joseph A Rothnagel**.** Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides. *BMC Genomics* 2006, **7**:16.

[25] Korber B. HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences, 2000, Chapter 4, pages 55-72. Allen G. Rodrigo and Gerald H. Learn, eds. Dordrecht, Netherlands: Kluwer Academic Publishers.

[26] Alicia A. Bicknelly, Can Ceniky, Hon N. Chua, Frederick P. Roth and Melissa J. Moore. Introns in UTRs: Why we should stop ignoring them. Bioessays 34: 1025–1034, 2012 WILEY Periodicals, Inc.

[27] Silva AL, Romão L. The mammalian nonsense-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision?. FEBS Lett. 2009 Feb 4;583(3):499-505.

[28] Rice,P. Longden,I. and Bleasby,A. *EMBOSS:* The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276-277 (2000).

[29] Ramana, V. D., et al. CART Classification of Human 5´UTR Sequences. *Genome Research* 10, 1807-1816 (2000)

[30] Avni, D., Biberman, Y., and Meyuhas, O.. The 5 terminal oligopyrimidine tract confers translational control on TOP mRNAs in a cell type- and sequence context-dependent manner. Nucleic Acids Res. **25:** 995–1001. 1997

[31] Kochetov, A.V., et al. Eukaryotic mRNAsencoding abundant and scarce proteins are statistically dissimilarin many structural features. FEBS Lett 440:351–355. 1988

[32] Pickering, B.M., Willis, A.E. The implications of structured 5´ untranslated regions on translation and disease. Semin Cell Dev Biol 16:39–47. 2005

[33] Kozak, M. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109-1123 (1978).

[34] Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986 Sep;3(5):418-26.

V

PAPEL REGULADOR DE LAS UTRs 3´ DE LIPOCALINAS DE MAMÍFEROS

1. - Objetivos

Los objetivos de este capítulo son los mismos que los expuestos para las UTRs 5´.

2. - Métodos

2.1.- Clasificación de las UTRs 3´de lipocalinas en función de la poliadenilación

Para realizar esta clasificación se determinó, utilizando patrones de búsqueda mediante "dreg" de EMBOSS, la presencia de las señales de poliadenilación (PAS) en las diferentes UTRs 3´, tanto la señal canónica "AAUAAA", como la señal alternativa "AUUAAA" y otra serie de señales alternativas menos frecuentes (AGUAAA, UAUAAA, etc), según las evidencias existentes [12]. Se anotó además la posición que estas señales ocupan en la UTR, estableciéndose en base a esta la siguiente clasificación de las PAS: *terminal* (últimos 50 nucleótidos de la UTR 3´), *intermedia* (en posición central de la UTR) o *proximal* (desde la mitad de la UTR hacia codón de STOP).

Con estas variables (PAS canónica o alternativa y posición que ocupa en la UTR 3´), y según las evidencias de como influyen estos factores en la eficiencia de la poliadenilación [12], se estableció una escala relativa de dicha eficiencia: ME (muy eficiente), E (eficiente), PE (poco eficiente). Este enfoque, aunque supone una simplificación, nos permite utilizar una escala relativa de comparación, con cierta objetividad, de las eficiencias de la poliadenilación entre las UTRs 3´alternativas de cada lipocalina o entre UTRs 3´de diferentes lipocalinas. La tabla de decisión elaborada (tabla 1) para establecer la escala de eficiencia se muestra en los resultados.

2.2.- Búsqueda de motivos validados, en las UTRs 3´

Se utilizó el mismo procedimiento que para las UTRs 5´. Ver métodos del capítulo IV

2.3.- Identificación de dianas de miARN

Se utilizó el mismo procedimiento que para las UTRs 5´. Ver métodos del capítulo IV.

2.4.- Determinación de oligonucleótidos sobrerrepresentados

Se utilizó el mismo procedimiento que para las UTRs 5´. Ver métodos del capítulo. La única excepción es que el modelo de fondo elegido en este caso fue la composición de nucleótidos de las propias secuencias UTRs 3´de las lipocalinas y no las regiones corriente arriba de los genes, por motivos evidentes. Dentro de esta opción se seleccionó un modelo de Markow de orden 2, de manera que sea tenida en cuenta la dependencia de orden superior (subpalabras dentro de palabras mayores) entre residuos vecinos.

3. - Resultados

3.1.- Regulación de la poliadenilación, primera aproximación a la clasificación de las UTRs 3´de lipocalinas

La región UTR 3´ de los ARNm se encuentra implicada en numerosos procesos que pueden afectar a la regulación de la expresión génica. Entre estos procesos destacan: el lugar de corte del transcrito en su región 3´ y su posterior poliadenilación, la presencia de dianas de miARN, la presencia de ciertos elementos de regulación como elementos ricos en "AU" (AREs) o diversos motivos estructurales que sirven de sitios de unión a proteínas reguladoras.

Uno de los mecanismos más conocidos de regulación en la región UTR 3´es la poliadenilación. Este proceso es fundamental para permitir la adecuada exportación del ARNm del núcleo al citoplasma, así como para dar estabilidad al ARNm, favoreciendo su traducción de forma eficiente. Es bien conocido que la poliadenilación viene determinada por señales de poliadenilación (PAS) en la región UTR 3´. La señal de poliadenilación más frecuente (señal canónica) es "AAUAAA", localizada a 10-30 nucleótidos al 5´del sitio de corte y posteriormente se encuentra una secuencia rica en G/U a unos 20-30 nucleótidos al 3´del lugar de corte [1]. Los dos complejos multiméricos, el de poliadenilación (que se une a la PAS) y el de corte (que se une a la señal rica en G/U) son necesarios simultáneamente para que ambos procesos ocurran adecuadamente.

La presencia de señales de poliadenilación alternativas es un fenómeno extendido en eucariotas [2], y debido a la relación entre el corte y la poliadenilación, suponen una oportunidad para generar ARNms con UTRs 3´ de diferentes longitudes. Dado que las regiones UTRs 3´son una fuente de señales regulatorias, como pueden ser las dianas de miARN, las diferentes UTRs 3´generadas son una oportunidad de ejercer una diferente regulación post-trancsripcional.

Alternativamente a la PAS canónica "AAUAAA" y en menor frecuencia se da la señal "AUUAAA" [11]. Se ha determinado, más recientemente, que pueden existir al menos otras 10 posibles variaciones de la señal de poliadenilación, todas ellas de menor frecuencia que la señal alternativa "AUUAAA" [12]. En este último estudio mencionado [12], con una muestra de más de 5000 ARNms, se determinó que la PAS que da lugar a una poliadenilación más eficiente es la señal canónica (AAUAAA), mientras que cualquiera de las otras señales no canónicas muestran una poliadenilación menos eficaz. En los casos en que existen PAS alternativas, las más distales (en dirección 3´) son preferentemente la señal canónica, mientras que las proximales (en dirección 5´) son preferentemente señales alternativas. También se determinó, en el mismo estudio el efecto de la posición de la señal de poliadenilación (en caso de PAS alternativas) sobre la eficiencia de la poliadenilación. Se observó que las posiciones más terminales (extremo 3´de UTR 3´) de la señal PAS tienen un efecto positivo en la eficiencia de la poliadenilación, siempre manteniéndose la superioridad de la señal canónica frente al resto de las señales alternativas, para una posición dada.

Como primera aproximación al papel que desempeñan las UTRs 3´de las lipocalinas se procedió a analizar las PAS que estas presentan y a establecer una clasificación de la eficiencia de las mismas en la poliadenilación, usando la información de las evidencias mencionada anteriormente, y elaborando con ellas una "tabla de decisión" (ver métodos) que se muestra en la tabla 1.

Los resultados de aplicar esta tabla de decisión a las UTRs 3´de las lipocalinas se muestran en las tablas 2 y 3.

Tipo de señal poliadenilación	Posición en UTR 3´	Eficiencia relativa de poliadenilación
Canónica “AAUAAA”	Terminal (últimos 50 nt)	Muy Eficiente: **ME**
	Intermedia	Eficiente: **E**
	Proximal (cercana al codón de stop)	Poco eficiente: **PE**
No canónica “AUUAAA” u otras minoritarias	Terminal (últimos 50 nt)	Eficiente: **E**
	Intermedia o Proximal	Poco Eficiente: **PE**

Tabla 1. Tabla de decisión sobre la eficiencia de la poliadenilación de la UTR 3´. Elaborada sintetizando los datos del estudio sobre la eficiencia de dicho proceso [12].

Observamos en las tablas 2 y 3 que, en las lipocalinas en las que las UTRs 3´ son de cierta longitud (generalmente > 200 nt), aparecen varias señales de poliadenilación, que mostrarían algunas diferencias en la eficiencia de la poliadenilación (según los criterios establecidos en la tabla de decisión, ver tabla 1). Cuando se han detectado varias PAS en una misma UTR 3´, se ha indicado en negrita (ver tablas 2 y 3) cual es la eficiencia de poliadenilación esperada, suponiendo que ha sido utilizada la señal más al extremo 3´de dicha UTR. En los casos en que puede haber duda debido a varias PAS próximas no se ha resaltado ninguna de ellas.

En los casos en los que se puede interpretar que una UTR 3´alternativa ha sido originada por corte alternativo de otra variante de mayor longitud, se ha indicado entre paréntesis la eficiencia de poliadenilación en el contexto de la de mayor longitud. En los casos de Ptgds humana y ratón y Lcn2 humana no puede interpretarse que las UTRs 3´ alternativas se originen por cortes alternativos, sino que se han utilizando ciertos exones de forma alternativa en cada una de ellas. Por

ello en estos casos se ha tomado el contexto de cada alternativa para determinar la eficiencia de la poliadenilación.

De forma general en las lipocalinas con UTRs 3´ más cortas cortas (generalmente < 200 nt) y que no muestran alternativas conocidas, estas poseen una sola PAS que presentan características (tipo de PAS y posición en UTR 3´), que nos permite presuponer que determinan una traducción muy eficiente o eficiente.

Lipocalina	Long UTR3´	Posición PAS	Tipo PAS	Eficiencia Poliaden
3utr_apod_a,b,c_hum	198	153	C	**ME**
" "	198	68	NC*	PE
3utr_PTGDS_c_hum	214	191	C	ME
3utr_PTGDS_g_hum	178	159	C	ME
3utr_PTGDS_j_hum	639	142	C	PE
" "	639	510	NC*	PE
" "	639	621	NC*	**E**
3utr_rbp4_b_hum	388	211	NC	PE
" "	388	360	NC	**E**
" "	388	112	NC*	PE
" "	388	130	NC*	PE
3utr_rbp4_c_hum	186	112	NC*	-
	186	130	NC*	(PE)
3utr_apom_d,e_hum	121	97	C	ME
3utr_LCN1_b,h_hum	185	166	NC*	E
3utr_LCN2_b_hum	153	130	C	ME
3utr_LCN2_b(2)_hum	334	315	C	ME
3utr_LCN8_e_hum	112	95	C	ME
3utr_LCN12_c,c(2)_hum	103	78	C	ME
3utr_OBP2A_b_hum	133	114	C	ME
3utr_C8G_a_hum	193	175	NC	E
3utr_ORM2_b_hum	122	94	C	ME

Tabla 2. Clasificación de las UTR 3´de las lipocalinas humanas en función del tipo de PAS y de la posición que ocupan dentro del UTR 3´. "C": señal poliadenilación canónica; NC: señal no canónica frecuente; "NC*": otras señales no canónicas poco frecuentes. "ME: poliadenilación muy eficiente; "E": poliadenilación eficiente; "PE": poliadenilación poco eficiente.

Lipocalinas	Long UTR3´	Posición PAS	Tipo PAS	Eficiencia Poliaden
3utr_apod_a,b,d,_mouse	223	203	C	(PE)
3utr_apod_c_mouse	1149	203	C	PE
" "	1149	672	NC	PE
" "	1149	1128	NC	**E**
3utr_PTGDS_d_mouse	159	139	C	ME
" "	159	135	NC*	E
3utr_PTGDS_e_mouse	614	594	C	ME
" "	614	590	NC*	E
3utr_rbp4_c_mouse	128	114	NC	(PE)
3utr_rbp4_a,d_mouse	252	114	NC	PE
" "	252	225	NC	**E**
3utr_apom_a_mouse	117	89	C	ME
3utr_VEGP1_rat_a	164	146	NC*	E
3utr_LCN2_b_mouse	237	212	C	ME
" "	237	216	NC*	E
3utr_LCN8_a_mouse	107	85	C	ME
3utr_LCN12_a_mouse	78	55	C	ME
3utr_LCN13_a_mouse	164	145	C	ME
3utr_C8G_b/c/d_mouse	154	136	NC	E
3utr_ORM2_a_mouse	113	84	C	ME

Tabla 3. Clasificación de las UTR 3´de las lipocalinas de ratón en función del tipo de PAS y de la posición que ocupan dentro del UTR 3´."C": señal poliadenilación canónica; NC: señal no canónica frecuente; "NC*": otras señales no canónicas poco frecuentes. "ME: poliadenilación muy eficiente; "E": poliadenilación eficiente; "PE": poliadenilación poco eficiente.

3.2.- Búsqueda de motivos validados, en las secuencias de UTRs 3´.

Para la búsqueda de posibles motivos reguladores en la región UTR 3´ de las lipocalinas humanas y de ratón se sometieron dichas regiones a un análisis mediante la herramienta UTRscan. El resultado

de este análisis se muestra en la tabla 4.

En dicha tabla se muestra las señales de poliadenilación detectadas por el algoritmo UTRscan en las UTRs 3´de lipocalinas. En la tabla se muestran a modo de ejemplo tres de ellas, aunque fueron detectadas PAS en todas las UTRs 3´de lipocalinas humanas y de ratón. En estos ejemplos observamos diferentes PAS: la canónica "AAUAAA", la no canónica y segunda en importancia "AUUAAA" y otra alternativa menos frecuente como "CAUAAA". Detrás de cada PAS se encuentra una secuencia rica en "GU", necesaria para el correcto procesado de corte y poliadenilación. Estos resultados confirman en una gran proporción las PAS detectadas mediante el uso del algoritmo de patrones "dreg" (EMBOSS), que fue el método utilizado en el apartado anterior, para poder detectar todas las posibles PAS alternativas y donde ya se trató más en profundidad este asunto.

UTRs 3´	MOTIVO	POSICIÓN	SECUENCIA
3utr_Apod_a/b/d_ratón	PAS	[191,214]	AAUAAAcuccggaagcaagucagu
3utr_Rbp4_b_hum	PAS	[360,388]	AUUAAAuacuggcuucucccaacuuc cug
3utr_LCN1_h_hum	PAS	[166,185]	CAUAAA gagcuucagcaguu
3utr_Apod_c_ratón	MBE	[255,259] [301,305] [1102,1106]	AUAGU GUAGU AUAGU
3utr_Rbp4_d_ratón	MBE	[74,80] [234,238]	GUUUAGU AUAGU
3utr_Apom_a_hum	MBE	[554,558] [914,918]	GUAGU GUAGU
3utr_Rbp4_b_hum	K-BOX	[234,241]	CUGUGAUU

Tabla 4. Motivos predichos por UTRscan en las UTRs 3´de lipocalinas humanas y de ratón

Un segundo motivo encontrado por UTRscan es el motivo MBE (*Musashi binding element*). Dicho motivo inicialmente identificado en Xenopus [3] ha sido identificado posteriormente en otros vertebrados. En mamíferos ha sido identificado en células madre del sistema nervioso [4]. En estas

células ciertos ARNm presentan varias repeticiones en tandem de la siguiente secuencia de consenso "(G/A)U(n)AGU (n = 1 to 3)" y así puede producirse la unión de la proteína "Msi-1", la cual ejerce un efecto negativo sobre la traducción de los mismos.

Este motivo MBE es predicho en las UTRs 3´de cinco lipocalinas, en la tabla 4 se muestran las tres que presentan repetida la secuencia consenso, si bien en ninguna de ellas se da la repetición en tandem que parece necesaria.

Por otra parte en los ARNm donde se ha comprobado la funcionalidad de dicha secuencia, ésta aparece situada en el bucle de estructuras en horquilla típicas [4], de forma que queda muy accesible a la proteína de unión "Msi-1". Se obtuvieron las estructuras secundarias (MFE) de las UTRs 3´de lipocalinas, donde es predicho el motivo MBE, y en ningún caso se encontró que la secuencia consenso estuviese situada, de forma nítida, en el bucle de estructuras en horquilla, como sería de esperar de ser funcional. En las figuras 1 y 2 se muestran ejemplos de las estructuras 2D de las UTR 3´de lipocalinas que contienen al motivo aludido.

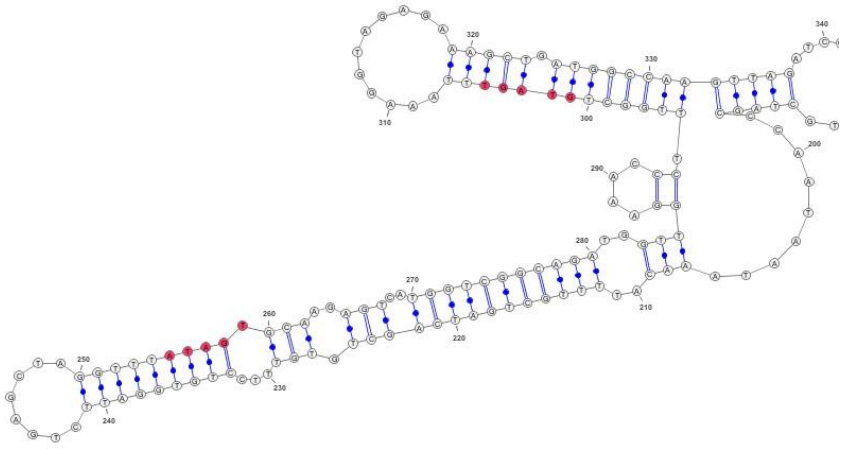


Figura 1. Estructura 2D MFE de la UTR 3´_apod_c_ratón predicha por RNAfold. En rojo se indica la posición del motivo MBE dentro de esta estructura. Sólo se muestra la región donde aparece el motivo.

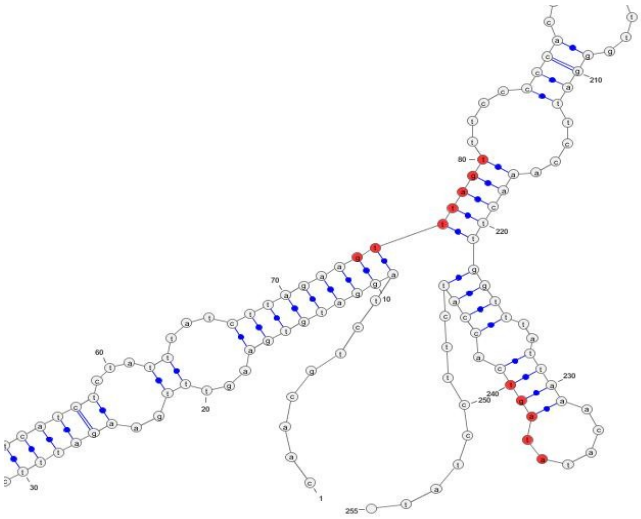


Figura 2. Estructura 2D (MFE) de la UTR 3´_rbp4_d_ratón predicha por RNAfold. En rojo se indica la posición del motivo MBE dentro de esta estructura. Sólo se muestra la región donde aparece el motivo.

Por último un tercer motivo predicho por UTRscan en la UTR 3´de Rbp4_b_humana es el motivo K-BOX. No parece esta una predicción correcta ya que el motivo funcional está formado por una o más copias de la secuencia "UGUGAU" que aparece, junto con otros motivos (Brd-box y GY-box), en las UTRs 3´de ciertos genes de Drosophila [5]. Este motivo parece implicar la formación de un duplex de ARN que contiene una secuencia complementaria a ciertos miARN, mediando así en una regulación pos-transcripcional negativa.

3.3.- Oligonucleótidos sobrerrepresentados

Como ya se comentó en el capítulo de las UTRs 5´el análisis de los oligonucleótidos sobrerrepresentados es una manera de encontrar posibles motivos implicados en alguna función reguladora. Para detectar oligonucleótidos sobrerrepresentados, en el conjunto de las UTRs 3´de lipocalinas, se procedió con el mismo algoritmo y método que para las UTRs 5´(ver métodos).

Los resultados muestran que hay un determinado número de oligonucleótidos sobrerrepresentados en ambas especies (ver tablas 5 y 6) alcanzando algunos de ellos elevados índices de significación.

UTR 3´ lipocalinas humanas			
(oligo 6 nt)			
Oligo	**Observado**	**Esperado**	**Ind. Signific.**
AAUAAA	8	0.1	9.48
AUAAAG	6	0.17	4.24
UAAAGA	6	0.23	3.44
AUAAAC	4	0.15	1.42
AGAUAA	4	0.18	1.13
CAUAAA	4	0.22	0.8
AAAUAA	3	0.1	0.56
AUUAAA	3	0.1	0.48
UAAACU	4	0.28	0.39
(oligo 7 nt)			
AUAAAGA	5	0.03	5.93
AAUAAAC	4	0.03	4.06
AAAUAAA	3	0.02	2.49
AAUAAAG	3	0.03	1.77
AUAAACU	3	0.04	1.47
CAUAAAG	3	0.08	0.69
UAAAGAG	3	0.08	0.61
UAUUAAA	2	0.01	0.59
UUAUUAA	2	0.01	0.54
CCAUAAA	3	0.09	0.51
AUUAAAU	2	0.02	0.37
GCACACA	5	0.5	0.28
GAAUAAA	2	0.02	0.07
(oligo 8 nt)			
AAUAAACU	3	0.01	3.53
CAUAAAGA	3	0.02	2.69
UAAAGAGC	3	0.02	2.05
UAUUAAAU	2	0	2.03
UUAUUAAA	2	0	1.86
UAAAGAUA	2	0	1.48
AAGAUAAU	2	0.01	1.25
AUAAAGAU	2	0.01	1.25
AAAUAAAC	2	0.01	1.24
AAUAAAGA	2	0.01	1.1
AUGUCUGU	3	0.05	1.07
UGAAUAAA	2	0.01	0.98
CCCUGCCC	8	1.21	0.84
GUCAGUGA	3	0.07	0.64
AUAAAGAG	2	0.01	0.6
CUGAAUAA	2	0.02	0.27
UCUCAGCC	4	0.27	0.19
AGAGCUUC	3	0.1	0.19
CCUGCCCC	7	1.21	0.01

Tabla 5. Oligonucleótidos sobrerrepresentados, de tamaño 6, 7 y 8 nt, en las UTRs 3´de lipocalinas humanas.

UTR 3´ lipocalinas ratón (oligo 6 nt)			
Oligo	**Observado**	**Esperado**	**Ind. Signific.**
AAUAAA	7	0.61	2.18
UAUAGU	4	0.26	0.55
AUUAAA	5	0.57	0.23
UAAACA	5	0.63	0.05
(oligo 7 nt)			
GUUUGUU	5	0.23	1.99
UAAACAU	4	0.13	1.6
UUUGUUU	5	0.32	1.27
UAAUAAA	3	0.1	0.43
UGUUUGU	4	0.26	0.39
AUUAAAC	3	0.12	0.23
UUGUUUG	4	0.31	0.11
(oligo 8 nt)			
GUUUGUUU	5	0.06	4.64
UGUUUGUU	4	0.07	2.48
UUGUUUGU	4	0.07	2.48
UUUGUUUG	4	0.09	2.2
CGGCAGAU	2	0.01	0.68
GCCCUGGC	3	0.09	0.54
GCCUCUGC	3	0.1	0.36
AUAAUAAA	2	0.02	0.31
ACGCCUCU	2	0.02	0.25
UUGUUCUU	3	0.11	0.23
UGUUCUUU	3	0.11	0.23
UCUGGAGG	3	0.12	0.18

Tabla 6. Oligonucleótidos sobrerrepresentados, de tamaño 6, 7 y 8 nt, en las UTRs 3´de lipocalinas de ratón.

Observamos, que tanto en humano como en ratón, aparecen entre los óligos más significativos de 6 nucleótidos los motivos “AAUAAA” Y “AUUAAA”, que se corresponden con las señales de poliadenilación (PAS) más frecuentes. Un gran porcentaje de los óligos de tamaños mayores (7 y 8 nucleótidos) sobrerrepresentados son óligos que incluyen o se solapan con los mencionados motivos de poliadenilación.

Como se observa en la tabla 7 el grado de significación de los motivos que contienen a “AAUAAA” en las UTRs 3´de lipocalinas humanas disminuye cuanto mayor es el tamaño del oligonucleotido, confirmando que la auténtica señal pertenece al óligo de 6 nucleótidos que se corresponde con dicha PAS canónica. Para ratón los resultados son semejantes.

Motivo	Longitud oligonucleótido	Indice de significación
AAUAAA	6	9,48
AAUAAAC	7	4,06
AAUAAAU	7	3,53
AAAUAAAC	8	1,24

Tabla 7. Motivos relacionados con la señal de poliadenilación "canónica" y su nivel de significación. En UTRs 3´de lipocalinas humanas

Respecto a la secuencia PAS no canónica "AUUAAA" no observamos, en humano, que haya disminución de la significación conforme aumenta el tamaño del óligo que la contiene o solapa con ella, más bien lo contrario (ver tabla 8). En ratón sin embargo tiene la misma significación el óligo de 6 y de 7 nt, no siendo significativo ningún óligo de 8 que la contenga (ver tabla 6).

Motivo	Longitud oligonucleótido	Indice de significación
AUUAAA	6	0,48
UAUUAAA	7	0,59
UUAUUAAA	7	0,54
UAUUAAAU	8	2,03

Tabla 8. Motivos relacionados con la señal de poliadenilación "no canónica", más frecuente, y su nivel de significación. En UTRs 3´de lipocalinas humanas

La explicación que podemos dar a este hecho es que esta PAS no canónica, que es utilizada con menor frecuencia en las UTRs 3´, debe ser una señal más débil, pudiendo ser esta la causa de los efectos observados en el índice de significación.

Respecto al resto de óligos, no relacionados con las señales de poliadenilación, cabe destacar el óligo "CCCUGCCC" en humano, que aparece en las UTRs 3´de cinco lipocalinas diferentes. La búsqueda de este motivo en TargetScan (versión 5.2) lo clasifica como diana de un miARN desconocido pero conservado en las UTRs 3´ ortólogas de 393 genes humanos.

En el caso de ratón observamos en la tabla 6 que aparecen óligos de tamaño 7 y 8 ricos en GU (GUUUGUUU y otros semejantes). Tras analizar las secuencias de UTRs 3´de ratón se comprobó

que están todos relacionados con una secuencia repetitiva (TTTG)n existente en Apo-D-c de ratón.

3.4.- Dianas de micro ARN en las UTRs 3´de lipocalinas

La región donde las dianas de miARN aparecen de forma preferente es la región UTR 3´[17], por ello el determinar la presencia de estos elementos en dicha región es un aspecto clave para determinar la regulación que esta ejerce sobre la expresión génica de las lipocalinas.

Tras proceder con el mismo algoritmo y método que en el caso de las UTRs 5´de lipocalinas (ver métodos) se obtuvieron los resultados que se muestran en las tablas 9 y 10.

UTR 3´	Variantes	Nº dianas accesibles (ddG < -10 Kcal/mol)	Nº dianas muy accesibles (ddG < -10 y dGopen > -10 Kcal/mol)
ApoD-Humana	a/b	23	13
Ptgds-Humana	c	14	0
	g	14	0
	j	33	0
Rbp4-Humano	b	70	59
Lcn1-Humano	h	50	20
Lcn2-Humano	b	9	0
	b(2)	29	8
Lcn8-Humano	e	7	1
Lcn12-Humano	c/c(2)	30	9
Obp2A-Humano	b	61	33

Tabla 9. Dianas de miARN predichas por el algoritmo PITA (Segal lab) en las UTRs 3´de lipocalinas humanas

UTR 3´	Variantes	Nº dianas accesibles (ddG < -10 (Kcal/mol))	Nº dianas muy accesibles (ddG < -10 y dGopen > -10 (Kcal/mol))
ApoD-Ratón	a	1	1
	c	7	2
Ptgds-Ratón	c/d	4	0
	e	8	4
Rbp4-Ratón	a/d	2	0
	c	1	0
ApoM-Ratón	a	3	2
Lcn1-Ratón	a	15	5
Lcn2-Ratón	b	8	5
Lcn8-Ratón	a	7	6
Lcn12-Ratón	a	7	5
Obp-2A-Ratón	a	10	6
C8G-Ratón	c	2	0
Orm2-Ratón	a	4	0

Tabla 10. Dianas de miARN predichas por el algoritmo PITA (Segal lab) en las UTRs 3´de lipocalinas de ratón

La comparación de los resultados obtenidos para las UTRs 3´, con los obtenidos para las UTRs 5´ (ver tablas 9 y 10 de capítulo x), nos muestra que, para ratón, hay un mayor número de lipocalinas que muestran dianas en su UTR 3´que en su UTR 5 ´ (11 frente a 5). Mientras que para humano el número de lipocalinas que presentan dianas de miARN es aproximadamente el mismo en ambas UTRs.

Por otra parte, en el conjunto de los datos, las dianas de miARN se muestran más abundantes en las UTRs 3´que en las 5´. Este hecho se hace más evidente si calculamos el número de dianas por cada 1000 nucleótidos de "secuencia total acumulada" de cada región UTR de las lipocalinas (se contabilizaron solo las dianas muy accesibles). Este valor es de 5,12 en las UTRs 5´ de humano y 9,7 para ratón, siendo para las UTRs 3´ de 68,4 y 13,6 respectivamente. Por lo que las dianas de miARN son, especialmente en humano, nítidamente más frecuentes en las UTRs 3´ que en las UTRs 5´.

Curiosamente se encontraron muy escasas coincidencias entre las dianas de humano y ratón, solo 6 dianas, de entre todas las dianas (más accesibles) que muestran en sus UTRs 3´ el conjunto de las lipocalinas en estas dos especies.

Observamos, que al igual que ocurría en las UTRs 5´, en las UTRs 3´que presentan formas alternativas, estas muestran un número diferente de dianas (bien en ambas clases de dianas: accesible y muy accesibles o al menos en las accesibles (ver tablas 9 y 10), dándose varios casos donde una UTR 3´alternativa presenta varias dianas muy accesibles y la otra ninguna. En el único caso donde hay posibilidades de comparar si existen dianas comunes, de entre las más accesibles, entre diferentes alternativas se reduce a Apo-D de ratón (ver tabla 10), y no se encontró ninguna coincidencia.

3.4.1.- Realidad biológica de las dianas de miARNs en las UTRs 3´de lipocalinas

Ya se ha comentado previamente, al analizar las UTRs 5´ (ver capítulo IV), la influencia que la longitud de la UTR y su MFE podrían tener en la probabilidad de aparición de dianas de miARN en estas regiones, si las mismas fuesen falsos positivos. Para comprobar en qué medida estos parámetros influyen en las diferencias encontradas en el número de dianas de miARN en las UTRs 3´ , se procedió al igual que con las UTRs 5´ (ver métodos), al cálculo de la correlación entre

dichas variables.

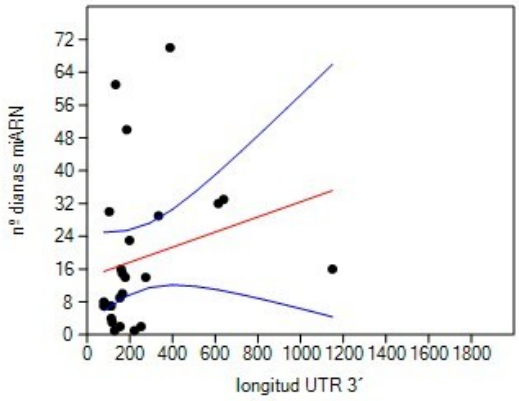


Figura 3. Relación entre el número de dianas de miARN encontradas y la longitud de la UTR 3´. En rojo línea de regresión, en azul intervalo del 95% confianza

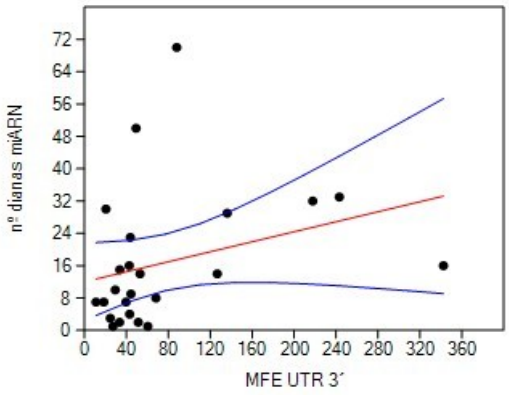


Figura 4. Relación entre el número de dianas de miARN encontradas y la MFE (Kcal/mol) de la UTR 3´. En rojo línea de regresión, en azul intervalo del 95% confianza Se han representado los valores de MFE en valor absoluto.

En las figuras 3 y 4 parece observarse una cierta correlación entre la longitud y la MFE de las UTRs 3´ frente al número de dianas de miARN encontradas, pero estas no son significativas (para longitud r = 0.233; P = 0.262 y para MFE r = 0.299; P = 0,154).

Otra forma de testar la realidad biológica de la dianas de miARN es comprobar si las mismas se corresponden con familias de miARNs que están profusamente conservados entre los vertebrados, o al menos conservados entre los mamíferos. Para ello se utilizó la base de datos TargetScan que dispone de datos de conservación de los miARN en las UTRs 3´de vertebrados. El cruce de los datos de las dianas (más accesibles) encontradas en las UTRs 3´de lipocalinas, con los disponibles en esta base de datos, nos muestra que hay cierto número de dianas, ubicadas en diferentes UTRs 3´, que se corresponden con miARNs que muestran alto grado de conservación (entre vertebrados) o al menos conservación entre mamíferos (ver tablas 11 y 12).

UTR 3´	miARN	Conservación
ApoD_a_hum	hsa-miR-185	1
	hsa-miR-202	1
Rbp4_b_hum	hsa-miR-125a-3p	1
	hsa-miR-127-3p	1
	hsa-miR-134	1
	hsa-miR-146a	2
	hsa-miR-185	1
	hsa-miR-296-3p	1
	hsa-miR-324-5p	1
	hsa-miR-363	2
Lcn1_b_hum	hsa-miR-24	2
	hsa-miR-296-3p	1
Lcn2_b2_hum	hsa-miR-296-3p	1
Lcn12_c_hum	hsa-miR-330-5p	1
Obp2a_b_hum	hsa-miR-125a-3p	1

Tabla 11. *miARNs que presentan dianas en las UTRs 3´ de lipocalinas humanas y que se encuentran conservados. 2: profusamente conservados en vertebrados; 1: conservados en mamíferos.*

UTR 3´	miARN	Conservación
ApoD_a_ratón	mmu-miR-383	2
ApoM_a_ratón	mmu-miR-124	2
Lcn1_a_ratón	mmu-miR-296-3p	1
Lcn8_a_ratón	mmu-miR-503	2
	mmu-miR-214	2
Ptgds_e_ratón	mmu-miR-202-3p	1
Lcn12_a_ratón	mmu-let-7b	2
	mmu-miR-449a	2
	mmu-miR-449b	2
	mmu-miR-34a	2
Obp2a_a_ratón	mmu-miR-125a-3p	1

Tabla 12. *miARNs que presentan dianas en las UTRs 3´ de lipocalinas de ratón y que se encuentran conservados. 2: profusamente conservados en vertebrados; 1: conservados en mamíferos.*

Observamos, en las tablas 11 y 12, que existe un número semejante de lipocalinas humanas y de ratón que muestran en sus UTRs 3´dianas que se corresponden con miARNs conservados, destacando Lcn12 de ratón y Rbp4 de humano en el número de ellas, especialmente esta última.

La proporción de dianas que se corresponden con miARNs conservados es mayor en ratón, donde se conservan un promedio del 40% de las dianas encontradas, mientras que en humano es del 12 % . Además las dianas que se encuentran más profusamente conservadas (indicadas con 2 en las tablas 11 y 12) son más abundantes en ratón que en humano.

Existen 2 miARNs (se muestran coloreados en tablas 11 y 12) que presentan dianas comunes en las UTRs 3´de diferentes lipocalinas humanas y así mismo estos dos miARNs presentan dianas en algunas UTRs 3´de ratón, dándose el caso que una de ellas es compartida por la UTR 3´ de las lipocalinas ortólogas de Obp2a humana y de ratón.

4. - Discusión

Es conocido que las señales de poliadenilación (PAS) alternativas en eucariotas son un fenómeno relativamente abundante [2]. En humanos ha podido estimarse que alrededor del 20% de los ARNms poseen señales de poliadenilación alternativa [12]. La presencia de estas señales alternativas suponen un mecanismo de regulación de la expresión génica importante, mediante la modulación del nivel de eficiencia de la poliadenilación.

El análisis de este fenómeno en las UTRs 3´de lipocalinas ha puesto de manifiesto que hay ciertas lipocalinas que parecen recurrir a este mecanismo de regulación. Ha podido detectarse la presencia de potenciales PAS alternativas en las UTRs 3´de lipocalinas de cierta longitud (generalmente > 200 nt) especialmente en las lipocalinas más ancestrales como Ptgds y Rbp4, en humano y ratón, y ApoD, solo en ratón. La presencia de estas PAS alternativas al ser clasificadas, según el procedimiento citado en métodos, tienen como resultado que exista la posibilidad de distintas eficiencias en la poliadenilación. En estos casos (con la excepción de Ptgds) se han encontrado que junto a la UTR 3´ larga existe otra alternativa más corta, que coincide con uno de los posibles cortes asociados a una de las PAS alternativas, hecho que corrobora que en estas lipocalinas el uso alternativo de PAS es una realidad biológica.

En las lipocalinas más recientes, que presentan UTRs 3´más cortas y sin formas alternativas (excepto Lcn2), no se detectan PAS alternativas, presentando estas generalmente la señal canónica y pudiendo clasificarse como de poladenilación muy eficiente o eficiente.

Tendríamos entonces por una parte lipocalinas evolutivamente más antiguas con la posibilidad de utilizar UTR 3´alternativos que permitirían, por la presencia de PAS alternativos, una regulación de la poliadenilación y por tanto de la eficiencia de la traducción, en función de la mayor o menor estabilidad del ARNm. Por otra parte las lipocalinas más recientes, sin UTRs 3´alternativas y que muestran en general una predisposición a una poliadenilación eficiente, darían lugar a una traducción eficaz y con menores posibilidades de regulación.

Además de la regulación que permite el uso de PAS alternativas, por lo comentado en el párrafo anterior, la elección de una de estas señales tiene también efectos sobre la longitud de la UTR 3 ´debido al corte asociado a la poliadenilación. Es conocido que una mayor longitud de la UTR 3 ´tiene influencia sobre la presencia de señales reguladoras, como pueden ser las dianas de miARN [13 y 14]. Por ello las PAS alternativas permiten también, al producir cortes alternativos en el

procesado del ARNm, que se puedan ejercer diferentes regulaciones debido a la presencia-ausencia de elementos como los miARN u otros elementos reguladores, consiguiéndose así que pueda llevarse a cabo un ajuste fino de la regulación post-transcripcional según las necesidades.

La búsqueda de motivos validados en las UTRs 3´de lipocalinas, mediante UTRscan, ha supuesto una corroboración de la mayoría de las PAS identificadas previamente por un método diferente. Así mismo el estudio sobre oligonucleótidos sobrerrepresentados también ofrece resultados que apuntan a que estas señales (al menos las PAS canónica y la señal alternativa más frecuente) tienen un significado biológico.

El motivo MBE también ha sido detectado por UTRscan en las UTRs 3´de algunas lipocalinas, si bien no muestra las características que están presentes en los motivos nativos. A pesar de esto, el hecho de que se presente en las UTRs 3´ cinco lipocalinas, mostrándose la secuencia consenso repetida entre 2 y 3 veces en tres de ellas, son indicios que nos hacen pensar en que podrían tener algún papel regulatorio, si bien no exactamente igual que en el motivo original. En la misma línea el estudio de sobrerrepresentación de nucleótidos nos indica que un motivo diferente identificado en cinco lipocalinas podría estar desempeñando alguna función reguladora en las UTRs 3´de las mismas. El análisis mediante TargetScan clasifica a dicho motivo como posible diana de un miARN desconocido, aunque no podemos descartar que sea un elemento regulador de diferente naturaleza.

El estudio realizado sobre la presencia de potenciales dianas de miARN en las UTRs 3´de lipocalinas demuestra que, incluso aplicando criterios muy restrictivos en los requerimientos energéticos de la interacción miARN-diana, hay un número considerable de lipocalinas que presentan dichas dianas en sus UTRs 3´. El hecho esperable de que estas dianas se muestren más abundantes en las UTRs 3´que en las 5´y que las mismas se muestren también más abundantes en las UTRs 3´de lipocalinas más ancestrales, que muestran signos de estar sometidas a mayor regulación, son indicios de que al menos parte de estas dianas pueden ser funcionales.

El hecho de que no haya una clara correlación entre la longitud de la UTR 3´y el número de dianas presentes es de difícil interpretación en el caso de estas regiones. Hay evidencias de que las UTRs 3´de mayor longitud suelen presentar mayor número de dianas de miARN frente a alternativas más cortas con menor número de dianas o sin ellas [16]. El hecho de que en las lipocalinas sólo algunas (entre las más ancestrales) muestren variabilidad en su UTR 3´ mostrando formas cortas o largas, podría explicar la falta de correlación encontrada. Por otra parte esta falta de correlación también nos indica que no debe de haber un número alto de falsos positivos en estas predicciones, ya que de

lo contrario la longitud si influiría en la probabilidad de aparición de las mismas y aparecería una clara correlación positiva longitud-nº de dianas. Por lo tanto hemos de interpretar, que en las lipocalinas, la presencia o no y el número de dianas de miARN obedece en gran medida a diferentes necesidades de regulación.

Se ha encontrado que la frecuencia de dianas de miARN muy accesibles alcanza valores mayores en las UTRs 3´de lipocalinas humanas que en las de ratón y por otra parte que la proporción de dianas que pertenecen a miARNs conservados es menor en humanos que en ratón. Podemos interpretar estos resultados como un indicio de diferentes necesidades de regulación de la expresión de lipocalinas en los diferentes taxones de mamíferos. Serían así las lipocalinas de primates las que, necesitadas de una mayor y más específica regulación por estos elementos, habrían evolucionado en el sentido de adquirir mayor cantidad de nuevas dianas en sus UTRs 3´.

A modo de conclusión podemos decir que se han encontrado evidencias de que en ciertas lipocalinas (tales como Ptgds, Rbp4 y ApoD) las UTRs 3´juegan un papel regulador mediante la presencia de señales de poliadenilación alternativas, las cuales según como sean utilizadas, pueden dar lugar a UTRs 3´ con diferencias en la eficiencia de la poliadenilación y por lo tanto en la establididad del ARNm. Además el uso de diferntes PAS, al producir UTRs 3´de diferente longitud, que determinan la presencia o ausencia de ciertos elementos, permiten que se produzca diferente regulación de la expresión génica. En este estudio ha podido corroborarse que hay diferencias entre UTRs 3´alternativas en la presencia-ausencia de dianas de miARN o en diferente número de ellas y esto podría estar dando lugar a una expresión diferencial entre las mismas.

No podemos descartar una regulación mediada por la presencia o no de circularización del ARNm, que podría producirse o no en función del UTR 3´alternativo utilizado, el cual permitiría o no la interacción entre la UTR 3´y 5´. Esta circularización puede tener un efecto tanto inhibitorio como potenciador de la traducción [15]. Por otra parte también se ha comprobado que el uso de UTRs 3 ´alternativas de diferente longitud puede dar lugar a una diferente localización de las proteínas expresadas [16]. Todos estas consideraciones complican la interpretación del posible papel que cumple una determinada UTR 3´. Si bien lo que nos dejan ver los datos aquí considerados para las UTR 3´de las lipocalinas más ancestrales, es que se constata en ellas la presencia de UTRs 3´ alternativas con diferentes longitudes y propiedades, existiendo así oportunidades para que se de en ellas una regulación más compleja y diversa de la expresión génica.

5. - Bibliografía

[1] Colgan, D.F. and J.L. Manley.. Mechanism and regulation of mRNA polyadenylation. *Genes. Dev.* **11:** 2755–2766 (1997).

[2] Graber, J.H., C.R. Cantor, S.C. Mohr, and T.F. Smith.. In silico detection of control signals: mRNA 3-end-processing sequences in diverse species. *Proc. Natl. Acad. Sci. USA* **96:** 14055–14060 (1999).

[3] Charlesworth, A., Ridge, J., King, L.A., MacNicol, M.C. and MacNicol, A.M. A novel regulatory element determines the timing of Mos mRNA translation during Xenopus oocyte maturation. *EMBO J.* **21**, 2798-2806 (2002).

[4] Imai T, Tokunaga A, Yoshida T, Hashimoto M, Mikoshiba K, Weinmaster G, Nakafuku M, Okano H. The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. *Mol Cell Biol.* **21**, 3888-900 (2001)

[5] Lai et al.The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of enhancer of split complex transcripts.*Development* **125**, 4077-4088 (1998).

[11]. Wahle, E. & W. Keller. The biochemistry of polyadenylation. *TIBS* **21**, 247–250 (1996).

[12] Beaudoing, E., et al. Patterns of Variant Polyadenylation Signal Usage in Human Genes. *Genome Research* **10,** 1001-1010 (2000).

[13] Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3´ untranslated regions and fewer microRNA target sites. *Science* **320,** 1643–1647 (2008).

[14] Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3´UTR evolution. *Cell* **123,** 1133–1146 (2005).

[15] Mazumder, B., Seshadri, V. and Fox, P.L. Translational control by the 3´-UTR: the ends specify the means. *TRENDS in Biochemical Sciences* **28,** 91-98 (2003).

[16] Berkovits, B. D.. & Mayr, C. Alternative 3´UTRs act as scaffolds to regulate membrane protein localization.. *Nature* **522**, 363–367 (2015)

[17] Barrett, L.W., Fletcher, S., Wilton, S.D.. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* **69,** 3613-3634 (2012)

(VI)

ESTRUCTURA SECUNDARIA DE LAS UTRS DE LIPOCALINAS

1.- Objetivos

Los objetivos de este capítulo son los siguientes:

- En primer lugar caracterizar, con un enfoque general, el repertorio de estructuras secundarias (2D) de las UTRs de las lipocalinas de mamíferos, mediante diversos análisis bioinformáticos. Se pretende conocer así la relevancia que dichas estructuras 2D tienen en la función reguladora de estas regiones, en esta familia de proteínas.

- En segundo lugar, a una escala más de detalle, utilizar la estrategia de buscar motivos 2D locales conservados en las UTRs 5´ y 3´ortólogas de las lipocalinas, para así determinar la existencia de posibles elementos reguladores en las mismas. Una vez encontrados someterlos a diversas pruebas de contraste que permitan respaldar su realidad biológica.

2.- Métodos

2.1.- Determinación del repertorio de estructuras 2D de las UTRs 5´ y 3´de las lipocalinas

Los modelos de predicción de estructuras 2D de ARN suelen ofrecer la estructura con menor energía libre de plegamiento (MFE) como la solución óptima. Pero es sobradamente conocido que no siempre la estructura MFE es la que se corresponde con la estructura nativa [1 y 2]. Sin embargo parece razonable que la estructura nativa debe encontrarse en un rango de energía subóptimo, no muy alejado respecto a la MFE. Surge aquí un problema ya que el número de estructuras subóptimas posibles, aún en un rango de energía reducido, crece exponencialmente con la longitud de la secuencia de ARN.

Existen diferentes enfoques para abordar este problema complejo. Uno de ellos es elaborar una clasificación abstracta de las posibles estructuras que puede adquirir un determinado ARN, obteniendo así estructuras representativas de los diferentes conjuntos de estructuras que se pliegan

con ramificaciones o derivaciones parecidas. Se reduce así enormemente el número de estructuras a considerar. Esto puede realizarse mediante el algortimo RNAshape (http://bibiserv.techfak.uni-bielefeld.de/rnashapes/) [2]. Como demuestran los autores, utilizando una amplia muestra de ARNs estructurales, el número de estructuras que habría que estudiar se reduce enormemente al utilizar RNAshape, frente al conjunto total de estructuras posibles. El hecho de obtener un número reducido de estructuras alternativas con este algoritmo no parece que lleve a perder información biológica relevante. Esto lo demuestra la constatación de que las predicciones con RNAshape sobre diferentes ARNs estructurales, cuyas estructuras nativas son conocidas, permiten obtener dichas estructuras dentro de las formas subóptimas próximas a la MFE (dentro de las 10 primeras estructuras predichas en un rango de energía libre de 5 Kcal/mol o incluso menor) [2].

Para conocer en que medida la estructura 2D de las regiones UTRs 5´de lipocalinas es relevante se procedió a estudiar el repertorio de plegamiento de estas secuencias para determinar si se ajusta a lo esperado para ARNs estructurales. Se procedió a analizar las secuencias UTRs 5´ con RNAshape en un intervalo de energía de 5 Kcal/mol, tal como los autores de RNAshape habían analizado una muestra de ARNs estructurales de la base de datos Rfam (http://rfam.xfam.org/) [3]. Una vez calculado el número de estructuras alternativas para las UTRs 5´de cada lipocalina se agruparon los datos con longitudes semejantes y se determinó el número promedio de dichas estructuras para cada categoría de longitud. Se procedió igualmente con las secuencias UTRs 3´de lipocalinas. Posteriormente se representó gráficamente el número promedio de estructuras alternativas observadas en las UTRs 5´ y 3´de lipocalinas, en función de su longitud, frente al número de estructuras esperadas para los ARNs estructurales (de la citada base de datos Rfam) de longitudes equivalentes.

Para conocer la semejanza entre las estructuras alternativas del repertorio posible de cada UTR de las lipocalinas (obtenidas con RNAshape) se utilizó el programa RNAforester (http://bibiserv2.cebitec.uni-bielefeld.de/rnaforester) [4]. Esta herramienta obtiene alineamientos múltiples de estructuras 2D de ARNs mediante un algoritmo de programación dinámica y basado en un modelo de alineamiento progresivo en forma de árbol. Dicho algoritmo se muestra muy eficiente en la comparación e identificación de elementos puramente estructurales [5]. De las diversas opciones de salida del programa se utilizó PseudoViewer para comprobar visualmente las estructuras que pueden considerarse semejantes.

2.2.- Identificación de motivos estructurales 2D " locales" de interés biológico en las UTRs 5´ y 3´de lipocalinas

Existen diferentes estrategias que pueden usarse para identificar nuevos candidatos a motivos reguladores en ARNs. Una de ellas es la "*estrategia basada en la secuencia",* dicha estrategia consiste en obtener un alineamiento de las secuencias de nucleótidos candidatas para identificar posibles motivos reguladores gracias a su grado de conservación. Posteriormente se utiliza un programa de plegamiento de ARN, para confirmar que dichas secuencias conservadas pueden formar un motivo estructural común. Esta estrategia falla si no se encuentran secuencias con el suficiente grado de conservación. Una estrategia alternativa es la "estrategia puramente estructural", que sin considerar el grado de conservación que puede haber entre secuencias candidatas y mediante diversos procedimientos, busca un consenso entre dichas secuencias que supuestamente comparten algún o algunos motivos estructurales.

Para la identificación de posibles motivos estructurales en las UTRs de lipocalinas se ha recurrido como estrategia principal a una "*estrategia puramente estructural"*. De forma complementaria para los casos en que dicha estrategia ha dado resultados positivos se ha aplicado la "*estrategia basada en el alineamiento de las secuencias"*, a modo de contraste.

2.2.1.- Estrategia puramente estructural

Recientemente se han desarrollado, dentro de este tipo de estrategias, enfoques alternativos en la búsqueda de motivos estructurales reguladores en ARNs, como es el caso de los modelos SCFG (*Stochastic context-free grammars*), una clase de modelos probabilísticos para predecir las estructuras comunes a unas secuencias de ARNs dadas. El enfoque de estos modelos sustituye a las consideraciones termodinámicas de la mayoría de herramientas de predicción de motivos en ARN. Un algoritmo de esta clase que se muestra muy eficiente y con tiempos de computación reducidos es Predict a Motif (*http://genie.weizmann.ac.il/pubs/rnamotifs08/rnamotifs08_predict.html*) [6]. Dicho algoritmo toma una serie de secuencias de ARN, no alineadas, que supuestamente comparten un motivo, con cierta estructura secundaria. El algoritmo primero identifica estructuras candidatas específicas y relativamente cortas, presentes en el mayor numero posible de entre las secuencias facilitadas, y estas son utilizadas luego como "semillas" para un modelo de inferencia probabilística que refina el motivo predicho usando estimaciones estadísticas.

Para llevar a cabo este análisis se tomaron las secuencias de UTRs 5´ortólogas previamente identificadas (ver capítulo sobre conservación), en las lipocalinas Apo-D, Rbp4, Ptgds y Apo-M, para diferentes especies de mamíferos. Así mismo se tomaron las secuencias de UTR 3´ ortólogas de Apo-D de diferentes mamíferos, previamente identificadas. Los conjuntos de secuencias ortólogas fueron analizados con Predict a Motif, y de los resultados obtenidos se seleccionaron los motivos estructurales que poseían mayor puntuación y que eran compartidos al menos entre tres especies de mamíferos. Posteriormente, de entre todos ellos, se eligieron los motivos que pudieron identificarse en las estructuras globales MFE o subóptimas de las UTRs 5´ y 3´de las lipocalinas citadas, que habían sido obtenidas previamente con RNAshape. Este último criterio se aplicó ya que, de no estar presentes, es probable que estos motivos no lleguen a formarse en la estructura funcional.

2.2.2.- Estrategia basada en alineamiento de secuencias

1. Se recurrió para este análisis al algoritmo RNAalifold [7]. Este algoritmo toma un alineamiento de ciertas secuencias relacionadas y a partir de este calcula una estructura consenso de mínima energía para ellas.

2.3.- Diseño de "patrones de búsqueda" de los motivos 2D identificados en las UTRs 5´ de lipocalinas y aplicación sobre bases de datos de secuencias UTRs 5´de mamíferos

La estrategia elegida para esta tarea ha sido el diseño gráfico de los motivos 2D y la posterior conversión de estos en patrones de búsqueda mediante algoritmos tipo ADP (*Algebraic Dynamic Programming*), que posteriormente se han aplicado a una muestra de 3000 secuencias de UTRs 5 ´de mamíferos, obtenida de forma aleatoria, de la base de datos UTRdbase [8]. Así mismo se obtuvo una muestra de secuencias aleatorias, mediante Shuffleseq de EMBOSS [9], del mismo tamaño, a partir de la muestra de secuencias de UTRs 5´de mamíferos, para ser utilizada como control.

Para realizar este análisis se ha utilizado el programa Locomotif (http://bibiserv.techfak.uni-bielefeld.de/locomotif/) [10]. Dicha herramienta permite un diseño gráfico de las estructuras deseadas, permitiendo realizar múltiples especificaciones en las mismas. Posteriormente permite convertir este diseño en un patrón de búsqueda mediante ADP, que es traducido posteriormente a lenguaje XML , el cual puede hacerse correr "via web" sobre la base de datos de secuencias

deseada. Los detalles de los diseños gráficos de los diferentes motivos y de las especificaciones de búsqueda de los mismos se muestran en el apartado de resultados correspondiente.

2.4.- Cálculo de la "robustez estructural" de los motivos 2D identificados en las UTRs 5´de lipocalinas

Para el cálculo de esta propiedad se utilizó el procedimiento que se detalla a continuación.

- Se eligió la secuencia de la UTR 5´de la lipocalina que contenía los motivos estructurales hipotéticamente funcionales, y se sometió al análisis mediante RNAshape [2], utilizando una ventana desplazante de 50 nt con un incremento de 5 nt. Del conjunto de subestructuras obtenidas se seleccionaron las subsecuencias que mostraban, sin ambigüedad, el motivo estructural en cuestión.
- La estructura (dot-bracket) de estas subsecuencias de 50 nt, fué utilizada para obtener secuencias de "ARN inversas" mediante "vrnainverse" de EMBOSS [9]. Dichas secuencias son secuencias aleatorias, pero que se pliegan con la misma estructura 2D, que se especifique al programa, en este caso las de las secuencias nativas que contienen a los motivos en estudio. Se seleccionaron así 10 secuencias inversas para cada una de las secuencias nativas, cuyo G+C fuese semejante al de la secuencia nativa correspondiente.
- Posteriormente se procedió a mutar tanto las secuencias nativas como las correspondientes aleatorias mediante el programa RNAmutants (http://bioinformatics.bc.edu/clotelab/RNAmutants/index.spy?tab=webserver) [11]. Esta aplicación obtiene todo un repertorio de estructuras para los mutantes (se seleccionó originar mutantes de 1 a 5 sustituciones) y ofrece entre sus resultados los que poseen estructuras más representativas, de todo el repertorio de plegamientos posibles. Esta característica es interesante por que excluye las secuencias mutadas cuyas estructuras que no sean representativas.
- Una vez obtenidos estos mutantes representativos, se seleccionaron al azar 100 de ellos, para cada uno de los diferentes niveles de mutación de 1, 3 y 5 sustituciones, tanto para las secuencias nativas como para las aleatorias, y se calculó la distancia entre la estructura de cada mutante y la estructura original correspondiente. Para esto se utilizó la aplicación RNApdist (Vienna RNA 1.8.4: http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::rnapdist) [12].
- Por último, los resultados de las distancias obtenidas para las secuencias nativas y aleatorias se trataron estadísticamente, mediante los recursos de la hoja de cálculo

"Calc" de Open Office.

2.5.- Búsqueda de *pseudoknots* entre los elementos de estructura secundaria de las UTRs 5´de lipocalinas. Predicciones sobre la estructura terciaria.

Para la determinación de posibles *pseudoknots* entre los motivos de estructura secundaria, identificados previamente en estas UTRs, se utilizó la herramienta pAliKiss (http://bibiserv2.cebitec.uni-bielefeld.de/palikiss) [13]. Esta herramienta es un híbrido entre un algoritmo que obtiene una estructura secundaria de consenso a partir de secuencias de ARN alineadas, y *otro* que predice la formación de *pseudoknots*. Se facilitó a esta herramienta el alineamiento múltiple entre las secuencias de UTRs 5´ ortólogas de lipocalinas de mamíferos y se analizaron los resultados en busca de la participación de los mencionados motivos en la formación de *pseudoknots*.

La predicción de la estructura terciaria de estas UTRs 5´ se realizó mediante RNAComposer (http://rnacomposer.ibch.poznan.pl/Home) [14]. Esta herramienta realiza una predicción totalmente automatizada de la estructura terciaria a partir de una secuencia de ARN y su estructura secundaria. El algoritmo descompone la estructura secundaria en elementos sencillos, calcula su estructura terciaria recurriendo a una librería de elementos de estructura terciaria (RNA Frabase) y posteriormente los ensambla. Se obtiene así una estructura terciaria inicial, la cual finalmente es sometida a un refinamiento por optimización energética, dando lugar a la predicción definitiva.

3.- Resultados

3.1.- Caracterización de las estructuras 2D de las UTRs 5´ y 3´de lipocalinas

En un capítulo previo habíamos realizado una primera aproximación para clasificar las distintas UTRs 5´de las lipocalinas utilizando el árbol de decisión del modelo CART [15]. Los resultados de aplicar dicho modelo a estas UTRs (ver tabla 1 y 2 del capítulo IV) pusieron de manifiesto que un gran número de ellas pueden considerarse que tienen función inhibidora de la traducción

(Clasificadas como categoría " I " según este modelo). Sin embargo el modelo CART solo considera parámetros como la longitud, el contenido en G+C y la presencia o no de uAUGs, no teniendo en cuenta aspectos relacionados con la estructura secundaria (2D) que adquieren las secuencias de las UTRs 5´. Dicha estructura, es conocido, que juega un importante papel a la hora de ejercer una determinada regulación de la traducción del ARNm, de la estabilidad de dicho ARNm o de la localización del mismo [16 y 17].

Así mismo algunas estructuras secundarias que adquieren las UTRs 3´desempeñan un papel importante en la regulación de la expresión génica [17].

La estructura nativa de un determinado ARN estructural no siempre es la estructura óptima o MFE, sino que puede ser una de las estructuras subóptimas del repertorio de plegamientos posibles [1 y 2]. Por ello conocer el repertorio de estructuras 2D alternativas, que en un intervalo de energía dado, puede presentar una determinada UTR 5´ es de gran interés. Así mismo conocer si estas estructuras presentan elementos característicos (como horquillas, o loops, entre otras), es fundamental para dilucidar la función reguladora de dicha región.

3.1.1.- Repertorio de estructuras 2D de las UTRs 5´ y 3´de lipocalinas

Existen algunos indicios de que los ARNs estructurales presentan un repertorio de estructuras 2D alternativas más reducido que otros ARNs no estructurales [2], por lo tanto conocer la magnitud del repertorio de estructuras de las UTRs 5´ y 3´de las lipocalinas puede darnos pistas sobre la relevancia que dichas estructuras tienen en la función reguladora de estas regiones.

Para determinar el repertorio de estructuras 2D alternativas de las UTRs se utilizo la herramienta bioinformática RNAshape [2], ya que su algoritmo es muy eficiente para obtener información biológica relevante (ver detalles en métodos). El valor medio del número de estructuras 2D alternativas, para las regiones UTRs (5´y 3´) de diferentes longitudes, se representó frente al número esperado para ARNs estructurales conocidos, de longitudes equivalentes (pertenecientes a una muestra de la base de datos de Rfam, ver métodos).

Como se observa en la figura 2, para las secuencias entre 75 y 200 nucleótidos el comportamiento es semejante entre las tres clases (UTR5´, UTR3´y ARNs estructurales de Rfam). A partir de un

tamaño aproximado de 200 nucleótidos el número medio de estructuras alternativas es apreciablemente menor en las secuencias UTR5´que en la muestra de ARNs estructurales, mientras que en las secuencias UTRs 3´ el número es mayor que las de dicha muestra de referencia.

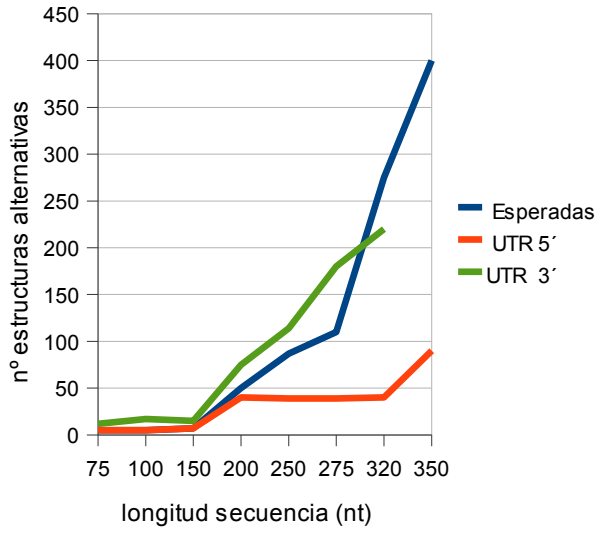


Figura 1. Representación del número promedio de estructuras alternativas obtenidas con RNAshape (intervalo de 5kcal/mol) para las UTRs 5´de lipocalinas (rojo), y las UTRs 3´ (verde), en función de su longitud, frente a los valores esperados para una muestra de ARNs estructurales (azul) de longitudes equivalentes .

Estos resultados de un menor repertorio de plegamientos alternativos en las UTRs 5´, respecto a los esperados, nos llevan a considerar un importante papel de la estructura secundaria global de dicha región en el desempeño de sus funciones reguladoras. Así mismo parecen confirmar la menor relevancia de la estructura 2D global de las regiones UTRs 3´. Al hacer estas consideraciones hemos de esperar que para las UTRs 5´, dado que su estructura secundaria parece estar muy definida, las diferencias entre la estructura MFE y las estructuras subóptimas sean pequeñas y que podamos tomar a esta (MFE) como la estructura biológicamente representativa en las UTRs 5`de lipocalinas. Este aspecto es analizado en el siguiente apartado.

3.1.2**.-** Semejanza entre las estructuras 2D alternativas del repertorio de plegamientos de las UTRs 5´ y 3´de lipocalinas

Para abordar esta cuestión se analizaron las diferentes UTRs 5´ y 3´de las lipocalinas humanas y de ratón, que previamente se habían obtenido con RNAshape en un intervalo de energía de 5 Kcal/mol, eligiendo entre las 15 primeras estructuras predichas y comparando entre ellas, tanto su estructura general como los principales elementos estructurales locales que las forman. Se utilizó para esta tarea la herramienta RNAforester [4], que permite obtener el alineamiento entre múltiples estructuras secundarias y obtener resultados gráficos fácilmente interpretables (ver métodos).

3.1.2.1- Región UTR 5´

Tras realizar el análisis con RNAforester se observó que existen diferencias entre las UTRs 5´ cortas (< 100 nt) y largas (>100 nt). Así para las UTRs 5´de más de 100 nucleótidos se encontraron escasas diferencias entre la estructura MFE y las estructuras subóptimas. Podemos observar en las figuras 2 a 4, ejemplos de como se mantiene la conformación general entre la MFE y las estructuras subóptimas, así como también se mantienen los principales elementos locales presentes (horquillas, bucles o *loops*, protuberancias o *bulges*, dobles hélices internas, etc). Comprobamos que la semejanza entre la MFE y estructuras subóptimas ocurre incluso en UTRs 5´ de gran tamaño (donde el número de estructuras alternativas posibles es más elevado) y que presentan una estructura compleja, como es el caso de Apom-a de ratón (ver figura 4). En función de estos resultados podemos admitir que, en las UTRs 5´ de cierta longitud de las lipocalinas (>100 nt), las estructuras MFE pueden elegirse como estructuras representativas de sus formas nativas.

En UTRs 5´de menor longitud (<100 nt), a pesar de que el número de estructuras posibles es como consecuencia de esto más reducido, se encuentran mayores diferencias entre su MFE y las estructuras subóptimas. En la figura 5 puede verse un claro ejemplo de esto, para el caso de la UTR 5´ Lcn12-b-humana, con una longitud de 72 nucleótidos. Existiendo en este caso solo tres estructuras alternativas en el intervalo de 5 Kcal/mol, apenas comparten elementos comunes. Otro ejemplo puede verse en la figura 6, que corresponde a la UTR 5´ Ptgds-b-ratón con 81 nucleótidos.

En este caso hay solo 5 estructuras subóptimas en el intervalo de energía de 5 Kcal/mol y sin embargo las formas subóptimas muestran diferencias importantes entre ellas.

Este es un resultado esperable debido a que las UTRs 5´cortas no parecen cumplir una función reguladora importante de la traducción, por lo que la selección natural no ha actuado en el sentido de mantener estructuras tan bien definidas como en las UTR 5´ de mayor longitud, que sí suelen desempeñar un importante papel regulador.

3.1.2.2.- Región UTR 3´

En el caso de las UTRs 3´ no se observan de forma clara las diferencias que se encontraron entre las secuencias de distintas longitudes de las UTRs 5´. Al comparar las estructuras de las secuencias MFE de las UTRs 3´ con las estructuras subóptimas y estas mismas entre sí, se observan mayores diferencias que las observadas para las UTRs 5´.

Como se observa en las figuras 7 y 8 hay diferencias apreciables entre las diferentes estructuras alternativas globales de las UTRs 3´(en el intervalo de 5 Kcal/mol). Respecto a las estructuras locales, si bien se conservan algunas de ellas entre las diferentes formas alternativas, lo hacen en menor medida que lo hacen en las UTRs 5´.

Estos resultados están en consonancia con el mayor repertorio de estructuras alternativas encontrado en las UTRs 3´ respecto al esperado, que ya es un indicio de una estructura menos definida. Estos resultados nos sugieren que en estas regiones no es tan importante la estructura secundaria global, como lo es en la región UTR 5´, sino que posiblemente son algunos elementos estructurales locales los que desempeñan un papel más importante.

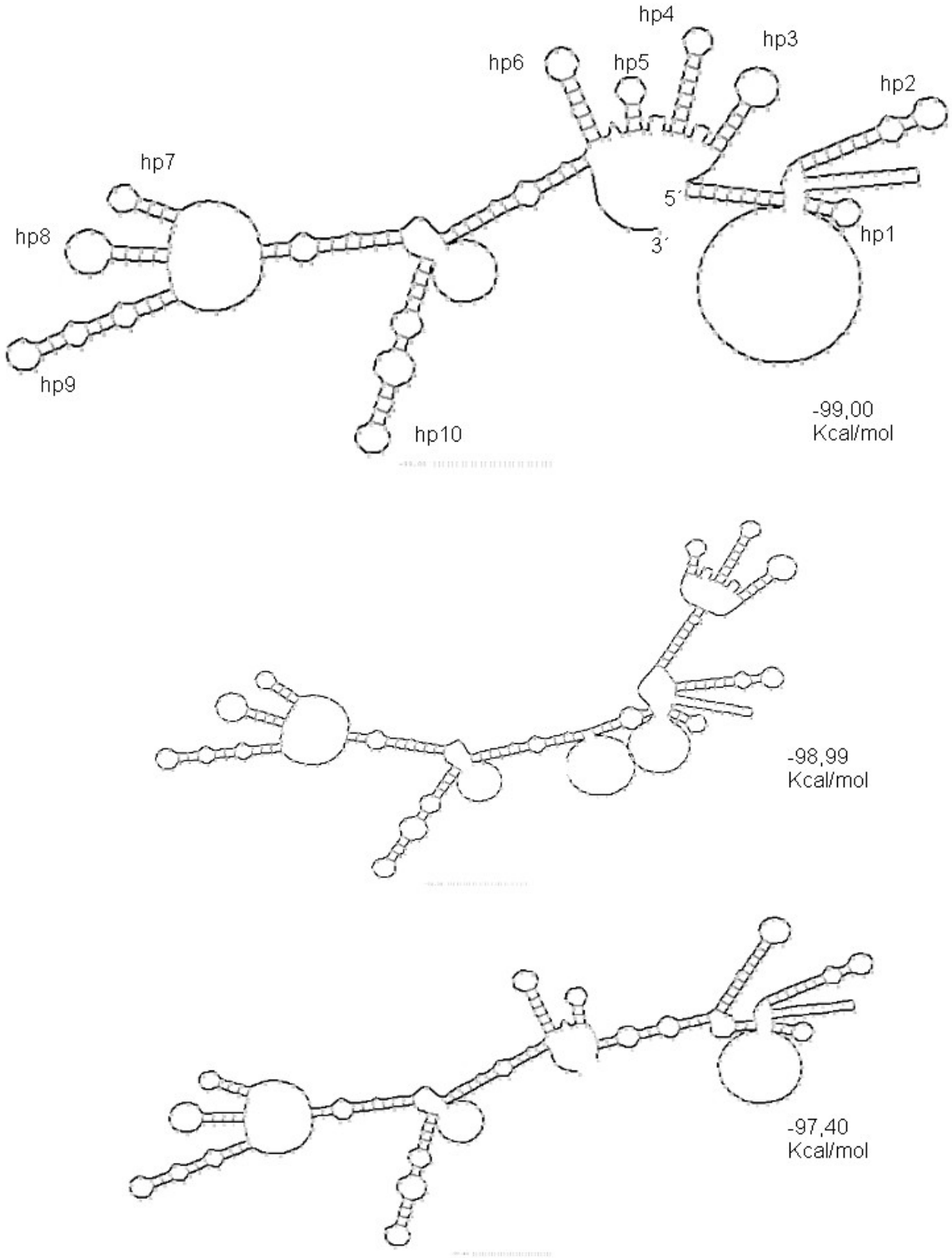


Figura 2. Estructura con MFE (arriba) y estructuras subóptimas (en un intervalo de 5 Kcal/mol) de la UTR 5´de Apo-D-a-humana. Obtenidas mediante RNAshape.

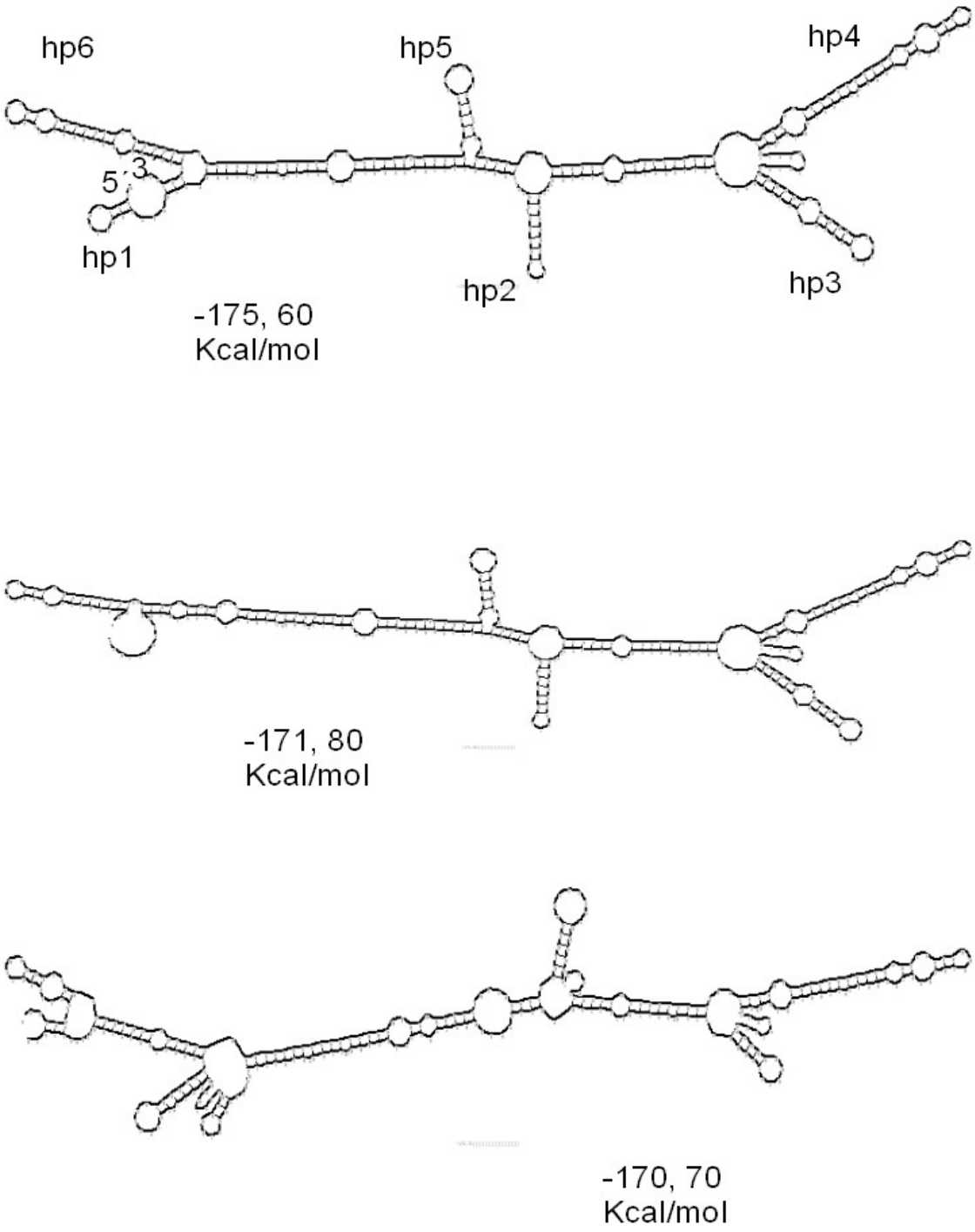


Figura 3. Estructura con MFE (arriba) y estructuras subóptimas (en un intervalo de 5 Kcal/mol) de la UTR 5´de Rbp4-b-humana. Obtenidas mediante RNAshape.

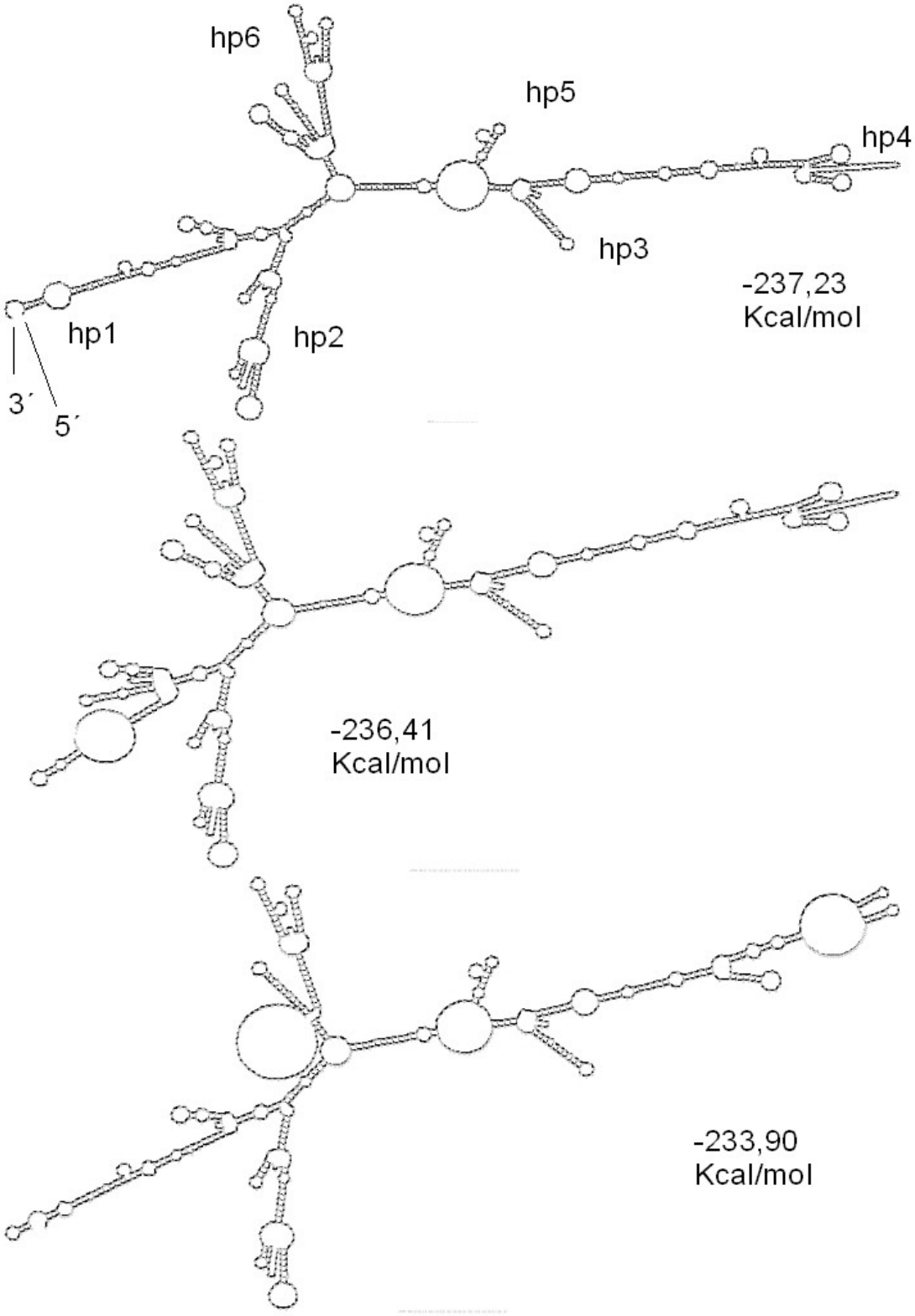


Figura 4. Estructura con MFE (arriba) y estructuras subóptimas (en un intervalo de 5 Kcal/mol) de la UTR 5´de Apom-a-ratón. Obtenidas mediante RNAshape.

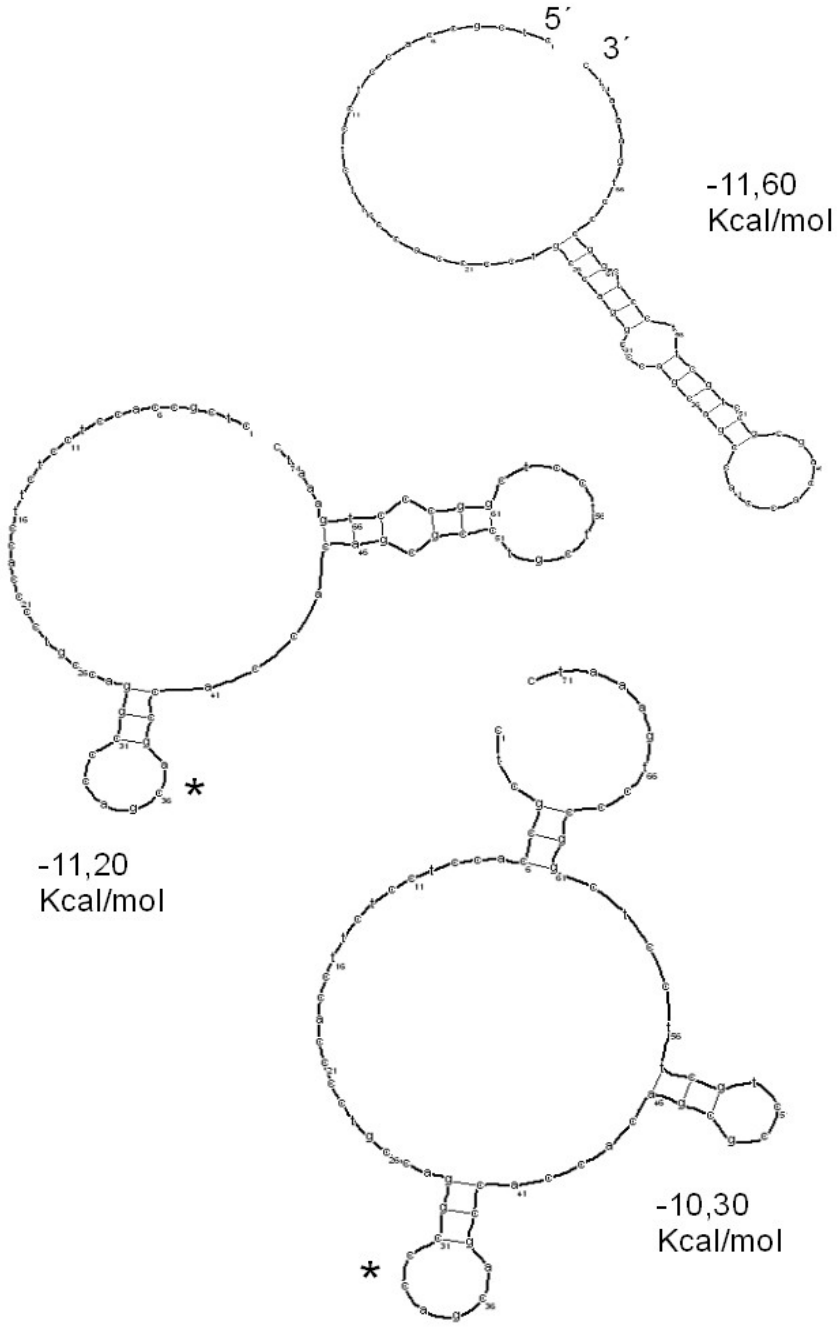


Figura 5. Estructura con MFE (arriba) y estructuras subóptimas (en un intervalo de 5 Kcal/mol) de la UTR 5´de Lcn12-b-humano. Obtenidas mediante RNAshape. Con un asterisco aparace marcada la única subestructura común entre dos de las formas de plegamiento alternativas.

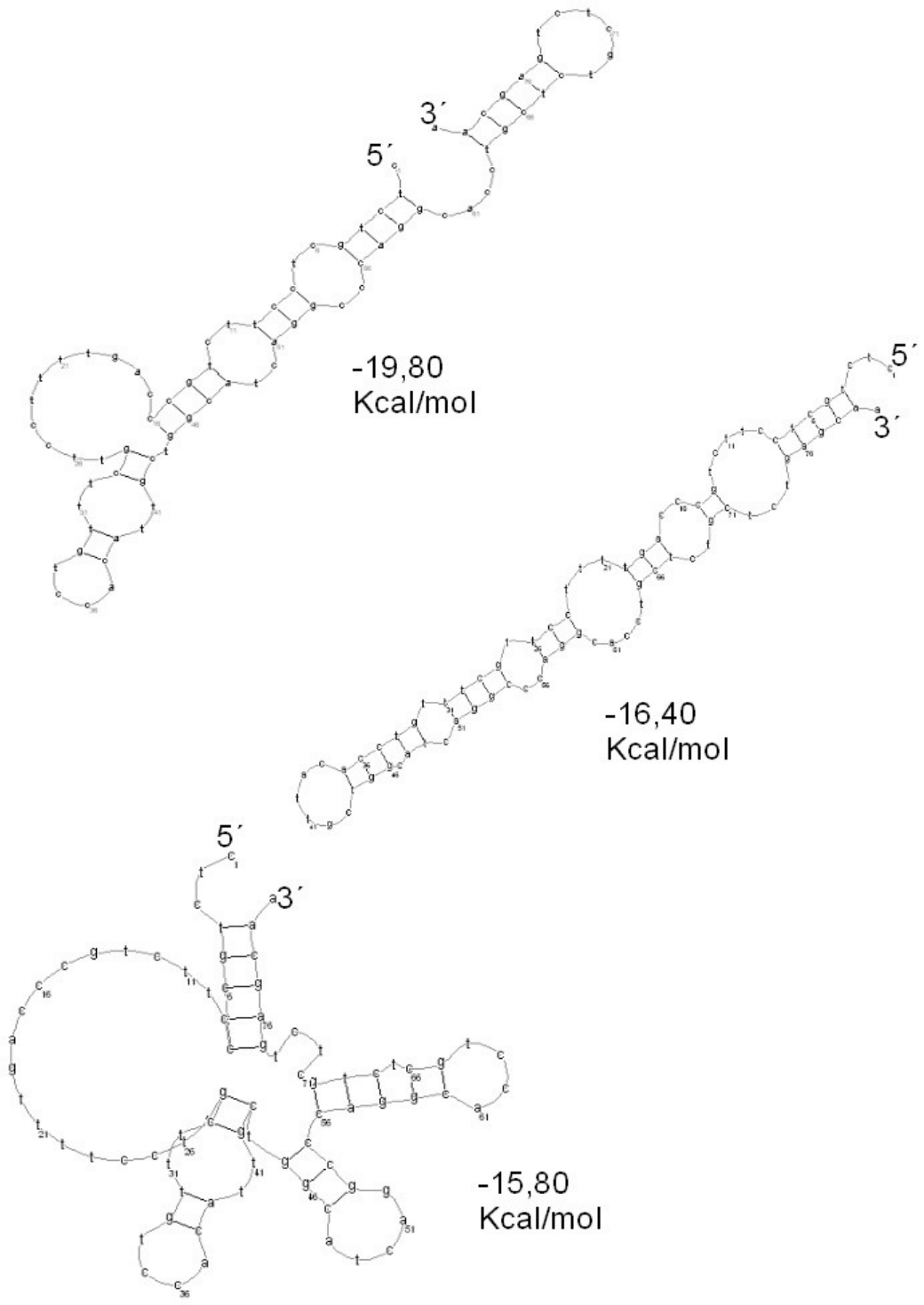


Figura 6 . Estructura con MFE (arriba) y estructuras subóptimas (en un intervalo de 5 Kcal/mol) de la UTR 5´de Ptgds-b-ratón. Obtenidas mediante RNAshape.

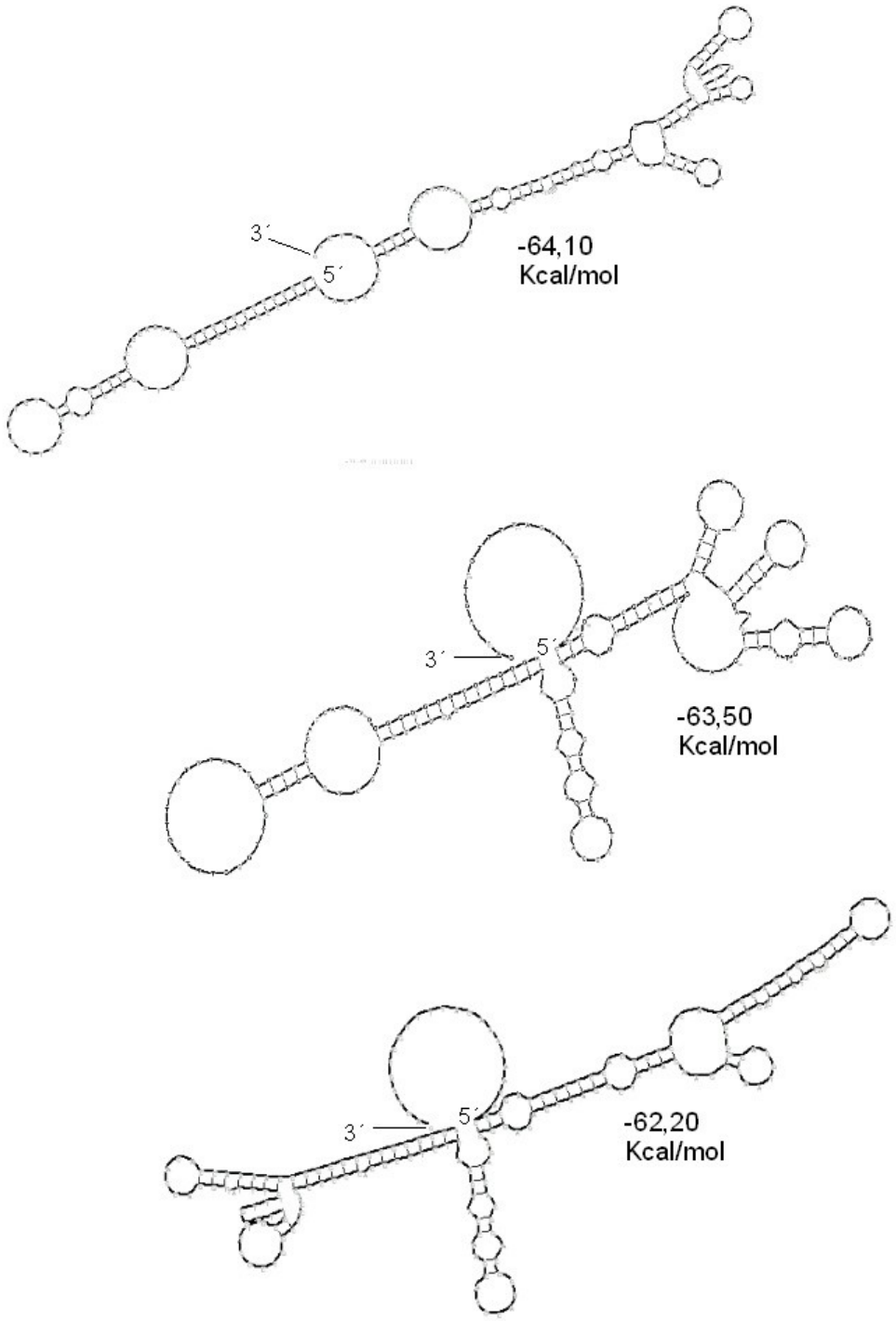


Figura 7 . Estructura con MFE (arriba) y estructuras subóptimas (en un intervalo de 5 Kcal/mol) de la UTR 3´de Apo-D-a-ratón. Obtenidas mediante RNAshape.

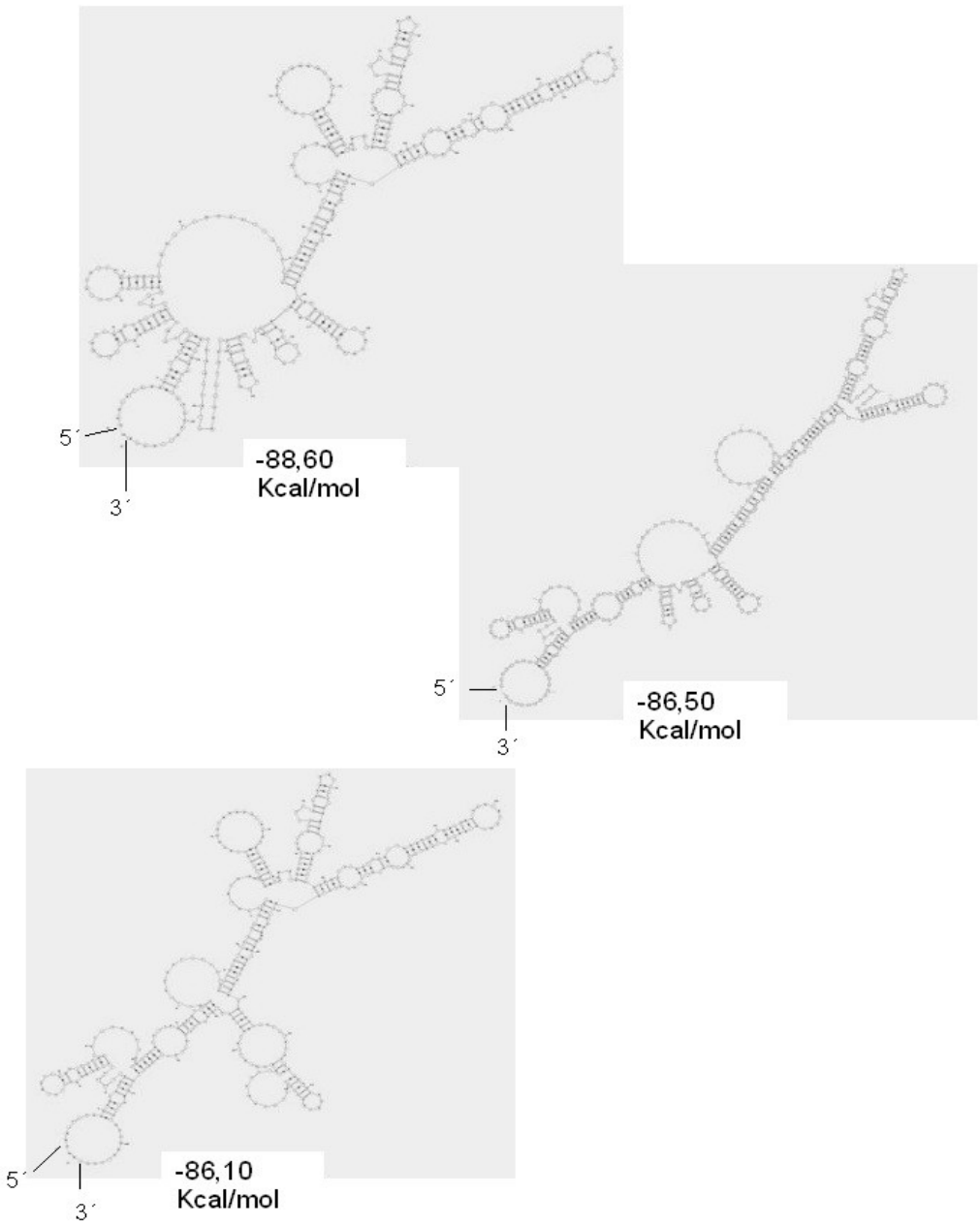


Figura 8 . Estructura con MFE (arriba) y estructuras subóptimas (en un intervalo de 5 Kcal/mol) de la UTR 3´de Rbp4--b-humana. Obtenidas mediante RNAshape.

3.2.- Motivos estructurales locales conservados en las UTRs ortólogas de lipocalinas.

El hecho de que en las UTRs 5´de lipocalinas de cierta longitud exista un repertorio relativamente reducido de estructuras alternativas y que las diferencias entre las formas subóptimas y las estructuras más estables (MFEs) de estas secuencias sean pequeñas, nos lleva a sugerir que dichas UTRs deben contener elementos estructurales importantes para la regulación de la expresión génica de esta proteína. De ser esto así, dichos elementos estructurales deben haberse conservado evolutivamente en las UTRs de lipocalinas ortólogas. Aunque en el caso de las UTRs 3´no se den estas mismas circunstancias, si se ha observado entre las formas alternativas de ellas, la presencia de ciertos elementos estructurales locales y es de esperar que haya habido también cierta conservación de los mismos.

Dado que previamente se habían identificado UTRs 5´ y 3´ortólogas en diversas especies de mamíferos (ver capítulo III de conservación de UTRs), se utilizaron éstas para llevar a cabo el estudio que permitiera identificar en cada conjunto de secuencias ortólogas posibles motivos estructurales locales conservados.

De entre las diferentes estrategias que pueden usarse para identificar nuevos candidatos a motivos reguladores en ARNs se recurrió, como método principal, a un método *"puramente estructural"* , ya que estos enfoques, que se basan exclusivamente en la estructura secundaria de los motivos, han mostrado dar buenos resultados [4, 5]. Para los casos que han dado resultado positivo con este método, se ha aplicado de forma complementaria un método basado en la *conservación de las secuencias* ortólogas mencionadas.

3.2.1.- Método "puramente estructural".

Se usó una herramienta que utiliza un algoritmo con enfoque estructural, combinado con cálculos probabilísticos (ver métodos), llamada "Predict a Motif" [6]. Se analizaron con dicha herramienta los diferentes grupos de UTRs 5´ ortólogas y tras filtrar los resultados mediante ciertos criterios (ver métodos), finalmente se obtuvieron algunos candidatos a motivos estructurales en las UTRs 5´de Apo-D y Apo-M.

3.2.1.1.- Resultados para la región UTR 5´

Motivos de Apo-D

En la figura 9 se muestran las estructuras 2D de consenso predichas por *"Predict a Motif"* , que cumplen los criterios previamente mencionados, en las UTRs 5´ ortólogas de Apo-D. Como se observa en dicha figura, hay buena conservación de la estructura en los diferentes motivos. Respecto a la secuencia, excepto para el motivo 2 con una secuencia consenso más definida, en los otros dos hay más flexibilidad, manteniéndose algunas posiciones más variables. Esta conservación de estructura con escasa o moderada conservación en la secuencia es un indicio de señal biológica ya que pone de manifiesto sustituciones compensatorias para el mantenimiento de una cierta estructura funcional [5].

En la figura 10 se muestra la ubicación de dichos motivos dentro de la estructura global de las UTRs 5´ortólogas de diversas especies (humano, toro y cerdo). Observamos que si bien las distintas UTRs 5` se pliegan globalmente de diferente forma, localmente si comparten todos o varios de los motivos mencionados. Puede observarse, que además de los motivos predichos por *"Predict a Motif"*, es posible identificar un cuarto motivo (en amarillo en la figura 10) que es fácilmente identificable en las estructuras de las UTRs 5´de algunas de las especies de mamíferos y que contiene en todos los casos la secuencia "UAUAAAAU" en el bucle (*loop*).

Motivos de Apo-M

En este caso sólo se detectó un motivo 2D conservado entre las UTRs 5´ortólogas de las especies humano, perro y nutria. Dicho motivo se muestra en la figura 11. Observamos un alto consenso en estructura y secuencia. Este motivo, al igual que los de Apo-D, también puede identificarse en las estructuras globales de MFE de las UTRs 5´ortólogas de Apo-M, ver figura 12. Puede observarse que, aunque las estructuras globales de las UTRs 5´ muestran diferencias en las diferentes especies, todas ellas muestran localmente el motivo identificado.

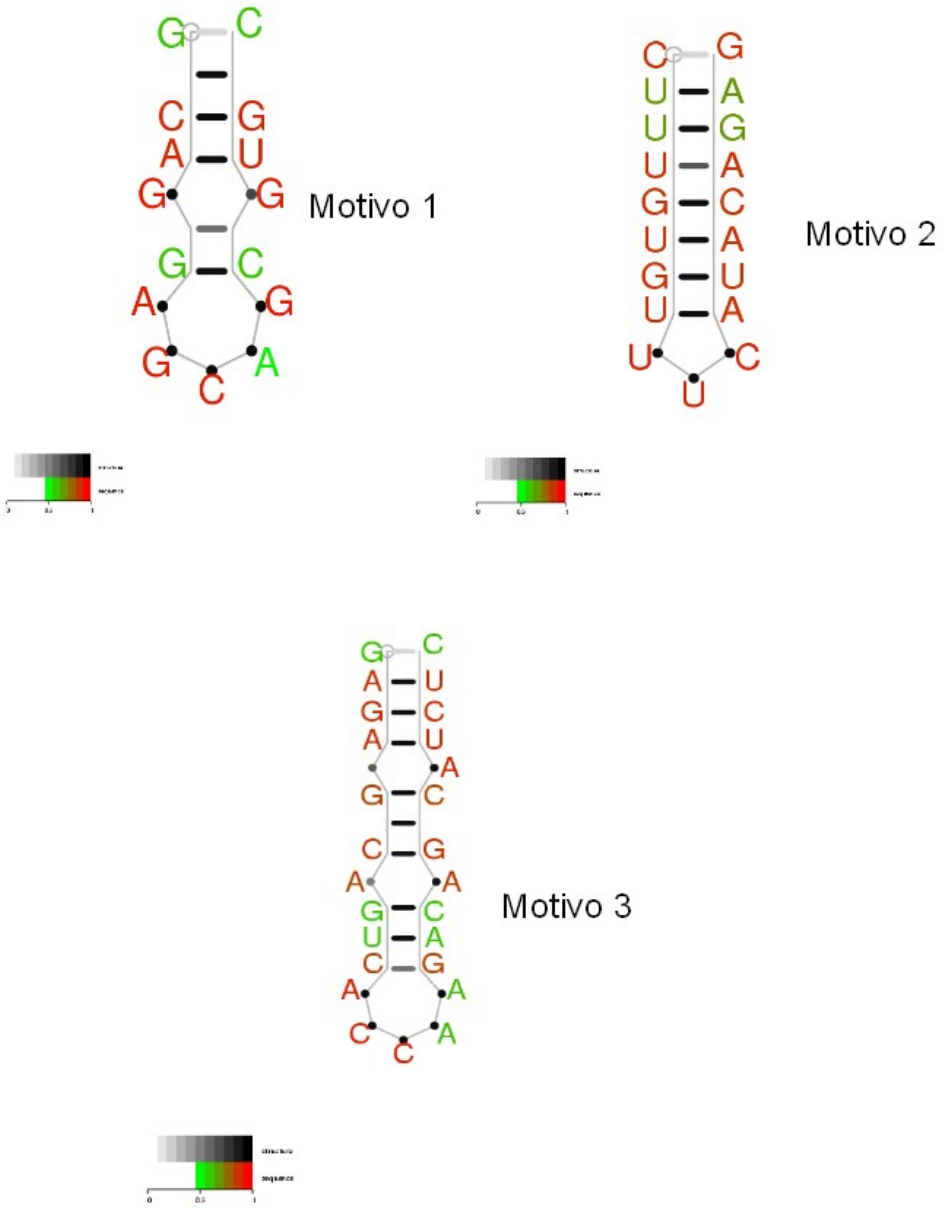


Figura 9. *Estructuras de consenso de los motivos conservados en las UTRs 5´ortólogas de Apo-D de mamíferos. Según predicciones de "Predict a Motif". El gráfico junto a cada motivo es una escala de la frecuencia con que aparece en el consenso cada elemento de la estructura (escala de grises, negro presente en todas las secuencia ortólogas), así como de cada una de las bases (verde a rojo, rojo base presente en todas las secuencias ortólogas). El extremo 5´es el extremo libre que contiene un círculo*

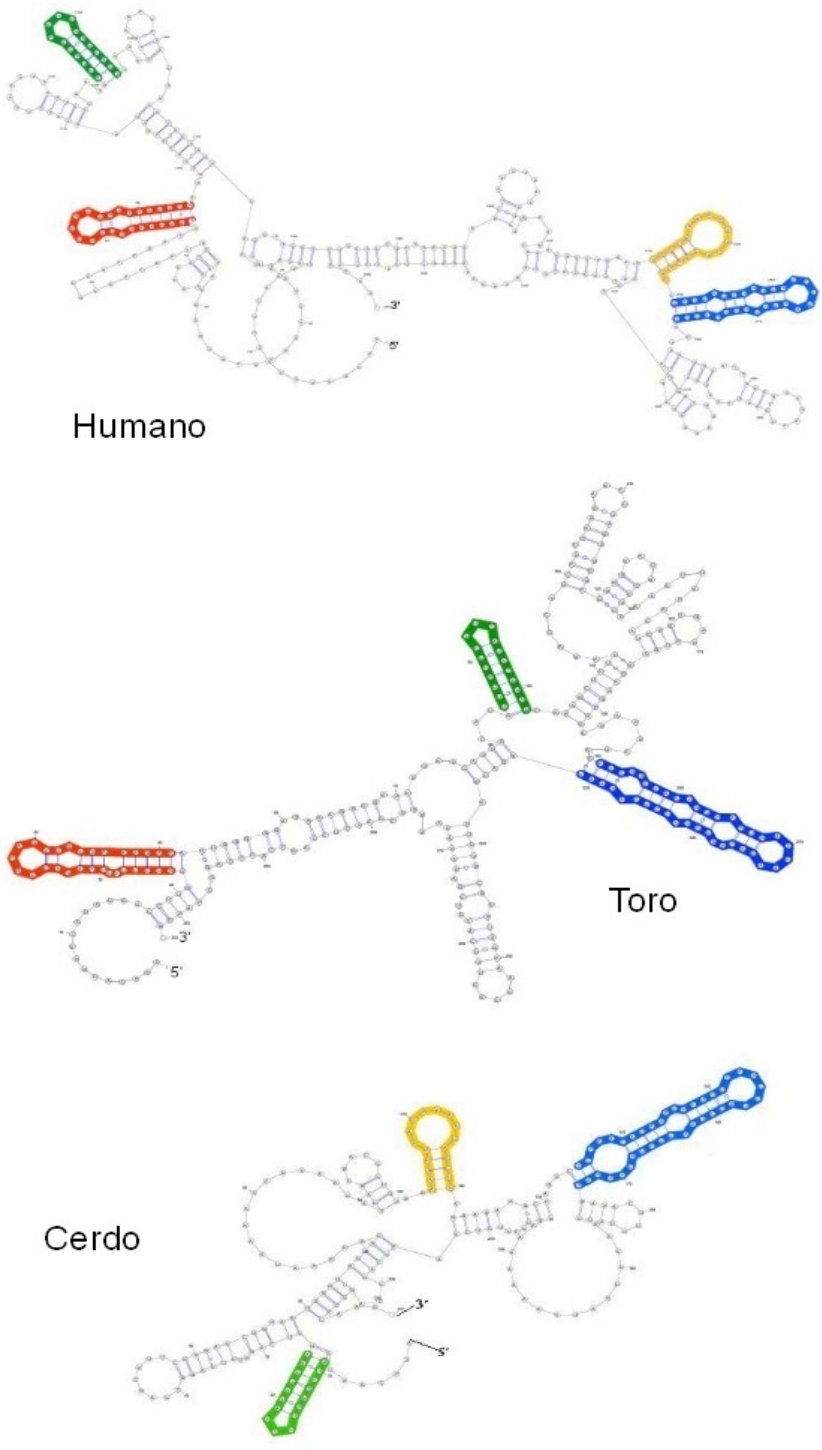


Figura 10. *Presencia de los diferentes motivos conservados en las UTR s 5´ ortólogas de Apo-D (variante humana "a"). Motivo 1 en rojo, motivo 2 en verde, motivo 3 en azul. En amarillo se muestra un motivo que no es predicho por "Predict a Motif", pero que aparece en la MFE de varias de las especies de mamíferos.*

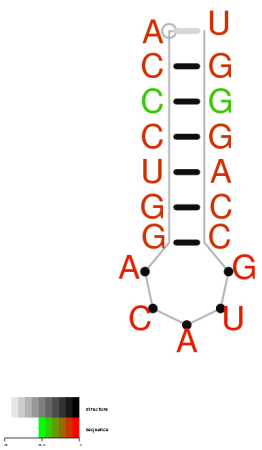


Figura 11. *Estructura de consenso del motivo conservado en las UTRs 5 ´ortólogas de Apo-M de mamíferos. Según predicciones de "Predict a Motif". El gráfico junto al motivo es una escala de la frecuencia con que aparece en el consenso cada elemento de la estructura (escala de grises, negro presente en todas las secuencia ortólogas), así como de cada una de las bases (verde a rojo, rojo base presente en todas las secuencias ortólogas). El extremo 5´es el extremo libre que contiene un círculo*

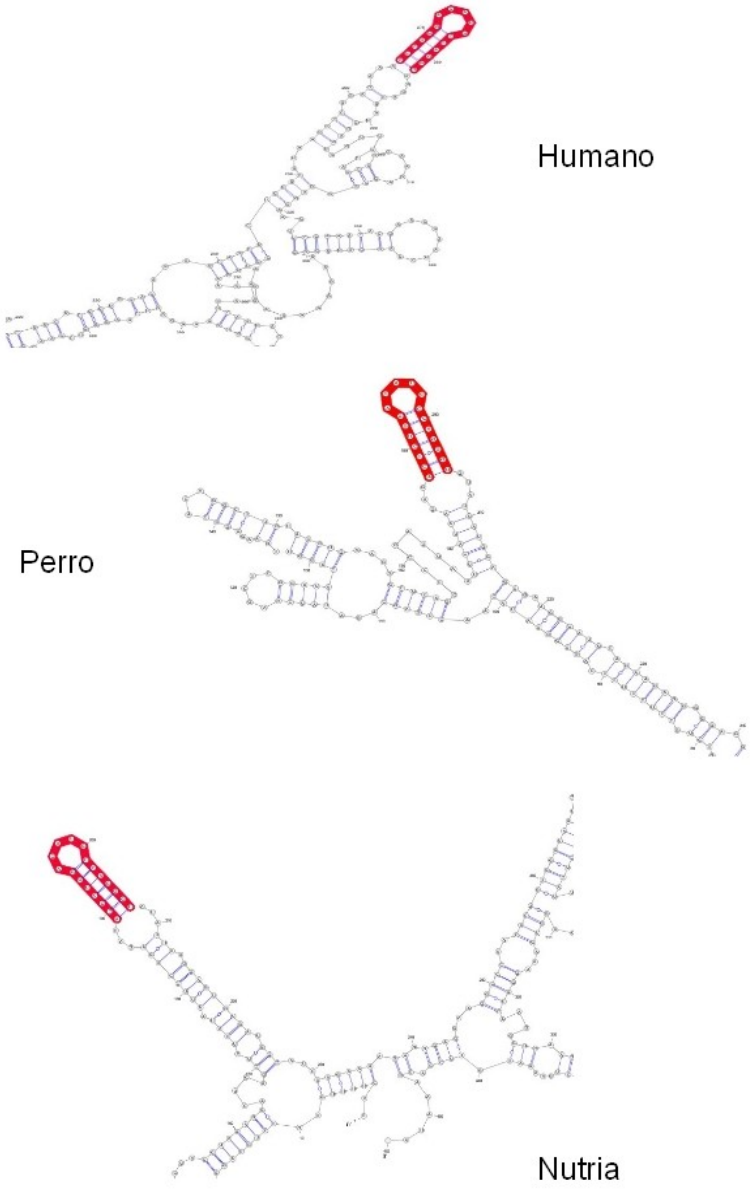


Figura 12. *Presencia del motivo conservado en las UTRs 5´ ortólogas de Apo-M de mamíferos. Motivo coloreado en rojo. Para una mejor representación gráfica sólo se muestra la región de la estructura de la UTR 5´, que contiene al motivo conservado.*

3.2.1.2.- Resultados en la región UTR 3´

Se aplicó el mismo procedimiento que en las UTRs 5´ a las secuencias UTRs 3´ortólogas de mamíferos, que previamente se habían encontrado. En este caso, solo en la UTR 3´de Apo-D fué predicho un motivo por Predict a Motif, que alcanzase los requisitos mínimos establecidos (ver métodos) para que este pueda considerarse candidato a tener un papel funcional. Pero este motivo no se observa en la estructura MFE ni en las estructuras subóptimas de esta UTR 3´, por lo que finalmente fue descartado.

3.2.2.- Método basado en "alineamiento de las secuencias".

Se utilizó este método como complemento al método *"puramente estructural"* y se aplicó solamente a los casos de la UTR 5´de Apo-D y de Apo-M, ya que son los que con dicho método han ofrecido resultados positivos. Para este análisis se utilizó el algoritmo RNAalifold [7]. Dicho algoritmo toma como partida el alineamiento múltiple de las secuencias relacionadas (ortólogas en este caso) y calcula una estructura de mínima energía de consenso para ellas.

El resultado de este análisis para Apo-D se muestra en la figura 13. En esta estructura de consenso se pueden identificar claramente los motivos estructurales 1, 2 y 3, previamente predichos por Predict a motif, estos aparecen señalados en dicha figura. Para Apo-M pudo identificarse igualmente el único motivo que había sido predicho previamente (no se muestra imagen).

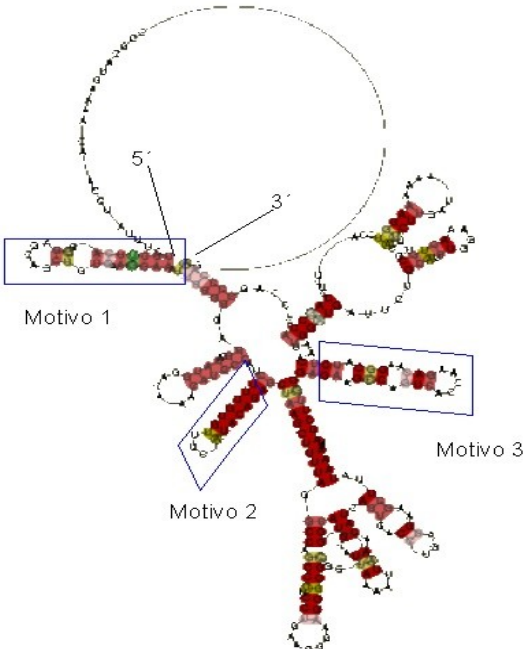


Figura 13. *Estructura 2d de consenso, de mínima energía, obtenida mediante RNAalifold a partir del alineamiento múltiple de las secuencias ortólogas de la UTR 5´de Apo-D de mamíferos.*

3.3.- Pruebas de contraste sobre la funcionalidad biológica de los motivos 2D identificados en la UTR 5´de Apo-D.

Se han realizado varias pruebas adicionales para obtener un respaldo a las predicciones sobre los motivos estructurales previamente identificados. Por motivos de mayor interés biológico, dado el mayor número de motivos encontrados, y de limitar la extensión de la tesis, estas pruebas se han aplicado solamente al caso de la UTR 5´de Apo-D.

En primer lugar se ha realizado un *"proceso inverso"*, en el que se han construido patrones de búsqueda de los motivos 2D previamente identificados. Dichos patrones se han aplicado posteriormente a una muestra de secuencias de UTRs 5´de mamíferos, para comprobar en qué medida se encuentran en estas motivos semejantes a los aquí tratados. En segundo lugar se ha realizado un análisis sobre la "robustez estructural" (o robustez genética) de los mismos motivos 2D. Para ello se procedió a comprobar si el comportamiento de las secuencias nativas de dichos motivos, frente a las mutaciones, es diferente del que muestran secuencias aleatorias, pero que presentan la misma estructura secundaria que las nativas, al ser mutadas igualmente.

3.3.1.- Búsqueda de los motivos 2D mediante patrones aplicados a bases de datos.

La estrategia de obtener patrones de búsqueda de hipotéticos motivos estructurales, encontrados en secuencias de diversos ARNs estructurales, es una estrategia comúnmente utilizada. Mediante este procedimiento es posible contrastar si dichos motivos pueden formarse, al menos teóricamente, en un conjunto dado de secuencias, más o menos semejantes a la secuencia o secuencias donde dicho motivo se ha identificado y así obtener un contraste, en términos probabilísticos de su realidad biológica.

En nuestro caso, para llevar a cabo este análisis, se procedió al diseño gráfico de los motivos 2D y la posterior conversión de estos en patrones de búsqueda, que posteriormente se aplicaron a una muestra de 3000 secuencias de UTRs 5´de mamíferos (obtenida de forma aleatoria de UTRdbase), y así mismo sobre una muestra de secuencias aleatorias (obtenida a partir de la misma muestra de UTRs 5´ citada), que sirve de control. Para obtener los patrones de búsqueda se ha utilizado el programa Locomotif [10]. El mismo programa permite correr dichos patrones sobre el conjunto de secuencias elegido (ver detalles en métodos).

En la figuras 14 a 17 se observan la representaciones gráficas de los motivos y las especificaciones del patrón de búsqueda para los mismos, que se diseñaron en Locomotif. Como se observa se ha especificado la secuencia consenso de los bucles (*hairpin loop*), mientras que en las dobles cadenas (*stem*) y los bucles internos (*internal loop*), solo se ha especificado el tamaño, con un mínimo nivel de flexibilidad, según la estructura consenso previamente obtenida por Predict a Motif para cada motivo.

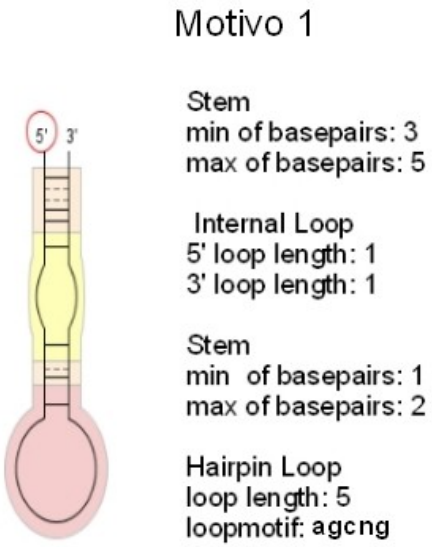


Figura 14

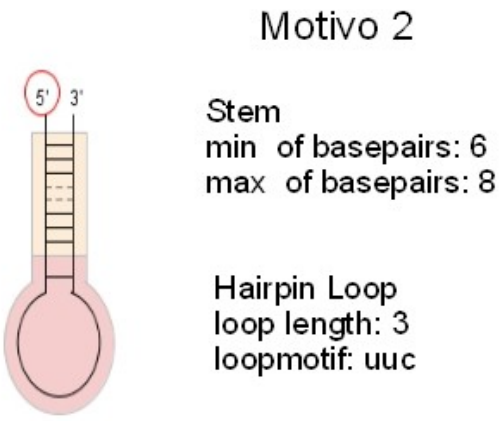
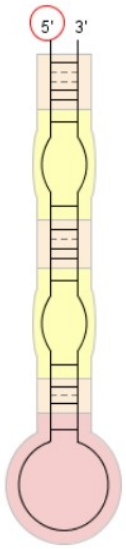


Figura 15

Motivo 3



Stem
min of basepairs: 3
max of basepairs: 5

Internal Loop
5' loop length: 1
3' loop length: 1

Stem
min of basepairs: 3
max of basepairs: 4

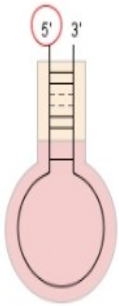
Internal Loop
5' loop length: 1
3' loop length: 1

Stem
min of basepairs: 2
max of basepairs: 3

Hairpin Loop
loop length: 5
Loop motif: accnn

Figura 16

Motivo 4



Stem
min of base pairs: 4
max of base pairs: 6

Hairpin Loop
loop length: 8
loop motif: uauaaaau

Figura 17.

Los resultados de aplicar estos patrones de búsqueda a las secuencias de UTRs 5´de mamíferos, y a la muestra aleatoria ya mencionada, se muestran en la tabla 1.

Motivo	Positivos en secuencias UTRs 5`mamíferos	Positivos en secuencias aleatorias
1	**PHTF**: Homoedominio de factor de transcripción. *E. caballus* (XM_001499447) **SET**: Dominio de proteína multidominio implicada en regulación de transcripción génica y estructura de cromatina. S. scrofa (XM_001927800)	–
2	**"Apo-D"**: Apolipoproteina D. *S. scrofa* (XM_001926063) **TIMM21**: Translocasa de membrana mitocondrial interna. *C. lupus* (XM_843471)	2
3	**TCEA2**: Factor de elongación A2. *H. sapiens* (NM_003195)	–
4	**"Apo-D"**: Apolipoproteina D. *S. scrofa* (XM_001926063)	–

Tabla 1. *Resultados obtenidos tras aplicar el patrón de búsqueda de los motivos 1 a 4 de la UTR 5´de Apo-D a una muestra de UTRs 5´de mamíferos y a estas mismas secuencias aleatorizadas.*

Observamos que los resultados positivos son claramente más frecuentes entre las secuencias UTRs 5´nativas que en las secuencias aleatorias. Sólo en el *"motivo 2"* se encuentran el mismo número de resultados positivos entre secuencias UTRs 5´ nativas y aleatorias, en el resto solo se encuentra en nativas. Comprobamos además que dos de los resultados positivos encontrados en las secuencias nativas (motivos 2 y 4) son en la UTR 5´de Apo-D de cerdo (*S. scrofa*). Esto demuestra que el algoritmo Locomotif es capaz de encontrar los motivos estructurales diseñados si se aplica a las secuencias adecuadas. El resto de resultados positivos obtenidos coinciden con genes que por su función (diversos factores de transcripción, elongación, etc) necesitan de una fuerte regulación y hemos de esperar que en sus UTRs 5´existan igualmente elementos estructurales funcionales, que parecen mostrar semejanza con los motivos identificados en Apo-D.

Para contrastar la hipótesis de que estos resultados positivos obtenidos se correspondan con una realidad biológica se llevó a cabo el siguiente test complementario. Se diseñaron los mismos motivos identificados en Apo-D (motivos 1 a 3), pero sin ninguna especificación de la secuencia y estableciendo alguna flexibilidad en el número de bases que compone los elementos de la estructura de cada uno de ellos. Se obtuvieron así unas estructuras más genéricas, pero guardando su configuración general original. Se obtuvo con Locomotif el patrón de búsqueda ADP de estos modelos y se aplicaron a una muestra de UTRs 5´ nativas y otra aleatoria, pero de igual composición de nucleótidos que la primera. Dado que es lógico pensar que a mayor longitud de secuencia mayor probabilidad de que se formen ciertos apareamientos, y por lo tanto de que se origine una cierta estructura, se quiso conocer cómo puede esto influir en los resultados, así que se realizó una clasificación de las secuencias por tamaño: 80-150 nt, 200-500 nt y > 600 nt. Se aplicaron los patrones de búsqueda mencionados a muestras de 1000 secuencias de UTRs 5`de mamíferos (obtenidas de UTR database), para cada uno de los tamaños mencionados. Se realizó el mismo procedimiento con una muestra de secuencias aleatorizada, obtenida a partir de la mencionada muestra de UTRs 5´nativas.

Los resultados de la frecuencia de aparición de estos motivos estructurales más genéricos se observan en la tabla 2.

Longitud de secuencias	**Motivo generico 1**		**Motivo generico 2**		**Motivo generico 3**	
	nativas	**aleatorias**	**nativas**	**aleatorias**	**nativas**	**aleatorias**
80-150	0.57	0.54	0.07	0.13	0.01	0.01
200-500	0.93	0.95	0.30	0.31	0.07	0.04
>600	1	1	0.62	0.60	0.24	0.18

Tabla 2. *Frecuencias de formación de modelos estructurales semejantes a los motivos 1, 2 y 3 de la UTR 5´de Apo-D. Se indican: "nº de casos positivos / total de secuencias de la muestra", para cada longitud de secuencia y según nativas o aleatorizadas.*

De los resultados obtenidos sacamos diversas conclusiones. En primer lugar se hace evidente el efecto que una mayor longitud de la secuencias tiene sobre la probabilidad de aparición de los motivos. En segundo lugar no parece haber diferencias en las probabilidades de aparición de esta clase de motivos, diseñados genéricamente, entre secuencias nativas (UTRs 5`) y secuencias aleatorias, ambas del mismo tamaño y composición de nucleótidos. En tercer lugar la probabilidad de formación de los diferentes motivos es diferente para cada uno de ellos, a igual longitud de secuencia, siendo el "motivo 3" el que muestra claramente una más baja probabilidad de formación.

El hecho de que, al ser diseñados genéricamente estos motivos sean igual de frecuentes en secuencias UTRs 5´nativas que en aleatorias, hemos de interpretarlo como un respaldo estadístico a los resultados positivos obtenidos previamente para los motivos específicos encontrados en la UTR 5´de Apo-D. Ya que en el caso de estos últimos, solo si su especificidad es una realidad biológica, podría explicarse que se encuentren éstos con más frecuencia en secuencias nativas que en aleatorias (ver tabla1). Especialmente relevante parece el motivo 3, ya que los resultados de este test complementario lo muestran como una clase de motivo de baja probabilidad de formación, incluso en secuencias de ARN de tamaño considerable.

3.3.2.- Pruebas de robustez genética de los motivos 2D identificados

Existen evidencias de que diversos ARNs estructurales, entre ellos miARNs [18] y elementos reguladores de la replicación del virus de la hepatitis C [11], muestran signos de robustez estructural (también llamada robustez genética). Dicho concepto puede definirse como la capacidad de que la estructura de estos ARNs se aleje poco de la estructura nativa, al sufrir éstos mutaciones.

Esta resistencia a los cambios de la estructura puede medirse si comparamos a los ARNs nativos, con el comportamiento que muestran secuencias aleatorias, que son elegidas artificialmente para que presenten una estructura secundaria como la de la secuencia nativa. Estas últimas cuando sufren dichas mutaciones, se ven afectadas por un mayor cambio respecto a su estructura original, ya que no se han visto sometidas a un proceso de selección que optimice dichas secuencias [18].

Para poder realizar estas pruebas con los motivos estructurales identificados en la UTR 5´de Apo-D, se obtuvieron subsecuencias de 50 nucléotidos que contuviesen completamente a los citados motivos. Solo se pudieron obtener 2 subsecuencias inequívocas, una para el “motivo “1 y otra para los “motivos 3 y 4” de forma conjunta, dada la proximidad de los mismos en la UTR 5´. Posteriormente se obtuvieron artificialmente secuencias aleatorias, del mismo tamaño y G+C que las nativas, pero que mantuviesen la estructura secundaria de cada una de las subsecuencias nativas mencionadas. A continuación se mutaron (mediante distintos niveles de sustituciones) las secuencias nativas y las correspondientes aleatorias y se calcularon las distancias, entre las estructuras 2D de los diferentes mutantes y la estructura original correspondiente. Finalmente se trataron estos datos estadísticamente, con objeto de comparar las secuencias nativas y aleatorias (ver métodos para los detalles).

En las gráficas de la figura 18 se muestran los resultados obtenidos. Los mismos parecen indicarnos que realmente existe robustez estructural en la secuencia nativa que contiene al *"motivo 1"* y en la que contiene al *"motivo 3-4"*, especialmente si observamos los valores para mutantes de 3 y 5 sustituciones. Para estos valores de mutación observamos que la distancia media entre estructura de secuencia original y mutada es mayor para las secuencias aleatorias. El test de diferencia de medias aplicado a estos datos dio como resultado el rechazo de la hipótesis nula (los valores medios de distancias de los dos conjuntos de datos son iguales), para los niveles de mutación de 3 y 5 sustituciones (nivel de significación $p < 0.05$). Estos resultados nos llevan a proponer que las secuencias nativas y aleatorias tienen distinto comportamiento frente a las mutaciones, siendo más robustas frente a dichas mutaciones las secuencias nativas. Esto solo puede haber ocurrido por acción de la selección natural, lo que es un argumento más, a favor de que los motivos estructurales identificados en la UTR 5´de Apo-D sean realmente funcionales.

En la figura 19 se muestran algunos ejemplos de cómo las estructuras de las secuencias nativas mutadas, correspondientes al *"motivo 1"*, se alejan menos de la estructura nativa original de lo que lo hacen las secuencias aleatorias mutadas, las cuales se muestran en la figura 20. Puede observarse, en esta figura, cómo se originan, en dichas secuencias aleatorias mutadas, nuevas horquillas no existentes en la estructura original. En las figuras 21 y 22 se muestra lo mismo para el caso del *"motivo 3-4"*. Comprobamos aquí como las secuencias nativas mutadas (Fig. 21) mantienen las dos horquillas con pequeñas modificaciones, mientras que las secuencias aleatorias mutadas (Fig.22) presentan estructuras en las que bien se pierde el bucle terminal o bien aparecen elementos inexistentes en la estructura original.

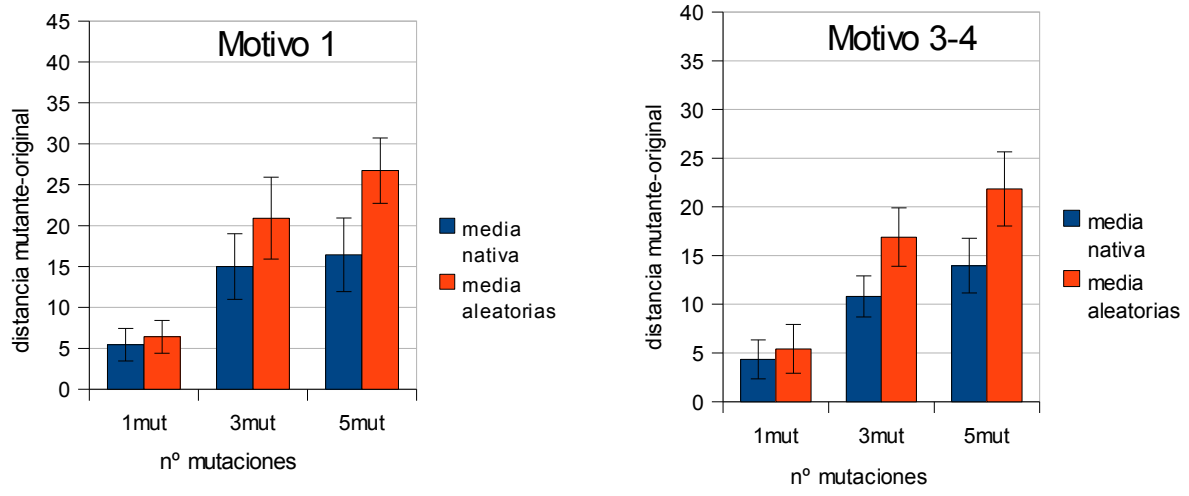


Figura 18. Representación de la distancia de la estructura 2d de los mutantes de 1, 3 y 5 sustituciones a la estructura 2d nativa original, de los dos motivos estructurales identificados en la UTR 5´de Apo-D. Las barras coloreadas representan el valor medio y los segmentos la desviación típica. Se muestran los datos de las secuencias nativas (azul) frente a los de las secuencias aleatorias(rojo).

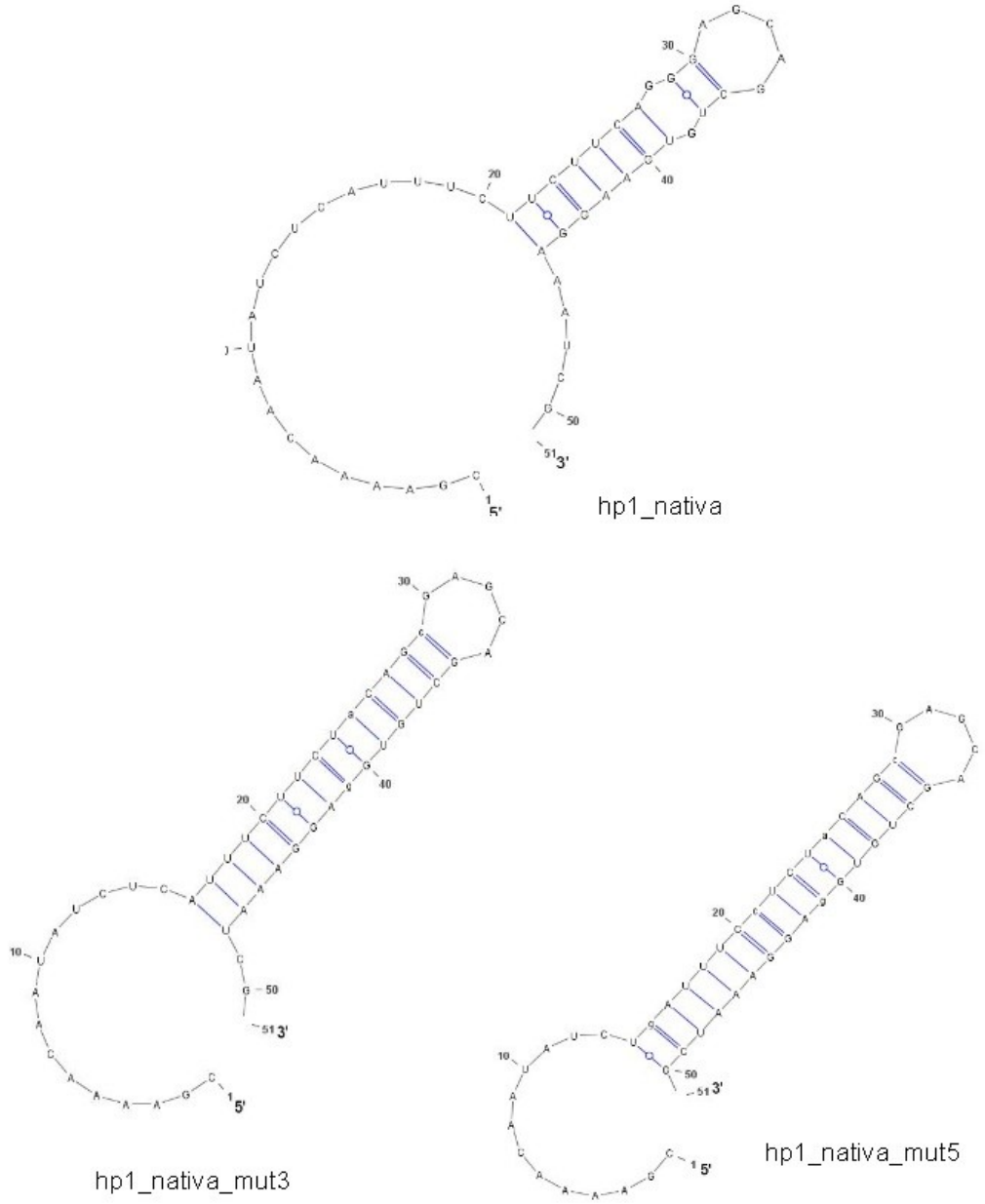


Figura 19. *Estructura original de la subsecuencia nativa que contiene al motivo 1 de Apo-D (arriba) y las estructuras de los mutantes de 3 y 5 sustituciones de esta misma secuencia (abajo).*

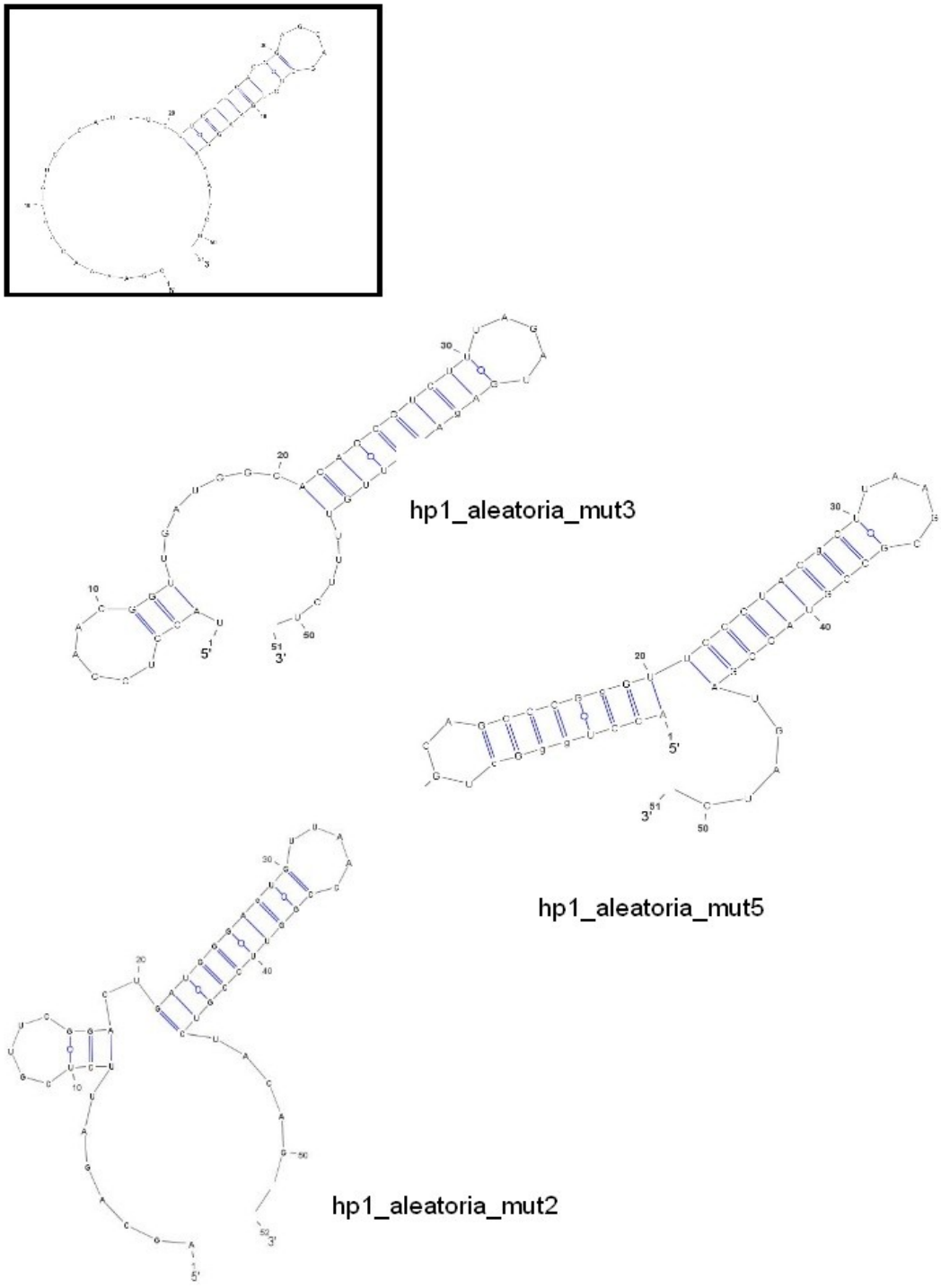
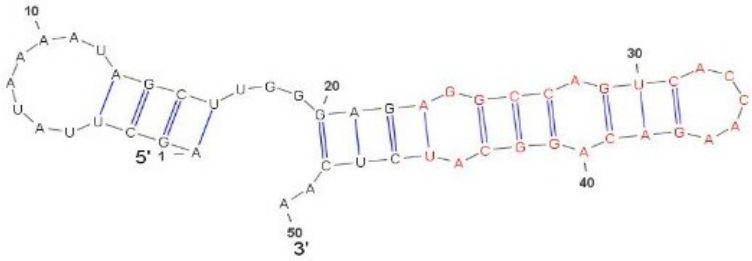
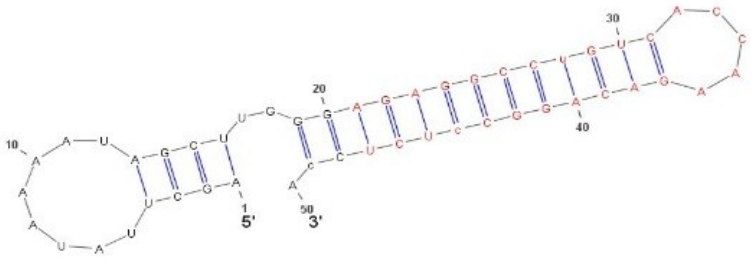


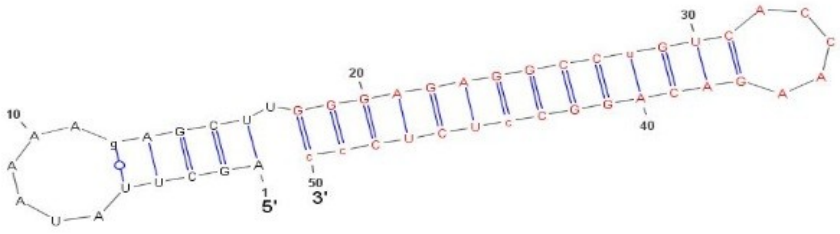
Figura 20. *Muestra de las estructuras que adquieren las secuencias aleatorias basadas en el motivo 1 de Apo-D, con distintos niveles de mutación. En recuadro estructura original.*



hp3/hp4_nativa

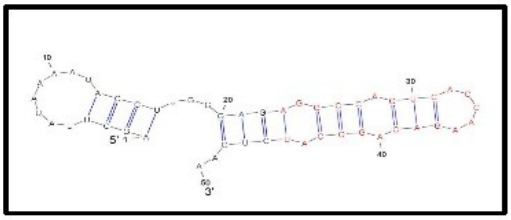


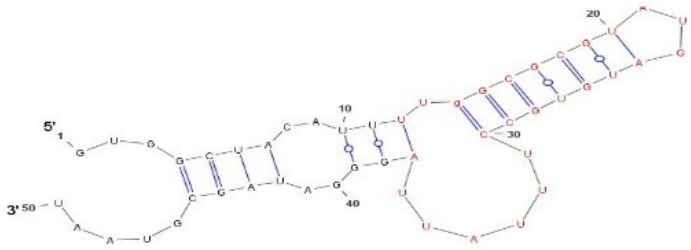
hp3/hp4_nativa_mut3



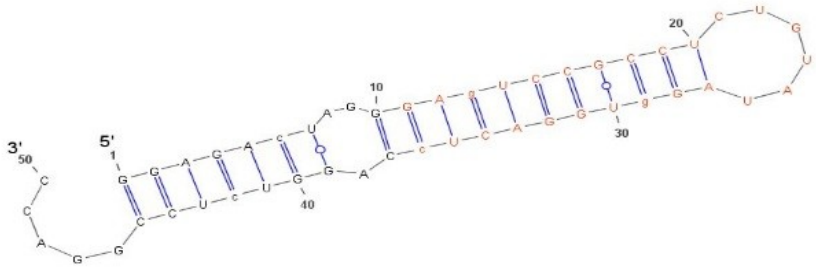
hp3/hp4_nativa_mut5

Figura 21. *Estructura original de la subsecuencia nativa que contiene a los motivos 3 y 4 de Apo-D (arriba) y las estructuras de dos mutantes de 3 y 5 sustituciones de esta misma secuencia (abajo). Motivo 3 representado con bases en negro, motivo 4 con bases en rojo.*

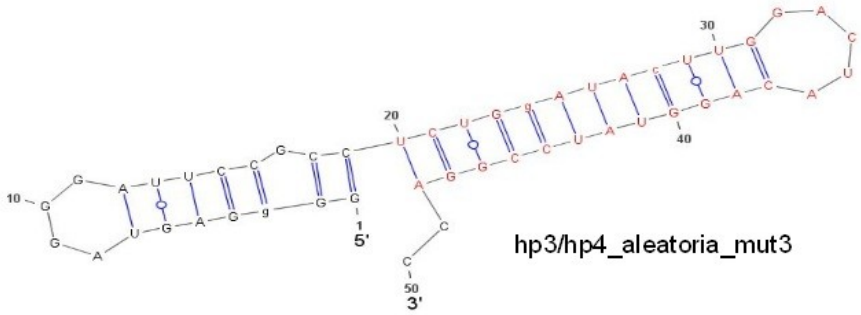




hp3/hp4_aleatoria_ mut1



hp3/hp4_aleatoria_mut5



hp3/hp4_aleatoria_mut3

Figura 22. *Muestra de las estructuras que adquieren las secuencias aleatorias basadas en las estructuras de los motivos 3 y 4 de Apo-D, con distintos niveles de mutación. Motivo 3 representado con bases en negro, motivo 4 con bases en rojo. En recuadro estructura original.*

3.4.- Interacciones entre los elementos de estructura secundaria de la UTR 5´de Apo-D. Aproximación a su estructura tridimensional.

La estructura terciaria de algunos tipos de ARN con actividades catalíticas o que desempeñan funciones de reconocimiento o regulación, descansa en cierta medida sobre estructuras como los *pseudoknots* [19-21]. Estas estructuras son el resultado de formación de pares de bases entre diferentes elementos estructurales 2D más o menos separados dentro de la estructura de un ARN dado.

El hecho de haber encontrado que varios elementos estructurales 2D se han conservado en una de las UTRs 5´ alternativas de Apo-D, entre diferentes especies de mamíferos, y que los mismos muestran signos de funcionalidad biológica, puede hacernos pensar en la posibilidad de que existan interacciones entre dichos elementos y que se originen *pseudoknots*. Dichos *pseudoknots* podrían condicionar la formación de una determinada estructura terciaria en dicha región, necesaria para ejercer sus funciones reguladoras de la expresión génica de la mencionada lipocalina.

La cuestión de predecir la existencia de *pseudoknots* es un problema computacional complejo, aunque se han diseñado algoritmos que mediante programación dinámica ofrecen una aproximación a la detección de dichos elementos. En este caso se recurrió al programa pAliKiss [13]. Este programa es el resultado híbrido de otros dos (ver métodos).

Se suministró al programa pAliKiss el alineamiento múltiple de las secuencias UTR 5´ortólogas de Apo-D de mamíferos y la salida del mismo ofrece una predicción de los posibles *pseudoknots* que pueden originarse por interacciones entre los diferentes elementos 2D, dentro de la estructura secundaria de consenso obtenida. Al estudiar esta predicción no se observa que los motivos estructurales 2D relevantes, previamente identificados (motivos 1 a 3), estén implicados en estas interacciones. Los motivos que la predicción muestra como formadores de *pseudoknots* no son motivos que muestren ser significativos en la estructura de la UTR 5´de Apo-D, además el número de bases apareadas en estos hipotéticos *pseudoknots* es escaso, por lo que su estabilidad sería baja.

A pesar de no poderse detectar la presencia de *pseudoknots* se quiso conocer la estructura terciaria que adquiere esta UTR 5´de Apo-D. Se utilizó el programa RNAComposer [14], el cual descompone la estructura secundaria en elementos más sencillos, determina su estructura terciaria y los ensambla realizando finalmente un refinamiento de la estructura terciaria global (ver métodos). Se analizó la UTR 5´de Apo-D humana (variante a) e igualmente, a modo de contraste,

se analizaron las siguientes secuencias:

- La UTR 5´de la lipocalina Rbp4 humana (variante b), la cual es de tamaño semejante a la de Apo-D, pero no muestra presencia de motivos 2D conservados.
- Secuencias aleatorizadas a partir de la UTR 5´ de Apo-D humana nativa.
- Secuencias UTRs 3´de lipocalinas, de tamaño semejante al de la UTR 5´ de Apo-D humana.

La estructura terciaria predicha para la UTR 5´de Apo-D(a) se muestra en la figura 23. Podemos observar que dicha UTR adquiere una estructura tridimensional compleja tipo globular. Al comparar esta estructura con las predichas para la UTR 5´de Rbp4 o con las secuencias aleatorias (figura 24) vemos una clara diferencia, presentando estas una estructura terciaria de menor complejidad y con una disposición más lineal. Respecto a las estructuras de las UTRs 3´ de lipocalinas observamos en la figura 25 que las predicciones ofrecen resultados similares a los de UTR 5´de Rbp4 y las secuencias aleatorias.

Figura 23. Estructura terciaria de la UTR 5´ humana(a) predicha por RNAComposer

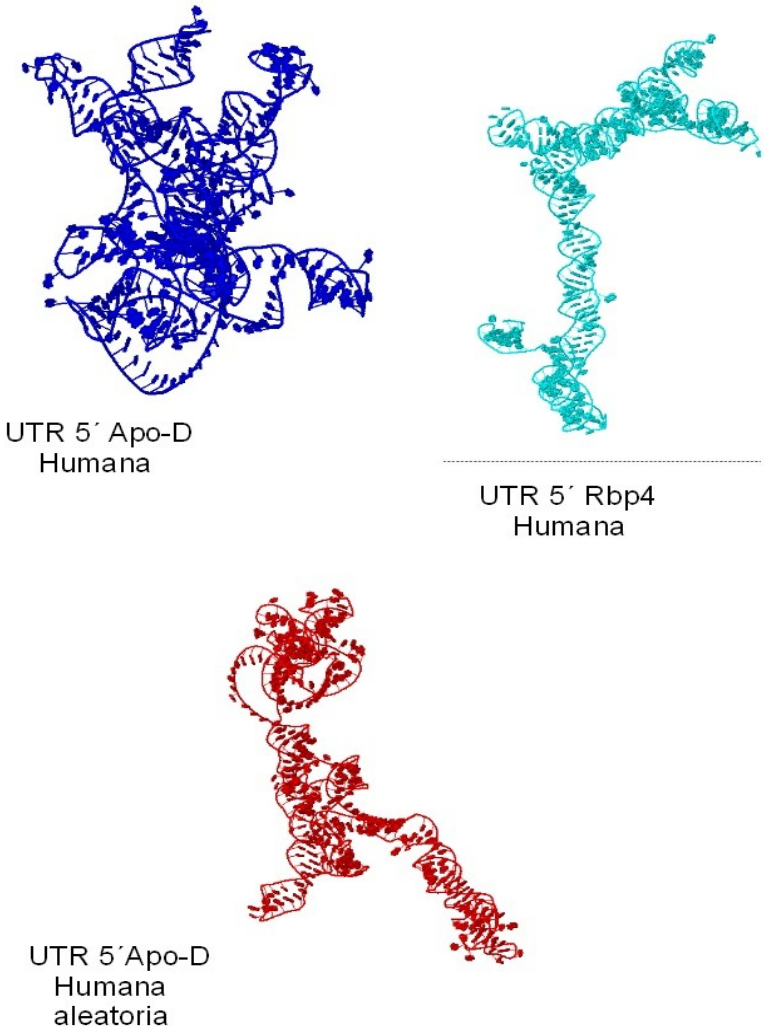


Figura 24. *Comparación de la estructura terciaria de la UTR 5´humana(a) predicha por RNAComposer frente a las estructuras predichas para la UTR 5´de rbp4 humana(b) y una secuencia aleatoria obtenida a partir de la UTR 5´de Apo-D humana(a).*

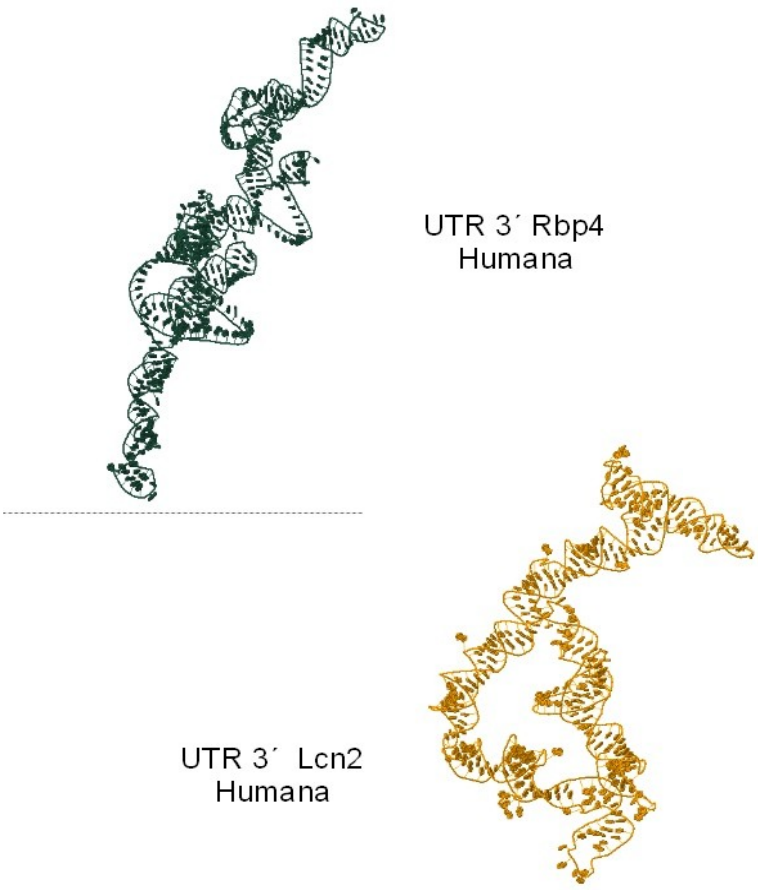


Figura 25. *Estructura terciaria de las UTRs 3´ humanas de las lipocalinas Rbp4 y Lcn2 predicha por RNAComposer*

4. - Discusión

El hecho de haber encontrado que las regiones UTRs 5´de las lipocalinas, de cierta longitud, muestren un repertorio de formas de plegamiento alternativas menor que las esperadas para una muestra de ARNs estructurales, no es una prueba definitiva de la importancia que puede tener la estructura secundaria en la función reguladora de estas regiones. Pero dado que las regiones UTRs 3 ´muestran, para tamaños comparables a las UTRs 5´, un repertorio mayor que la citada muestra de ARNs y dado que al comparar entre sí las formas alternativas del repertorio de plegamiento de las diferentes UTRs, estas son más semejantes entre sí para el caso de las UTRs 5´, hemos de admitir

que estas evidencias apuntan en la dirección de que la estructura secundaria está muy bien definida en las regiones UTRs 5´de las lipocalinas y esto es un fuerte indicio de la mayor importancia de la estructura secundaria global en dicha región.

Por otra parte ha quedado de manifiesto que en las UTRs 5´de ciertas lipocalinas (Apo-D y Apo-M) existen motivos locales de estructura secundaria que muestran signos de relevancia biológica, por lo que demuestran los resultados de las diversas pruebas de contraste a las que estos han sido sometidos. Dichos motivos son desconocidos pero deben ejercer una función semejante a las de otros bien caracterizados. Probablemente estos motivos sean sitios de unión de proteínas afines al ARN que ejerzan alguna función de regulación en la traducción del ARNm. No podemos descartar que los citados motivos sean directamente responsables, sin mediación de proteínas, de ejercer algún tipo de regulación de la expresión génica, a modo de como actúa el conocido motivo IRES [22].

Aunque el estudio de la estructura terciaria de las UTRs, por su extensión y complejidad, no era uno de los objetivos de esta tesis, el hecho de haber encontrado en las UTRs 5´ ortólogas de Apo-D(a) de mamíferos varios elementos relevantes de estructura secundaria, ha llevado a considerar al menos una aproximación a dicho nivel de estructura en este caso concreto. Los resultados obtenidos muestran que dicha UTR 5´muestra una complejidad tridimensional claramente superior a las UTRs 5´de otras lipocalinas de semejante tamaño, así como a secuencias aleatorias de semejante tamaño y composición de nucleótidos. Estos indicios no hacen sino sugerirnos el importante y complejo papel regulatorio que debe ejercer la UTR 5´ de la mencionada lipocalina.

El hecho de no haber encontrado motivos estructurales en el resto de UTRs 5´ de lipocalinas ni en ninguna de sus UTRs 3´, aún habiendo evidencias de secuencias ortólogas entre diferentes especias de mamíferos, no debemos interpretarlo como que estos están ausentes. Esta afirmación podemos hacerla ya que la mayoría del ARN no codificante parece ser específico de cada especie [23], al menos a nivel de la secuencia primaria. En este sentido ciertos estudios han puesto de manifiesto que en el genoma de mamíferos se encuentran cientos de elementos de estructura secundaria conservados, si bien no muestran evidencias de conservación en sus secuencias primarias [24 y 25]. De manera que la identificación de elementos reguladores estructurales en las regiones UTRs de las lipocalinas, donde estos no se han podido identificar con los procedimientos aquí utilizados, requeriría estudios más amplios y con una metodología adecuada a este propósito.

5. - Bibliografía

[1] Ding, Y., Chan, C. Y. & Lawrence, C. E. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **11,** 1157-1166 (2005).

[2] Giegerich, R., Voss, B. & Rehmsmeier, M.. Abstract Shapes of RNA, *Nucleic Acids Research* **32,** 4843-4851 (2004).

[3] Burgue, S. W. et al. Rfam 11.0: 10 years of RNA families. *Nucl. Acids Res. doi: 10.1093/nar/gks1005* (2012)

[4] Hochsmann, M; Voss, B. ; Giegerich, R. Pure multiple RNA secondary structure alignments: a progressive profile approach. IEEE/ACM Transactions on Computational Biology and *Bioinformatics* **1**, 53-62 (2004)

[5] Hochsmann, M . The Tree Alignment Model: Algorithms, Implementations and Applications for the Analysis of RNA Secondary Structures. Thesis, Bielefeld University. (2005)

[6] Rabani, M., Kertesz, M. & Segal, E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *PNAS* **30**, 14885–14890 (2008)

[7] Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics. 2008 Nov 11;9:474.

[8] Grillo, G., et al. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res. Jan 2010; 38(Database issue): D75–D80.

[9] *EMBOSS:* The European Molecular Biology Open Software Suite (2000) Rice,P. Longden,I. & Bleasby,A. Trends in Genetics 16, (6) pp276--277

[10] Reeder, J. ; Reeder, J. & Robert Giegerich. Locomotif: from graphical motif description to RNA motif search. Bioinformatics Vol. 23 ISMB/ECCB 2007, pages i392–i400

[11] Waldispühl1, J. et al. Efficient Algorithms for Probing the RNA Mutation Landscape. PLoS Computational Biology, August 2008, Volume 4, Issue 8, e1000124

[12] Bonhoeffer S, McCaskill J S, Stadler P F, Schuster P, (1993) RNA multistructure landscapes, Euro Biophys J:22,13-24

[13] Janssen, Stefan & Giegerich, Robert The RNA shapes studio, Bioinformatics, 2015. Vol. 31, Issue 3: 423-425.

[14] Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K.J., Lukasiak, P., Bartol, N., Blazewicz, J., Adamiak, R.W. Automated 3D structure composition for large RNAs, Nucleic Acids Research, 2012, 40(14):e112

[15] Ramana, V. D., et al. CART Classification of Human 5´UTR Sequences. Genome Research 10: 1807-1816. 2000

[16] Hughes, T.A. Regulation of gene expression by alternative untranslated regions. Trends Genet 22:119–122. 2006

[17] Pesole, G., et al. Structural and funtional features of eukaryotic mRNA untranslated regions. Gene.276, 73-81. 2001

[18] Borenstein, E. & Ruppin, E. Direct evolution of genetic robustness in microRNA. PNAS April 25, 2006, vol. 103 no. 17, 6593–6598

[19] Rastogi T, Beattie TL, Olive JE, Collins RA A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *EMBO J* **15,** 2820–2825. (1996).

[20] Theimer CA, Blois CA, Feigon J Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* **17,** 671–682 (2005).

[21] Nixon PL, Rangan A, Kim YG, Rich A, Hoffman DW, et al. (2002) Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. J Mol Biol 322: 621–633.

[22] Le, S.Y., & Maizel, J.V., Jr. A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs , Nucleic Acids Res, 1997, 25: 362-69.

[23] Hawkins PG, Morris KV (2008) RNA and transcriptional modulation of gene expression. Cell Cycle 7:602–607.

[24] Torarinsson E, Sawera M, Havgaard JH, Fredholm M, Gorodkin J Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res* **16,** 885–889 (2006).

[25] Torarinsson E, Yao Z, Wiklund ED, Bramsen JB, Hansen C, Kjems J, Tommerup N, Ruzzo WL, Gorodkin J Comparative genomics beyond sequence-based alignments: RNA structures in the ENCODE regions. *Genome Res* **18,** 242–251 (2008).

CONCLUSIONES FINALES

Conclusiones finales

En este trabajo se ha pretendido caracterizar las UTRs 5´y 3´de lipocalinas de mamíferos. Así mismo se ha tratado de dilucidar el papel que las misma desempeñan en la regulación postranscripcional de estas proteínas. Las principales conclusiones obtenidas son las siguientes:

- Las UTRs 5´de las lipocalinas de mamíferos estudiadas muestran valores de longitud y composición en G+C que se encuentran en consonancia con los valores medios de la globalidad de las UTRs 5´de mamíferos. Sin embargo las UTRs 3´de esta familia de proteínas muestran tener una longitud menor y, por otra parte un contenido en G+C claramente superior, que la media de la globalidad de UTRs 3´de mamíferos. Se ha comprobado, que al contrario de lo que ocurre con las UTRs 5´, las UTRs 3´ no reflejan el contenido G+C de la región genómica donde se ubican. Esto podría ser consecuencia de algún mecanismo de adaptación relacionado con las necesidades de regulación de la expresión génica propia de lipocalinas.

- Las UTRs de las lipocalinas de mamíferos presentan cierta diversidad, especialmente las UTRs 5´, como es de esperar por el mayor número de exones alternativos existentes en estas regiones. Esta variabilidad en las UTRs 5´ se da especialmente en las lipocalinas más ancestrales como son Apo-D, Ptgds y Rbp4. Este hecho podría explicarse por la mayor necesidad de regulación de estas, debido a que cumplen funciones más homeostáticas.

- Los mecanismos que originan las diferentes formas de UTR 5´ mencionadas resultan de una combinación de promotores alternativos junto a splicing alternativo (barajado de exones, omisión de exón y retención de intrón, entre otros). Estos mecanismos parecen estar siendo sometidos a una fina regulación. La presencia o ausencia de los oportunos factores reguladores de los promotores y del splicing, en ciertos tipos celulares o condiciones fisiológicas dadas, estaría dando lugar a la expresión UTRs 5´ alternativas, según las necesidades.

- Los estudios experimentales llevados a cabo con ApoD de ratón confirman la realidad biológica de las UTRs 5´alternativas encontradas para dicha proteína en las bases de datos. Los resultados ponen de manifiesto además que hay diferencias en la expresión de diferentes alternativas en diferentes tejidos e incluso en diferentes condiciones fisiológicas. Lo que indica que las UTRs 5´alternativas deben cumplir funciones reguladoras diferentes sobre la expresión de ApoD.

- El estudio de la conservación de las UTRs 5´ nos indica que existe un grado de conservación considerable (alrededor del 80% de identidad entre pares de secuencias ortólogas) de parte de la arquitectura de la UTR 5´ entre las lipocalinas ortólogas más ancestrales. Por otra parte hay nula conservación en las lipocalinas más recientes, a pesar de tener una estructura genómica más sencilla. Aunque hay conservación de parte de la organización genómica de las UTRs 5´, se ha producido igualmente divergencia en ella entre los diferentes linajes de mamíferos, hemos de suponer que en función de las diferentes necesidades de regulación en cada uno de estos linajes. Respecto a las UTRs 3´ se han obtenido resultados semejantes. El grado de conservación encontrado en los casos mencionados nos está indicando la importancia de la función reguladora ejercida por estas regiones.

- Respecto a los elementos reguladores que pueden estar presentes en las UTRs 5´ de lipocalinas destacan los uAUGs y los uORFs. Estos se muestran abundantes en las lipocalinas más ancestrales, siendo frecuente que exista más de un uAUG y uORF en cada UTR 5´. Las características que muestran dichos uORFs, así como las evidencias de ciertos uORFs ortólogos entre mamíferos apuntan a la funcionalidad de los mismos y a su importancia por haberse mantenido en la evolución. La presencia en las UTRs 5´alternativas de una misma lipocalina, como el caso de ApoD, de una diferente combinación de uORFs podría dar cuenta de la diferente capacidad de regulación postranscripcional de estas formas alternativas. Aunque los miARNs actúan generalmente sobre la UTR 3´ se han encontrado indicios de que en las UTRs 5´de lipocalinas de mamíferos estos elementos podrían desempeñar algún papel regulador, especialmente en las lipocalinas humanas más ancestrales.

- En cuanto al papel regulador que ejercen las UTRs 3´de lipocalinas hemos de destacar el papel que parecen desempeñar las señales de poliadenilación alternativa (PAS). Dichas señales alternativas se han encontrado especialmente en las lipocalinas más ancestrales como Ptgds, Rbp4 y ApoD. Los análisis realizados ponen de manifiesto que la presencia de estas PAS alternativas pueden dar lugar a distintas eficiencias en la poliadenilación y por lo tanto a la estabilidad de los ARNm correspondientes, pudiendo ejercer así diferentes niveles regulación de la expresión génica. Por otra parte los análisis llevados a cabo sobre los miARNs sugieren que estos podrían desempeñan un papel importante en la regulación ejercida por las UTRs 3´de las lipocalinas más ancestrales, especialmente las humanas. Las diferencias que muestran las formas largas o cortas de las UTRs 3´ de estas lipocalinas

respecto a la presencia/ausencia o la diferente composición de dianas de miARNs ofrece la posibilidad de ejercer una regulación alternativa de la expresión génica.

- Los estudios realizados para conocer las estructuras 2D globales de las UTRs han revelado que dicha estructura global está más definida en las UTRs 5´ que en las UTRs 3´. Este hecho sugiere la gran importancia que dicha estructura tiene para la acción reguladora que ejercen estas regiones UTRs 5´. Al menos para las lipocalinas ApoD y ApoM han podido identificarse elementos estructurales locales (horquillas) que muestran signos de relevancia biológica y que se han conservado en la evolución. La complejidad que muestra la estructura terciaría de una de las UTRs 5´ alternativas de ApoD de mamíferos (variante "a" humana y sus ortólogas) pone de relieve la complejidad de la regulación ejercida por las UTRs 5´de esta lipocalina, que según todos los indicios se encuentra muy regulada a nivel postranscripcional.

- Ya hemos mencionado que parte de la diversidad de los mecanismos reguladores ejercidos por las UTRs de las lipocalinas se ha conservado dentro del clado de los mamíferos, si bien se ha dado cierta divergencia entre los diferentes linajes. Los resultados obtenidos ponen de manifiesto que la complejidad y diversidad encontrada en estos mecanismos de regulación es mayor en humano que en ratón, como era de esperar por la mayor complejidad orgánica del primero. Estos resultados son especialmente interesantes para lipocalinas como ApoD, que interviene en diversos procesos en sistema nervioso, por lo que podrían tener implicaciones en los procesos cognitivos que nos diferencian respecto a otros linajes de mamíferos.