

APROXIMACIONES COMPUTACIONALES PARA LA IDENTIFICACIÓN DE PROMOTORES EUCARIOTAS TIPO II

M. Mejía-Guerra, L. Lareo

*Departamento de Nutrición y Bioquímica, Facultad de Ciencias,
Pontificia Universidad Javeriana, Cra. 7ª No. 40-62, Bogotá, Colombia
llareo@javeriana.edu.co*

RESUMEN

La identificación de segmentos funcionales dentro del DNA es uno de los campos más estudiados en la actualidad, debido a la generación de grandes volúmenes de información a partir de los proyectos genómicos, las predicciones computacionales han atraído la atención de la comunidad científica en general.

En este marco la identificación de secuencias reguladoras se erige como un importante campo del análisis de genomas; sin embargo, dada la enorme complejidad del proceso de regulación de la transcripción aunque promisorios el problema de la predicción de las zonas reguladoras no se ha solucionado en su totalidad.

Este artículo presenta una descripción de algunas aproximaciones desarrolladas para la identificación de promotores y sus elementos cis-reguladores, también se describen las estructuras de señales funcionales y estructurales que son diana de reconocimiento de algunos de estos programas de predicción aplicados al gen que codifica para la subunidad NR1 del receptor de glutamato sensible a N-metil-D-aspartato (iGluR-NMDA) el cual ha sido objeto de estudio desde la década de los noventa.

Palabras clave: genes, GRIN1, NMDA, promotores, reguladores.

ABSTRACT

The identification of functional segments in DNA sequences is one of the more heavily studied fields at the present time. Due to the generation of large volumes of information from genome projects, the usefulness of computational predictions has attracted the attention of the general scientific community.

At this point the identification of regulatory sequences has emerged as an important area in the field of genome analysis; however, given the enormous complexity of the transcription regulation process, the problem concerning the prediction of regulatory areas, although promising, has not been completely solved.

This paper presents a description of some approaches developed for the identification of promoters and their elements, cis-regulators. The structures of functional and structural signals are also described that are recognition targets for some of these prediction programs as applied to the gene that codes for the subunit, NR1, from the glutamate receptor sensitive to N-methyl-D-aspartate (iGluR-NMDA), which has been an object of study since the nineties.

Key words: genes, GRIN1, NMDA, promoters, regulator.

INTRODUCCIÓN

La identificación de elementos reguladores es hoy uno de los problemas más relevantes en los estudios de biología molecular. En los últimos años grandes avances se han hecho desde el campo experimental, esto ha permitido el acercamiento a muchos de los aspectos de la regulación de la transcripción. Además dada la importancia actual de los métodos computacionales en el análisis de datos de biosecuencias, se han desarrollado variadas aproximaciones, *in silico*, para detectar motivos reguladores dentro de diferentes genomas, como herramienta para identificar genes, en grupos de genes co-regulados, y genes ortólogos entre otros.

Si bien algunos algoritmos son promisorios, las aproximaciones computacionales para la identificación de promotores se encuentran en su infancia debido a las siguientes razones entre otras:

- Si bien experimentalmente en los últimos años se ha avanzado, eventos tales como el control del inicio de la transcripción no están totalmente comprendidos. (Fickett y Hatzigeorgiou, 1997).
- La carencia de definición estructural de las regiones promotoras, la cual tan sólo se encuentra definida a nivel funcional. (Rombauts *et al.*, 2003).
- Carencia de aproximaciones que contemplen todos los niveles (por lo menos la mayoría) de complejidad estructural y funcional de las secuencias promotoras. (Glazko *et al.*, 1998). Los modelos actuales, además, presentan diferentes resultados entre especies distintas e incluso dentro de la misma especie. (Rombauts *et al.*, 2003).
- El nivel más frecuentemente abordado para la predicción de promotores *in*

silico es la búsqueda de elementos *cis*-reguladores organizados modularmente dentro de los promotores; sin embargo, los genes pueden ser regulados en muchas formas involucrando interacciones entre varios elementos reguladores. (Hu, 2000).

- El patrón de nucleótidos que constituyen las regiones promotoras, tanto proximal (núcleo) o distal, es muy complejo y como resultado un solo programa puede tener un pobre desempeño. (Bajic, 2000).
- La identificación de homólogos aun en el caso de homólogos cercanos no garantiza que su regulación sea igual de cercana. (Rombauts *et al.*, 2003).
- La mayoría de la identificación de secuencias promotoras se ha centrado en el extremo 5'-UTR del sitio de inicio de la transcripción, aun cuando la evidencia actual apunta a la existencia de patrones con carácter regulador tanto en el extremo 3'-UTR como dentro del primer intrón. (Burset *et al.*, 2001).

Organización de las regiones promotoras

La región promotora de un gen está constituida por un grupo de secuencias específicas (elementos *cis*-reguladores) que pueden interactuar con un conjunto de proteínas específicas denominadas factores de transcripción (elementos *trans*-reguladores), con el fin de regular la transcripción de un gen. Esta región se encuentra por lo general al extremo 5'-UTR del sitio de inicio de la transcripción y se ha dividido en varias partes, la parte que contiene al sitio de inicio y la que al parecer controla la transcripción a nivel basal se denomina región promotora núcleo, en ésta se encuentran una serie de motivos de DNA más o menos conservados en todos los eucariotes que se

unen a distintos factores de transcripción (TFs) para lograr así la unión de la RNA polimerasa II. Es necesario tener en cuenta que la identificación de secuencias reguladoras a la que se refiere este artículo corresponde únicamente a los genes tipo II de eucariotes. En esencia la región núcleo del promotor corresponde al inicio del gen, la región que sigue a la región núcleo es llamada región promotora proximal, también contiene múltiples motivos de DNA que se unen a distintos TFs, es la región que usualmente regula, ya no el inicio de la transcripción, sino la velocidad a la que ésta se da, por último esta la región distal la cual no está presente en todos los genes tipo II, puede contener tanto amplificadores como represores de la transcripción, esta es la parte más variable de toda la región promotora y está menos descrita que la parte núcleo y la parte proximal.

La anterior es una consideración del nivel más básico de la región promotora, por lo tanto, se ha de tener en cuenta que el proceso de transcripción también depende de la estructura de orden superior de la cromatina, de la posición no aleatoria de los nucleosomas y de la torsión del DNA, entre otros. Sin embargo, estos aspectos no están lo bastante descritos como para generar modelos que permitan la predicción de las zonas reguladoras, por lo tanto, la mayoría de las aproximaciones actuales se basan en la identificación de los promotores a partir de la aproximación lineal antes presentada.

Los motivos cis-reguladores componentes de los elementos promotores son motivos cortos que usualmente están entre los 5 y los 20 nucleótidos de longitud (Wingeder, 1998; Tjian, 1995; Fickett y Hatzigeorgiou, 1997), varias bases de datos que relacionan los elementos cis-reguladores con Factores de Transcripción se encuentran disponibles, algunas de ellas son TRANSFAC, COMPEL, EPD y TRRD; sin embar-

go, si se está interesado en ahondar sobre un factor de transcripción, no hay ningún sustituto de la revisión de la literatura.

Dentro de los factores cis-reguladores, más conservados en la región núcleo del promotor, están la secuencia TATA localizada corriente arriba del sitio de inicio de la transcripción, y la secuencia iniciadora (Inr), la cual incluye el sitio de inicio. La región nuclear del promotor incluye una caja TATA, una secuencia Inr, o ambos elementos, un tercer elemento conservado en el promotor es la secuencia DPE el cual fue inicialmente descrito en *Drosophila* y está localizada alrededor de 30 pb corriente abajo del sitio de inicio de transcripción. El DPE aparece funcionando en conjunto con el elemento Inr, como un sitio de unión para el factor de transcripción TFIID en promotores carentes de caja TATA (TATA -).

Caja TATA

La caja TATA en eucariotes varía su localización entre especies por ejemplo en *S. cerevisiae* está localizada entre 40 y 120 pb corriente arriba del sitio de inicio de la transcripción, en otro eucariota *Shizosaccharomyces pombe*, el elemento TATA está localizado entre 25 a 80 pb del sitio de inicio de transcripción, al igual que en el hombre.

TATA es el sitio de unión para la proteína de unión a TATA (TBP por sus iniciales en inglés) la cual hace parte del complejo de preiniciación de la transcripción, la secuencia consenso del elemento TATA es TATAAA en levaduras; sin embargo, muchos elementos derivados de esta secuencia tales como TGTA AAA funcionan como caja TATA, algunos estudios sugieren que esto ocurre debido a otros factores que pueden unirse también a TBP.

Algunos genes carecen de una caja TATA aproximadamente entre 30 y 50% del total

de los genes humanos se presume son TATA -, al parecer dichos genes presentan características y niveles de expresión diferentes a aquellos genes que poseen caja TATA; sin embargo, estos genes no se encuentran tan estudiados como los TATA +.

Elemento iniciador (Inr)

El elemento Inr, como está definido en eucariotas superiores, no es simplemente la secuencia que incluye el sitio de inicio de la transcripción, el elemento Inr fue visto por primera vez en el promotor TATA - del gen de la deoxinucleotidiltransferasa como un elemento del promotor núcleo de dicho gen, el cual puede ensamblar el complejo de preiniciación nuclear. Posteriormente el elemento Inr fue identificado en muchos promotores, tanto TATA + como TATA -, y ha sido implicado en el control transcripcional por dirigir la iniciación adecuada de manera independiente de la caja TATA.

Al parecer el elemento Inr facilita el control transcripcional y podría ser una característica universal de la transcripción por la RNA polimerasa II.

Otros elementos tales como las islas CpG, UAS y URS también son elementos conservados.

La combinación de los motivos simples de los promotores distales y proximales genera la diversidad de la regulación en la expresión génica. Los factores de transcripción y las secuencias reguladoras están compuestos por componentes modulares que otorgan un alto nivel de especificidad por un número relativamente pequeño de diferentes TFs.

Para el entendimiento apropiado de la función génica, se debe concentrar la atención en los patrones de regulación de secuencias y no en los elementos individuales. La

búsqueda de tales secuencias debería ser más efectiva en la anotación de nuevas secuencias en comparación con el reconocimiento de motivos individuales.

En la actualidad numerosas bases de datos relacionales de los elementos cis-reguladores y factores de transcripción entre estas bases de datos están:

EPD: contienen información acerca de promotores, como ellos están definidos experimentalmente proveen sitios de inicio de la transcripción y su especificidad de tejido.

TRANSFAC: provee información sobre estructura, función, secuencia y clasificación de TFs, acerca de los sitios de unión.

TRRD: base de datos de regiones reguladoras transcripcional colecta información acerca de la estructura total de regiones reguladores de genes eucariotas.

COMPEL: colectan la información sobre elementos regulatorios compuestos. Éstos son definidos como parejas de sitios de unión situados cercanamente, corresponde a factores de transcripción e interacciones proteína-proteína entre ellas, y patrones de expresión provista por esta regulación combinatorial.

Receptor ionotrópico de glutamato activado por N-metil-D-aspartato (iGluR-NMDA)

El glutamato aparece como el principal transmisor excitador y sólo los receptores ionotrópicos de glutamato sensibles a N-metil-D-aspartato (iGluR-NMDA) juegan un importante papel en la plasticidad sináptica, la inducción de potenciación a largo término y la excitotoxicidad entre otros (Bai y Kusiak, 1995). Además es un hecho que el receptor iGluR-NMDA al igual que los receptores de glutamato no sensibles al

N-metil-D-aspartato (GluRs) se expresa de manera diferencial tanto espacial como temporalmente, confiriendo propiedades específicas en diferentes tejidos y etapas del desarrollo, (Bai *et al.*, 2003, Pellegrini-Giampietro *et al.*, 1991). La subunidad NR1 (NMDAR1) en el iGluR-NMDA es clave en el funcionamiento del receptor, tanto en humanos como en muridos se encuentran isoformas diferentes, lo que sugiere una compleja regulación para el gen encargado de codificar esta proteína (GRIN1), las otras subunidades (codificadas a su vez por otros GRIN) se combinan con NR1 y juegan papeles importantes aunque se ha determinado que la ausencia de ésta (NR1) causa la no funcionalidad del receptor.

Las regiones reguladoras de los genes que codifican para el iGluR-NMDA tanto en humano como en sus homólogos en muridos se han estudiado desde hace algunos años de manera experimental; sin embargo, la investigación al respecto permanece poco desarrollada y muchos aspectos están por dilucidarse aun, por esto, es necesario abordar este problema desde la perspectiva bioinformática con el fin de proponer un modelo de las regiones reguladoras para el gen GRIN1 en humanos y roedores.

GRIN1

El gen que codifica para la subunidad NR1 del iGluR-NMDA ha sido objeto de estudio desde la década de los noventa. El mapeo de este gen se ha realizado de manera independiente por varios investigadores desde el trabajo de Karp y colaboradores, en 1993, cuando llevaron a cabo análisis de hibridización del DNA de un grupo de células somáticas híbridas de humano/hámster y por hibridización *in situ* mediante fluorescencia; otros dos grupos de investigadores motivados por la probable asociación de este gen con enfermedades como Huntington; fue mapeado en humanos en la región 9q34.3 sugiriendo ambos

trabajos la participación de GRIN1 en la distonía de torsión. (Collins *et al.*, 1993; Takano *et al.*, 1993). Brett, *et al.*, en 1994, realizaron el mapeo del gen en humanos en la posición 9q34.3 usando un clon genómico que demostró la presencia de una secuencia polimórfica (VNTR) hacia la región 5' del gen y utilizó esta secuencia como marcador para análisis de ligamiento en familias de caucásicos determinó la tasa de recombinación en cero (0) para machos y 0,03% para hembras.

En 1995 Zimmer, *et al.*, clonaron el gen y determinaron que consiste en 21 exones distribuidos aproximadamente en 31 Kb, el equipo secuenció un total de 24.485 pb del gen y las dividió en tres segmentos llamados A, B y C; a estas secuencias se les asignaron los números de acceso Z32772, Z32773 y Z32774 en la base de datos de nucleótidos del GenBank éstas con GI's 807892, 807893 y 807894 respectivamente.

Los exones 4, 20 y 21 son idénticos en la secuencia de aminoácidos para la cual codifican y son susceptibles de corte y empalme alternativo tanto en rata como en humano.

En la región 5'-UTR del gen se han identificado una caja TATA putativa a 629 pb del sitio de inicio de la transcripción y dos sitios de unión a Sp1. Tres sitios adicionales de unión a Sp1 fueron encontrados en orientación inversa en el primer intrón. Una secuencia consenso localizada 33 pb corriente arriba del exón 2 y un sitio putativo de unión a la proteína HS en el segundo intrón y dos secuencias de reconocimiento para la proteína homeótica similares a los "even-skipped" de *Drosophila*.

MÉTODOS

Aproximaciones desarrolladas para la identificación de promotores y sus elementos cis-reguladores

La disponibilidad de secuencias de genomas completos y datos de expresión de mRNA para muchos genes crea nuevas oportunidades y metas para la identificación de motivos de DNA que controlan la expresión génica.

Un vasto número de métodos computacionales han sido diseñados para el reconocimiento de regiones reguladoras, este reconocimiento va desde la detección del sitio de inicio de la transcripción, reconocimiento de motivos a partir de matrices de peso. Modelos de cadenas de Markov o redes neurales. Muchos de estos métodos han sido revisados desde inicio de los noventa.

En uno de los primeros intentos de predecir promotores eucarióticos uso la densidad de los sitios de unión de los factores de transcripción en combinación con la matriz de peso para la caja TATA construida por Bucher en 1990. (Prestridge, 1995).

Las matrices de peso se derivan de un grupo de secuencias de sitios funcionales de longitud L. La matriz de peso $F = (i, j)$, ($i = A, T, G, C$; $j = 1, L$) es la matriz de frecuencia de nucleótidos y $f(i, j)$ es la frecuencia absoluta de ocurrencia del tipo i -ésima del nucleótido en la posición j -ésima. La matriz optimizada de peso del sitio funcional puede ser construida en la siguiente forma:

$$W = |w(i, j)|, w(i, j) = \ln(f(i, j) / e(i, j) + s / 100 + c(i))$$

Donde $e(i, j)$ es la frecuencia nucleotídica expresada del tipo i -ésimo en la posición j -ésima, $c(i)$ es la columna específica y S es el porcentaje suavizante. La matriz de peso es luego usada para discriminar y reconocer los sitios funcionales por encima del umbral del puntaje de apareamiento.

A partir de otros procedimientos distintos se pueden generar otras matrices de peso más complicadas. El intento de Prestridge

(1995) hace parte de la tendencia principal del reconocimiento de promotores combinando el reconocimiento individual de los sitios de unión y alguna descripción de cómo estos sitios podrían estar arreglados espacialmente. Algunas veces la especificidad de unión está caracterizada usando secuencias consenso, en vez de matrices de peso, lo que muestra la base más preferida en cada posición dentro de una secuencia.

Esta aproximación pierde mucha información y es de utilidad marginal ya que no tiene en cuenta la especificidad de tejido de los diferentes valores, entre las aproximaciones de predicción a partir de consensos, se usan diferentes aproximaciones para construir consensos, algunas de éstas han sido analizadas por Day y McMorris (1992).

Debido a que el consenso habla de una señal que presenta variabilidad puede ser descrita usando el código estándar IUPAC-IUB el cual usa un único carácter para describir los nucleótidos ambiguos presentes en el consenso. Este código se presenta en la tabla 1.

Tabla 1
Nomenclatura IUPAC-IUB para las bases de las secuencias nucleotídicas

Código estándar IUPAC-IUB
A, T (U), G, C
R= (A o G)
Y= (T o C)
S= (G o C)
W= (A o T)
K= (G o T)
M= (A o C)
B= (T o G o C)
V= (A o G o C)
H= (A o T o C)
D= (A o T o G)
N= (A o T o G o C)

Otra aproximación metodológica es la asignación de un peso o valor a cada nucleótido posible en cada posición de un sitio putativo de unión y da un puntaje para dicho sitio a partir de la suma de los valores o pesos. Es de conocimiento generalizado que la información a partir de dicha metodología describe más información de la especificidad de unión de los diferentes factores de transcripción que las secuencias consenso.

La metodología PWM está sustentada sobre la hipótesis de que diferentes posiciones dentro del sitio hacen contribuciones independientes a la unión. Si bien, un número de casos conocidos corresponden a dicha hipótesis, PWM en algunas ocasiones muestra resultados de baja calidad.

Esto podría atribuirse a muchas razones, por ejemplo a la existencia de múltiples isoformas de la proteína, para construir un modelo de la unión de un factor de transcripción al DNA, se necesita un número de sitios conocidos y es muy valioso tener los valores de la fuerza de unión del factor de transcripción a dichos sitios, esta información es poco usual.

En 1997, Fickett y Hatzigeorgiou revisaron varias de las herramientas computacionales que existían en ese momento para la predicción de promotores. Los programas que se probaron no identificaban confiablemente promotores en una secuencia genómica, había una gran cantidad de falsos positivos, aproximadamente 1 falso positivo por cada Kbp analizado. En este trabajo se consideró, que esta revisión presenta la razón para que estos programas no tuvieran un mejor desempeño debido a que están enfocados en la búsqueda de señales usando una o dos características, pero descartando propiedades estructurales y propias de la secuencias que caracterizan elementos promotores.

Las aproximaciones evaluadas por Fickett y Hatzigeorgiou incluían oligonucleótidos basados en contenidos (Hutchinson, 1996; Audic y Claverie, 1997), redes neurales (Guigo *et al.*, 1992) y aproximaciones de discriminación lineal (Solovyev y Salamov, 1997), estos programas podrían ser usados para encontrar la posición del promotor en una región 5'-UTR dada o para ayudar a seleccionar al exón correcto en una predicción génica previa.

Burset *et al.*, (2001) sugieren que los programas que basan su búsqueda de señales en la presencia de la caja TATA son insensibles a los promotores TATA - ya que éstos tienen características muy diferentes, y dichos programas no son entrenados con secuencias de este tipo.

El programa TSSW es uno de los pocos entrenados con secuencias TATA -, este programa tiene una función lineal que le permite discriminar los TATA + de los TATA -; sin embargo, este programa muestra una pobre calidad de reconocimiento de promotores TATA - contrario a lo que sucede con los TATA +.

La detección de señales también se lleva a cabo a partir de algoritmos de alineamientos múltiples locales que identifican motivos reguladores, los cuales están basados simplemente en propiedades estadísticas. Tales algoritmos buscan patrones en el DNA que estén con más frecuencia en un grupo de secuencias "relacionadas" frente a las "no relacionadas".

La confiabilidad de tales predicciones depende en mucho del grupo de genes escogidos.

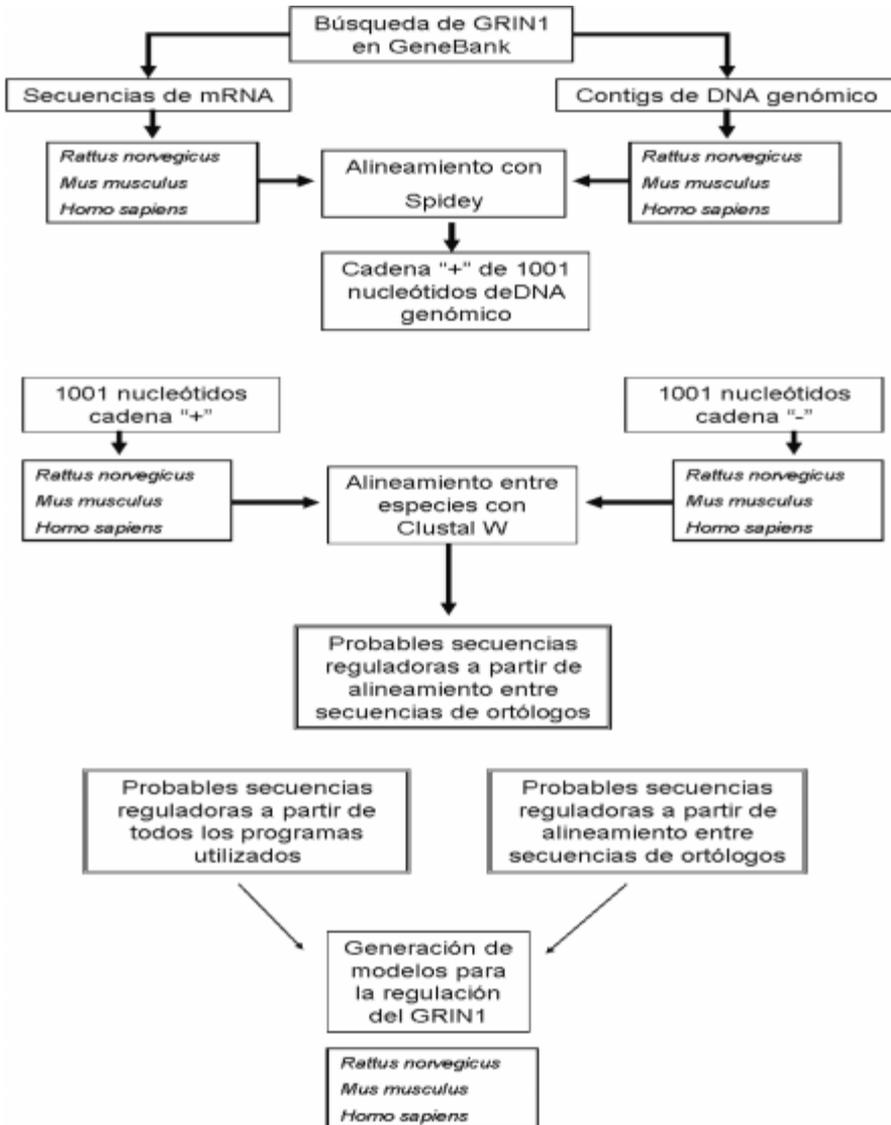
Para la realización del trabajo con el gen GRIN1 del iGluR-NMDA se seleccionaron las secuencias de mRNA correspondientes al gen para rata (*Rattus norvegicus*), ratón

(*Mus musculus*) y humano (*Homo sapiens*) disponibles en la base datos del GenBank. Tales genes se encontraron reportados dentro de "contigs" de DNA genómico y para identificar la región UTR (región no transcrita) se realizaron alineamientos de dichos "contig" con mRNA utilizando el

software Spidey (<http://www.ncbi.nlm.nih.gov/Spidey>) disponible en el sistema Entrez de NCBI.

En la figura 1 se presenta un diagrama de flujo de la secuencia del trabajo sobre los genes GRIN1.

Figura 1. Diagrama de flujo para los procesos hechos con el gen GRIN1



Luego de acceder a esta región se tomaron 1001 nucleótidos de región 5'-UTR y de región transcrita de manera que del sitio +1 al +200 estén contenidos en la secuencia a trabajar. La secuencia obtenida se denominó cadena "+" a partir de ésta se escribió una segunda secuencia a trabajar la cadena "-" que es la complementaria.

Las cadenas + y - de cada gen se alinearon entre sí con el fin de identificar secuencias conservadas entre ellos por identificación de secuencias reguladoras entre genes ortólogos.

La cadena + fue sometida a predicción con diferentes software, todos los software arrojan datos tanto para la cadena + como para

la cadena -. Los resultados fueron revisados teniendo en cuenta lo reportado en la literatura de regulación de dichos genes y la expresión de tejido de los diferentes factores de transcripción que podrían estar involucrados.

Por último a partir del consenso de los alineamientos y las predicciones se esquematizaron las propuestas de regulación para cada uno de los genes GRIN1.

RESULTADOS

Los elementos reguladores del GRIN1 identificados con los programas TSS, ConSite y TFSEARCH en los genomas de humano, rata y ratón se presentan en las tablas 2, 3 y 4.

Tabla 2
Localización genómica de los elementos reguladores del gen GRIN1 identificados con el programa TSS para humano, rata y ratón

Posición	<i>Homo sapiens</i>		<i>Rattus norvegicus</i>		<i>Mus musculus</i>	
	+	-	+	-	+	-
-800 to -700	Lyf-1, En, Sp1, NIT2	Sp1, GCF, MEF-2	SRF	SRF, ANF, USF	Kr	MEF-2
-700 to -600	RAR α , Lyf-1	RAR α (4 copies), E12	Pit-1A		Pit-1A	
-600 to -500	c-Ets-1	Ap-2	TBP, D1	ATF, RAR β	C/EBP α , D1	FtZ, C/EBP, H4TF-2
-500 to -400	D1, Sp1(4 copies), Olf-1	NF-ODC1, RXR α , XPF-1	Sp1	E12	H4TF-2	E12
-400 to -300	Sp1 (2 copies)	D1, H4TF-1, Sp1 (7 copies)	SRF, Sp1 (2 copies)	SRF, CArG	SRF, Sp1 (2 copies)	SRF, CArG
-300 to -200	Pur-factor, USF		c-Ets-1, MTB-Zf	Zn-15, MTB-Zf	c-Ets-1, MTB-Zf	Sp1, Dfd, MTB-Zf
-200 to -100	TF68, Sp1	CREB-BP1, Sp1			Dof2, Sp1	
-100 to 0	Sp1		Sp1 (2 copies), Sryδ (2 copies)	SRF	PU.1, Sryδ (2 copies)	MIG-1
0 to +100	MZF-1	GAGA	NF-Ins, TGAI, Sp1	Zn-15	NF-Ins, TGAI, Sp1	Zn-15
+100 to +200			Eve	GR, GCM, NRF-1	Eve	NRF-1 (2 copies)

Los resultados obtenidos por estos tres métodos no resultaron idénticos, como era de esperarse, pero tampoco son contradictorios entre ellos. Los resultados reportados en las tablas son los que presentaron mayor significancia estadística según los indicadores de cada programa, por ejemplo, para TSS se seleccionaron aquellas predicciones con puntajes superiores a 18, mientras que para ConSite se seleccionaron los resultados con puntajes mayores o iguales al 10%.

Los resultados obtenidos con el programa Promoterscan, que no se presentan en nin-

guna tabla, reportaron predicciones sólo para los humanos e identificaron sitios reguladores solamente entre las posiciones 700 y 720 de ambas cadenas. Se identificaron, particularmente los sitios de los motivos Sp1, EARLY SEQ1 y para el decámero TCR β. Cada uno de los factores de transcripción que potencialmente interactúan con el GRIN1 fueron analizados individualmente y sus descripciones se obtuvieron de las bases de datos TRANSFAC, EPD, GeneCards, Lion, COMPEL y OMIM (Schmid *et al.*, 2004; Kel-Margoulis *et al.*, 2002; Zhang, 2003) así como en la literatura especializada respectiva.

Tabla 3
Localización genómica de los elementos reguladores del gen GRIN1 identificados con el programa ConSite para humano, rata y ratón.

Posición	<i>Homo sapiens</i>		<i>Rattus norvegicus</i>		<i>Mus musculus</i>	
	+	-	+	-	+	-
-800 to -700	MEF2	Bsap, HMG-IY, SQUA	SRF-AGL3, Max-Myc-Max, ARNT-nMyc	ARNT-nMyc		
-700 to -600	Spz1	Snail				
-600 to -500	RREB-1 (2 copies)			E4BP4, RORα1, Staf	E4BP4	E4BP4, Staf
-500 to -400	Snail	p50				
-400 to -300	RREB-1			Snail		Snail
-300 to -200	SU_h, Staf	CFI-USP				
-200 to -100					Thing1-E47	
-100 to 0	TEF-1		RORα2	Snail	RORα2	Staf, Snail
0 to +100		Bsap		NF-Y		NF-Y
+100 to +200			Tal1β-47S, CFI-USP	Tal1β-47S	Tal1β-47S, CFI-USP	

En negrilla aparecen los elementos comunes con localización aproximada, pero idéntica. En itálica aparecen los elementos comunes a las tres especies.

Tabla 4
Localización genómica de los elementos reguladores del gen GRIN1 identificados con el programa TFSEARCH para humano, rata y ratón.

Posición	<i>Homo sapiens</i>		<i>Rattus norvegicus</i>		<i>Mus musculus</i>	
	+	-	+	-	+	-
-800 to -700	Lyf-1 (2 copies)	deltaE	CdxA	USF	CdxA , HNF-3b	SKY (2 copies)
-700 to -600	GATA1, GATA2, GATA3	GATAX, GATA1	CdxA	SRY	Mkx-2	
-600 to -500	GATA1	Sp1	Mkx-2			
-500 to -400	E47	c-Ets	STATx	E12	STATx	
-400 to -300	AML-1a		SRF	SRF	AML-1a	HFS-2
-300 to -200	<i>GATA1</i>		<i>GATA1</i> , <i>GATA2</i>	GATAX	<i>GATA1</i> , <i>GATA2</i>	GATAX
-200 to -100	AML-1a	CREB				
-100 to 0			HSF2	MZF1		
0 to +100			Tal-1a			
+100 to +200			Tal-1b			

En negrilla aparecen los elementos comunes con localización aproximada, pero idéntica. En itálica aparecen los elementos comunes a las tres especies.

DISCUSIÓN

Las regiones reguladoras de los genes para las subunidades del iGluR-NMDA, en humanos y sus homólogos, los roedores han sido estudiados experimentalmente desde hace varios años pero no se ha logrado identificar muchos de los sitios potenciales. Krainc *et al.*, (1998) sugiere que es posible encontrar muchos más sitios potenciales y factores adicionales a los ya reconocidos.

La identificación de las secuencias reguladoras del GRIN1 en humanos y roedores se hace sobre la premisa de que estos genes son ortólogos.

Los genes de humano, rata y ratón presentan, en promedio, un 85% de similitud mientras que sus secuencias proteicas tienen una similitud promedio de hasta 98%. Los alineamientos de las regiones 5'-

UTR revelan similitudes de tan sólo 48% entre las tres especies aunque alcanza valores de 90,0% entre los roedores. Estos resultados revelan la especificidad de la regulación entre las especies.

El GRIN1 se ha identificado como uno de los genes más conservados, con respecto a los otros GRIN dentro de las especies consideradas, en su región reguladora, sin embargo, presentan pocos sitios en común y menos aun en las mismas posiciones relativas. Por ejemplo, es posible encontrar el elemento Sp1 en todas las especies pero no en las mismas posiciones relativas al codón de iniciación. El elemento Pit-1A es común entre los roedores y conserva, aproximadamente la misma posición así como los elementos SRF, CArG pero no son comunes con las regiones reguladoras en los humanos.

Estos resultados pueden tener amplias implicaciones y permitir explicar muchas de las variaciones funcionales del complejo heteromérico iGluR-NMDA, y por lo tanto, deben ser validados experimentalmente.

También estos resultados conllevan la implicación limitar la aplicación de los modelos animales para explicar los procesos neurofisiológicos a del iGluR-NMDA en los humanos dada su significativa diferencia de regulación y posiblemente, como consecuencia de esto, de estequiometría de los complejos entre las especies.

En negrilla aparecen los elementos comunes con localización aproximada, pero idéntica. En itálica aparecen los elementos comunes a las tres especies.

LITERATURA CITADA

- AUDIC, S. y CLAVERIE, J.M. 1997. Detection of eukaryotic promoters using Markov transition matrices. *Comput Chem* 21: 223-227.
- BAI, G. y KUSIAK, J.W. 1995. Functional analysis of the proximal 5'-flanking region of the N-methyl-D-aspartate receptor subunit gene, NMDAR1. *J Biol Chem* 270: 7737-7744.
- BAI, G.; ZHUANG, Z.; LIU, A.; CHAI, Y. y HOFFMAN, P.W. 2003. The role of the RE1 element in activation of the NR1 promoter during neuronal differentiation. *J Neurochem* 86: 992-1005.
- BAJIC, V.B. 2000. Comparing the success of different prediction software in sequence analysis: a review. *Brief Bioinform* 1: 214-228.
- BRETT, P.M.; LE BOURDELLES, B.; SEE, C.G.; WHITING, P.J.; ATTWOOD, J.; WOODWARD, K.; ROBERTSON, M.M.; KALSJ, G.; POVEY, S. y GURLING, H.M. 1994. Genomic cloning and localization by FISH and linkage analysis of the human gene encoding the primary subunit NMDAR1 (GRIN1) of the NMDA receptor channel. *Ann Hum Genet* 58: 95-100.
- BURSET, M.; SELEDTSOV, I.A. y SOLOVYEV, V.V. 2001. SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res* 29: 255-259.
- COLLINS, C.M.; GUTMAN, D.M. y LAMAN H. 1993. Identification of a nitrogen-regulated promoter controlling expression of *Klebsiella pneumoniae* urease genes. *Mol Microbiol* 8: 187-198.
- FICKETT, J.W. y HATZIGEORGIOU, A.G. 1997. Eukaryotic Promoter Recognition. *Genome Res* 7, 861-878.
- DAY, W.H. y McMORRIS, F.R. 1992. Critical comparison of consensus methods for molecular sequences. *Nucleic Acids Res* 20: 1093-1099.
- GUIGO, R.; KNUDSEN, S.; DRAKE, N. y SMITH, T. 1992. Prediction of gene structure. *J Mol Biol* 226: 141-157.
- GLAZKO, G.B.; MILANESI, L. y ROGOZIN, I.B. 1998. The subclass approach for mutational spectrum analysis: application of the SEM algorithm. *J Theor Biol* 192: 475-487.
- HU, Y.; PAN, Q.; PARDALI, E.; MILLS, F.C.; BERNSTEIN, R.M.; MAX, E.E.; SIDERAS, P. y HAMMARSTROM, L. 2000. Regulation of germline promoters by the two human Ig heavy chain 3' alpha enhancers. *J Immunol* 164: 6380-6386.
- HUTCHINSON, G.B. 1996. The prediction of vertebrate promoter regions using differential hexamer frequency analysis. *Comput Appl Biosci* 12: 391-398.

- KARP, S.J.; MASU, M.; EKI, T.; OZAWA, K. y NAKANISHI, S. 1993. Molecular cloning and chromosomal localization of the key subunit of the human N-methyl-D-aspartate receptor. *J Biol Chem* 268: 3728-3733.
- KEL-MARGOULIS, O.V.; KEL, A.E.; REUTER, I.; DEINEKO, I.V. y WINGENDER, E. 2002. TRANSCompel: a database on composite regulatory elements in eukaryotic genes. *Nucleic Acids Res* 30: 332-334.
- KRAINC, D.; BAI, G.; OKAMOTO, S.; CARLES, M.; KUSIAK, J.W.; BRENT, R.N. y LIPTON, S.A. 1998. Synergistic activation of the N-methyl-D-aspartate receptor subunit 1 promoter by myocyte enhancer factor 2C and Sp1. *J Biol Chem* 273: 26218-26224.
- PELLEGRINI-GIAMPIETRO, D.E.; BENNETT, M.V. y ZUKIN, R.S. 1991. Differential expression of three glutamate receptor genes in developing rat brain: an in situ hybridization study. *Proc Natl Acad Sci U S A*. 88: 4157-4161.
- PRESTRIDGE, D.S. 1995. Predicting Pol II promoter sequences using transcription factor binding sites. *J Mol Biol* 249: 923-932.
- ROMBAUTS, S.; FLORQUIN, K.; LESCOT, M.; MARCHAL, K.; ROUZE, P. y VAN DE PEER, Y. 2003. Computational Approaches to Identify Promoters and cis-Regulatory Elements in Plant Genomes. *Plant Physiology* 132: 1162-1176.
- SCHMID, C.D.; PRAZ, V.; DELORENZI, M.; PERIER, R. y BUCHER, P. 2004. The Eukaryotic Promoter Database EPD: the impact of *in silico* primer extension. *Nucleic Acids Res* 32: D82-D85.
- SOLOVYEV, V. y SALAMOV, A. 1997. The Gene-Finder computer tools for analysis of human and model organisms genome sequences. *Proc Int Conf Intell Syst Mol Biol* 5: 294-302.
- TAKANO, H.; ONODERA, O.; TANAKA, H.; MORI, H.; SAKIMURA, K.; HORI, T.; KOBAYASHI, H.; MISHINA, M. y TSUJI, S. 1993. Chromosomal localization of the epsilon 1, epsilon 3 and zeta 1 subunit genes of the human NMDA receptor channel. *Biochem Biophys Res Commun* 197: 922-926.
- TJIAN, R. 1995. Molecular machines that control genes. *Sci Am* 272: 54-61.
- ZHANG, M.Q. 1998. Identification of Human Gene Core Promoters *in silico*. *Genome Res* 8: 319-326.
- ZIMMER, M.; FINK, T.M.; FRANKE, Y.; LICHTER, P. y SPIESS, J. 1995. Cloning and structure of the gene encoding the human N-methyl-D-aspartate receptor (NMDAR1) *Gene* 159: 219-223.

Recibido: 12.05.2005

Aceptado: 14.03.2006

