

UNIVERSIDAD DE CHILE
Facultad de Ciencias
Ingeniería en Biotecnología Molecular

**Búsqueda de regiones de baja identidad
entre secuencias de genes relacionados filogenética o metabólicamente
utilizando estrategias de programación dinámica**

Mónica Ester Cisternas Fierro



Santiago, Chile
2006

UCH-FC
Biotecnología
C-579
C-1



“BÚSQUEDA DE REGIONES DE BAJA IDENTIDAD ENTRE SECUENCIAS DE GENES RELACIONADOS FILOGENÉTICA O METABÓLICAMENTE UTILIZANDO ESTRATEGIAS DE PROGRAMACIÓN DINÁMICA”

Seminario de Título entregado a la Universidad de Chile en cumplimiento parcial de los requisitos para optar al Título de Ingeniero en Biotecnología Molecular.

MÓNICA CISTERNAS FIERRO

Dr. Mauricio González Canales
Director Seminario de Título

Mauricio González

Comisión de Evaluación de Seminario de Título

Dr. Osvaldo Álvarez Araya
Presidente Comisión

Dra. Verónica Cambiazo Ayala

Verónica Cambiazo



Santiago de Chile, _____ 2006

*A mis papás por su esfuerzo.
A mis amores, Ernesto y Diego,
por hacerme muy feliz.*



AGRADECIMIENTOS



Al proyecto FONDECYT 1000852.

A Mauricio, por su paciencia a toda prueba después de todos estos años.

A Verónica por enseñarme muchas cosas del trabajo de laboratorio, conocimiento que sólo da la experiencia.

A mis amigas Miriam y Lucía, mis chicas súper poderosas.

A mis compañeros de lab: Hans, Chodar, Talía, Negro, Alejandro, por todos los días juntos y la buena onda en las jornadas de trabajo.

A Lorena, Mireya, Mónica y Jessiça por los conversados almuerzos de mujeres.

A Lucho por la desinteresada ayuda que me dio para terminar esta Memoria de Título.

A Camilo y Pilar, por compartir conmigo la tranquilidad de su casa.

A mi tía Norma por ayudarme a custodiar cariñosamente al loco de la casa.

A Felipe por compartir conmigo y unas galletitas a JCS y poner de su parte para entender algo de biología.

A mi amiga Paula por su tiempo y sus comentarios desde la distante tierra de las cervezas.

A los chicos del Laboratorio de Bioinformática y Expresión Génica por el apoyo moral en el momento crítico y por recoger mis pedacitos cuando fue necesario.

ÍNDICE DE MATERIAS



Título	i
Dedicatoria	ii
Agradecimientos	iii
Índice de Materias	iv
Lista de Tablas	vii
Lista de Figuras	viii
Lista de Símbolos, Abreviaturas o Nomenclatura	ix
Resumen	x
Abstract	xii
INTRODUCCIÓN	1
Una alternativa para estudiar cambios de expresión génica a escala genómica	1
Construcción y factores que afectan el ensayo de hibridación	3
Fuentes de error asociadas al ensayo de hibridación y su análisis	6
Programas de análisis de identidad entre secuencias	9
Sistema de estudio	11
Hipótesis	13
Objetivo General	13
Objetivos Específicos	13
Estrategia Experimental	14

MATERIALES Y MÉTODOS	15
Materiales	15
Métodos	17
Diseño de un programa de búsqueda de dominios de baja identidad	17
Características del programa	17
Análisis preliminares para optimizar uso del programa	20
Selección de clones	23
Diseño de los partidores	25
Amplificación por PCR de los fragmentos	26
Clonación en el vector	28
Transformación para la obtención de clones con el fragmento	28
Cultivo de colonias positivas	29
Verificación de clones positivos	29
Stock de clones positivos	29
Obtención del DNA a sembrar	29
Obtenciones de las muestras	30
Tratamiento de las membranas	33
Hibridación	34
Lavados y exposición	35
Revelado	35
Análisis de los puntos	35
Cultivo celular	36
RESULTADOS	37
Programa Redigen	37
Ensayos Experimentales	49

DISCUSIÓN

Análisis de las características de operación del programa

Estrategias actuales de selección de sondas

Análisis de resultados en ensayos de hibridación

CONCLUSIÓN

REFERENCIAS BIBLIOGRÁFICAS



59

60

65

66

69

70

LISTA DE TABLAS

- Tabla 1: Genes involucrado en el metabolismo celular de cobre utilizados en este trabajo.
- Tabla 2: Partidores utilizados para la amplificación de sondas.
- Tabla 3: Cuantificación del desplazamiento de las curvas de comparación de *atp7a* y *atp7b*.
- Tabla 4: Comparación de fragmentos seleccionados como sondas con los genes involucrados en el metabolismo celular de cobre mediante ClustalW.
- Tabla 5: Porcentaje de identidad dentro del grupo de fragmentos de alta y baja identidad de *atp7a* y *atp7b*.



LISTA DE FIGURAS

- Figura 1: Participación de proteínas involucradas en el metabolismo de cobre en mamíferos.
- Figura 2: El programa Redigen compara la secuencia del gen de interés con la de otro(s) gen(es) mediante fragmentos.
- Figura 3: Desplazamiento de las curvas generadas por el programa Redigen al comparar las secuencias de *atp7a* y *atp7b* entre sí.
- Figura 4: Efecto del tamaño de ventana en el valor de unidades de identidad de las secuencias analizadas.
- Figura 5: Efecto de los parámetros porcentuales en las curvas de identidad de *atp7a* y *atp7b*.
- Figura 6: Efecto del tamaño de ventana relacionado con los parámetros porcentuales en las curvas de identidad de *atp7a* y *atp7b*.
- Figura 7: Comparación de la secuencia de *sod1* con las de otros genes relacionados con la vía metabólica del cobre.
- Figura 8: Especificidad de los fragmentos versus concentración de las sondas sembradas.
- Figura 9: Expresión de genes relacionados con el metabolismo de Cu en fibroblastos wt.
- Figura 10: Sensibilidad de detección de las sondas de *atp7a*.
- Figura 11: Sensibilidad de detección de las sondas de *atp7b*.
- Figura 12: Sensibilidad de las sondas de alta y baja identidad de *atp7a* y *atp7b* en células HepG2 tratadas con cobre.
- Figura 13: Aumento en los valores de UA al aumentar la longitud de las secuencias comparadas en las curvas generadas con el programa Redigen.

LISTA DE SÍMBOLOS, ABREVIATURAS O NOMENCLATURA

DNA:	Ácido desoxirribonucleico.
cDNA:	DNA de hebra simple que es complementario al RNA mensajero, sintetizado a partir de éste.
RNA:	Ácido ribonucleico.
PCR:	Reacción en cadena de la polimerasa.
Empalme alternativo de intrones:	Remoción de intrones y unión de exones en el RNA, los intrones se eliminan y los exones se unen (en inglés: splicing).
Isoformas:	Familias de proteínas que tiene formas heterogéneas. Las isoformas tienen variación en la vida media y en el peso molecular.
Relación filogenética:	Relación evolutiva entre dos organismos o moléculas, sugiriendo un origen común.
Relación metabólica:	Relación entre un grupo de genes que participa de un mismo proceso celular.
ATPasas:	Familia de proteínas transmembrana que utilizan ATP como fuente de energía para transportar un elemento contra gradiente electroquímico.
Ventana:	Región de una secuencia nucleotídica que toma <i>Redigen</i> para compararla con otra secuencia y determinar el número de diferencias y el grado de identidad entre ambas.
cds:	Región de la secuencia de DNA que codifica para proteína.
Gen 'house keeping':	Gen expresado constitutivamente en la célula.
Oligonucleótidos (oligos):	Pequeñas secuencias de DNA (25 a 100 pb) utilizados para hibridar con RNA mensajero o DNA genómico para el reconocimiento de ciertas secuencias.

RESUMEN

La tecnología de microarreglos de ADN se ha convertido en una poderosa herramienta para monitorear la expresión génica en los organismos vivos, pues permite analizar cientos o miles de genes en forma simultánea con un gran nivel de reproducibilidad. A pesar de su utilidad, la técnica presenta limitaciones debido a que la expresión de muchos genes no puede ser cuantificada directamente a partir de la información entregada por los microarreglos, ya sea por factores que modifican la abundancia de un RNA transcrito (como el empalme alternativo de intrones) o por similitudes en la secuencia nucleotídica de algunos genes. Por esta razón muchas veces se confunde la expresión de genes estructuralmente relacionados (ortólogos, parálogos) o que comparten dominios comunes (genes que codifican proteínas funcionalmente relacionadas). En este trabajo, hemos abordado dicho problema técnico creando un método para encontrar secuencias nucleotídicas complementarias específicas (sondas). Esta Memoria de título propone un método para identificar regiones de baja identidad de un transcrito, que sean útiles como sondas de identificación o cuantificación en técnicas de biología molecular. Adicionalmente, se entregan antecedentes para diseñar sondas específicas para la identificación de una secuencia determinada dentro de un conjunto de genes involucrados en el metabolismo de cobre mediante ensayos de hibridación en microarreglos de cDNA. Con este propósito, se generó un programa (*Redigen*) que cuantifica el número de diferencias entre dos o más secuencias nucleotídicas para delimitar una región de baja identidad que permite buscar allí sondas específicas para evaluar la expresión de ese gen en un microarreglo. Para optimizar la detección de regiones de baja identidad, analizamos el efecto de variaciones en dos de los parámetros ajustables de *Redigen* en las curvas generadas por dos secuencias estructuralmente relacionadas (*atp7a* y *atp7b*). Estos parámetros fueron tamaño de ventana de comparación y de los porcentajes para clasificar el grado de identidad. Las comparaciones se realizan mediante fragmentos de una longitud que establece el usuario y que

denominamos 'tamaño de ventana'. Los porcentajes de identidad se calculan con respecto a la longitud del fragmento y se clasifican dentro de tres rangos que delimita el usuario con los parámetros j y k. A cada uno de estos grupos corresponde un valor en unidades arbitrarias (UA) que penaliza las identidades entre secuencias. De esta manera, mientras más similitudes tengan las secuencias, más alto será el valor en UA asignado. Para cada fragmento, se realiza la suma de los valores expresados en UA y se almacena en una planilla Excel. Esto permite graficarlos y visualizar el patrón de identidad entre las secuencias analizadas. Nuestros resultados indican que los parámetros porcentuales 50-70 generan curvas de fácil análisis y permiten descartar secuencias con más de un 50% de identidad. En cuanto al tamaño de la ventana, un largo de 300 pb tiene la ventaja de entregar una curva que conserva el patrón de identidad en un rango de variación de ± 50 pb, por lo que permite selección de fragmentos para amplificación mediante PCR de entre 250 y 350 pb. La presentación gráfica de *Redigen* ha sido diseñada de tal modo que nuestro método de identificación de regiones de baja identidad para el diseño de sondas específicas resulte accesible para cualquier laboratorio de investigación. Finalmente, se realizó algunas hibridaciones experimentales para evaluar la sensibilidad y especificidad de las sondas seleccionadas teóricamente con el programa Redigen. Los resultados de estos ensayos no permitieron evaluar adecuadamente el comportamiento de las sondas. Las posibles causas de ello se discuten en este trabajo.

ABSTRACT

DNA microarray technology is currently a powerful tool to determine gene expression in living organisms. It allows the simultaneous analysis of hundreds or thousands of genes with high reproducibility. In spite of its utility, this technique still presents limitations. Thus, gene expression of particular genes cannot be directly quantified from microarray data due to factors that change the abundance of a given transcript (e.g., alternative splicing) or to the occurrence of nucleotide sequence similarities between some genes. This explains why expression of structurally related genes (orthologs, paralogs) or genes sharing common domains are frequently confused. This work addressed such technical problem developing an identification method for specific probes. This research report proposes a method to identify low identity regions of a transcript in order to be used as identification/quantification probes in molecular biology assays. In addition, this study provides useful information for the design of specific probes to identify a given sequence within a set of copper-related genes using cDNA microarray hybridization assays. For this purpose we developed a software (*Redigen*) that quantifies the number of differences between two or more nucleotide sequences. To optimize the detection of low identity regions, we analyzed the effect of variation in two *Redigen* parameters – window size and identity classification percentages– on curve characteristics of two structurally related sequences (*atp7a* and *atp7b*). These parameters may be set by users. This defines a low identity region where to search for specific probes to evaluate the expression of a given gene in microarrays. Comparisons are made using sequence fragments of user-defined length. This fragment length is termed 'window size'. Identity percentages are calculated related to the fragment length and are classified into three groups defined by j – k parameters. Parameters j – k are also set by the user. Each group corresponds to a number of arbitrary units (AU) that penalizes sequence identities. Thus, the more identities the higher number of AU assigned. For each fragment, *Redigen* sums UA values and stores the outcome in an Excel spreadsheet, where a plot is automatically obtained. This allows an easy inspection of the identity pattern of the analyzed sequences. Our results indicate that percentage parameters 50–70 generate curves that can be

readily analyzed and allow to discard sequences showing more than 50% identity. With regard to window size, a 300 bp length has the advantage of rendering curves that maintain their aspect within ranges of ± 50 bp. This allows the selection of 250-350 bp fragments for PCR amplification. The graphical layout of *Redigen* is designed in such a way that our low identity region identification method for specific probe design can be shared with any research laboratory. Finally, we carried out some experimental hybridizations to test the sensitivity and specificity of *Redigen* selected probes. Results of these tests did not allow us to evaluate the probe performance properly. Possible causes of this outcome are discussed in this work.

1. INTRODUCCIÓN

En la interacción de los organismos con su ambiente es necesario que éstos reconozcan como señales los cambios de su entorno y realicen las modificaciones que permitan adaptarse a una nueva condición. Parte importante de la respuesta a estas señales se encuentra a nivel de los mecanismos que regulan la expresión génica y definen la vida media de sus productos (RNA y proteínas). Por esta razón, el estudio de dichos mecanismos de regulación puede ser de gran importancia en la comprensión de los cambios asociados a fenómenos de adaptación. En una célula, estas variaciones de estado pueden significar cambios tan importantes como diferenciación, división, motilidad, proliferación y muerte.

TECNOLOGÍA DE MICROARREGLO/MACROARREGLO: Una alternativa para estudiar cambios de expresión génica a escala genómica.

1.1. ESTADO DEL ARTE

Uno de los métodos tradicionales más utilizados en la determinación de cambios de expresión génica es el Northern blot (Alwine y col., 1977) que, al igual que el Dot blot (Kafatos y col., 1979), tiene la desventaja de permitir analizar solo un gen por vez. Otros métodos que permiten el análisis diferencial de expresión génica son: SAGE (Serial Analysis of Gene Expression), TOGA (Total Gene-expression Analysis) y análisis diferencial de mRNA mediante PCR (Polymerase Chain Reaction) (Velculescu y col., 1995; Liang y Pardee, 1992; Sutcliffe y col., 2000), cuya aplicación es limitada debido a que aún no permiten el análisis de un gran número de genes. Desde hace algunos años, se ha desarrollado una nueva tecnología, llamada microarreglo de DNA, que ha generado un gran interés entre los biólogos, pues permite a los investigadores evaluar la expresión de miles de genes activos en el genoma simultáneamente y tener

una mejor idea de las interacciones en y entre dos diferentes condiciones experimentales (Gross y col., 2000). Una guía para la definición de microarreglo es: 'arreglo microscópico y ordenado de ácidos nucleicos, proteínas, moléculas pequeñas, células u otras sustancias que permitan el análisis en paralelo de muestras bioquímicas complejas' (Scherer y col., 1995; 1998). Además de las ya conocidas secuencias de DNA, se ha sembrado tejidos congelados sobre matrices de vidrio (Schoenberg y Salmón, 2001) y se ha diseñado arreglos con células animales incubadas sobre pequeñas matrices de vidrio en las que previamente se ha sembrado cDNAs clonados en vectores de expresión (Ziauddin y Sabatini, 2001).

1.2. APLICACIONES DE LOS ARREGLOS

Hasta hoy, se ha descrito un sinnúmero de aplicaciones para los microarreglos, siendo una de las más utilizadas el análisis de fenómenos biológicos, a través de la identificación de grupos de genes con perfiles de expresión génica coordinada asociados en forma específica a dichos fenómenos. Esto es posible mediante el análisis comparado de las variaciones en los niveles de expresión génica a través del tiempo o entre distintas situaciones experimentales (Duggan y col., 1999; Iyer y col., 1999). Aplicaciones alternativas utilizando esta metodología son: secuenciación de cadenas nucleotídicas (Pease y col., 1994; Ju y col., 1995), establecimiento de líneas filogenéticas (McGonigle y col., 2000), detección de variaciones alélicas entre cepas de microorganismos (Lipshutz y col., 1999), clasificación molecular de diferentes tipos de cáncer (Golub y col., 1999) e identificación de polimorfismos en diferentes organismos (Jaccoud y col., 2001).

1.3. CLASIFICACIÓN DE LOS ARREGLOS

Dependiendo del diámetro del punto sembrado, los arreglos se clasifican en *macro* (>300 micrones) y *microarreglos* (<200 micrones). Es posible, además, clasificarlos según la longitud de la sonda fijada en la matriz (Ekins y Chu, 1999): arreglo de cDNA, si el fragmento inmovilizado tiene entre 300 y 5000 bp de longitud y arreglo de oligonucleótidos si se trata de sondas de 20 a 80 pb (McGall y col., 1996; Lipshutz y col., 1999). Si bien las sondas elaboradas *in situ* (Pease y col., 1994; Gao y col., 2001) son mucho más precisas para seleccionar su blanco en la muestra, la producción de fragmentos de cDNA tiene la ventaja de

ser bastante menos costosa y más flexible (Southern y col., 1999). Dicha flexibilidad no se refleja sólo en la cantidad de protocolos disponibles para sintetizar los fragmentos y juzgar su calidad, sino también en la posibilidad de seleccionar su secuencia (regiones específicas dentro de cada gen a evaluar) (Duggan y col., 1999), por lo que es posible obtenerlos en casi cualquier laboratorio. En general, los productos de PCR (fragmentos de DNA de 200 – 700 pb) o plasmidios, que contienen clonada la secuencia de interés, se imprimen sobre membranas de nylon y los oligonucleótidos (DNA de 25 – 100 bp de longitud) sobre matrices de vidrio.

TECNOLOGÍA DE MICROARREGLO/MACROARREGLO: Construcción y factores que afectan el ensayo de hibridación.

1.4. CONSTRUCCIÓN DE ARREGLOS DE DNA

Las etapas en el proceso de construcción de un arreglo de DNA son: 1) elección de la secuencia de DNA de genes sonda (el que será sembrado en la matriz) incluyendo las de genes que sirvan como referencia para determinar los cambios de expresión; 2) amplificación de ésta por PCR; 3) preparación del cDNA para siembra, así como del soporte físico (matriz) sobre las cuales se inmovilizará; 4) depósito o siembra de la sonda sobre la matriz, formando un punto y 5) fijación del cDNA a la matriz (Eisen y Brown, 1999).

1.5. FIJACIÓN DE LAS SONDAS A LA MATRIZ

En la mayoría de los arreglos las sondas son sembradas sobre las matrices. Las sondas utilizadas mayoritariamente son secuencias de DNA (Lashkari y col., 1997; McGonigle y col., 2000). Un factor importante a considerar en la siembra es la viscosidad de la muestra, ya que ésta puede afectar la cantidad de sonda depositada sobre la membrana. Mientras más diluida esté la muestra, menor será la dificultad para depositarla en la matriz de siembra.

1.6. OBTENCIÓN DE LAS MUESTRAS

En la preparación de la muestra, se utiliza protocolos ya establecidos para extracción de RNA total, purificación de RNA mensajero y/o síntesis de cDNA mono o doble hebra, que posteriormente hibridará con las sondas sembradas en la matriz del arreglo. La utilización de RNA total y poliadenilado ha sido exitosa en análisis de variaciones en los niveles de expresión génica (Eisen y Brown, 1999), sin embargo, la preparación de mRNA permite que aún los transcritos con un bajo número de copias (incluso menores al 0.01% del total) puedan ser detectados in vitro (Bertucci y col., 1999)

1.7. MARCACIÓN DE LAS MUESTRAS

Una vez obtenidas las muestras, son marcadas con trazadores que emiten fluorescencia, radioactividad o quimio-luminiscencia, siendo los dos primeros los métodos más usados. El uso de fluorescencia tiene la ventaja de permitir la comparación de dos situaciones experimentales en una misma matriz (vidrio) al marcar cada muestra con un fluoróforo distinto. Los radioisótopos, por su parte, tienen la ventaja de no alterar en absoluto la estructura de las moléculas y se utilizan principalmente en matrices de nylon. Por otra parte, el ruido (o background) generado en los ensayos de hibridación en membranas de nylon es generalmente menor al ruido producido en matrices de vidrio. Además, la sensibilidad de los arreglos vidrio-fluorescencia es cien veces menor que la de los arreglo nylon-radiomarcaje. Esto se debe principalmente a que en la hibridación de macroarreglos de nylon se utiliza un mayor volumen de muestra, lo que supone una mayor cantidad de molécula blanco disponible (Bertucci y col., 1999).

1.8. INTERFERENCIA ENTRE MOLÉCULAS SEMBRADAS

Se ha realizado algunos estudios de estabilidad de la doble hebra de DNA durante el ensayo de hibridación en relación con el largo y con el grado de complementariedad entre ellas. Éstos incluyen estudios fisicoquímicos (Maskos y Southern, 1993; Fotin y col., 1998) y diversas adaptaciones de los protocolos, tales como: optimización del largo de las sondas sembradas y amplificación de la señal emitida para disminuir la cantidad necesaria de RNA en la muestra (Schwarzacher y Heslop-Harrison, 2000). Los

estudios fisicoquímicos han permitido establecer que mientras menor sea la longitud de un fragmento de DNA, mayor será el efecto de un error de apareamiento (mismatch) sobre la estabilidad de la doble hebra (Fotin y col., 1998). En el caso de las moléculas en la muestra, si éstas son muy largas, podrían formar estructuras secundarias que dificulten la hibridación con las sondas, aunque este problema es menor al trabajar con blancos de cDNA y productos de PCR en lugar de moléculas de RNA. Además de las mentadas estructuras secundarias, si las moléculas sembradas son muy largas, podría generarse interferencia entre ellas dependiendo de la densidad de siembra (Southern y col., 1999).

La densidad de siembra en un área determinada dependerá del diámetro del punto y del límite de resolución de detección de los equipos de análisis. La cantidad a sembrar puede normalizarse por concentración o masa, dependiendo de la cantidad de sonda que se disponga y de la longitud de los fragmentos. Las moléculas sembradas en alta densidad pueden causar problemas de interferencia para el análisis al interactuar entre ellas, lo que disminuye la cantidad efectiva de sonda en el proceso de hibridación. En arreglos de oligonucleótidos, en que la sonda se ha fijado a la matriz por uno de sus extremos, esta interacción hace que los más cercanos a la matriz estén menos accesibles que los extremos más alejados de ella. En estos casos, esta interferencia estérica puede solucionarse agregando 'espaciadores' de distinta longitud entre la matriz y el extremo más cercano a ella, de manera que las moléculas se alejen de sus vecinas y de la superficie, facilitando de esta forma la interacción con su blanco (Southern y col., 1999).

TECNOLOGÍA DE MICROARREGLO/MACROARREGLO: Fuentes de error asociadas al ensayo de hibridación y su análisis

1.9. FUENTES DE ERROR EN EL PROCESO DE HIBRIDACIÓN

La capacidad de interpretar los resultados obtenidos en ensayos de hibridación en microarreglo requiere conocer el grado de similitud entre las secuencias presentes en la matriz. En teoría, genes que codifican proteínas de estructura similar o que contienen dominios estructurales comunes –por relación funcional o filogenética– pueden presentar hibridación cruzada en los ensayos de hibridación (Lyne y col., 2003). En este contexto, las variantes de empalme alternativo de intrones en eucariontes tienen un notable impacto en el análisis de los microarreglos (Yeakley y col., 2002), pues contribuyen a ignorar la estructura génica correcta si las sondas no representan las distintas variantes de corte y empalme y las diferentes moléculas aparecen juntas en el análisis. Sabiendo que los procesos de regulación de la expresión génica son estrechamente controlados, el conocer la secuencia de las sondas sembradas permitiría definir cuándo las sondas utilizadas representan las variantes de splicing, lográndose una información más real de lo que genera el cambio de expresión del gen que produce dicho(s) mensajero(s). De otra forma, sólo sería posible concluir que cambios de abundancia relativa del mensajero de un gen representarían alguna de las múltiples posibilidades de combinación de las distintas isoformas en las proteínas traducidas.

Cuando se busca una buena sonda para evaluar la expresión de un gen dentro de un grupo de genes relacionados en una determinada situación experimental, se vuelve necesario contar con una herramienta que nos permita hallar dominios particulares de su secuencia. Estos fragmentos de DNA dentro del gen de interés deben tener una secuencia que no genere hibridación cruzada bajo las condiciones de hibridación. Se ha reportado una baja incidencia con identidades menores al 50% (Rouillard y col., 2001), sin embargo con identidades superiores podría generarse hibridación cruzada principalmente entre secuencias de genes relacionados.

1.10. FUENTES DE ERROR EN EL ANÁLISIS POST EXPERIMENTAL

Uno de los principales problemas al momento de analizar los datos es el ruido producido por una señal inespecífica en la matriz de hibridación. Tanto las matrices de vidrio como las de nylon requieren extremo cuidado, pues impurezas, deshidrataciones, burbujas y desniveles en ella pueden incrementar el ruido, así como causar deformaciones en la forma del punto sembrado de cada cDNA y debilitamiento en señales locales (Jaccoud y col., 2001). El registro y clasificación de una señal como tal se vuelve importante durante la adquisición de datos, ya que es posible que la señal de genes poco representados sea clasificada como ruido, generando un error en la interpretación de los datos. No se ha reportado la influencia de la posición del punto en la superficie. Sin embargo, se ha recomendado al menos dos réplicas en una misma matriz al diseñar arreglos de cDNA, ya que una sola matriz está sujeta a una gran variabilidad y las réplicas disminuyen los errores de cuantificación de la expresión de los genes (Jenssen y col., 2002).

Se ha realizado numerosos estudios tendientes a optimizar estrategias de normalización y análisis computacional de los datos (Eickhoff y col., 1999; Epstein y Butow, 2000; Brazma y Vilo, 2000; Li y Wong, 2001; Khitrov, 2001; Schuchhardt y col., 2001). Se entiende como normalización la expresión de los valores de intensidad obtenidos en todos los puntos de una matriz en unidades comparables.

1.11. OTROS FACTORES QUE AFECTAN INTENSIDAD DE LA SEÑAL DE HIBRIDACIÓN

Cuando una de las moléculas a hibridar se encuentra fija a una matriz, son varios los factores que influyen en el nivel de hibridación que tendrá con su homólogo. Este nivel de hibridación se reflejará en la intensidad de la señal de las sondas sembradas.

Para comprobar que se tiene la sonda efectivamente sembrada en la matriz es necesario comprobar la unión efectiva de la sonda. Puede que el mecanismo utilizado para unir covalentemente el DNA a la membrana no haya sido efectivo o puede haber fragmentado las moléculas (por ejemplo, sobre exposición a luz UV daña el DNA). Si la sonda se fijó de forma eficiente, es necesario chequear que existe suficiente DNA sembrado en cada punto. Ya se ha mencionado la concentración de la molécula de interés en la muestra como un factor influyente en la intensidad de la señal de hibridación.

Las condiciones de hibridación también afectan el apareamiento entre hebras. La estrictez de las condiciones de hibridación puede determinarse variando en la solución la concentración de formamida (agente desnaturante) y la temperatura a la que se realiza el proceso. Mientras mayor sea el valor de estos parámetros, más exigentes serán las condiciones para el apareamiento de las moléculas homólogas. Aumentos en los tiempos de hibridación también aumentan la intensidad de la señal obtenida (Bertucci y col., 1999). Finalmente, condiciones de lavado muy fuertes (alta concentración de sales) pueden debilitar las uniones establecidas entre moléculas, lo que implicará una señal de menor intensidad para el análisis.

Una de las principales fuentes de error en los análisis de microarreglos de DNA es la similitud entre las secuencias presentes en la muestra que se analiza, debido a relaciones funcionales o filogenéticas o a las variantes de empalme alternativo de intrones generadas durante el procesamiento del RNA mensajero. Una propuesta interesante para disminuir la generación de falsos positivos y aumentar la sensibilidad en la detección, sería el desarrollo de un protocolo que considere la utilización de regiones génicas que posean baja identidad con otras secuencias presentes en la muestra. Esto permitiría el estudio de las variaciones de expresión génica en un sinnúmero de rutas metabólicas. Considerando que el largo de la secuencia es un factor clave en la estabilidad de la hibridación, la utilización de los dominios de baja identidad requiere el establecer un compromiso entre la similitud estructural y la longitud de cada gen para optimizar el grado de especificidad en el reconocimiento de su blanco. La utilización de fragmentos de cDNA como sondas sembradas en membranas de nylon puede ser una buena alternativa dentro de las posibles soluciones al problema. Entre las ventajas de la utilización de estas secuencias en el proceso de hibridación en macroarreglos se cuentan: la posibilidad de elegir la secuencia de los genes de interés, su fácil obtención en laboratorio (incluyendo purificación y certificación de calidad) y la longitud de los fragmentos, que hace más estable la doble hebra una vez que la sonda ha reconocido su blanco en la muestra.

1.12. PROGRAMAS DE ANÁLISIS DE IDENTIDAD ENTRE SECUENCIAS

En general, los programas desarrollados para la comparación de secuencias (de nucleótidos o de aminoácidos) persiguen establecer en forma cuantitativa las diferencias y/o el grado de similitud entre ellas. Dos de los parámetros más influyentes en la elección de dichos programas son la sensibilidad en la detección y el tiempo que toma el proceso.

Los métodos basados en programación dinámica implican una comparación nucleótido a nucleótido. Los métodos exactos exploran el espacio de soluciones buscando una solución óptima que conlleva tiempos de ejecución muy largos. La elección del óptimo en estos modelos se toma en etapas múltiples, con un total de n etapas. La primera etapa corresponde a la fijación de la variable x_1 , sobre la que se ejecutarán las siguientes etapas. Sus resultados son exactos, pero mucho más lentos y requieren un tiempo de cálculo proporcional al cuadrado de la longitud (número de pares de bases) de las secuencias que se comparan.

En cambio, los métodos de tipo heurístico (aproximación de longitud de fragmentos) privilegian la velocidad a cambio de sensibilidad y selectividad. En estos modelos, el utilizar algoritmos paralelos permite mejorar la relación calidad:tiempo de las soluciones obtenidas. Dichas soluciones pueden ser obtenidas de diversas técnicas: buscando en los alrededores de una solución inicial tomada al azar (búsqueda local), dentro de una población inicial de soluciones (algoritmos genéticos) o determinando valores a partir de gráficas (métodos espectrales), siempre y cuando respeten las restricciones del problema. Utilizando métodos heurísticos, la revisión completa de secuencias requiere de un tiempo equivalente sólo a la suma de las longitudes de las dos secuencias comparadas.

Algunos programas que utilizan métodos heurísticos en la búsqueda de sus soluciones son FASTA y BLAST. Éstos buscan fragmentos que permiten identificar diagonales con el mayor número de identidades que luego analiza utilizando un sistema de puntuación (Lipman y Pearson 1985, Pearson y Lipman 1988). Es decir, buscará las regiones más parecidas entre dos secuencias sin importar que el mejor alineamiento corresponda solo a un fragmento de la secuencia completa y permitiendo el desplazamiento de la posición relativa de las regiones alineadas entre secuencias.

En forma similar proceden los programas que permiten comparar simultáneamente múltiples secuencias, no obstante, la complejidad aumenta exponencialmente con el número de secuencias involucradas en el proceso. Esta complejidad está dada por la dificultad de alinear uno a uno una gran cantidad de nucleótidos que supone un grupo de secuencias, por el gran número de iteraciones y matrices que sería necesario realizar para cuantificar su similitud y, también, por el tiempo que tomaría dicho proceso. Una solución clásica para reducir los tiempos de alineamiento de múltiples secuencias, consiste en formar 'clusters' o grupos de secuencias en la fase inicial del proceso, que luego se resuelven progresivamente (Corpet, 1988). La formación de estos clusters no tiene un criterio establecido, muchas veces obedece a la necesidad de reducir el número de secuencias a comparar, para hacer más rápida la comparación y obtención de una secuencia consenso entre ellas (resolución). De esta forma una vez definido un cluster de secuencias, éste será tratado como una sola secuencia. Uno de los programas más conocidos para realizar esta comparación es ClustalW. Métodos alternativos para el alineamiento múltiple de secuencias utilizan 'matrices consenso' para identificar la preferencia de cada aminoácido o de cada deleción en la secuencia alineada (Alexandrov, 1992), o incorporan información sobre: a) las estructuras secundarias (Smith y Smith, 1992; Godzik y Skolnick, 1994) o b) las características fisicoquímicas de los aminoácidos en cada posición (Livingstone y Barton, 1993).

Una aproximación complementaria utilizada en la búsqueda y comparación de secuencias homólogas es la construcción de bases de datos que contienen motivos conservados entre secuencias de proteínas (Boguski y col., 1992). Estas bases de datos permiten el hallazgo de secuencias de interés a pesar de tener una similitud global muy baja, las cuales normalmente no son detectadas con los métodos que utilizan aproximaciones heurísticas como BLAST y FASTA. La idea que sustenta esta aproximación es que aquellas regiones que codifican dominios que definen la función molecular de una proteína (sitio activo de enzimas, dominios de interacción proteína-proteína, dominios de unión a metales, entre otras) sufren fuertes restricciones estructurales a lo largo de la evolución, de manera que secuencias que tienen un origen común

pueden ser reconocidas mediante análisis de motivos y permiten tener una idea de la función que una proteína desconocida tiene en la célula.

La existencia de dominios comunes entre secuencias con baja homología global puede, de esta manera, promover la hibridación cruzada entre ellas, lo cual introduce un factor que afecta la interpretación de los resultados obtenidos en estudios de expresión génica utilizando microarreglos/macroarreglos de DNA.

1.13. SISTEMA DE ESTUDIO

La descripción del programa diseñado en esta Memoria se realizó con los genes de uno de los sistemas estudiados en el Laboratorio de Bioinformática y Expresión Génica del INTA, Universidad de Chile. Estos genes están relacionados con el metabolismo celular de cobre. Dentro de este conjunto, las secuencias *atp7a* y *atp7b* tienen un alto grado de identidad y ambas codifican ATP-ases tipo-P.

En los sistemas biológicos, el cobre participa de la catálisis oxidativa facilitando la transferencia de electrones en varias enzimas que lo utilizan como cofactor, teniendo un rol esencial para muchos procesos celulares tales como respiración celular, respuesta al estrés oxidativo –que él mismo puede generar– (Uauy y col., 1998) y biosíntesis de neurotransmisores. Los organismos eucariontes, requieren de una estricta regulación en la homeostasis del cobre (procesos de captación, acumulación –cantidad y lugar– y eliminación) dado que, si bien es esencial, un exceso del metal resulta altamente tóxico.

En *Saccharomyces cerevisiae* se ha descrito los genes involucrados en dichos procesos y estudios realizados en este modelo señalan que cambios en la expresión génica son fundamentales para mantener la homeostasis de cobre (Winge y col., 1998; Labbé y Thiele, 1999) y que es el mismo metal el que regula la expresión de estos genes a través de activadores transcripcionales que responden específicamente a su déficit y exceso (Labbé y Thiele, 1999). En mamíferos, las células cuentan con diversos mecanismos fisiológicos destinados a mantener el suplemento adecuado de cobre y controlar su exceso. En la Figura 1 se esquematiza la ruta metabólica que da cuenta de la homeostasis de cobre en mamíferos.

En células de mamíferos, el mecanismo celular de captura de cobre incluye los transportadores de Cu^{+1} y 2 (Ctr1/2) (Lee y col., 2002) y el transportador de metales divalentes-1 Nramp (Gunshin y col. 1997). En el

citoplasma, el cobre libre existe raramente (Rae y col., 1999) porque es transferido inmediatamente a las chaperonas de cobre, como Atox1, Cox17 y Ccs (Harrison y col., 2000, Mercer, 2001), glutatión (GSH) o metalotioneínas (MT) (Mattie y Friedman, 2004). La última corresponde a la mayor familia de proteínas citosólicas que unen cobre reversiblemente para distribución intracelular y almacenamiento. Además, su expresión es sensible a los cambios en los niveles de concentración de Cu^+ , aumentando en respuesta a la exposición a Cu^{++} extracelular (Palmiter, 1998, Tapia y col., 2004), y de esta manera provee vías específicas seguras y eficientes para el almacenamiento intracelular y el transporte de este metal. Finalmente, dos ATPasas tipo-P: Atp7a (proteína Menkes) es expresada por tejidos extra-hepáticos (Mercer y col., 1993, Vulpe y col., 1993) y Atp7b (proteína Wilson) expresada principalmente en el hígado (Bull y col., 1993, Yamaguchi y col., 1993). Atp7b se localiza en la red trans-Golgi y provee de las bases moleculares para la excreción del metal. Ambas son posibles de trasladar a la membrana celular en condiciones de exceso de cobre (Mercer y col., 2003, Petris y col., 1993, Pascale, y col., 2003, Cater y col., 2004). Sin embargo, en condiciones normales de niveles de cobre estas ATPasas cargan ceruloplasmina con cobre y este cobre unido a ceruloplasmina es secretado a través de transporte vesicular (Linder y col., 1998, Cater y col., 2004).

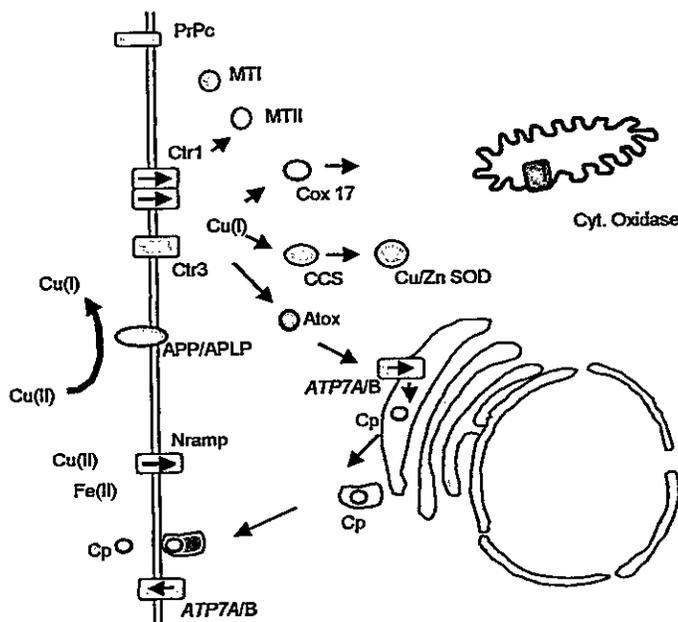


Figura 1: Participación de proteínas involucradas en el metabolismo de cobre en mamíferos. Subconjunto de proteínas que interactúan con el ión metálico.

HIPÓTESIS

La selección de una región en un gen particular que tenga un bajo nivel de identidad con el resto de las secuencias de un grupo de genes metabólica o filogenéticamente relacionados, permitirá la utilización de dicha región como sonda específica para ese gen.

OBJETIVO GENERAL

En atención a lo expuesto, el objetivo general de este trabajo es la creación de un software que permita encontrar regiones de baja identidad en cada una de las secuencias de un grupo de genes (filogenética o funcionalmente relacionados) para ser utilizadas como sondas específicas en un macroarreglo. En este caso, se seleccionó un grupo de genes que codifican proteínas que participan en el metabolismo celular de cobre (relación funcional dentro del grupo y estructural entre las ATPasas).

OBJETIVOS ESPECÍFICOS

1. Diseñar un algoritmo y crear un programa que realice la búsqueda de regiones de baja identidad dentro de un gen al compararlo con un grupo de genes previamente seleccionados, para ser utilizadas como sondas en un microarreglo.
2. Describir el efecto de los cambios en los parámetros de búsqueda del programa en pares y grupos de secuencias (caracterización).
3. Evaluar las funciones alternativas para la búsqueda de secuencias nucleotídicas con dicho programa.
4. Seleccionar los partidores que permitan amplificar fragmentos relativamente pequeños dentro de las regiones de baja identidad.
5. Confeccionar una biblioteca con los fragmentos amplificados.
6. Evaluar la expresión de un grupo de genes relacionados con el metabolismo celular de cobre utilizando las sondas obtenidas.

ESTRATEGIA EXPERIMENTAL

Con este propósito, realizamos las siguientes actividades:

1. Diseño y caracterización de un programa computacional que permita cuantificar el nivel de similitud de regiones génicas discretas (100–300 bp) que posean un bajo nivel de identidad con la secuencia de otros genes.
2. Establecimiento de un protocolo de uso del programa, evaluando los efectos de los cambios en los parámetros establecidos.
3. Descripción de usos alternativos del programa diseñado para el análisis de secuencias génicas.
4. Diseño de partidores, clonamiento de los fragmentos seleccionados y creación de una biblioteca de cDNA con dichos fragmentos.
5. Estandarización de un protocolo de hibridación que permita el reconocimiento entre los fragmentos utilizados como sonda y su blanco en una muestra compleja.
6. Utilización de las sondas generadas en la evaluación de expresión de genes relacionados con el metabolismo celular de cobre.

Dado que es una situación común a varias líneas de investigación, abordamos los objetivos propuestos seleccionando un grupo de genes vinculados en un problema biológico, pues resulta útil contar con una herramienta que permita analizar de manera confiable la expresión de genes con algún grado de similitud en su secuencia en un mismo experimento.

2. MATERIALES Y MÉTODOS

I. MATERIALES

El reactivo X-gal 50 mg/ml fue elaborado por Sigma. Los partidores fueron sintetizados por BiosChile. En todos los ensayos se utilizó una concentración de 50 pmol/ μ l excepto en el PCR anclado en el cual se usó el partidior bsWs a una concentración de 5 pmol/ μ l. La enzima EcoRI 10 U/ μ l se adquirió en Fermentas. La enzima transcriptasa reversa 200 U/ μ l pertenece a Gibco. El DNA de espermatozoos de salmón 10 mg/ml, el partidior oligo dT 1 μ g/ml, la agarosa, el bromuro de etidio (EtBr) 10 mg/ml, el kit Random Primer, los reactivos para la amplificación por PCR (Buffer 10x, MgCl₂ 50 mM, dNTP's 10 mM y Taq DNA polimerasa 5 U/ μ l) y el buffer Prehybridation / Hyb 2x para la preparación del buffer de prehibridación / hibridación fueron adquiridos en Gibco. El kit RNawiz™, pertenece a Ambion. Los kits PolyATtract mRNA para la extracción de mRNA, Wizard® DNA Clean-Up System utilizado en la limpieza del DNA marcado previo a la hibridación con las membranas fueron encargados a Promega. En la recuperación o resuspensión de DNA se utilizó buffer TE (Tris-EDTA) 8.0 (1 mM Tris pH 8.0 y 1 mM EDTA pH 8.0). El nucleótido marcado α -³²-dCTP fue proporcionado por la Comisión Chilena de Energía Nuclear. Para la medición de radiactividad, las muestras fueron resuspendidas en líquido de centelleo EcoScint. La librería de cDNA de ratón MT3 fue solicitada al Centro Alemán de Recursos para la Investigación Genómica RZPD, Berlín (www.rzpd.de).

En el protocolo para extracción de DNA plasmidial (Miniprep I, II, III) se utilizaron las siguientes soluciones: 'solución I' (50 mM Tris pH 8.0, 10 mM EDTA pH 8.0), 'solución II' (1% SDS, 0.2 M NaOH), 'solución III' (5 M KAc, HAc) y una solución de fenol:cloroformo:isoamilalcohol 25:24:1. Para la precipitación de DNA se utilizó EtOH absoluto, 4M LiCl y/o 3M AcNa. La solución usada para desnaturar las sondas recién sembradas estaba compuesta de 0.5 M NaOH y 1.5 M NaCl, mientras que la solución de

neutralización estaba compuesta por 1.5 NaCl, 0.5 Tris pH 7.4 y 1 mM EDTA pH 8.0. La solución de prehibridación / hibridación de las membranas estaba compuesta por buffer de Prehibridación/hibridación 1x, 20 mM buffer fosfato de sodio pH 6.5, 40% formamida y 1% SDS. Las soluciones de lavado de las membranas ya hibridadas fueron: 'solución I' (2x SSC), 'solución II' (2x SSC, 0.1% SDS), 'solución III' (1x SSC, 0.1% SDS) y 'solución IV' (0.1% SSC).

Los reactivos de cultivo celular fueron obtenidos en Gibco. La línea Caco2 (células de epitelio intestinal de ratón) fue cedida por Miguel Arredondo, del Laboratorio de Fisiología de Membranas, Facultad de Ciencias, Universidad de Chile; y la línea HepG2 (células hepáticas de ratón), fue proporcionada por la Dra. Ana María Pino, Laboratorio de Hormonas y Receptores, INTA, Universidad de Chile. Las líneas 805 (fibroblastos de ratón) y 804 (fibroblastos de ratón con la expresión suprimida de las isoformas I y II de metalotioneína) se utilizan habitualmente en el Laboratorio de Bioinformática y Expresión Génica.

La membrana de nylon (amina cuaternaria en alta densidad) Zeta-probe Blotting membranes es de BioRad. Los líquidos de revelado y fijado de señal en las placas radiactivas son Kodak al igual que las placas radiográficas BioMax™ Film ML. Las reacciones de amplificación se realizaron en un termociclador PTC – 100™ (MJ Research Inc). El transiluminador CL-1000 Ultraviolet Crosslinker (UVP) y el horno de hibridación Smart Check™ pertenecían, respectivamente a Vilber-Lourmat y Labnet 211 DHS.

El programa utilizado en la captación de las imágenes y cuantificación de las señales por densitometría es el Kodak Digital Science 1D (Kodak). El código de fuente del programa de procesamiento de señales (extracción de la señal neta para cada clon, eliminando el ruido) fue ejecutado por Pablo Moreno, Ingeniería Civil en Biotecnología, Universidad de Chile.

II. MÉTODOS

2.1. DISEÑO DE UN PROGRAMA DE BÚSQUEDA DE DOMINIOS DE BAJA IDENTIDAD

REQUERIMIENTOS

En este trabajo de investigación la búsqueda de dominios de baja identidad se realizó en las secuencias derivadas de un grupo de genes en atención al nivel de similitud entre ellas. En la web es posible hallar programas de comparación de secuencias con acceso gratuito, como BLAST y ClustalW, sin embargo y dadas las necesidades más frecuentes de los usuarios (estimar relaciones funcionales o filogenéticas entre genes), estos programas privilegian las similitudes entre las secuencias al realizar los alineamientos, y no incorporan en sus algoritmos de búsqueda la detección de regiones que presenten escasa similitud al ser comparadas con otras secuencias. Por lo tanto, nuestro objetivo se centró en diseñar un sistema de búsqueda capaz de cuantificar las diferencias entre secuencias e identificar en qué regiones de las mismas se hallaban concentradas estas diferencias.

2.2. CARACTERÍSTICAS DEL PROGRAMA

El programa Redigen (Regiones de Identidad Génica), fue diseñado en conjunto con Felipe Hoffa (Ingeniería Civil en Computación, Universidad de Chile) para comparar dos secuencias nucleotídicas y cuantificar el número de diferencias que exista entre ellas. El programa Redigen está escrito en Java 1.4 y funciona como aplicación en cualquier plataforma que tenga disponibles las librerías Java. Como mínimo requiere un procesador Pentium 2 con 128 Mb de memoria RAM y 300 Mb de disco duro. Los límites del programa no fueron explorados, ya que soportaba sin inconvenientes todos los procedimientos ejecutados. Teóricamente los límites dependen más que nada de las capacidades de la máquina, que progresan constantemente.

El programa opera de la siguiente forma: Sean A y B dos secuencias nucleótídicas en dirección 5'-3'. De la secuencia B se toma un fragmento de longitud establecida por el usuario y que se inicia en el primer nucleótido de ella. Para referirnos al largo de este fragmento al momento de ejecutar el programa, decimos tamaño de ventana de comparación. El nucleótido central de este fragmento se alinea con el primer nucleótido de la secuencia A y se registra el número de diferencias. Luego, se desplaza una posición hacia la derecha sobre la secuencia A y cuenta el número de bases diferentes. Así, hasta que el nucleótido central de este fragmento alcance el último nucleótido de la secuencia A. En este momento, toma la secuencia complementaria del fragmento de la secuencia B y se invierte la secuencia para poder realizar la comparación en dirección 5'-3'. Luego, se ejecuta la comparación a través de la secuencia A tal como se ha descrito (Figura 2). El número de diferencias encontradas en cada posición de la secuencia A para las hebras sentido y antisentido del fragmento se ubican en una matriz. Esta matriz tiene tantas columnas como el número de pares de bases la secuencia A. El número de filas está dado por dos veces el número de pares de bases de la secuencia B menos el tamaño de ventana. La comparación se inicia y termina en el nucleótido central de la ventana porque, eventualmente, medio fragmento podría tener homología suficiente para estabilizar el alineamiento y generar falsos positivos por hibridación cruzada.

Para diferenciar fácilmente las regiones de interés del usuario, el programa calcula el porcentaje de identidad con respecto al fragmento que significa el número de diferencias encontrado. Estos valores porcentuales son clasificados dentro de tres grupos cuyos límites se fijan con los parámetros j y k. A cada intervalo se le asigna un valor numérico en unidades arbitrarias (UA) dependiendo del nivel de identidad de los fragmentos. Entre 0 y j % de diferencias se asignan 1000 UA, entre j y k % de diferencias se le asigna 1 UA y entre k y 100 % de diferencias recibe 0 (cero) UA. Por ejemplo, para menos de 50% de diferencias se asigna 1000 UA, ya que ese fragmento de B tiene más de un 50% de identidad con la secuencia A en esa posición. De usar ese fragmento como sonda, éste podría hibridizar de manera estable con la secuencia A en un ensayo de hibridación, que es lo que se busca evitar. Los valores ya transformados en unidades arbitrarias se suman y ese valor final es entregado en un archivo Excel asociado a la posición a la que

corresponde ese valor. La representación gráfica de los valores expresados en UA la llamaremos curvas de identidad.

Debido a los puntajes arbitrarios asignados, las regiones de baja identidad quedan representadas en los gráficos como valles en la curva, mientras que las regiones de alta identidad, están representadas por crestas. Cada punto en la curva, corresponde a la posición del nucleótido en que comienza la comparación con el fragmento de la otra secuencia.

pancreática de ratón y nodulina de *A. thaliana* usando ventanas de 150 y 300 pb con parámetros porcentuales 50-70 (el efecto del tamaño de ventana y de los parámetros porcentuales j-k se observa más adelante).

- Efecto del tamaño de ventana: Dada la necesidad de un rango de soluciones dentro de los cuales seleccionar los mejores partidores posibles, se evaluó el cambio en los patrones gráficos al cambiar el tamaño de la ventana. Para el par de secuencias *atp7a-atp7b* se utilizó ventanas de 50, 100, 150, 200, 245, 250, 255, 300, 350, 400, 500 y 600pb. Para comparar el conjunto de genes relacionados metabólicamente (incluidas *atp7a-atp7b*) se utilizó ventanas de 50, 100, 300 y 600pb.
- Forma de las curvas de identidad al comparar genes con un alto grado de homología: Considerando su estrecha relación filogenética, se analizó los patrones de la comparación de las secuencias completas (incluyendo 5' y 3'-UTR) y las regiones codificantes (cds) del par *atp7a-atp7b* y se evaluó el desplazamiento entre las curvas. Al comparar ambas secuencias completas éstas mostraban formas similares aunque aparecían ligeramente 'desplazadas' sobre el eje x (Figura 3). Ese desplazamiento motivó la comparación de las secuencias codificantes de *atp7a* y *atp7b*. Este cotejo se realizó con ventanas de 150 y 300pb y con los parámetros porcentuales 30-40, 50-70 y 60-80. Para cuantificar el desplazamiento de las curvas de identidad generadas se anotó la posición del nucleótido en que se produce el primer cambio en la forma de ellas (ver Tabla 3, Resultados).

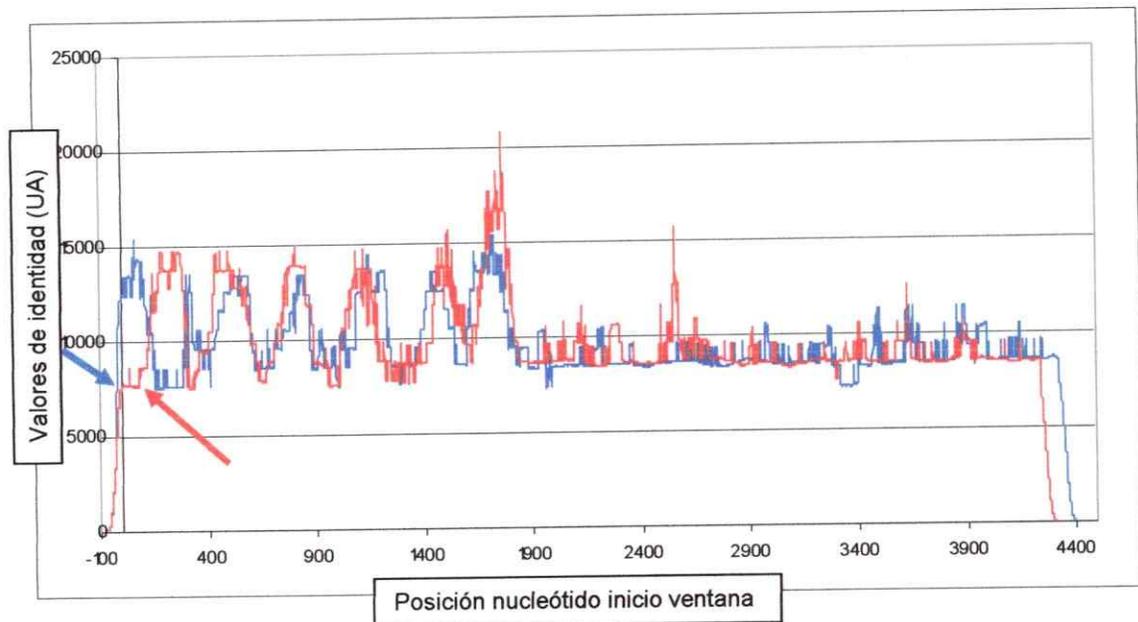


Figura 3: Desplazamiento de las curvas generadas por el programa Redigen al comparar las secuencias de *atp7a* y *atp7b* entre sí. Los nucleótidos en que se produce la primera variación en las curvas de identidad están indicados por flechas. La curva roja representa la comparación de secuencias con *atp7a* fija y la línea roja con la secuencia *atp7b* fija. La comparación de estas secuencias se realizó con un tamaño de ventana de 250 pb y parámetros porcentuales 60-80.

- Efecto de parámetros porcentuales: Con el fin de tener en la secuencia una región de mayor amplitud en la que encontrar sondas para las regiones de baja identidad, se evaluó el efecto en la forma de las curvas al utilizar distintos valores para los parámetros porcentuales j-k (30-40, 50-70 y 60-80). Se comparó las secuencias de la *atp7a* y *atp7b* alternando el orden de ingreso al programa Redigen, y con ventanas de 100 y 600 pb.
- Efecto del número de secuencias ingresadas a la comparación: Se comparó el efecto en las curvas de identidad al ingresar al programa una secuencia con sólo otra como contraparte o con un grupo de secuencias.
- Similitudes entre secuencias de un grupo de genes metabólicamente relacionados: Para evaluar la similitud entre secuencias de este conjunto de genes, se comparó tres grupos que incluían todos los genes, pero que cambiaban la secuencia de la cual se tomaba la ventana (*sod*, *atp7b* y *atox*), con ventanas de 50, 100, 300 y 600 pb y parámetros 30-40, 50-70 y 60-80.

2.4. SELECCIÓN DE CLONES

2.4.1. Selección de la región a amplificar

Para un par de genes (*atp7a* y *atp7b*) se comparó las secuencias completas de ellos, incluyendo las regiones 5' y 3'-UTR para encontrar las regiones de alta y baja identidad. Para la comparación se utilizó el programa Redigen con una ventana de 250 pb y un rango porcentual entre 50 y 70%. En los resultados gráficos, se escogió las zonas más amplias de baja identidad para tener una región más extensa en la secuencia donde seleccionar un buen par de partidores específicos.

Para un grupo de doce genes metabólicamente relacionados, la elección de las regiones de baja identidad se realizó ingresando al programa Redigen las secuencias completas (incluyendo las regiones 5' y 3'-UTR) de cada uno de ellos (Tabla 1). En las curvas se escogió como regiones de baja identidad aquellas en que todas las curvas tuvieran valores inferiores a 1000 UA.

Genes involucrados en metabolismo celular de cobre utilizados en este trabajo.

Gen	ID GeneBank	Largo gen (bp)	Largo sonda (bp) ^(a)	Proteína	Especie
<i>atp7a</i>	AB007134	4575	202(aMk)-293(bMk)	Transportadora-Cu tipo-P ATPasa	Mus musculus
<i>atp7b</i>	NM_007511	4854	207(aWs)-262(bWs)	Transportadora-Cu tipo-P ATPasa	Mus musculus
<i>ccs</i>	AF173379	1052	220(b-ccs)	Chaperona de Cu para superóxido dismutasa	Mus musculus
<i>sod1</i>	M35725	535	261(b-sod)	Cu-Zn superóxido dismutasa	Mus musculus
<i>app</i>	NM_000484	3058	200(b-app)	Precursor del beta amiloide	Mus musculus
<i>mt1</i>	J00605	391	134(b-mt)	Metalotioneína-I	Mus musculus
<i>nramp</i>	AF029758	2261	289(b-nramp)	Resistencia natural assoc. prot-2 macrófago	Mus musculus
<i>rCTR</i>	AF268030	1423	258(b-rctr)	Transportador Cu 1	Rattus norvegicus
<i>cp</i>	U49430	3363	269(b-cp)	Ceruloplasmina	Mus musculus
<i>atox1</i>	AF004591	432	251(b-atox)	Chaperona Cu para <i>Atp7a</i>	Mus musculus
<i>prpc</i>	NM_011170	2153	206(b-prpc)	Proteína priónica	Mus musculus
<i>act</i>	NM_007393	1892	522	Melanoma x-actina	Mus musculus
<i>tub</i>	NM_011655	1591	269	Tubulina, beta 5	Mus musculus
<i>g3p</i>	M32599	1228(mRNA)	452	Gliceraldehído-3-fosfato deshidrogenasa	Mus musculus

Tabla 1: La tabla lista algunos de los genes involucrados en el metabolismo celular de cobre utilizados como sonda en el presente trabajo, señalando el código de acceso GeneBank, nombre del gen y longitud del gen, longitud de la sonda y nombre de la proteína. Entre paréntesis (°) se indica el nombre de la sonda para ese gen, seleccionada utilizando el programa Redigen.

Para los tres genes 'house keeping' (o de expresión constitutiva) –tubulina, actina y gliceraldehído fosfato deshidrogenasa– no se seleccionó las secuencias en función de su nivel de identidad, debido a que en nuestro estudio no se deseaba evaluar su identidad sino utilizarlos en la normalización de los valores obtenidos por la hibridación de los genes en estudio. Al momento de la comparación no se tenían antecedentes de la abundancia de los transcritos de estos genes ni si existían variaciones en su expresión en las condiciones experimentales como sugieren Eisenberg y Levanon (2003). Sin embargo, aún sin tener esta información, se decidió utilizarlos en la normalización del nivel de expresión considerando que estos genes no están involucrados en el metabolismo celular de cobre. Se realizó una comparación de estas secuencias con la del grupo de genes en estudio con el programa Redigen para descartar hibridación cruzada con las moléculas de nuestro interés. El grupo completo de genes sometidos a este estudio se muestra en la Tabla 1.

2.4.2. Verificación del nivel de identidad de las regiones definidas por el programa Redigen.

Se evaluó las regiones de alta y baja identidad previamente seleccionadas entre cada una de las secuencias de las ATPasas (*atp7a* y *atp7b*), utilizando otros programas de alineamiento de secuencias disponibles en red: BLAST y ClustalW. Esta comparación es sólo aproximada, puesto que las secuencias ingresadas al análisis fueron de mayor longitud que aquellas que serán sembradas en las membranas, pero nos dio una idea de la similitud entre las secuencias antes de comenzar con el diseño de los partidores. Las regiones de alta y baja identidad se obtuvieron de curvas generadas con ventanas de 250 pb.

Sabiendo que cada punto en la curva corresponde a la posición del nucleótido en que comienza la ventana que se ha comparado, se tomó una secuencia desde el nucleótido de inicio del valle o cresta hasta el final de éste y se sumó 250 pb (puesto que la ventana que allí se inicia tiene esa longitud). Esta secuencia fue ingresada a los programas BLAST y ClustalW junto con las secuencias completas sentido y antisentido de la otra ATPasa, separadamente. Dado que deseamos estimar la calidad de la hibridación de ambas secuencias, en ambos programas se ajustó los parámetros de análisis de tal forma que la apertura y extensión de gaps estuviese fuertemente sancionada.

El programa ClustalW no entrega resultados porcentuales, sino que ubica un asterisco cada vez que encuentra una identidad entre secuencias de DNA. Para estimar el porcentaje de identidad obtenido en las comparaciones con ClustalW, se contó el número de asteriscos y se dividió este valor por el tamaño del fragmento ingresado. En todas las comparaciones realizadas con este programa se penalizó la apertura y extensión de gaps.

Para el conjunto de genes relacionados con el metabolismo celular del cobre, se procedió de la misma forma, ingresando a los programas el fragmento de la región de baja identidad en pareja con cada una de las secuencias sentido y antisentido del conjunto.

2.4.3. Verificación de regiones de alta/baja identidad

Utilizando el programa ClustalW, se comparó la secuencia delimitada por los partidores seleccionados para cada ATPasa con las secuencias sentido y antisentido de la otra ATPasa y se estableció el porcentaje de identidad con la relación 'número de asteriscos/longitud de ventana'. Este análisis se realizó para el resto de los genes del grupo y en la comparación se incluyó a las secuencias de los genes 'house keeping'.

2.5. DISEÑO DE LOS PARTIDORES

Para obtener los partidores de las ATPasas, se ingresó la secuencia completa de cada una de ellas (incluidos los extremos 5' y 3' no codificantes), en el programa Primer Premier 5.0. Dos de los parámetros fueron ajustados antes de la búsqueda: longitud de los productos (200 – 300 pb) y de los oligos generados (20 ± 2 pb) y se dejó los demás parámetros sin alterar, según los propone por defecto el programa. En la elección de los partidores de los otros genes relacionados con el metabolismo celular de cobre se procedió de la misma forma.

2.6. AMPLIFICACIÓN POR PCR DE LOS FRAGMENTOS

a) *atp7a* (Menkes)

Una vez definidos los partidores, los fragmentos de alta y baja identidad de *atp7a* fueron amplificados mediante PCR. Se utilizó el fragmento g3p (gliceraldehído-3-fostato deshidrogenasa) como control positivo de amplificación. Las reacciones de amplificación por PCR fueron realizadas en el termociclador PTC-100 (MJ Research). Los tubos contenían 1.5 µg cDNA de una línea silvestre de fibroblastos de ratón, 1x PCR Buffer (Invitrogen), 1,5mM MgCl₂, 2.5 mM dNTPs y 2.5 U de Taq DNA polimerasa en un volumen final de 50µl. Se utilizó un programa de 30 ciclos (desnaturalización 30 segundos a 95°C, alineamiento 30 segundos a 55°C y extensión 3 minutos a 68°C) con una extensión final a 68°C por diez minutos.

Se verificó el tamaño de los productos de PCR corriendo una electroforesis en un gel de agarosa al 1% y 1µg/µl EtBr en buffer TAE (Tris, ácido acético, EDTA) a 100mV.

b) *atp7b* (Wilson)

Para obtener los fragmentos de *atp7b* se intentó una amplificación con PCR anclado desde cDNA de hígado de ratón (órgano en que se expresa mayoritariamente) utilizando el fragmento g3p como control positivo. El PCR anclado (o 'single-sided') permite la amplificación específica de DNA cuando la región 5' de la molécula de interés es desconocida. En este caso, aunque se conoce la esta región, se realizó un PCR anclado para aumentar la representación de la región de la que se extraería el producto de PCR, pues no se consiguió amplificar el fragmento con los partidores directamente a partir de cDNA.

Como sustrato para el PCR anclado se utilizó el 10% del producto de una transcripción reversa. Para la transcripción reversa se desnaturó 100ng de mRNA de hígado de ratón y 20 µg/ml oligo dT₁₂₋₁₈ a 70°C por 10 minutos. Luego se agregó First Strand Buffer 1x, DTT 0.01M y dNTPs 0.5mM y se puso a 42°C por dos minutos antes de agregar 200 U de Superscript RT (Invitrogen). La reacción se efectuó en un baño a 42°C por 50 minutos y se detuvo neutralizando la enzima a 70°C por 15 minutos.

El cDNA obtenido de la reacción anterior se desnaturó a 95°C por cinco minutos y luego se agregó PCR Buffer 1x, MgCl₂ 1.5 mM, dNTPs 0.2 mM, oligo dT 0.02 µg/ml y Taq DNA polimerasa 5 U/µl. Se agregó

además un partidor antisentido para *atp7b* (bWs) 3.5 pmol/ μ l para aumentar en el producto la representación de la región que contiene las sondas de alta y baja identidad de *atp7b*. La amplificación por PCR anclado constó de 35 ciclos de 1 minuto de desnaturación a 95°C, alineamiento 1 minuto 30 segundos a 42°C y extensión 1 minuto 30 segundos a 72°C. El programa tenía una extensión final de 10 minutos a 72°C.

Del producto de PCR anclado se tomó 1.5 μ g y se agregó PCR Buffer 1x (Invitrogen), MgCl₂ 1,5mM, dNTPs 2.5 mM y 2.5 U de Taq DNA polimerasa y se llevó a un volumen final de 50 μ l. Para el fragmento de alta identidad de *atp7b* se utilizó un programa de 30 ciclos (desnaturalización 30 segundos a 95°C, alineamiento 30 segundos a 55°C y extensión 3 minutos a 68°C). Para el fragmento de baja identidad de *atp7b* se utilizó el mismo programa de amplificación, pero se ajustó la temperatura de alineamiento a 58°C para evitar la síntesis de un fragmento inespecífico.

En ambos casos, el tamaño de los fragmentos amplificados se verificó corriendo electroforesis en geles de agarosa al 1% a 100mV.

c) Grupo de genes relacionados

El protocolo utilizado para la amplificación de los fragmentos de alta y baja identidad de *atp7a* y *atp7b* se conservó para la amplificación de los fragmentos correspondientes al resto de los genes involucrados en el metabolismo celular de cobre. Los tiempos de cada etapa en los ciclos de amplificación se mantuvieron constantes y se modificó la temperatura de alineamiento y de extensión según se muestra en la Tabla 2.

En cada caso, el producto de la amplificación se fraccionó mediante electroforesis en geles de agarosa al 1% y 100 mV hasta lograr una buena definición.

Partidores utilizados para la amplificación de sondas.

Nombre	T _m (°C)	T _{elong} (°C)	P. sentido (5' - 3')	P. a-sentido (5' - 3')
<i>aMk</i>	55	68	TGAAGCCACTATTGTAAGTC	AGGTTTCTTAGCCACAGG
<i>bMk</i>	55	68	CAGGCTGTGGTATTAGCT	AGTCGACCTCTTCTTTC
<i>aWs</i>	55	68	CGCAGATGCTGTCAAACC	GCGACTTTCITCCCTTCA
<i>bWs</i>	58	68	GGAGACCTTGGGATACAG	CTGTCTTTCATCTCGTG
<i>bCes</i>	55	68	TCCCTATCCAAGATCACAG	CGAGGAOCAAATAACCTGA
<i>bSOD</i>	55	68	TGACTGCTGGAAAGGACG	AGTTTAATGGTTTGAGGGTA
<i>bPpc</i>	58	68	TGTTCTCTCGCATCTCG	TGTTAGCACTGGCTGATG
<i>bMI</i>	56	68	CCGTGGGCTGTCCAAAT	TAGGAAGACGCTGGGTTG
<i>Act(HK)</i>	65	72	GTCGTCGACAAAGGCTCCGGCATGTG	GCCAGCCAGGTCCAGACGCAGGATGG
<i>G3P(HK)</i>	55	68	ACCACAGTCCATGCCATCAC	TCCACCACCCTGTGCTGTGA
<i>Tub(HK)</i>	65	72	CATTGCCAACCATCAAGACCA	TCAAGCOCAAACCTCTCATA

Tabla 2: La tabla muestra las temperaturas de alineamiento y elongación para los partidores utilizados en la síntesis de los fragmentos que servirán como sonda de los genes que representan. T_m, temperatura de alineamiento; T_{elong}, temperatura de extensión.

2.7. CLONACIÓN EN EL VECTOR (pGEM-T™, Promega)

Se clonó los fragmentos (previamente verificados mediante electroforesis) utilizando los extremos poliA generados por la polimerasa (plásmido tiene extremos poliT). La reacción se efectuó a 4°C ON con las condiciones de ligación estándar sugeridas, que corresponden a 50 ng de DNA del plásmido, 1 unidad de T4 DNA ligasa. La reacción se detuvo desnaturalizando la enzima a 65°C por 10 minutos.

2.8. TRANSFORMACIÓN PARA OBTENCIÓN DE CLONES CON EL FRAGMENTO

Se transformó las células de la cepa DH5- α de *Escherichia coli* según el protocolo establecido por el Inoue y col, 1990, y se seleccionó las transformantes que llevan el vector por la resistencia al antibiótico (ampicilina) codificada en el vector. Las placas de medio LB-agar con ampicilina 50 μ g/ml en las que se realizó esta selección se sembraron además con IPTG (isopropyl-beta-D-thiogalactopyranoside) y X-gal, lo que permitió discriminar entre las que tenían y las que no tenían el inserto. Las colonias que llevan el vector más el inserto no expresan la enzima β -galactosidasa que codifica el vector y tienen un color blanco ya que no pueden degradar el X-gal.

2.9. CULTIVO DE COLONIAS POSITIVAS

Se picó las colonias transformadas con el vector más el fragmento clonado sembrándolas en placas de medio LB-agar con ampicilina 50 µg/ml. para aislarlas. Una vez separadas, se replicó las colonias positivas creciéndolas en medio LB líquido con ampicilina 50 µg/ml a 37°C ON y 220 rpm.

2.10. VERIFICACIÓN DE CLONES POSITIVOS

Se extrajo el DNA plasmidial de las colonias positivas utilizando el protocolo 'Miniprep I, II, III' (Bimboim, H., 1979; Sambrook, J. y col, 1989). Se digirió con la enzima que flanquea el inserto agregando 10µg de DNA a digerir, Buffer 1x correspondiente a la enzima, y 1 U de la enzima de digestión EcoRI. La reacción de digestión se detuvo desnaturalando la enzima por calor a 65°C por 5 minutos.

El tamaño de los fragmentos producidos se verificó mediante electroforesis a 100 mV en geles de agarosa al 1%. Como marcador de peso molecular se utilizó 100 bp DNA ladder (Promega).

2.11. STOCK DE CLONES POSITIVOS

Para cada uno de los fragmentos, se seleccionó tres colonias positivas con inserto verificado y se crecieron ON en medio LB líquido con ampicilina 50 µg/ml a 220 rpm. Los stocks fueron congelados en glicerol 15% a -80°C. En adelante, los inóculos para recuperar el inserto fueron tomados de estos stocks.

2.12. OBTENCIÓN DEL DNA A SEMBRAR

2.12.1. OBTENCIÓN DE LOS FRAGMENTOS

Para la extracción de DNA plasmidial se utilizó el protocolo de 'Miniprep I, II, III'. Se cuantificó las muestras midiendo la absorbancia a 260nm en cubetas de cuarzo y se estableció su nivel de pureza utilizando la relación 260/280.

El DNA sembrado en las membranas se obtuvo por amplificación por PCR de los fragmentos a partir de DNA plasmidial. Las condiciones de amplificación para cada uno de los fragmentos se explican en la sección 2.6. El volumen final preparado para cada uno de los fragmentos fue 200µl.

Los fragmentos amplificados fueron precipitados ON en 0.1M LiCl. Luego se centrifugó a 4°C por 30 segundos a 15000 rpm y el pellet se lavó con EtOH 70% frío. Se centrifugó nuevamente a 4°C por 30 segundos a 15000 rpm. El pellet se secó a 37°C y se resuspendió en dH₂O sin nucleasas. Se cuantificó midiendo su absorbancia en cubetas de cuarzo a 260 nm.

2.12.2. DILUCIÓN DE LAS SONDAS

Se sembró distintas concentraciones de los fragmentos en las membranas de hibridación. La unidad que se utilizó en todos los casos para hacer las diluciones fue ng DNA/punto (los que se obtendrían en las membranas al finalizar el proceso de siembra). En el cálculo, se consideró que el sembrador manual que se usó deposita en la membrana 0.1 µl de la muestra cada vez y que el proceso de sembrado se realizaría cuatro veces por membrana. Dependiendo de la cantidad de DNA obtenido después de la precipitación, se generó dos grupos de diluciones en placas ELISA de 96 pocillos. Desde allí tomó las muestras el sembrador manual. Las cantidades de DNA (expresadas en ng) sembradas al final del proceso de siembra se detallan a continuación. Los genes a los que corresponde cada fragmento están listados en la Tabla 1.

- 200, 100, 10, 2 y 0,2 ng DNA/punto para los clones b-ws, b-mt, b-sod, act, g3p, tub y plr.
- 100, 10, 2, 1,5 y 0,2 ng DNA/punto para los clones a-mk, b-mk, a-ws y b-ccs.

2.13. OBTENCIÓN DE LAS MUESTRAS

2.13.1. CULTIVO CELULAR

Llamamos muestra al cDNA de hebra simple suspendido en la solución de hibridación que estará en contacto con la sonda sembrada en la membrana. Las muestras se obtuvieron a partir de RNA total de células de epitelio intestinal Caco2 y células hepáticas HepG2, crecidas a 37°C y con 5% CO₂ en medio DMEM con 10% suero fetal bovino. Se trató un grupo de cada línea celular con 100µM CuHis durante 24 horas, mientras que el tratamiento control permaneció sólo con el medio de cultivo.

2.13.2. EXTRACCIÓN DE RNA TOTAL

La recuperación de células se realizó mediante un tratamiento suave con Tripsina – EDTA (Gibco BRL). Inicialmente las células fueron lavadas con tampón fosfato salino (PBS) y se incubaron en 0,01% de Tripsina EDTA a 37°C por 5 minutos. En adelante, todo el procedimiento se realizó manteniendo las células en hielo, para evitar la acción de RNAasas. La digestión se detuvo con la adición de DMEM-SFB y el medio con las células fue centrifugado a 3000 rpm por 5 minutos. El pellet se lavó una vez con PBS y se centrifugó nuevamente a 3000 rpm por 5 minutos. Las células lavadas se resuspendieron en 1 ml de reactivo RNAwiz (Ambion) por cada 10^7 células y se prosiguió con el protocolo de extracción según las indicaciones del proveedor. El RNA obtenido se cuantificó midiendo su absorbancia en cubetas de cuarzo a 260nm.

La calidad del RNA se revisó en un gel desnaturante (1,2% de agarosa en MOPS 10x/formaldehído 12,3M). De las muestras de RNA se tomó 3µl y se llevó a un volumen final de 25 µl (MOPS 1x, formaldehído 6,7%, formamida 50%) que se incubó a 55°C por 15 minutos. Las muestras se chequearon mediante electroforesis en gel desnaturante por 50 minutos a 80V. Luego se tiñó las bandas en una solución de EtBr 0,2 µg/ml por 45 minutos con agitación suave.

2.13.3. DIGESTIÓN DNA GENÓMICO

En los casos en que las muestras de RNA total estaban contaminadas con DNA genómico, éste se eliminó antes de la extracción de mensajero incubando a 37°C por 30 minutos con 1 unidad de DNAasa I por µl (Buffer DNAasa 1x). La reacción se detuvo con Termination mix 1x (0,1M EDTA pH8,0, glicógeno 1mg/ml). La solución se llevó a 1x NaOAc pH 4,5 e isopropanol 70% y se incubó 10 minutos en hielo. Se centrifugó a 14000 rpm por 15 minutos a 4°C y el sobre nadante se lavó con 0,1 volúmenes de EtOH 80%. Se centrifugó a 14000 rpm por 5 minutos a 4°C y se removió el sobre nadante. Se dejó secar el pellet al aire y se resuspendió en agua sin nucleasas.

2.13.4. EXTRACCIÓN DE RNA MENSAJERO

Para la extracción de mRNA a partir de RNA total, se utilizó el kit PolyATtract mRNA siguiendo las indicaciones del proveedor (Promega). Se cuantificó midiendo la absorbancia en cubetas de cuarzo a 260nm. Se obtuvo un rendimiento inferior al 1%.

2.13.5. SÍNTESIS DE cDNA

Para la síntesis de la hebra simple de cDNA de las células Caco2 y HepG2, tratadas y no tratadas con 100 μ M CuHis, se siguió dos protocolos: el tradicional, que utiliza un oligo dT como partidor para sintetizar la hebra de cDNA; y una variación propuesta para este trabajo, en que se pretendió aumentar la representación de las secuencias de interés utilizando una mezcla de los partidores antisentido de las sondas.

Para el protocolo tradicional se incubó todo el RNA total extraído durante 5 minutos a 70°C con 3pmol/ μ l de primer oligo dT. El tubo con la reacción se puso en hielo y luego se agregó buffer 1x MMLV RT, dNTP's, 5U de enzima transcriptasa reversa MMLV y agua hasta completar los 25 μ l. Luego se incubó a 42°C por 1 hora y se detuvo la reacción desnaturalizando la enzima a 70°C por 15 minutos.

En el protocolo alternativo se siguió el mismo procedimiento, pero en lugar del partidor oligo dT se agregó 35pmol/ μ l de cada uno de los partidores (sentido y antisentido) de los genes involucrados en la ruta metabólica del cobre listados en la Tabla 1.

2.13.6. MARCACIÓN DE MUESTRAS DE DNA

El marcaje de las muestras de DNA se realizó siguiendo el protocolo de Gibco para el kit Random Primer, que utiliza α^{32} P-dCTP suministrado por la Comisión Chilena de Energía Nuclear. Para ello, se desnaturalizó 25ng de DNA en 20 μ l de agua destilada por diez minutos en agua hirviendo. Luego se agregó 0,01mM dATP, dTTP y dGTP, 15 μ l de Random Primer Buffer, 50mCi de α^{32} P-dCTP, agua destilada y 3U de Polimerasa Klenow. El tiempo necesario para obtener una marca visible varió según la longitud de las moléculas a marcar: se marcó los fragmentos obtenidos por PCR durante una hora y el cDNA durante tres horas. La reacción se detuvo con 5 μ l de Stop buffer y se purificó en columnas del kit Wizard DNA Clean Up

de Promega siguiendo las instrucciones del proveedor. Se cuantificó la marca incorporada midiendo sus cuentas por minuto (cpm) un contador de centelleo Packard 1600PR.

2.14. TRATAMIENTO DE MEMBRANAS

2.14.1. SIEMBRA

Se cortó manualmente las membranas de nylon y el proceso de sembrado en cuadruplicado siguió las instrucciones del distribuidor. Las membranas tenían 8 cm. de ancho y 12 cm. de largo.

El sembrador manual consiste en un soporte metálico con 96 puntas de acero (que llamaremos sembrador), un soporte para la placa de ELISA con las diluciones de la sonda (matriz de carga) y un soporte para la base sobre la que se fijan las membranas (matriz de siembra). Se debe inmovilizar las membranas para evitar desplazamientos que alteren la forma del punto.

2.14.2. PREPARACIÓN DEL DNA SEMBRADO

Una vez sembradas las membranas, se desnaturó el DNA sumergiendo las membranas en una solución de 0,5M NaOH y 1,5M NaCl por 7 minutos sin agitación. Luego, se las lavó dos veces en una solución 1,5M NaCl, 0,5M Tris pH 7.4 y 1mM EDTA con agitación durante 3 minutos y se las dejó secar al aire. Se fijó las sondas a la membrana por entrecruzamiento con luz ultravioleta durante 45 segundos en el equipo CL-1000 Ultraviolet Crosslinker (UVP) a 254nm.

2.15. HIBRIDACIÓN

2.15.1. PREHIBRIDACIÓN

Se dispuso las membranas en botellas de hibridación para ser prehibridadas en un horno de hibridación a 42°C durante una hora con 2ml de una solución de 1x buffer de prehibridación, 100mM buffer fosfato de sodio pH 6.5, 20% formamida y 5% SDS.

2.15.2. HIBRIDACIÓN

Se realizó tres ensayos de hibridación para tratar de mejorar los resultados, estimados a través de la intensidad de la señal de hibridación.

Primero, para evaluar la calidad de la sonda en condiciones que favorecen la hibridación con su homólogo (descartar, por ejemplo, que la formación de estructuras secundarias disminuye la concentración efectiva de sonda), se hibridó las membranas:

- con sólo un fragmento marcado en la muestra.
- con diluciones del fragmento marcado (*) con otro sin marca (1*/10 y 1*/100 con respecto a la masa de los fragmentos).

Por ejemplo, para generar la dilución 1/10 del fragmento aMk se agregó a una botella 50 ng de fragmento marcado con $\alpha^{32}\text{P}$ -dCTP y 500 ng de la sonda aWs sin marca. Para la dilución 1/100 se agregó 5000 ng de aWs sin marca radiactiva a otros 50 ng de aMk marcado con $\alpha^{32}\text{P}$ -dCTP. De manera equivalente, se preparó las diluciones para las sondas bMk (diluido con bWs frío), aWs y bWs (preparadas con aMk y bMk frío, respectivamente). El agregar fragmento de la otra atpasa tuvo como objetivo sólo aumentar levemente la complejidad de la muestra. No constituyó un elemento de competencia a nivel de identidad de secuencias porque, como se observa en la Tabla 5, los fragmentos tienen un bajo nivel de identidad entre ellos.

Seguidamente, las muestras utilizadas en los experimentos de hibridación del conjunto de genes en estudio fueron de cDNA de hebra simple sintetizado a partir de RNA total.

Finalmente, se realizó una hibridación con muestras sintetizadas con los partidores de los genes en estudio, pero se cambió la presentación de la sonda en la membrana para poder sembrar una mayor cantidad de ella. Se sembró las sondas de *mt*, *sod*, *ccs* y *prpc* clonadas en el vector pGEM-T^{IM}Easy (200, 50, 20, 4 y 0,4 ng DNA/punto). En estas hibridaciones, se sembró cinco diluciones del vector como control de hibridación inespecífica y fragmentos, también clonados, de los genes 'house keeping': *g3p*, *act* y *tub*.

El DNA marcado, se mezcló previamente con 50µg/ml de DNA de esperma de salmón y se desnaturó en agua a 100°C por diez minutos. Esta mezcla se agregó a la solución de las membranas en prehibridación y se procedió con la hibridación a 42°C ON.

2.16. LAVADOS Y EXPOSICIÓN

Concluida la hibridación se recuperó la solución con la sonda y se lavó las membranas. Primero, 10 minutos a temperatura ambiente (TA) con solución I (2x SSC), luego 10 minutos a TA con solución II (2x SSC, 0,1% SDS). Después dos lavados de 15 minutos a 65°C con solución III (1x SSC, 0,1% SDS) y finalmente dos lavados de 15 minutos a TA con solución IV (0,1x SSC). Las membranas lavadas fueron inmovilizadas en una película plástica dentro de un casete de revelado. Sobre ellas se puso dos placas radiográficas y la exposición se realizó a -80°C por un tiempo que dependió de la marca radiactiva detectada en las membranas (al menos dos horas).

2.17. REVELADO

Se descongeló los casete de revelado a temperatura ambiente antes de revelar las placas. La señal se obtuvo con las soluciones de revelado y fijación utilizadas para el proceso. Se lavó las películas con abundante agua y se las dejó secar al aire. El tiempo de revelado dependió de la intensidad de la marca en la membrana y de la resolución y la calidad de los puntos en la matriz, considerando su posterior análisis computacional.

2.18. ANÁLISIS DE LOS PUNTOS

Para el análisis de los datos de hibridación se obtuvo una imagen digital utilizando el programa Kodak Digital Science 1D, con el que además se obtienen datos de intensidad de la señal de los puntos. Para estimar el valor promedio de la señal considerando el ruido alrededor de ella, se utilizó un programa diseñado por Pablo Moreno (Ingeniería Civil en Biotecnología, Universidad de Chile). Este programa calcula la densidad de color de cada uno de los puntos de sonda sembrados, realiza un promedio de ellos y le resta a este valor el promedio de cinco puntos imaginarios que creaba alrededor de la siembra para estimar la señal de ruido de la placa en esa zona. Además realiza el análisis de la desviación de los valores de estos dos grupos de puntos, de

manera que si alguna irregularidad en la membrana alteraba el valor esperado de la señal y alejándolo del promedio, la excluía del análisis. Los resultados del programa de Moreno evaluaron la calidad del resultado con tres tipos de cálculo para la señal de hibridación:

- i) Tomando el promedio de los cuatro puntos sin excluir ninguno de ellos.
- ii) Tomando sólo el máximo valor dentro del grupo.
- iii) Tomando los tres mejores valores del cuadruplicado (es decir, aceptando como valor real el promedio de las intensidades de tres puntos que tuviera la menor desviación estándar).

Se realizó un pequeño estudio de los resultados de hibridaciones realizadas por otros usuarios del Laboratorio de Bioinformática y Expresión Génica (Mauricio González Agüero trabajó con las líneas celulares 804 y 805) para evaluar cuál de estas opciones de análisis entregaba resultados más confiables y con menores variaciones entre réplicas. En todos los casos se evaluó además la calidad de la información tomando como valor de intensidad la diferencia entre señal y ruido (S-B) o la razón entre señal y ruido (S/B). Finalmente, se decidió trabajar con el promedio de los tres mejores puntos.

2.19. CULTIVO CELULAR

Se cultivó las líneas celulares de epitelio intestinal (Caco2) y hepáticas (HepG2) a 37 °C y 5% CO₂ en medio DMEM suplementado con 10% suero fetal bovino (SFB).

3. RESULTADOS

Los programas de alineamiento de secuencias disponibles en la red suelen comparar secuencias (aminoacídicas o nucleotídicas) buscando las regiones que presenten mayor similitud dentro de ellas. En este trabajo de Memoria de título, se evaluó el desempeño de un algoritmo implementado en un programa computacional diseñado para discriminar las regiones de baja identidad entre las secuencias (nucleotídicas) que se comparan. Nuestro propósito fue identificar secuencias que sirvan como sondas específicas en ensayos de hibridación en micro o macroarreglos de DNA.

3.1. PARTE I: Programa Redigen

3.1.1. Análisis de los parámetros de búsqueda del programa Redigen para la identificación de dominios de baja identidad.

En la mayor parte de los estudios de las variaciones en los parámetros de búsqueda del programa Redigen se utilizó las secuencias de los genes *atp7a* y *atp7b* (Menkes y Wilson). La principal razón para escoger estos genes fue el alto grado de similitud que presentan sus secuencias nucleotídicas (53.7% de identidad), lo que se traduce en un buen nivel de exigencia para examinar la sensibilidad del programa en evaluación. Información sobre las secuencias utilizadas en las comparaciones, así como los tamaños de ventana, parámetros porcentuales y orden y cantidad de dichas secuencias se describe con mayor detalle en Materiales y Métodos (secciones II y III). Las secuencias de *atp7a* y *atp7b* fueron comparadas recíprocamente en todos los casos, es decir se evaluó a cada una de ellas tanto como secuencia móvil así como secuencia fija. A modo de guía de nomenclatura, es pertinente aclarar que al mencionar dos secuencias en comparación, la primera de ellas siempre corresponderá a la secuencia fija y la segunda a la secuencia móvil.

a) Efecto del tamaño de ventana: Se comparó las secuencias de *atp7a* y *atp7b* con el programa Redigen variando la longitud del fragmento móvil. En los tres tamaños de ventana examinados se observó diferencias en el grado de oscilación y en la magnitud de los valores de identidad. Utilizando la ventana de 100 bp, se registró los mayores niveles de identidad, aunque estos valores se alternan muy frecuentemente con dominios en los cuales el nivel de identidad presenta una marcada disminución. Es decir, la frecuencia de oscilación de los valores es alta. En contraposición, al utilizar ventanas de 300 y 600 bp la magnitud y frecuencia de oscilaciones en los valores de las unidades de identidad se redujo notablemente. En cuanto a las zonas en que se ubican las regiones de baja identidad, tanto en *atp7a* como en *atp7b*, se observó que estas son más o menos constantes en su ubicación aunque mayor es su extensión mientras mayor es el tamaño de la ventana (Figura 4).

Es necesario aclarar que los bajos valores de UA de todos los gráficos en el eje Y antes de la posición 1 en el eje X, se producen por un artefacto del procedimiento, ya que los nucleótidos no apareados se cuentan como diferencias y eso genera valores bajos. Recordemos que la primera comparación se realiza alineando el nucleótido de la mitad de la ventana con el primer nucleótido de la secuencia fija para luego ir desplazándose en dirección 5'→3', por lo tanto antes de alinear el primer nucleótido de la ventana con el primero de la secuencia fija siempre habrá al menos un nucleótido sin alinear y que será considerado como una diferencia.

En general en los análisis de las curvas de identidad se observó que mientras menor es el tamaño de la ventana, mayor es la variación del nivel de identidad. Probablemente, esto se debe a que cuanto menor es la longitud de la secuencia a alinear, mayor es el peso porcentual de una diferencia entre las secuencias comparadas. Por otra parte, se observó que la forma de las curvas se mantiene relativamente constante al cambiar el tamaño de ventana de comparación. Esto puede atribuirse a que las secuencias tienen regiones bien definidas de alta y baja identidad que manifiestan pequeñas variaciones al cambiar el tamaño de ventana con que se comparan. El cambio en el tamaño de ventana tiene que ser muy drástico para afectar su porcentaje de identidad con la secuencia que se compara (reflejado en las unidades arbitrarias de identidad).

El análisis permitió detectar además que el tamaño de la ventana a utilizar debe estar relacionado con el tamaño de los genes que se comparan. Con el uso de ventanas pequeñas (con relación a las secuencias ingresadas a la comparación) el programa entrega curvas que representan altos niveles de identidad.

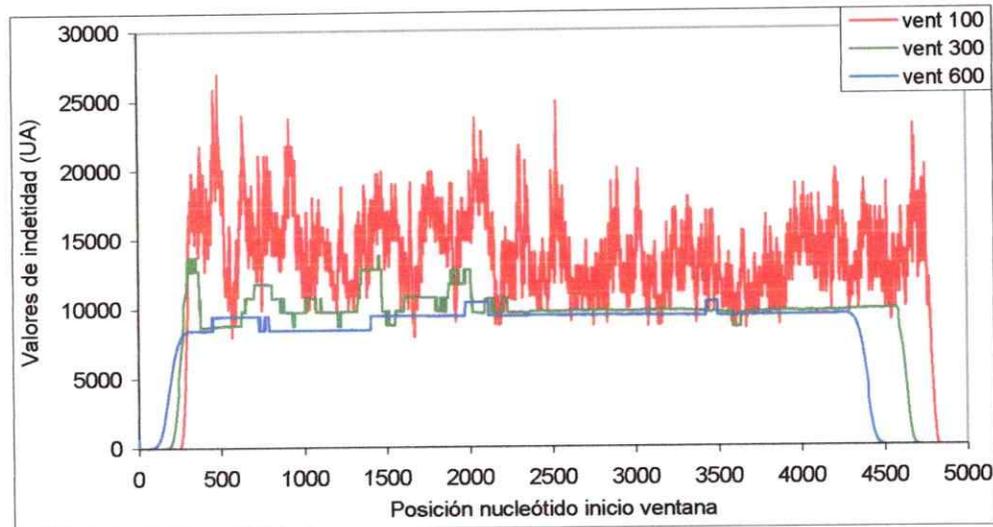


Figura 4: Efecto del tamaño de ventana en el valor de unidades de identidad de las secuencias analizadas. Se graficó la comparación de *atp7a* con *atp7b* con tamaños de ventana 100 (vent 100), 300 (vent 300) y 600 pb (vent 600) con parámetros *j* y *k* de 60 y 80, respectivamente. A los valores porcentuales que indican de cero a 60% (*j*) de diferencias se les asignó 1000 UA, a aquéllos entre 60 y 80% (*k*) les correspondió 1 UA y a los valores porcentuales correspondientes a más de un 80% de diferencias, 0 UA. Los saltos discretos en las curvas de identidad generadas por las ventanas de 300 y 600 pb indican que esas regiones de *atp7b* tenían muchos eventos de identidades superiores al 40% con la secuencia de *atp7a*. Eje X, posición en *atp7a* del primer nucleótido del fragmento de *atp7b* con que se compara; eje Y, unidades arbitrarias correspondientes al porcentaje de identidad entre secuencias.

b) Efecto de los parámetros porcentuales: Para estudiar el efecto de los parámetros porcentuales se realizó previamente una comparación del grado de identidad de tres secuencias al azar utilizando ClustalW. La comparación de las secuencias de *atp7a*, lipasa pancreática de ratón (*Pnlip*) y raíz de *Arabidopsis thaliana* (*rAt*) sugiere que las secuencias no relacionadas se parecen azarosamente alrededor de un 45% (*atp7a* - *rAt* = 47.8%; *atp7a* - *Pnlip* = 43.2%; *Pnlip* - *rAt* = 46.2%).

Gráficamente, en la misma Figura 4 podemos observar para las posiciones alrededor del nucleótido 270 (eje X) el valor calculado es 14,000 UA. Esto quiere decir que los fragmentos iniciados en estos nucleótidos tienen a lo más 14 posiciones con un nivel de identidad superior al 40% (entre 0 y 60% de diferencias). Ya

que sabemos que dos secuencias escogidas al azar tienen cerca un 45% de identidad, la elección de estos parámetros porcentuales no nos permitirá seleccionar una sonda específica.

Luego, se analizó la fluctuación de valores de identidad al variar los parámetros porcentuales (j-k), es decir, manteniendo fijo el tamaño de ventana y cambiando sólo la exigencia en el porcentaje para asignar la identidad. Mientras más pequeña era la diferencia entre j y k (variaciones de 5 ó 10 puntos porcentuales), menos evidente fue la distinción de las regiones comparadas. Es decir, las fluctuaciones de los valores de UA en el eje Y disminuyen. Sin embargo las tendencias en la forma de las curvas se mantienen (Figura 5). Si ambos parámetros tenían valores superiores a 70, todos los valores de la curva se desplazaron a la región donde los valores de UA son $>10^6$, es decir, todas las secuencias móviles tienen más de un evento en el que se parecen más de 30% a cualquier región de la secuencia fija (Figura 6).

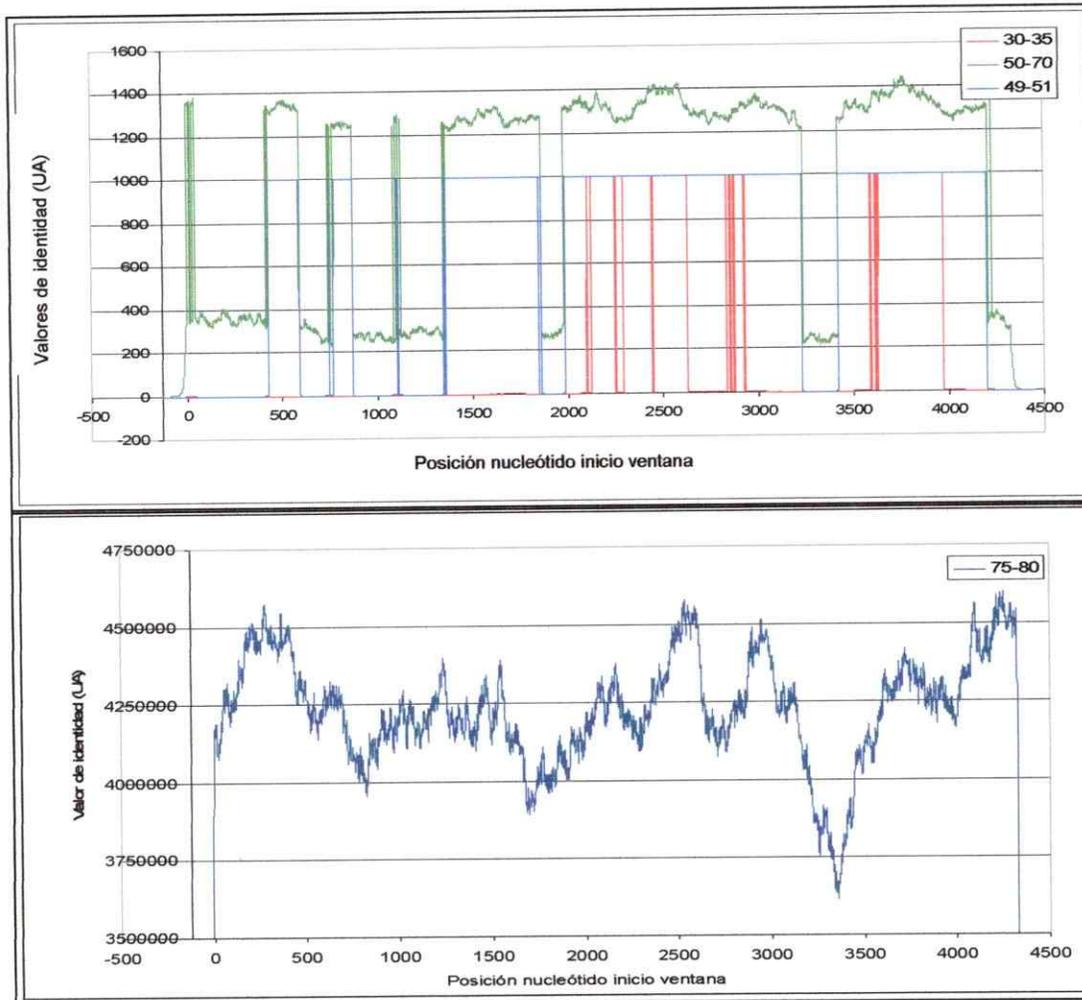


Figura 5: Efecto de los parámetros porcentuales en las curvas de identidad de *atp7a* y *atp7b*. Las secuencias de ambas *atp*-asas fueron comparadas con una ventana de 300 pb (*atp7a* fija) y el número de diferencias fue clasificado según cuatro pares de parámetros porcentuales: 30-35, 49-51, 50-70 y 75-80. Este último par se grafica separadamente porque el orden de magnitud de sus valores era muy superior a los otros tres. Las curvas mantienen la tendencia de su forma, independiente del porcentaje con que se clasificó el número de diferencias. Por ejemplo, alrededor del nucleótido 1900 existe una región con bajos valores de UA en las cuatro curvas, lo que indica que aún con parámetros muy exigentes, esa región de la secuencia encuentra pocas regiones con las que hibridaría de forma cruzada.

Las numerosas fluctuaciones que se observan para una ventana de 100 pb en el gráfico A de la Figura 6 muestran que el peso de las diferencias entre secuencias es mucho mayor cuando la ventana es pequeña. Por otro lado, en el gráfico B de la figura 6 la curva de una ventana de 600 pb muestra escasas variaciones. Al analizar el efecto de los parámetros porcentuales se observa que sin importar el tamaño de ventana, no se producen entre las secuencias muchos eventos de identidad superior al 60% (línea de color azul) y, en todos

los casos, la identidad es superior al 20% (línea de color rojo). Se graficó valores extremos tanto en tamaño de ventana como en parámetros porcentuales con el fin de mostrar más claramente el comportamiento de las curvas. Las curvas de parámetros porcentuales y tamaños de ventana con valores intermedios a los aquí utilizados se ubican entre las líneas presentadas en esta figura. Debido a la diferencia de magnitud en los valores de identidad de las curvas, se utilizó su logaritmo en base 10 para poder comparar los resultados en un solo gráfico.

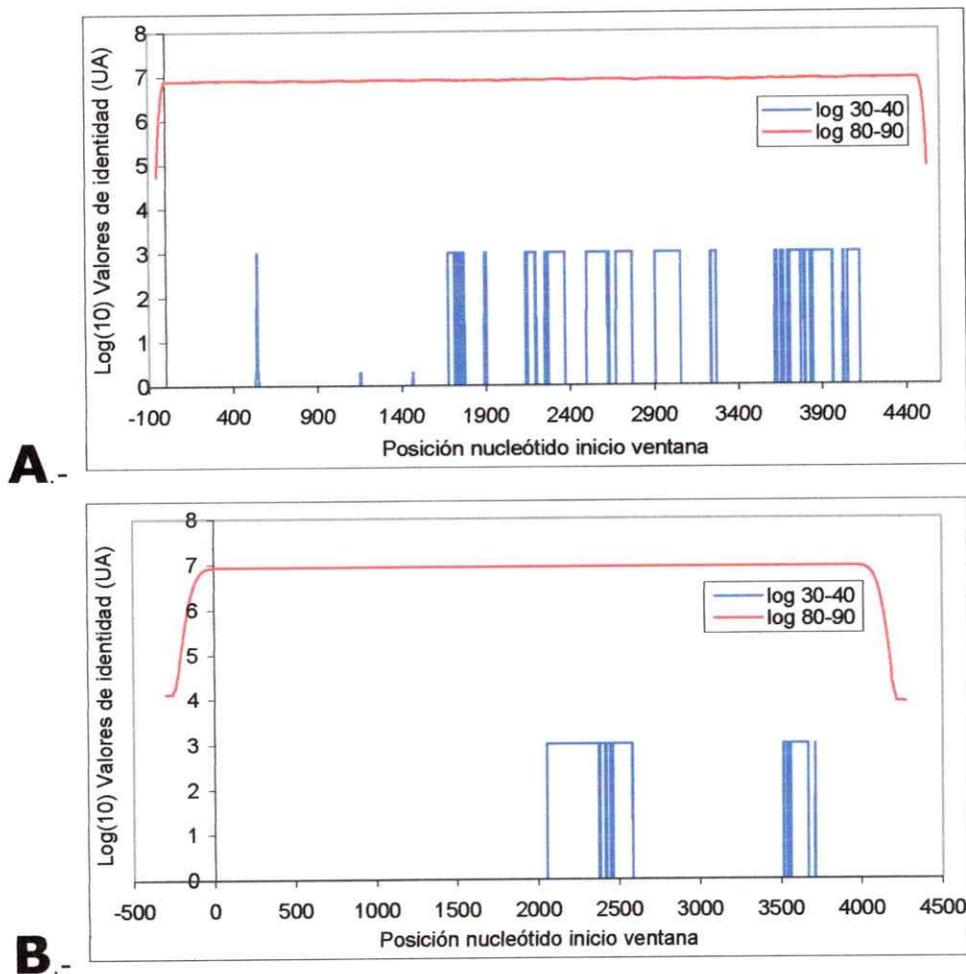


Figura 6: Efecto del tamaño de ventana relacionado con los parámetros porcentuales en las curvas de identidad de *atp7a* y *atp7b*. **A.** ventana de comparación 100 pb; **B.** ventana de comparación 600 pb. La secuencia de *atp7a* se mantuvo fija. Debido a las grandes diferencias de orden de magnitud entre ellas, se graficó el resultado del logaritmo en base 10 de las unidades arbitrarias asignadas a los porcentajes de identidad para cada secuencia. La línea azul representa el grado de identidad con parámetros 30 – 40, mientras que la línea roja, parámetros 80 – 90. En la curva azul, los puntos de valor 3 (log1000) indican que en esa posición se encuentra una secuencia de *atp7a* que tiene un evento de identidad superior a 70% (entre 0 y 30% de diferencias).

Usando los parámetros porcentuales 30-40, la comparación de las secuencias de cualquier par de genes de la ruta metabólica del cobre entrega líneas rectas de valor cero sin importar el largo de la ventana (50, 100, 300 ó 600 pb). Sin embargo, entre los pares de secuencias más parecidas surgen algunas variaciones: *atp7a-atp7b* y *sod1-ccs* que muestran regiones bien marcadas en las curvas con valores de 1 ó 10^3 UA, lo que significa que existen regiones entre estas secuencias que tienen identidades por sobre el 70%.

En la comparación de la secuencia de lipasa pancreática de ratón con la secuencia de raíz de *A. thaliana* (secuencias no relacionadas) utilizando Redigen se observa que el máximo valor mostrado por las curvas bordea las 300 UA. El promedio de los valores de identidad era todavía menor aún cuando se utilizó una ventana pequeña para la secuencia móvil.

Finalmente, los resultados obtenidos en los análisis comparativos, indican que las curvas generadas con los parámetros porcentuales 50-70 entregan una información clara y resuelven bien las regiones de baja identidad. Con el propósito de ilustrar más claramente la utilización de los parámetros porcentuales se ha ideado el siguiente ejemplo. Si un fragmento posee hasta un 50% de diferencias (es decir, son idénticos en más de la mitad de las bases) entonces el fragmento tiene una alta probabilidad de hibridar en forma cruzada y se le entrega un alto puntaje (1000 UA). En cambio, si el número de diferencias está entre 50 y 70%, ambas secuencias se parecen entre 30 y 50%, la probabilidad de que ocurra hibridación cruzada sería menor, por lo tanto se le asigna un puntaje menor (1 UA). La diferencia de magnitud en los valores asignados permite ver claramente en los gráficos los saltos en identidad, de ser menor esta distancia (por ejemplo 3, 2 y 1 UA), los números se mantienen dentro de un rango pequeño que no permite discriminación. Finalmente, si las diferencias son mayores al 70%, las secuencias son idénticas en menos de un 30% y se considera que una hibridación azarosa entre ambas moléculas no es estable y no permanecerá en las condiciones de hibridación, otorgándosele un puntaje igual a 0 (cero) UA. Las hibridaciones serán más estables mientras más puntos de unión tengan entre sí, por lo tanto, secuencias con un alto porcentaje de diferencias se disociarán hasta encontrar el alineamiento ideal.

c) Efecto del orden de comparación de las secuencias: Para realizar las comparaciones entre pares de secuencias, el programa Redigen toma una ventana de la que llamamos 'secuencia móvil' (segunda secuencia) manteniendo la primera secuencia fija ingresada. Considerando esto, se realizó una comparación entre los patrones de identidad al invertir el orden de comparación de dos secuencias estrechamente relacionadas, en este caso *atp7a* y *atp7b*. Las diferencias entre ambos patrones son mínimas y reflejan el grado de similitud entre estas secuencias, sin embargo las curvas parecen desplazadas una con respecto a la otra.

Cuando se comparan las secuencias codificantes de *atp7a* y *atp7b* se observa que, independiente del parámetro porcentual o el tamaño de ventana, cuando la ventana móvil corresponde a *atp7a* las curvas están casi totalmente sobrepuestas y las diferencias son mínimas. Cuando la ventana es de *atp7b*, los valores de los desplazamientos son mayores (en número de nucleótidos, ver Tabla 3). Si bien los valores de la siguiente tabla fueron tomados sólo en una región de las curvas, es necesario mencionar que el patrón se mantuvo a lo largo de ésta. El significado de esta diferencia de desplazamiento será discutido más adelante.

Quantificación del desplazamiento de las curvas de comparación de *atp7a* y *atp7b*.

secuencia a/secuencia b-tamaño de ventana	parámetros porcentuales		
	30/40	50/70	60/80
<i>atp7a/atp7b-v150</i>	13	14	0
<i>atp7a/atp7b-v300</i>	13	13	0
<i>atp7b/atp7a-v150</i>	163	163	151
<i>atp7b/atp7a-v300</i>	163	163	163

Tabla 3: La comparación de las secuencias codificantes ('cds') de las secuencias de *atp7a* y *atp7b* entrega curvas con formas tan similares que sugerían un desplazamiento de la misma curva (ver Figura 3). La tabla muestra la magnitud -en número de nucleótidos- de dicho desplazamiento. Estos valores fueron tomados sólo en una región de la curva, sin embargo el desplazamiento parece tener un valor constante a lo largo de toda la curva.

d) Efecto del aumento en el número de secuencias comparadas: Uno de los usos esperados para el programa Redigen es el análisis de las diferencias entre secuencias de un grupo de genes relacionados. Por lo tanto, se evaluó la relevancia del número de secuencias ingresadas. El programa Redigen toma fragmentos de la segunda ingresada y los compara con la primera secuencia completa, que se mantiene fija. Una vez

concluida esta comparación, toma la secuencia complementaria del fragmento en dirección 5'-3' y lo comprara con la secuencia completa de la primera secuencia. Así hasta tomar fragmentos en toda la segunda secuencia. Cuando concluye el proceso entre las dos primeras secuencias del conjunto de genes ingresados, realiza la comparación entre la primera y la tercera secuencia de la misma forma. Luego del análisis, se observó que la ventaja de ingresar simultáneamente las secuencias se limita al ahorro del tiempo que significa ingresarlas de una en una. La forma de cada una de las curvas es idéntica a aquella de la comparación que se realizó ingresando sólo el par de secuencias que generan dicha curva (Figura 7). En la comparación del grupo de secuencias de la Figura 7, los valores de las curvas oscilan muy por debajo de las 1000 UA para nueve de las diez comparaciones. Esto significa que con estas secuencias no se produjeron alineamientos con más de 50% de identidad. En la curva resultante de la comparación de *sod1* con *ccs* se observa que una región de *sod1* posee alta identidad con la secuencia de *ccs* (gen que codifica para la proteína chaperona de Sod1).

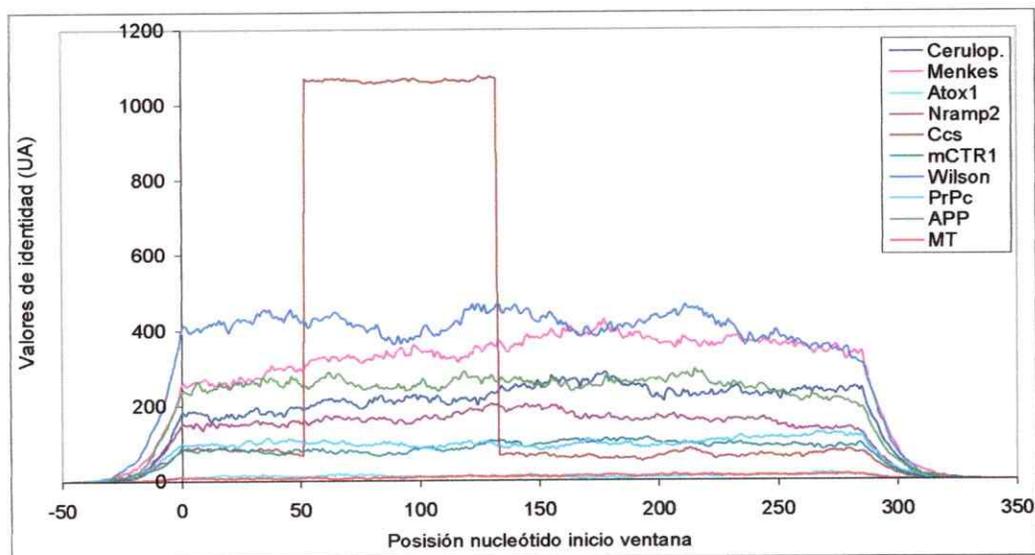


Figura 7: Comparación de la secuencia de *sod1* con las de otros genes relacionados con la vía metabólica del cobre. Las comparaciones se realizaron utilizando una ventana de 300 pb y parámetros porcentuales j-k = 50-70.

3.1.2. Establecimiento de las identidades entre fragmentos

La estimación del porcentaje de identidad entre fragmentos de ATPasas se realizó en dos etapas: En la primera aproximación se utilizó los programas BLAST y ClustalW, ingresando la secuencia completa de la región a analizar de una de las ATPasas (previamente seleccionada como de alta o baja identidad, según las curvas hasta ahora analizadas) más la longitud de la ventana que generó esa curva. Este fragmento extra se ingresa debido a que las curvas obtenidas con el programa Redigen representan la posición del nucleótido en que se inicia la ventana, y quisimos incluir en este análisis aproximado a todas las posibles sondas de esa región de alta o baja identidad. Esto se comparó con las hebras completas sentido y antisentido de la otra atpasa. Luego de los primeros análisis con este sistema y de aquí en adelante, BLAST fue desechada como herramienta para estimar valores de identidad por diferentes razones: (1) sólo entrega resultados si detecta alineamientos sobre un valor de probabilidad ("e value") según sus parámetros de búsqueda (ellos privilegian las similitudes y no muestran los resultados si las secuencias son muy diferentes) y (2) alteraba la longitud de las secuencias ingresadas para lograr un mejor alineamiento, sin importar la penalización asignada a esto por el usuario. Como se explicó en la introducción, BLAST (blastn) utiliza métodos heurísticos en la búsqueda de identidades entre secuencias sin importar que el mejor alineamiento corresponda sólo a un fragmento de la secuencia completa. Sin embargo, se utilizó la base de datos BLAST Mouse Sequences para estimar hibridación cruzada con una secuencia del genoma de ratón que no pertenezca al grupo de genes relacionados con el metabolismo celular de cobre.

El porcentaje de identidad calculado a partir de los alineamientos realizados con el programa ClustalW se enmarcó dentro de los parámetros porcentuales de identidad obtenidos mediante el programa Redigen. Este resultado sugiere que el programa Redigen arroja aproximaciones confiables de los porcentajes de identidad entre secuencias, ya que sus resultados coinciden con los entregados por un algoritmo ya probado que calcula las similitudes entre secuencias. Por lo tanto, se aceptó como válida la búsqueda de partidores para regiones de alta o baja identidad (según correspondiese) dentro de dichas regiones. Con el programa ClustalW se

realizó la segunda etapa de la estimación del porcentaje de identidad, ingresando a la comparación la región ya seleccionada para su amplificación mediante PCR (ver Métodos sección 2.4.3).

3.1.3. Búsqueda de regiones específicas en genes relacionados

a) Selección de regiones de alta y baja identidad en *atp7a* y *atp7b*: Se analizó las curvas obtenidas de la comparación de ambas secuencias con el programa Redigen con una ventana de 300 pb y $j-k = 50-70$. Se consideró como regiones de baja identidad aquellas en que el valor de la curva fuera menor a 1000 UA. Se tomó como regiones de alta identidad a aquellas con valores por sobre las 1000 UA.

b) Selección de regiones de baja identidad en las secuencias de genes relacionados metabólicamente: Se ingresó al programa Redigen las secuencias nucleotídicas de los genes involucrados en la ruta metabólica del cobre (ver Métodos, Tabla 1) y se las comparó con una ventana de 300 pb y parámetros porcentuales $j-k = 50-70$. Se consideró como regiones de baja identidad a aquellas zonas en que todas las curvas coincidían con valores inferiores a 1000 UA (ver Figura 7).

c) Evaluación de las regiones de baja identidad en los genes: Se estimó el grado de identidad entre las secuencias de los genes 'house keeping' (tubulina, *tub*; actina, *act* y gliceraldehido-3-fosfato deshidrogenasa, *g3p*) y de éstos con las secuencias de los genes involucrados con la ruta metabólica del cobre. Se analizó las curvas resultantes de dicha comparación. No se observó valores mayores a 700 UA en ninguna de las curvas obtenidas.

d) Aproximación al porcentaje de identidad: Con el programa ClustalW se evaluó el porcentaje de identidad aproximado de las regiones de alta identidad (para *atp7a* y *atp7b*) y de baja identidad (para todas las secuencias, Tabla 1) ya seleccionadas, comparándolas con las hebras sentido y antisentido de las secuencias de los genes involucrados en la ruta metabólica del cobre. Las regiones de alta identidad de *atp7a* y *atp7b* presentaron valores superiores al 60%. Las regiones de baja identidad de todos los genes estudiados presentaron identidades inferiores al 40%.

e) Selección de partidores: Se utilizó el programa Primer Premier v. 5.0 en la búsqueda de pares de partidores para las regiones de baja identidad seleccionadas. El par de partidores escogido fue aquel mejor calificado dentro del conjunto de resultados que entregó el programa Primer Premier.

La nomenclatura diseñada para los partidores consistió en una **a** o **b**, si correspondían a partidores para regiones de alta o baja identidad respectivamente, una letra **s** o **a** referida a la dirección del partidor (**sentido** y **antisentido**) y una abreviación de dos o tres letras del nombre del gen. Las secuencias de los partidores seleccionados se muestran en la Tabla 2 (ver Métodos, sección 2.6).

f) Verificación de alta y baja identidad de las sondas: Se calculó el porcentaje de identidad de las secuencias determinadas por los partidores seleccionados con las secuencias sentido y antisentido de los genes involucrados en la ruta metabólica del cobre. Coincidiendo con la aproximación realizada en la sección 2.4, los valores para los fragmentos de baja identidad aparecían bajo el 40%, alrededor del 30%. En la Tabla 4 se muestran los valores máximos de identidad obtenidos con el programa ClustalW para los fragmentos de baja identidad del grupo de genes involucrados en la ruta metabólica del cobre. En el caso de las secuencias de *atp7a* y *atp7b* este valor se obtiene de la comparación con la otra atp-asa.

Comparación de fragmentos seleccionados como sondas con los genes involucrados en el metabolismo celular de cobre mediante ClustalW.

Fragmento	Porcentaje
<i>a-atp7a</i>	71%
<i>b-atp7a</i>	39%
<i>a-atp7b</i>	70%
<i>b-atp7b</i>	40%
<i>b-ccs</i>	32%
<i>b-sod1</i>	40%
<i>b-app</i>	40%
<i>b-mt1</i>	37%
<i>b-nramp</i>	27%
<i>b-ctr</i>	35%
<i>b-cp</i>	34%
<i>b-atox</i>	35%
<i>b-prpc</i>	28%

Tabla 4: Los valores de la tabla corresponden al máximo porcentaje de identidad de en la comparación de los fragmentos de baja identidad con el grupo de genes relacionados con el metabolismo celular de cobre. Se incluyen además, los fragmentos de alta identidad para las secuencias de *atp7a* y *atp7b*. Los resultados se corresponden con los parámetros ingresados al programa Redigen para la selección de las regiones de alta y baja identidad.

3.2. PARTE II: Ensayos experimentales.

3.2.1. Obtención y evaluación experimental de fragmentos seleccionados con Redigen.

Se amplificó los fragmentos por PCR utilizando los partidores diseñados y estableciendo las condiciones óptimas de temperatura y tiempo de elongación. Las diluciones de los productos de PCR sembrados en las membranas se realizaron dependiendo de las cantidades obtenidas de los fragmentos amplificados.

a) Especificidad de los fragmentos versus concentración de sonda sembrada: Las membranas fueron sembradas con las cuatro sondas de *atp7a* y *atp7b* de alta y baja identidad en cuatro diferentes concentraciones y cada una de ellas en triplicado (Figura 8). Cada una de estas membranas se hibridó con los fragmentos purificados de alta y baja identidad de *atp7a* y *atp7b* marcado con ^{32}P (cuatro eventos de hibridación independientes). Se expuso las membranas ya hibridadas a placas radiográficas y de estas últimas se midió la intensidad de la señal. En la Figura 8 se muestra una imagen representativa de estas hibridaciones.

Como se observa para cada uno de los fragmentos marcados con ^{32}P , la señal de hibridación mas intensa se observa con su respectiva sonda. En algunas de las concentraciones utilizadas, el blanco en la muestra es detectado sólo por la sonda específica (por ejemplo en la tercera columna de aWs y bWs). El análisis densitométrico indica que en ningún caso se detectó señal neta de hibridación con valores superiores al 40% del valor máximo (el cual corresponde a la hibridación del fragmento sembrado como sonda con el mismo fragmento presente en la muestra para cada concentración de siembra). Esto coincide con las hibridaciones teóricas mostradas en la Tabla 5. Para cada uno de los cuatro fragmentos, se observó que dentro de una muestra simple, el fragmento es altamente específico.

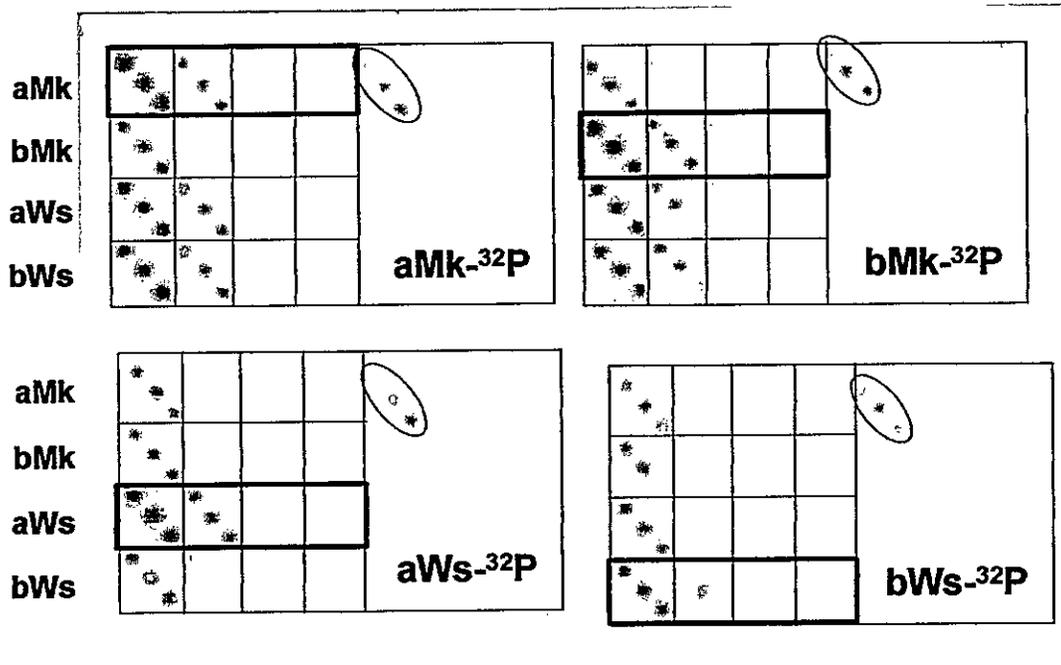


Figura 8: Especificidad de los fragmentos versus concentración de las sondas sembradas. Se hibridó las membranas con las cuatro sondas marcadas con $\alpha\text{-}^{32}\text{P}\text{-dCTP}$. El nombre del fragmento marcado en la muestra se observa dentro de cada rectángulo, que corresponden a eventos de hibridación independientes. El nombre de la sonda sembrada se observa a la izquierda. Se sembró por triplicado cuatro concentraciones de las sondas de alta y baja identidad de *atp7a* y *atp7b*: 100, 10, 2 y 0.2 ng DNA/punto. La cantidad de sonda sembrada disminuye hacia la derecha (separadas en cuadrados). Se sembró como DNA heterólogo las mismas concentraciones de vector pGEM-T, pero las cantidades sembradas por punto disminuyen hacia abajo. Sólo se observa hibridación cruzada en 100ng DNA plasmidial/punto (puntos contenidos en óvalo). Para cada sonda, su máximo de hibridación se obtiene cuando encuentra su blanco en la muestra (rectángulo en negrita) y se aparece relacionado con la cantidad de sonda sembrada en la membrana.

Porcentaje de identidad dentro del grupo de fragmentos de alta y baja identidad de *atp7a* y *atp7b*.

<i>clon id</i>	aMk	bMk	aWs	bWs
aMk	100,0	35,1	32,7	31,2
bMk	24,2	100,0	23,9	35,5
aWs	31,9	33,8	100,0	34,8
bWs	24,0	39,7	27,5	100,0

Tabla 5: Comparación con el programa ClustalW entre las secuencias de los fragmentos sonda de alta y baja identidad de *atp7a* y *atp7b*. Los números corresponden a valores porcentuales expresados en función del porcentaje máximo de identidad generado por el alineamiento entre dos secuencias idénticas. El porcentaje máximo de identidad (100%) corresponde al alineamiento de una secuencia consigo misma.

En algunas de las señales de hibridación, se observó que existe un óptimo entre el límite de detección de la sonda y calidad de la señal (intensidad de señal versus ruido asociado). En la mayoría de los casos no se obtuvo información de la sonda sembrada en mayor concentración posiblemente porque la señal alcanzó el nivel de saturación. Sondas sembradas en mayor concentración tenían un buen grado de hibridación con su homólogo en esta muestra simple (que contenía sólo un fragmento en la muestra) y las menores cantidades tenían una señal muy baja. Resultados de un experimento posterior a este trabajo de tesis utilizando membranas sembradas con las sondas diseñadas por Redigen, sugieren que la cantidad óptima de sonda sembrada podría estar asociada de manera inversa con el largo ésta (Figura 9, realizado por Mauricio González-Agüero). En dicho ensayo se hibridó cDNA de fibroblastos wild type (wt) de ratón en situación control y expuestos a 100 μ M Cu-His. Las membranas utilizadas fueron sembradas con sondas de genes cuya expresión se ve afectada por la exposición a cobre, entre ellas algunas de las sondas generadas en este trabajo (b-sod, a-ws7, b-ws5, a-mk9 y b-mk9). Se sembró además sondas de genes que se utilizaron como control de expresión.

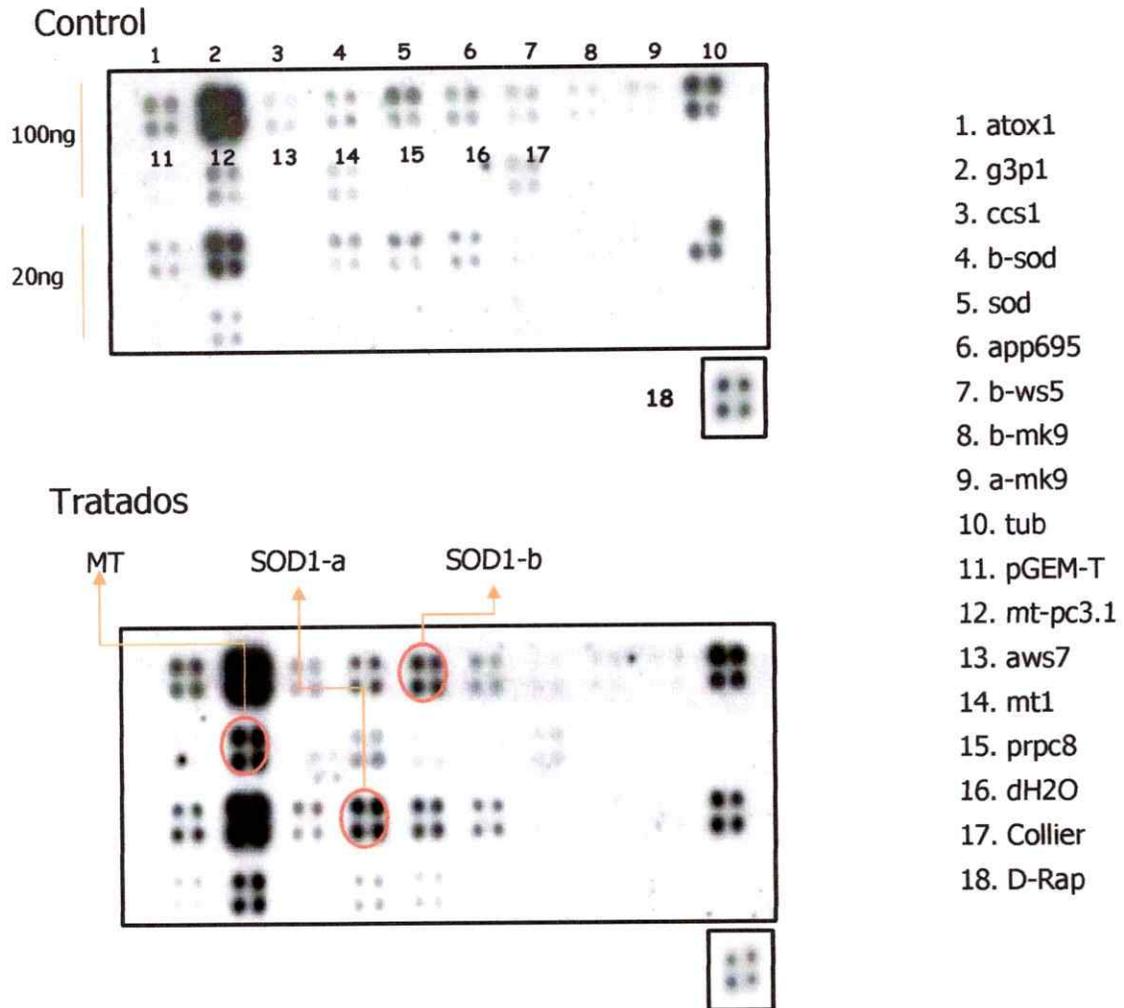


Figura 9: Expresión de genes relacionados con el metabolismo de Cu en fibroblastos wt. Fibroblastos wt (wild type) de ratón fueron tratados o no con 100 μ M Cu-His. cDNA de estas células se hibridó con membranas idénticas sembradas con sondas de genes involucrados en el metabolismo celular de cobre. La lista de estos genes se observa a la derecha de las imágenes. En el panel superior se observa el resultado de hibridación de las células control y en el inferior el resultado de hibridación de las células tratadas. Cada una de las sondas fue sembrada en dos cantidades (100 y 20 ng DNA/punto) y se indica su posición con el número que le corresponde en la columna de la derecha. *atox*, chaperona de *atp7a*, *g3p*, gliceraldehído-3-fostato deshidrogenasa, *ccs*, chaperona de *sod*, *b-sod*, baja identidad superóxido dismutasa, *sod*, súper óxido dismutasa, *app*, precursor del β -amiloide, *a-ws*, alta identidad *atp7b*, *b-ws*, baja identidad *atp7b*, *a-mk*, alta identidad *atp7a*, *b-mk*, baja identidad *atp7a*, *pGEM-T*, vector de clonamiento, *mt pc3.1*, metalotioneína inserta en vector *pc3.1*, *mt*, metalotioneína, *prpc*, proteína priónica, *dH₂O*, agua destilada, *collier* y *d-rap*, genes de *Drosophila melanogaster*.

Una sonda larga necesitaría más espacio para encontrar a su blanco (por eso es más eficiente en concentraciones menores), pero una vez que lo encuentra es más eficiente en el proceso de 'acomodarse' hasta un alineamiento perfecto. Esto, porque su longitud le permitiría anclarlo por alguna región hasta que se

estabilicen las hibridaciones correctas. Por otra parte, una sonda más pequeña necesitaría una mayor cantidad sembrada en la matriz para generar una señal que entregue información confiable. Este aumento de moléculas sembradas por punto podría, sin embargo, generar cierta interferencia entre ellas al momento de encontrar a su blanco en la muestra.

Nota: Dado que algunas de las hibridaciones no mostraron resultado alguno para analizar, se realizó otro experimento para evaluar en qué paso del procedimiento estaba la causa. Se evaluó una de las muestras a la que se había agregado un exceso del fragmento de *actina* marcada radiactivamente por un tiempo tres veces mayor al utilizado para marcar fragmentos. Se hibridó y lavó la membrana sin cambiar las condiciones establecidas y luego de cuatro horas de exposición a -80°C se obtuvo una señal perfectamente analizable en la posición correspondiente a *actina* en cuatro de sus cinco diluciones sembradas. Con este antecedente, se realizó las siguientes hibridaciones aumentando el tiempo de marcaje de las muestras de cDNA de una a tres horas.

b) Evaluación del límite de detección de diferentes cantidades de sonda versus distintas concentraciones del fragmento en la muestra: Se sembró las membranas con las cuatro sondas de *atp7a* y *atp7b* de alta y baja identidad en cuatro diferentes cantidades y cada una de ellas en triplicado. Se hibridó dichas membranas con una muestra algo más compleja: sólo un fragmento marcado y luego el fragmento de una atp-asa marcado diluido con el fragmento no marcado de la otra atp-asa (una botella por cada fragmento marcado). En cada una de las botellas se dispuso un fragmento de atp-asa marcado con $\alpha^{32}\text{P}$ -dCTP y se agregó la cantidad necesaria de otro fragmento no marcado para variar las proporciones de 1/10 o 1/100 con respecto a la masa del fragmento marcado (ver sección 2.15.2.). El análisis densitométrico de la intensidad de los puntos fue graficado como porcentaje de la señal de máxima intensidad (en cada membrana, de cada fragmento consigo mismo). Se observó señal aún en las cantidades más pequeñas de sonda sembrada y no se observó una relación entre la intensidad de la señal y la cantidad sembrada de ella. Esto indicaría que el límite de detección está por sobre la mayor dilución del fragmento en la muestra (1/100) de cualesquiera de las sondas de las atp-asas (dato no mostrado).

En hibridación de las sondas con su blanco en la muestra cuando es la única molécula con la que puede hibridar se observan diferencias en la intensidad de la señal de hibridación según la cantidad de sonda sembrada en la membrana. Cuando el fragmento en la muestra no es su homólogo, la intensidad de la señal de hibridación es considerablemente menor (30%) que la intensidad de la señal de las sondas con un fragmento homólogo.

La Figura 10 muestra que las sondas de *atp7a* son capaces de discriminar su blanco en la muestra aunque con diferencias de sensibilidad que no parecen estar asociadas a la cantidad de sonda en la membrana. La hibridación de las sondas de *atp7a* de alta y baja identidad sembradas en 10 ng de DNA sugiere una mayor señal de hibridación cuando la molécula marcada en la muestra está más concentrada. Las otras cantidades de sonda sembrada entregan resultados que no muestran una relación entre la cantidad de sonda sembrada y la señal de hibridación con la molécula marcada en la muestra. La Figura 11 muestra los resultados de hibridación de los experimentos equivalentes con las sondas de *atp7b* de alta y baja identidad sembradas en membranas de nylon en cuatro diferentes cantidades. Los mínimos valores de intensidad de la señal de hibridación observada para la sonda de alta identidad de *atp7b* no pueden asociarse ni a cantidad sembrada ni a las diluciones de *atp7a* en la muestra. Por otro lado, los altos valores de las intensidades de la sonda sembrada a 100 ng DNA por punto sugerirían que con esa cantidad de moléculas en la membrana se genera un alto porcentaje hibridación cruzada. Decimos esto, porque para las diluciones de *batp7a* se obtienen señales de hibridación cuya intensidad corresponde a alrededor del 60% del máximo de la señal. En un ensayo de hibridación, esta intensidad de señal podría interpretarse expresión de *atp7b* cuando en realidad corresponde a la abundancia de *atp7a* (como se observa en este gráfico). Sin embargo, la casi nula señal detectada en los puntos con menores cantidades de sonda sembrada no permiten concluir la especificidad de detección de las sondas de *atp7b*. Las posibles causas de estos resultados en la intensidad de la señal de hibridación de las sondas de alta y baja identidad de *atp7a* y *atp7b* se comentan en la Discusión.

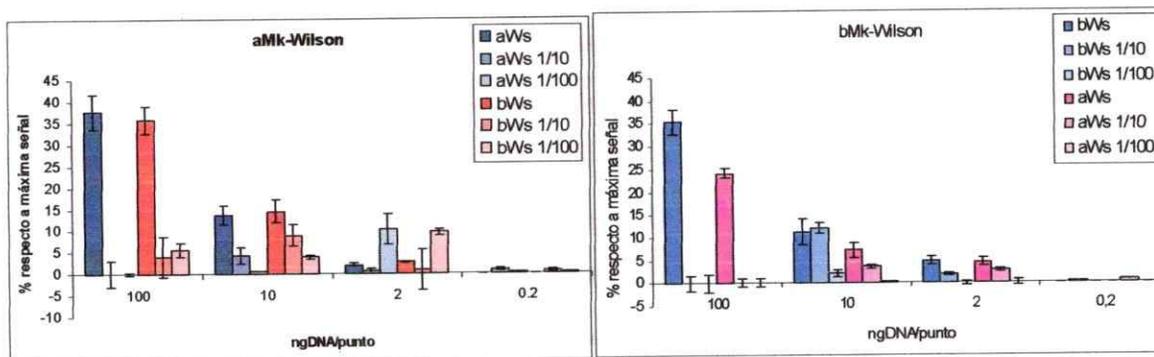


Figura 10: Sensibilidad de detección de las sondas de *atp7a*. Se graficó el porcentaje de hibridación de las sondas de alta y baja identidad de *atp7a* (Mk) con fragmentos de la secuencia de *atp7b* (Wilson, Ws) presentes en la muestra. El porcentaje de hibridación fue calculado en función del máximo valor de hibridación, obtenido al hibridar cada sonda de *atp7a* consigo misma (en puntos sembrados en la misma membrana). El fragmento de *atp7b* marcado con $\alpha^{32}\text{P}$ -dCTP fue diluido con el mismo fragmento de *atp7b* frío, según las proporciones indicadas en cada gráfico. Izquierda: Grandes cantidades sembradas del fragmento de alta identidad de *atp7a* detectan la secuencia de *atp7b* en la muestra sólo cuando es la única molécula presente en la muestra, mientras que las cantidades pequeñas de sonda sembrada (2 y 0,2 ng DNA/punto) no muestran diferencias de sensibilidad ante las distintas concentraciones de *atp7b* en la muestra. Derecha: La hibridación del fragmento de baja identidad de *atp7a* no permite realizar comentarios concluyentes. Las intensidades de la señal de hibridación en 100 como con 0,2 ng DNA/punto no parecen relacionarse con la concentración de molécula marcada en la muestra. Las cantidades intermedias (10 y 2 ng DNA/punto) sí permiten observar diferencias y sugieren relacionarse con la dilución de *atp7b* en la muestra. En los resultados de ambas sondas, los porcentajes de hibridación (con respecto a su homólogo en la membrana) son menores que el valor que indica hibridación cruzada y no parece haber diferencias en la especificidad de las los fragmentos de alta y baja identidad.

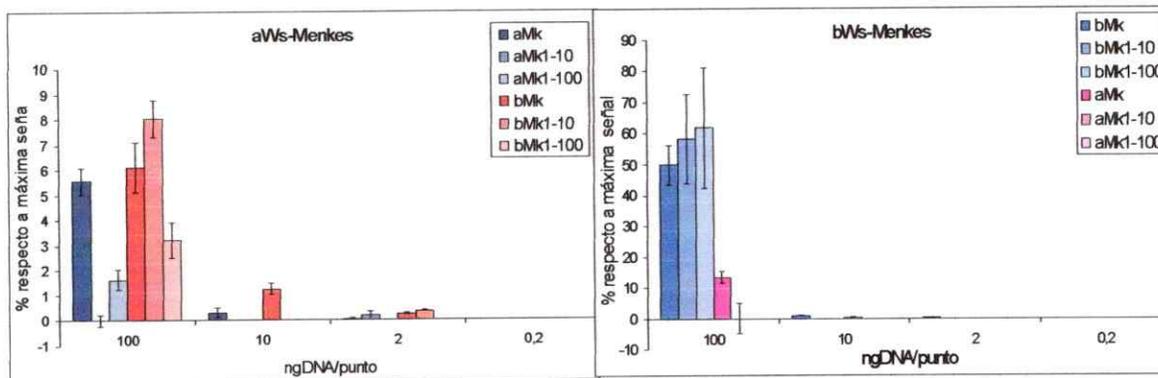


Figura 11: Sensibilidad de detección de las sondas de *atp7b*. Se graficó el porcentaje de hibridación de las sondas de alta y baja identidad de *atp7b* (Ws) con fragmentos marcados de la secuencia de *atp7a* (Menkes, Mk) en la muestra. El porcentaje de hibridación fue calculado con respecto al máximo valor de hibridación, obtenido al hibridar cada sonda de *atp7b* consigo misma. El fragmento de *atp7a* marcado con $\alpha^{32}\text{P}$ -dCTP fue diluido con *atp7a* frío según las proporciones indicadas en cada gráfico. Izquierda: No se observa consistencia en los datos de hibridación de sonda de alta identidad de *atp7b*, los pequeños valores de intensidad de la señal no se relacionan con la cantidad de sonda sembrada. Derecha: A pesar de que no se observa señal de hibridación en cada una de las cantidades de sonda sembradas, este fragmento de *atp7b* parece ser poco específico pues muestra un alto nivel de hibridación con el fragmento marcado de *atp7a* en la muestra.

c) Primera hibridación a partir de RNA total: Se sembró membranas en cuadruplicado y en cuatro distintas cantidades las sondas de los genes *atp7a*, *atp7b*, *ccs*, *sod1*, *mt* y *prpc* clonadas en el vector pGEM-TTM Easy. Se hibridó estas membranas con cDNA de hebra simple marcado con $\alpha^{32}\text{P}$ -dCTP sintetizado con oligo dT a partir de RNA total de células Caco2 y HepG2 tratadas y no tratadas con 100 μM de Cu-His. Luego de una exposición toda la noche a $-80\text{ }^\circ\text{C}$ no se observó señal de hibridación en ninguna de las membranas para ninguno de los fragmentos sembrados, aún cuando se expuso una nueva placa a estas membranas por 15 días (vida media del isótopo).

d) Segunda hibridación a partir de RNA total: Con el fin de aumentar la representación de los genes blanco en la muestra, se sintetizó cDNA de hebra simple con los partidores antisentido de las sondas sembradas en la membrana a partir de RNA total de células HepG2 y Caco2 tratadas y no tratadas con 100 μM de Cu-His. Este cDNA se hibridó con membranas idénticas a las del paso anterior. No fue posible observar señales luego de una exposición toda la noche y luego de una exposición de 15 días, se advierte una señal débil correspondiente a los 'house keeping' (*act*, *tub* y *g3p*).

e) Tercera hibridación a partir de RNA total: Se ajustó el protocolo de hibridación seguido en el paso d), aumentando el tiempo de marcaje de las muestras de una a tres horas y se realizó un nuevo ensayo. Se hibridó durante toda la noche membranas sembradas con cDNA de hebra simple sintetizado con partidores antisentido de las sondas. Se las sembró en cuadruplicado y en cuatro concentraciones (100, 10, 2 y 0,2 ng DNA/punto). En las placas radiográficas fue posible observar señal en algunas de las concentraciones de algunas de las sondas como las de los genes 'house keeping', *mt* y *sod1* de HepG2. Se observó señal en las sondas de *atp7a* y *atp7b* y se realizó un breve análisis para estos dos genes aunque la señal de hibridación no era intensa (Figura 12).

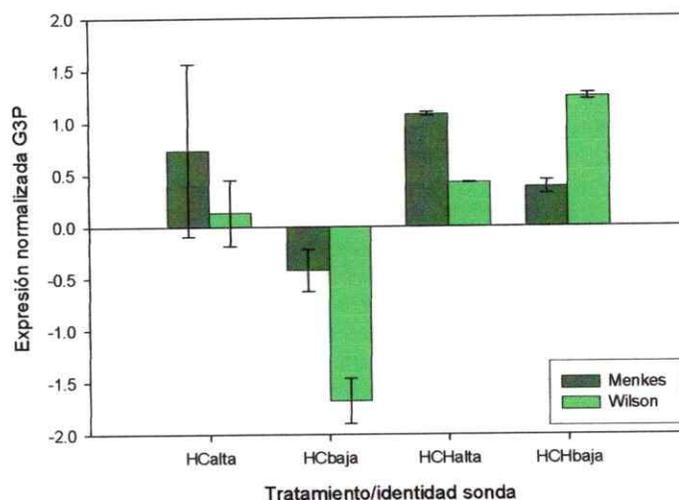


Figura 12: Sensibilidad de las sondas de alta y baja identidad de *atp7a* y *atp7b* en células HepG2 tratadas con cobre. Las células HepG2 fueron tratadas o no por 24h con 100 μ M CuHis. Con partidores antisentido de las sondas sembradas en la membrana, se sintetizó cDNA a partir de RNA total de estas células marcándolo con α^{32} P-dCTP. Este cDNA se hibridó toda la noche con membranas que tenían las sondas de baja identidad sembradas en cuadruplicado en cuatro diferentes cantidades. Si bien la intensidad de las señales de hibridación para la mayoría de las sondas no fue acentuada, en el gráfico se muestra la intensidad de señal de los fragmentos de alta y baja identidad de *atp7a* y *atp7b* sembrados en 100ng DNA por punto. Esta señal fue normalizada con respecto a la expresión del 'house keeping' gliceraldehído-3-fosfato deshidrogenasa. En las células control (HC) el comportamiento de la señal no permite evaluar la sensibilidad de ninguno de los fragmentos ya que los valores de expresión normalizada de las sondas de alta y baja identidad de ambas *atp*-asas no son congruentes. En las células expuestas a cobre (HCH) advertimos que la sonda de baja identidad de *atp7b* evidencia un aumento de la expresión de esta *atp*-asa (responsable de la excreción del metal). Considerando que el criterio de selección de las sondas privilegiaba su especificidad con el gen blanco, esperamos que el nivel de expresión observado con las sondas de baja identidad represente la situación real. HC, células HepG2 control; HCH, células HepG2 expuestas a cobre; alta, sonda de alta identidad de las *atp*-asas; baja, sonda de baja identidad.

3.2.2. Aplicación de sondas generadas.

Una vez concluidos los experimentos diseñados inicialmente para este trabajo de Memoria de título, el Laboratorio de Bioinformática y Expresión Génica (INTA, Universidad de Chile) utilizó las sondas aquí diseñadas para estudiar los niveles de expresión de los genes involucrados con el metabolismo celular de cobre en las líneas celulares de fibroblastos de ratón 804 (MT I/II) y 805 (wt). Para realizar estos ensayos, sembraron en cuadruplicado dos cantidades de sondas en membranas de nylon (algunas generadas en este trabajo y otras incluidas por ellos). Estas membranas fueron hibridadas con cDNA de las líneas celulares recién mencionadas. Los datos obtenidos se clasificaron de tres formas: M3 (el promedio de las tres

intensidades con menos desviación estándar), Prom. (el promedio de las cuatro intensidades) y Máx. (el máximo valor de intensidad de entre los puntos sembrados).

Lucía Tapia (Tapia y col. 2004) estudió los cambios de expresión en genes relacionados con el metabolismo celular de cobre en respuesta a la exposición a cobre en las líneas celulares 804 y 805. Observó que la expresión de tres de estos genes (*ccsI*, *mtI* y *sodII*) aumenta en células silvestres ante la exposición a cobre, sin embargo su expresión se mantiene o disminuye ligeramente en células mutantes.

4. DISCUSIÓN

Como se ha mencionado, los métodos de alineamiento en general favorecen la detección de secuencias o de dominios en ellas que presenten un alto nivel de similitud con otras secuencias presentes en la base de datos. Para ello, utilizan aproximaciones (métodos heurísticos) que permiten el hallazgo de soluciones ahorrando tiempo de búsqueda. Por ambas razones, resulta infructuoso recurrir a estos programas para encontrar regiones que representen específicamente a una única secuencia. Con estas regiones específicas se podría evitar que la secuencia seleccionada genere un reconocimiento cruzado por más de una secuencia presente en la mezcla de secuencias a hibridizar al ser utilizada como sonda en microarreglos/macroarreglos de DNA.

Considerando entonces que la utilización de métodos heurísticos no permite la elección de regiones que puedan servir como sonda en microarreglos/macroarreglos de DNA (porque privilegia la velocidad ante la sensibilidad) y que ClustalW busca el mejor alineamiento global para una secuencia o grupo de ellas, se hizo necesario desarrollar y automatizar un algoritmo (software) que seleccionara las regiones distintivas de una secuencia para con el resto de ellas. Para esto, se privilegió el hallazgo de diferencias penalizando las identidades entre secuencias de, en este caso, DNA.

Con el propósito de cuantificar el porcentaje real de identidad entre los fragmentos ya seleccionados con el programa desarrollado en esta Memoria de título (Redigen) y los genes listados en la Tabla 1, se decidió usar ClustalW. Este procedimiento permitió además descartar alguna(s) posición(es) en la(s) que la sonda completa pudiese tener hibridación cruzada y producir resultados de difícil interpretación. Aún cuando sabemos que la comparación es sólo aproximada y no representa la hibridación real de la sonda en las condiciones experimentales, este procedimiento permite estimar con mayor precisión la posibilidad de hibridación cruzada de las secuencias de interés, información que, estimamos, puede contribuir a incrementar

la calidad de las sondas que se utilizaran en el arreglo. El aporte del procedimiento que utilizamos en la interpretación de resultados obtenidos en ensayos de hibridación utilizando microarreglos ha sido discutido por Evertsz y cols. (2001).

4.1. ANÁLISIS DE LAS CARACTERÍSTICAS DE OPERACIÓN DE PROGRAMA.

Para una mejor comprensión de la selección de parámetros porcentuales, tal vez sea necesario aclarar que al aumentar estos valores los eventos de alta identidad ocurren con mayor probabilidad. Por ejemplo: un punto en la curva de porcentajes 80-90 con un valor de 15.000 UA tiene a lo más 15 posiciones en las que tiene entre 20 y 100% de identidad. El límite inferior es un valor bajo y es muy probable que dos secuencias cualesquiera tengan ese grado de similitud, por eso en las curvas generadas por estos parámetros los valores son frecuentemente altos.

Ya que en general dos secuencias no relacionadas pueden llegar a parecerse hasta alrededor de un 45%, decidimos no utilizar parámetros porcentuales sobre el 60% en la comparación de secuencias puesto que, de ser así, siempre se obtendrán valores altos que no permiten discriminar regiones de alta o baja identidad en las curvas. Por otra parte, valores inferiores al 40% generan regiones estrechas que dificultan la selección de fragmentos lo cual podría generar un grado importante de hibridación cruzada, pues se pasan por alto identidades inferiores al 60%. Por lo tanto y para obtener un rango de resultados que resolviese bien regiones de baja identidad, se decidió utilizar los parámetros 50-70 en la comparación de secuencias.

En una primera fase, se decidió comenzar con el análisis de patrones de identidad entre secuencias utilizando solamente *atp7a* y *atp7b* debido a su alto grado de similitud (53,7% identidad en secuencia nucleotídica, 73% homología en secuencia aminoacídica) pues los posibles problemas de hibridación cruzada en hibridaciones experimentales podrían entregar resultados más interesantes de estudiar.

Las curvas generadas por las secuencias completas de *atp7a* y *atp7b* aparecen con formas similares y ligeramente desplazadas una con respecto a la otra, insinuando una relación entre ambas secuencias. Esto llevó a una comparación sólo de las regiones codificantes de dichas secuencias, lo que generó curvas que aparecen casi superpuestas. El resultado sugiere que las mayores diferencias entre dichas secuencias se han

generado en la región 5' no codificante, ya sea por reiteradas inserciones en la secuencia de *atp7b* o deleciones en *atp7a*. Esto, porque cuando es *atp7b* la secuencia móvil, la curva de las secuencias completas aparece después de la curva de las regiones codificantes y cuando la secuencia móvil es *atp7a*, las curvas aparecen casi superpuestas. Es decir, antes de la región codificante las variaciones de *atp7a* son mínimas mientras que las variaciones en la región 5' de *atp7b* son mucho mayores.

Este desplazamiento observado (cuantificado en número de pb) coincide con los nucleótidos de inicio de la región codificante mostrados en la base de datos (*atp7a*, 14.4489; *atp7b*, 163.4551). Los valores observados en los desplazamientos de dichas curvas coinciden con los registrados en la Tabla 3 para la ventana de 300pb y parámetros porcentuales 50-70. Los valores de desplazamiento se mantienen constantes dentro de un rango de tamaños de ventana y de parámetros porcentuales, aunque con $j-k = 60-80$ se observa una tendencia a la superposición de las curvas de identidad. La finalidad de estos análisis no fue definir regiones de alta o baja identidad con el programa Redigen, sino describir el comportamiento del programa y rescatar la información adicional posible de obtener de las curvas que genera, pues podrían ser útiles para expresar gráficamente las similitudes entre secuencias.

En la comparación con distintos tamaños de ventana se observó que con ventanas pequeñas se generan curvas con valores de identidad altos sin importar el largo de las secuencias ingresadas. Algo parecido ocurre con las secuencias muy largas, como *cox17* (acceso: AB047323) que aún con ventanas de 300pb muestra valores altos de UA en las curvas generadas. Esto sugiere que existe una mayor probabilidad de que la ventana de búsqueda encuentre alguna posición con un alto grado de identidad cuando las secuencias son extensas. Este fenómeno se observa más claramente en la Figura 13. Al comparar los valores de las curvas (expresados en UA) con la longitud de la secuencia con la que se compara (ver tabla adjunta), se ve que existe una relación directa entre ambas variables aún cuando se hubiese utilizado una ventana de gran tamaño (600 pb).

Gen	Longitud Decrec.(pb)
<i>atp7b</i>	4854
<i>atp7a</i>	4575
<i>cp</i>	3363
<i>app</i>	3058
<i>nramp</i>	2261
<i>prpc</i>	2153
<i>m-ctr</i>	1423
<i>ccs</i>	1052
<i>sod</i>	535
<i>atox</i>	432
<i>mt</i>	391

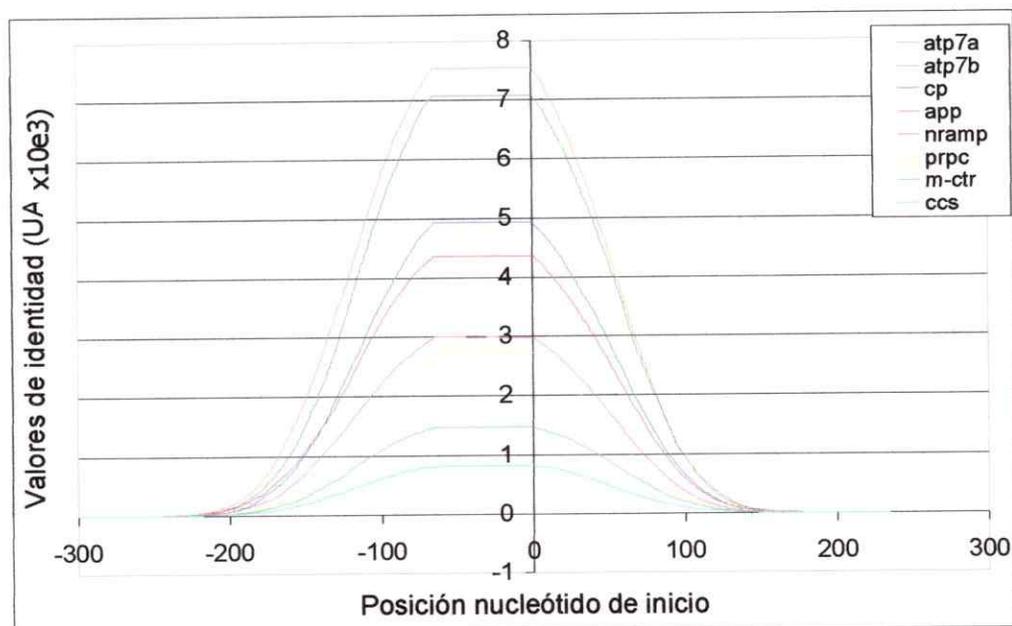


Figura 13: Aumento en los valores de UA al aumentar la longitud de las secuencias comparadas en las curvas generadas con el programa Redigen. Se comparó *sod1* con las secuencias completas de los otros genes involucrados en el metabolismo celular de cobre utilizando una ventana de 600pb y parámetros porcentuales 60 – 80. Mientras mayor es la longitud de la secuencia fija (ver tabla a la izquierda) mayores son los valores que se observan en las curvas de identidad. Esto es independiente del largo de la secuencia móvil y del largo de la ventana utilizada en la comparación. Las secuencias de *atox* y *mt* no aparecen en este gráfico porque la longitud de estos genes es menor al tamaño de ventana escogido para la comparación.

El estudio de las curvas indica que las secuencias que no poseen relación funcional o filogenética exhiben curvas con valores bastante bajos (150 UA en promedio) incluso cuando se utilizan ventanas relativamente pequeñas (100pb). Estos valores indican que las similitudes entre estas secuencias (en este caso relacionadas con manejo celular de cobre) no son mayores a las generadas por secuencias que no tienen ninguna relación funcional entre si. De este modo, las sondas pueden ser seleccionadas desde cualquier región de la secuencia de interés pues tienen una baja probabilidad de generar hibridación cruzada.

De la gran cantidad de curvas de identidad analizadas concluimos que la comparación de secuencias con ventanas de 300pb brinda un margen de ± 50 pb para la selección de partidores, pues las curvas casi no se modifican en ese rango de tamaño de ventana. Esto significa que a partir de una curva generada con 300pb de tamaño de ventana permite seleccionar partidores que amplifiquen fragmentos con un tamaño entre 250 y 350 pb.

Experimentalmente, las sondas no mostraron resultados que permitieran concluir la utilidad del programa Redigen en la selección de las mismas. Las razones de ello son principalmente experimentales, pues no se consideró en todo el proceso de hibridación algunos criterios que en ese momento no estaban documentados. Algunos de esos factores son el cambio de concentración y viscosidad de la solución de sonda al momento de la siembra y verificación de la calidad de la siembra una vez fijado el DNA en las membranas. Aunque el tiempo requerido para poner a punto las condiciones de hibridación (tiempo de marcaje) y de evaluación de las opciones de siembra de las sondas en la matriz (si purificadas o clonadas en el vector) fue superior al calculado inicialmente para realizar un número mayor de pruebas, nuestros resultados permiten presumir que un aumento en la concentración de sonda sembrada en la membrana incrementa la probabilidad de detectar a su homólogo en la muestra a pesar de los posibles impedimentos estéricos causados por su longitud. Esto, porque tiene una mayor probabilidad de encontrarlo en la muestra. Por otro lado, la menor cantidad de sonda sembrada localiza igualmente a su blanco en la muestra, pero la cantidad no parece ser suficiente como para mostrar una señal en las placas radiográficas expuestas toda una noche a la marca de $\alpha^{32}\text{P}$ -dCTP.

Muchos de los ajustes al procedimiento de hibridación afectaron el análisis de la expresión de los genes involucrados en el metabolismo celular de cobre en HepG2 (Figura 12). Además de las causas antes mencionadas, la carencia de señales que puedan interpretarse en el análisis de expresión génica de las células control puede atribuirse a un error de procedimiento en la obtención de RNA total. Si bien el ajuste de los tiempos de marcaje (según el tipo de muestra) produjo diferencias en los resultados de los ensayos de hibridación, aún deben ajustarse los parámetros de obtención de sondas y muestras y producción de membranas de certificada calidad. Antes de ser utilizadas en ensayos de hibridación, se debe comprobar que las sondas fueron efectivamente fijadas a la membrana y si la cantidad real en la membrana es la que el usuario deseaba sembrar inicialmente.

Los resultados experimentales de este trabajo no entregaron resultados que permitieran evaluar la sensibilidad de las sondas de baja identidad seleccionadas con el programa Redigen. Sin embargo, sí pueden ser empleados como antecedente que ayude a mejorar el procedimiento de las futuras hibridaciones con

micro o macroarreglos al nivel de laboratorio, ya que han quedado en evidencia aspectos que es importante mejorar.

4.2. ESTRATEGIAS ACTUALES DE SELECCIÓN DE SONDAS

Actualmente, son múltiples las estrategias desarrolladas para seleccionar la mejor sonda para detectar variaciones en la expresión génica en distintas condiciones experimentales. El diseño de sondas para los microarreglos construidos por Affymetrix y el software OligoArray diseñado por Rouillard y col. (2001) utilizan métodos heurísticos. En ambos casos se evalúa tanto la longitud de las sondas como la posibilidad de que generen hibridaciones cruzadas. Estas sondas no superan los 50 pb. En particular, Affymetrix (cuyas sondas no superan las 25pb) toma varias sondas de distintas regiones de un gen para tener una detección más precisa, tanto así que pueden diferenciar miembros de una familia, variantes de empalme alternativo de intrones y detectar la presencia de alelos.

Por su parte, el programa OligoArray inicia la búsqueda de oligos en dirección 3'→5' exigiendo que cumplan con un cierto porcentaje de identidad. Si el fragmento cumple con las exigencias de identidad, deja de buscar y estudia la posible formación de estructuras secundarias fuertes que puedan interferir con la hibridación.

La gran ventaja del reducido tamaño de estas sondas es su alta especificidad. En este contexto, el programa Redigen se convierte en una herramienta masiva para el hallazgo de regiones de baja identidad ya que busca optimizar la relación sensibilidad/tamaño de las sondas, de manera que éstas puedan ser obtenidas sin grandes problemas en los laboratorios tradicionales con técnicas ya conocidas (por ejemplo, obtención de sondas como productos de amplificación vía PCR).

Existe, eso sí, otro algoritmo que busca regiones específicas llamado LALIGN (Evertsz y col., 2001). El programa LALIGN también busca identidades entre las secuencias que compara y entrega la longitud del mejor alineamiento y el porcentaje de identidad en este segmento. Los autores que lo utilizan consideran que la combinación de ambos parámetros puede ser un buen predictor de hibridaciones cruzadas (Southern y col., 1999). Sin embargo, ya que el uso de este algoritmo comprueba que se produce hibridación cruzada

entre secuencias de genes de la misma familia, no nos permite discriminar la expresión de los genes que poseen las secuencias seleccionadas como sondas.

La principal desventaja del programa Redigen es que no se aplica fácilmente a genomas y no es posible aplicarlo a grupos muy numerosos de genes, básicamente porque la búsqueda de identidades nucleótido por nucleótido toma demasiado tiempo aunque se utilice un procesador de gran capacidad de cálculo. Comparando la velocidad de análisis del programa OligoArray (oligos de 50pb en 6343 genes en 24 horas con un procesador de 700 MHz), con la del programa Redigen (fragmentos de 300pb en 12 genes en 1½ hora con un procesador de 800 MHz) en términos del número de genes explorados por hora, OligoArray es más 30 veces más veloz. El programa Redigen no estima la formación de estructuras secundarias que sí realiza OligoArray y que en el caso de Affymetrix no se producirían, por la escasa longitud de sus sondas. Sin embargo nuestro programa revisa la secuencia completa del gen sin importar su longitud, mientras que OligoArray se detiene apenas encuentra un oligonucleótido que tenga las características requeridas.

Dentro del grupo de genes estudiados no se observan regiones con un nivel de identidad superior al 50% que permitiera sugerir que las secuencias codifican proteínas relacionadas metabólicamente utilizando dominios comunes. Sin embargo, al comparar las secuencias nucleotídicas de dos proteínas que interactúan directamente se obtuvo un resultado interesante. Se comparó las secuencias de los genes que codifican proteínas (*sod1*, *citocromo c oxidasa*, *atp7alb*) que interactúan con chaperonas de cobre (*ccs*, *cox17*, *atox1*, respectivamente). Sólo el par de secuencias correspondientes a *sod1* y su chaperona *ccs* presentan un exclusivo par de regiones de alta identidad cuando se comparan con una ventana de 250pb y parámetros $j-k=50-70$. La comparación de las secuencias proteicas de *Sod1* y *Ccs* con el programa Clustal W mostró un 27% de identidad y una 43% de similitud en residuos aminoacídicos (calculados con respecto a la longitud de *Ccs*). Esta región de alta homología está ubicada casi al final de la secuencia de *Ccs* (residuos 85 a 200). Con el propósito de determinar si existe alguna relación entre altos niveles de identidad en los gráficos entregados por el programa Redigen y la interacción entre las proteínas codificadas por las secuencias que se

comparan, se revisó si el sitio que lleva cobre en ambas proteínas está codificado por dicha región de alta identidad con el programa de análisis de estructura de proteínas de SwissProt. En ambas proteínas (relativamente pequeñas) los residuos que unen cobre se encuentran dispersos en la secuencia aminoacídica lineal y no se ubican necesariamente dentro de las regiones de alta identidad que comparten las secuencias nucleotídicas de *sod1* y *ccs*. Se desprendió de esto que altos valores de identidad en las curvas generadas por el programa Redigen no se asocian necesariamente con una interacción entre las proteínas para las que las codifican ni con una conservación de los sitios de unión al metal.

4.3. ANÁLISIS DE RESULTADOS EN ENSAYOS DE HIBRIDACION

Con los resultados obtenidos del análisis experimental de las sondas seleccionadas con el programa Redigen pudimos rescatar principalmente la importancia de la siembra en distintas cantidades. En estudios posteriores realizados en laboratorio con fibroblastos de ratón, se observó que para dos sondas del mismo gen (*sod1*, ver Figura 8) había distintos grados de detección en las distintas cantidades sembradas y esto parecía tener relación inversa con el tamaño de la sonda, ya que una de ellas es el doble del tamaño de la otra. En este ensayo la sonda más larga tenía una mejor detección con una menor cantidad en la membrana, mientras que la sonda más corta tenía una mayor intensidad de señal en la mayor cantidad de siembra. En estos ensayos se tenía la cantidad de sonda desecada efectivamente unida a la matriz de nylon, por lo tanto se descartaban el efecto de la calidad del punto sembrado sobre la intensidad de la señal de hibridación. Si bien la alta densidad de siembra puede generar interferencia entre las moléculas sembradas, ésta se reduce si la sonda es más corta. De esta forma el incremento en la cantidad de sonda en la membrana tiene la ventaja de reconocer su blanco en la muestra incluso cuando está pobremente representada, como se observó en los experimentos de este trabajo en los que se diluyó la muestra.

Al realizar la hibridación con sólo un fragmento marcado en la muestra, las intensidades de la señal de hibridación tuvieron un máximo de 30% con respecto a la intensidad de la señal de hibridación de la sonda con su homólogo (máximo). El porcentaje sugiere que las sondas se comportaron como una sonda de cualquier gen y que el grado de hibridación cruzada que se generó no debiera diferir de la hibridación

cruzada producida entre dos secuencias no relacionadas. Esto, a pesar de que en este ensayo se cuantificó el nivel de hibridación entre moléculas de dos genes con un alto nivel de homología.

En los experimentos en que la sonda sembrada correspondía a los fragmentos clonados no se observó señal en las membranas reveladas. Podemos atribuir este problema tanto a la sonda como a la muestra. Al sembrar las moléculas clonadas en el vector pudo producirse una disminución importante de la concentración efectiva de sonda. Esta situación se previó al diluirlas, puesto que se sembró una mayor cantidad de DNA por punto, sin embargo parece ser que este aumento no fue suficiente y la concentración efectiva de la sonda cayó por debajo de los límites de detección de nuestro ensayo. No se sembró una cantidad aún mayor de sonda para evitar interferencias entre moléculas en el proceso de hibridación que finalmente disminuirían aún más la concentración efectiva de la sonda. Entre las posibilidades analizadas previamente, se verificó mediante alineamientos con ClustaW que el vector no generaba hibridación cruzada con ninguno de los genes bajo estudio. Por otro lado, si consideramos los efectos de la muestra en la ausencia de señal de estos experimentos, es posible que los genes estudiados estén escasamente representados en el RNA total por lo que su detección no es posible en estas condiciones de hibridación (recordemos que el mRNA corresponde a un 1 – 2% del total). Estudios realizados en el laboratorio con posterioridad al término de este trabajo de Memoria de título confirmaron la baja proporción de los transcritos estudiados en las líneas celulares utilizadas.

En todos los ensayos pudo tener un gran efecto sobre la intensidad de la señal de hibridación alguna falla en el proceso de fijación de las sondas a la membrana de nylon. En ninguno de los experimentos realizados en esta Memoria de Título se verificó que los fragmentos estuvieran covalentemente unidos a la membrana ni que la cantidad de sonda fijada fuera la que esperábamos sembrar. Tampoco se realizó un estudio de los efectos de la concentración de agentes desnaturantes en el proceso mismo de hibridación, ni de la estrictez de los lavados sobre la remoción de la sonda que se hubiera fijado a las sondas durante la hibridación.

Cuando se decidió aumentar la representación de las secuencias blanco en la muestra sintetizando cDNA a partir de RNA total con los partidores antisentido de las sondas, se aumentó también el tiempo de marcaje

pero no se obtuvo resultados que permitieran evaluar diferencias de expresión. Se evaluó entonces la posibilidad de realizar hibridaciones con cDNA sintetizado a partir de mRNA, para aumentar la concentración efectiva de las moléculas blanco en la muestra. Sin embargo, la masa de RNA total extraída no permitió la obtención de mensajero suficiente para la síntesis de cDNA. Finalmente, ya que no se pudo extraer suficiente mRNA para mejorar la muestra, se decidió mejorar la presentación de la sonda y se preparó una nueva serie de diluciones con fragmentos purificados (no insertos en el vector).

En los primeros revelados de las membranas hibridadas tuvimos un importante nivel de señal en zonas sembradas con altas diluciones de la sonda. Desconocemos a que se debió esta señal, pero la atribuimos a una deformación en la membrana causada por las puntas del sembrador al momento de sembrar. En estas cavidades se acumularía muestra marcada que no se remueve totalmente con los lavados posteriores.

La imposibilidad de evaluar diferencias de expresión principalmente en las hibridaciones de las sondas de *atp7b* sugieren un error bien en la amplificación de las sondas o de la dilución de ellas. También pudo deberse a defectos en el procedimiento de producción de las membranas, ya sea en el proceso de siembra o de fijación de las sondas, defectos que no alcanzaron a ser corregidos en el tiempo disponible para realizar los experimentos de este trabajo.

A pesar de no poder evaluar sensibilidad ni diferencias de expresión en ensayos de hibridación que utilizaron las sondas seleccionadas teóricamente con el programa Redigen, se consiguió una estrategia de clonamiento de ellas, buscando partidores restringidos a una región de baja identidad y a un tamaño de fragmento. Con dicha estrategia fue posible generar una librería con las secuencias de algunos de los genes involucrados en el metabolismo celular de cobre.

5. CONCLUSIÓN

El programa Redigen es de fácil manejo para los usuarios y se puede aplicar a cualquier grupo de genes que se estudie. Se originó pensando en una manera de seleccionar sondas específicas para estudios de expresión génica que fuera accesible a muchos laboratorios. El punto en el que diverge nuestro programa de los otros programas disponibles para comparación de secuencias es que hace que el número exacto de diferencias sea el que establezca el grado de identidad entre secuencias y no una aproximación del número de identidades. El uso de aproximaciones en la búsqueda descartaría secuencias que podrían ser buenas sondas y que no descamos perder.

Las proyecciones de uso del programa Redigen apuntan a la difusión y utilización en los laboratorios donde sea posible analizar genes con sondas que sean fragmentos de PCR y no con sondas sintetizadas *in situ* por maquinas que elevan su costo de producción. Para el futuro y como mejoras al programa podrían eliminarse algunos 'bugs' que requiere se requieren para reconocer las secuencias y analizarlas, hacer una interfase más amigable y más atractiva en su presentación.

6. REFERENCIAS BIBLIOGRÁFICAS

1. Alexandrov, N. 1992. Local multiple alignment by consensus matrix. *CABIOS* 8:339-345
2. Alwine, J.C., Kemp, D.J. y Stark, G.R. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Nat. Acad. Sci. USA* 74: 5350-5354.
3. Barton, G.J. y Sternberg, M.J.E. 1987. Evaluation and improvements in the automatic alignment of protein sequences, *Protein Eng.* 1:89-94
4. Bertucci, F., Bernard, K., Liorod, B., Chang, Y.C., Granjeaud, S., Birnbaum, D., Nguyen, C., Peck, K. y Jordan, P.R. 1999. Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum. Mol. Genet.* 8(9):1715-22. Errata en *Hum. Mol. Genet.* 8(11):2129.
5. Birnboim, H.C. y Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513.
6. Birnboim, H. 1983. A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol.* 100: 243 - 255.
7. BLAST: www.ncbi.nlm.nih.gov/blast
8. Boguski, M.S., Hardison, R.C., Schwarz, S. y Miller, W. 1992. Analysis of conserved domains and sequence motifs in cellular regulatory proteins and locus control regions using new software tools for multiple alignment and visualization. *New Biol.* 483:247-260
9. Brazma, A. y Vilo, J. 2000. Gene expression data analysis. *FEBS Letters* 480:17-24.
10. Bull, P. C., Thomas, G. R., Rommens, J. M., Forbes, J. R., and Cox, D. W. 1993. *Nat Genet* 5, 327-337
11. Cater, M. A., Forbes, J., La Fontaine, S., Cox, D., and Mercer, J. F. 2004 *Biochem J Pt*
12. Caviani Pease, A., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P. y Fodor, S.P.A. 1994. Light generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Nac. Acad. Sci. USA* 91:5022-5026.
13. Chelly, J., and Monaco, A. P. 1993. *Nat Genet* 5, 317-318
14. Clontech: www.clontech.com
15. ClustalW: align.genome.jp
16. Corpet, F. 1988. Multiple sequence alignments with hierarchical clustering. *Nucleic Acids Res.* 16:10881-10890
17. Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. y Trent, J.M. 1999. Expression profiling using cDNA microarrays. *Nature Genetics* 21:10-14
18. Eickhoff, B., Kom, B., Schick, M., Poutska, A. y van der Bosh, J. 1999. Normalization of array hybridization experiments in differential gene expression analysis. *Nucleic Acids Res.* 27:e33.
19. Eisenberg, E. y Levanon, E. Y. 2003. Human housekeeping genes are compact. *Trends in Genetics* 7: 362 - 365.
20. Eisen, M.B. y Brown, P.O. 1999. DNA arrays for analysis of gene expression. *Methods in Enzimology* 303:179-205.
21. Ekins, R. & Chu F.W. Microarrays: their origins and applications. 1999. *Trends Biotechnol.* 19:217-218.
22. Epstein, C.B. y Butow, R.A. 2000. Microarray technology-enhanced versatility, persistent challenge. *Current Op. Biotech.* 11:36-41.
23. Everst, E.M., Au-Young, J., Ruvolo, M.V., Lim, A.C. y Reynolds, M.A. 2001. Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. *Biotechniques* 5:1182-1186.

24. Fejzo, M.S. y Slamon, D.J. 2001. Frozen tumor tissue microarray technology for analysis of tumor RNA, DNA and proteins. *American Journal of Pathology* 159:1645-1650.
25. Fotin, A.V., Drobyshev, A.L., Proudnikov A.N. y Mirzabekov, A.D. 1998. Parallel thermodynamic analysis of duplexes oligodeoxirribonucleotides microchips. *Nucleic Acids Res.* 26:1515-1521.
26. Gao, X., LeProust, E., Zhang, H., Srivannavit, O., Gulari, E., Yu, P., Nishiguchi, C., Xiang, Q. y Zhou, X. 2001. A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res.* 29:4744-4750.
27. Godzik, A y Skolnick, J. 1994. Flexible algorithm for direct multiple alignment of protein structure and sequences. *Comput. Appl. Biosci.* 10(6): 587-596
28. Golub T.R., Slonim D.K., Tamayo P., Huard C., Gaasenbeek M., Mesirov J.P., Coller H., Loh M.L., Downing J.R., Caligiuri M.A., Bloomfield C.D., Lander E.S. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286(5439):531-7
29. Gross, C., Kelleher, M., Iyer, V.R., Brown, P.O. y Winge, D.R. 2000. Identification of the copper regulon in *Saccharomyces cerevisiae* by DNA microarrays. *J. Biol. Chem.* 275:32310-32316.
30. Gunshin, H., Mackenzie, B., Berger, U. V., Gunshin, Y., Romero, M. F., Boron, W. F., Nussberger, S., Gollan, J. L., and Hediger, M. A. 1997. *Nature* 388, 482-488
31. Harrison, M. D., Jones, C. E., Solioz, M., and Dameron, C. T. 2000. *Trends Biochem. Sci.* 25: 29-32
32. Inoue, H., Nojima, H. Y Okayama, H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96: 23 – 28.
33. Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C., Trent, J.M., Staudt, L.M., Hudson, J. Jr., Boguski, M.S., Lashkari, D., Shalon, D., Botstein, D. y Brown, P.E. 1999. The transcriptional program in the response of human fibroblast serum. *Science* 283(5398):83-87.
34. Jaccoud, D., Kaiman, P., Feinstein, D. y Kilian, A. 2001. Diversity arrays: a solid-state technology for sequence information independent genotyping. *Nucleic Acids Res.* 29:e25.
35. Jenssen, T., Langaas, M., Kuo, W.P., Smith-Sørensen, B., Myklebost, O. y Hoving, E. 2002. Analysis of repeatability in spotted cDNA microarrays. *Nucleic Acids Res.* 30:3235-3244.
36. Ju, J., Ruan, C., Fuller, C.W., Glazer, A.N. y Mathies, R.A. 1995. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Nac. Acad. Sci. USA* 92:4347-4351.
37. Kafatos, F.C., Jones, C.W. y Efstratiadis, A. 1979. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res.* 7:1541-1552.
38. Khitrov, G. 2001. Use of inexpensive dyes to calibrate and adjust your microarray printer. *BioTechniques* 30:748.
39. Labbé, S. y Thiele, D.J. 1999. Pipes and wiring: the regulation of copper uptake and distribution in yeast. *Trends in Microbiology* 7:500-505.
40. Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. y Davis, R.W. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Nac Acad. Sci. USA* 94:13057-13062.
41. Lee, J., Pena, M. M., Nose, Y. y Thiele, D. J. 2002. *J Biol Chem* 277, 4380-4387
42. Li, C. y Wong, W.H. 2001. Model-based analysis oligonucleotide arrays: expression index computation and outlier detection. *Proc. Nac. Acad. Sci. USA* 38:31-36.
43. Liang, P. y Pardee, A.B. 1992. Differential display of eukaryotic messengers RNA by means of the polymerase chain reaction. *Science* 257:967-970.
44. Linder, M. C., Wooten, L., Cerveza, P., Cotton, S., Shulze, R., and Lomeli, N. 1998. *Am J Clin Nutr* 67, 965S-971S
45. Lipman D.J. y Pearson W.R. 1985. Rapid and sensitive protein similarity search, *Science*, 227:1435-1444
46. Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R. y Lockhart, D.J. 1999. High-density synthetic oligonucleotide arrays. *Nature Genetics* 21:20-24.

47. Livingstone, C.D. y Barton, G.J. 1993. Protein sequence alignments: a strategy for hierarchical análisis of residue conservation. *CABIOS* 9:745-756
48. Maskos, U. y Southern, E.M. 1993. A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesized on a glass support. *Nucleic Acids Res.* 21:4663-4669.
49. Mattie, M. D., and Freedman, J. H. 2004. *Am. J. Physiol. Cell* 286, C293-301.
50. McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T. y Hinsberg, W. 1996. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresist. *Proc. Natl. Acad. Sci. USA* 93:13555-13560.
51. McGonigle, B., Keeler, S.J., Lau, S.M., Koeppe, M.K. y O'Keefe, D.P. 2000. A genomics approach to the comprehensive analysis of the glutathione S-transferase gene family in soybean and maize. *Plant Physiol.* 124:1105-1120.
52. Mercer, J. F., Livingston, J., Hall, B., Paynter, J. A., Begy, C., Chandrasekharappa, S., Lockhart, P., Grimes, A., Bhavé, M., Siemieniak, D., and et al. 1993. *Nat. Genet.* 3, 20-25
53. Mercer, J. F. 2001. *Trends. Mol. Med.* 7, 64-69
54. Mercer, J. F., Barnes, N., Stevenson, J., Strausak, D., and Llanos, R. M. 2003. *Biometals* 16, 175-184
55. Palmiter, R. D. 1998. *Proc. Natl. Acad. Sci. USA* 95, 8428-8430
56. Pascale, M. C., Franceschelli, S., Moltedo, O., Belleudi, F., Torrisi, M. R., Bucci, C., La Fontaine, S., Mercer, J. F., and Leone, A. 2003. *Exp Cell Res* 291, 377-385
57. Pearson W.R. y Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc. Nat. Acad. Sci. USA* 85:2444-2448
58. Petris, M. J., Mercer, J. F., Culvenor, J. G., Lockhart, P., Gleeson, P. A., and Camakaris, J. 1996. *Embo J* 15, 6084-6095
59. PubMeb: www.ncbi.nlm.nih.gov/entrez
60. Rae, T. D., Schmidt, P. J., Pufahl, R. A., Culotta, V. C., and O'Halloran, T. V. 1999. *Science* 284, 805-808
61. Rouillard, J.M., Herbert, C.J. y Zuker, M. 2002. Oligoarray: genome-scale oligonucleotide design for microarrays. *Bioinformatics* 18:486-487.
62. Sambrook, J., Fritsch, E. y Maniatis, T. 1989. *Molecular cloning: A laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory, NY.
63. Schena, M., Shalon, D., Davis, R.W. y Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-470.
64. Schena, M., Shalon, D., Heller, R., Cha, A., Brown, P.O. y Davis, R.W. 1996. Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Nac. Acad. Sci. USA* 93:10614-10619.
65. Schena, M., Heller, R.A., Theriault, P., Konrad, K., Lachenmeier, E. y Davis, R.W. 1998. Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* 16:301-306.
66. Schuchhardt, J., Beule, D., Malik, A., Wolsky, E., Eickhoff, H., Lehrach, H. y Herzog, H. 2001. Normalization strategies for cDNA microarrays. *Nucleic Acids Res.* 28:e47.
67. Schwarzhacher, T. y Heslop-Harrison, P. 2000. *Practical in situ hybridization*. Capítulos 3 y 7. BIOS Scientific Publishers Limited, The Bath Press, Bath, UK.
68. Smith, R.F. y Smith, T.F. 1992. Pattern-induced multi-sequence alignment (PIMA) algorithm employed secondary structure-dependent gap penalties for use in comparative protein modeling. *Protein Eng.* 5:35-41
69. Southern, E., Mir, K. y Shchepinov, M. 1999. Molecular interactions on microarrays. *Nature Genetics* 21:5-9.
70. Sutcliffe, J.G., Foye, P.E., Erlander, M.G., Hilbush, B.S., Bodzin, L.J., Durham, J.T. y Hasel, K.W. 2000. TOGA: an automated parsing technology for analyzing expression of nearly all genes. *Proc. Nac. Acad. Sci. USA* 97:1976-1981.
71. Swiss prot: geno3d-pbil.ibcp.fr

72. Tapia, L., González-Agüero, M., Cisternas, M.F., Suazo, M., Cambiazo, V., Uauy, R. y González, M. 2004. Metallothionein is crucial for safe intracellular copper storage and cell survival at normal and supra-physiological levels. *Biochem. J.* 378:617-62
73. Taylor, W.R. 1987. Multiple sequence alignment by a pairwise algorithm. *Comput. Appl. Biosci.* 3:81-87
74. Tiquia, S.M., Wu, L., Chong, S.C., Passovets, S., Xu, D., Xu, Y. y Zhou, J. 2004. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *BioTechniques* 36:664-75.
75. Uauy, R., Olivares, M. y González, M. 1998, suppl. Essentiality of copper in humans. *Am. J. Clin. Nutr.* 67:952-959.
76. Velculescu, V.E., Zhang, L., Vogelstein, B. y Kinzler, K.W. 1995. Serial analysis of gene expression. *Science* 270:484-487.
77. Vulpe, C., Levinson, B., Whitney, S., Packman, S., and Gitschier, J. 1993. *Nat Genet* 3, 7-13
78. Winge, D.R., Jensen, L.T. y Srinivasan, C. 1998. Metal-ion regulation of gene expression in yeast. *Current Opinion in Chemical Biology* 2:216-221.
79. Yamaguchi, Y., Heiny, M. E., and Gitlin, J. D. 1993. *Biochem. Biophys. Res. Commun.* 197, 271-277
80. Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S. y Fu, X.D. 2002. Profiling alternative splicing on fiber-optic arrays. *Nat Biotechnol.* 20(4):353-8.
81. Ziauddin, J. y Sabatini, D.M. 2001. Microarrays of cells expressing defined cDNAs. *Nature* 411:107-110.



Mónica Cisternas Fierro (junio, 1978). Nací y viví en Santiago hasta el 2002; siempre con mis padres, mi hermano menor Rodrigo y la mascota de turno. Con ellos pasé un año en Argentina, movidos por el trabajo de mi padre, en donde encontré la parte sociable de mi personalidad. En mi paso por la enseñanza básica pensé que mi buen rendimiento de entonces me permitiría entrar sin problemas a Arquitectura o Diseño, pero en la educación media descubrí la maravilla de la biología. Hice algunos

intentos por entrar al mundo de la farándula a través del modelaje, pero quiso mi buena estrella alejarme del mal camino y llevarme por la senda de la biología en la Universidad de Chile. El año que entré a la universidad llegó la última integrante de la familia, Constanza. En mi muy agradable paso por la Facultad de Ciencias encontré algunos grandes amigos, presenté en un par de congresos (Sociedad de Biología, 2001; International Workshop on Iron and Copper Homeostasis, 2001), aporté a dos publicaciones (Revista Chilena de Entomología, 2000; Biochemical Journal, 2004) y conocí a Ernesto, con quien me casé el 2002. Con él nos mudamos a Concepción, donde actualmente vivimos muy felices con nuestro pequeño hijo Diego, de tres años y medio.