



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE QUIMICA

CARACTERIZACION MOLECULAR DE UN NUEVO GEN DEL SISTEMA NERVIOSO DE MAMIFERO

TESIS

QUE PARA OBTENER EL TITULO DE
QUIMICA FARMACEUTICA BIOLOGA

PRESENTA:

ROSALBA MEDINA BENITEZ



EXAMENES PROFESIONALES
MEXICO, D. F. FACULTAD DE QUIMICA 2005

m341633



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Jurado asignado:

Presidente **JESÚS FERNANDO MONTIEL AGUIRRE.**

Vocal **MARTHA LETICIA JIMÉNEZ PARDO.**

Secretario **BENITO ANTON PALMA.**

1er.Suplente **NORMA TREJO MEDINA.**

2°. Suplente **MARCO VELASCO VELÁSQUEZ.**

**Sitio en donde se realizó la tesis: Laboratorio de Neurobiología Molecular y
Neuroquímica de Adicciones. Subdirección
de Investigaciones Clínicas. Instituto
Nacional de Psiquiatría Ramón de la Fuente
Muñiz**

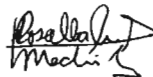
Asesor del tema: Dr. Benito Antón Palma.



Supervisor técnico: Q.F.B. Maura E. Matus Ortega.



Sustentante: Rosalba Medina Benítez.



AGRADECIMIENTOS

Al Instituto Nacional de Psiquiatría por las facilidades otorgadas para la realización de mi tesis de licenciatura.

A la fundación Gonzalo del Río Arronte por el apoyo financiero.

A mi asesor por participar en mi proceso de aprendizaje.

Al Dr. Leff por su apoyo para el registro de la secuencia clonada en la base de datos del GenBank, al Dr. Arreola por su apoyo en el uso de programas especializados de computadora para el análisis de la secuencia primaria de aminoácidos de la Antolefinina.

A mi asesor técnico Maura Matus por su paciencia para enseñarme los métodos de Biología Molecular empleados en la parte experimental de esta tesis.

A mis compañeros de laboratorio por su apoyo en todo momento.

A mis amigos de la facultad que me brindaron su cariño y apoyo en los momentos difíciles.

A mi padres Amadeo Medina M. Y Rosalba Benítez L. por su amor incondicional que me ha proporcionado siempre las fuerzas suficientes para lograr todas la metas que me propongo.

A mis hermanos Katia y Amadeo que me impulsan a superarme siempre con los consejos oportunos para vencer los obstáculos del camino.

ÍNDICE

	Página
I. Resumen	1
II. Introducción	
Antecedentes generales	
a) La información genética y su organización.	3
b) Estrategias experimentales empleadas usualmente en la caracterización de genes de interés.	4
c) Técnicas auxiliares para la caracterización de nuevos genes y sus productos.	9
d) Programas especializados de computadora para localizar secuencias consenso para las cuales ya hay funciones identificadas.	11
Antecedentes particulares	
e) Poros nucleares y señales para importación y exportación de proteínas al núcleo.	15
f) Importancia de las proteínas que contienen NLS y NES.	19
g) Recorte de las cadenas polipeptídicas.	21
i) La α -amidación.	23
III. Planteamiento del problema y justificación del trabajo experimental	24
IV. Hipótesis del trabajo experimental	25
V. Objetivos del trabajo experimental	26

VI. Materiales y métodos

- | | |
|---|----|
| a) Rastreo inmunológico de la biblioteca de ADNc de cerebro total de ratón. | 28 |
| b) Purificación del plásmido que contiene el ADNc que da origen a la proteína identificada con los anticuerpos. | 32 |
| c) Tratamiento con las enzimas de restricción <i>Xba</i> I y <i>Xho</i> I. | 32 |
| d) Análisis electroforético del inserto de ADNc en gel de agarosa. | 32 |
| e) Caracterización estructural del ADNc clonado a través de secuenciación nucleotídica automatizada con la ayuda del método de marcaje de nucleótido terminal fluorescente. | 33 |
| f) Búsqueda de secuencias homólogas a la secuencia nucleotídica del ARNm clonado con ayuda del programa BLAST (NCBI) de la base de datos del GenBank. | 34 |
| g) Búsqueda de secuencias homólogas a la secuencia de aminoácidos codificada por el ARNm clonado. | 35 |
| h) Análisis estructural de la secuencia primaria de aminoácidos obtenida para la identificación de la presencia de secuencias parciales de aminoácidos (secuencias consenso) con función identificada en otras proteínas. | 35 |

VII. Resultados

- | | |
|--|----|
| Resultado 1: Obtención de una clona inmunopositiva a partir del rastreo inmunológico efectuado. | 37 |
| Resultado 2: Obtención de un inserto de ADNc menor a 1kb de tamaño molecular purificado de la clona inmunopositiva en el ensayo. | 37 |
| Resultado 3: Caracterización de la secuencia primaria de nucleótidos obtenida del ADNc clonado. | 39 |
| Resultado 4: Resultado de la búsqueda de secuencias homólogas a la secuencia del ARNm clonado. | 41 |

Resultado 5: Análisis estructural de la secuencia primaria de aminoácidos de la Antolefinina.	44
Resultado 6: Resultado de la búsqueda de secuencias peptídicas homólogas a la Antolefinina.	46
Resultado 7: Predicción de los parámetros fisicoquímicos de la Antolefinina.	49
Resultado 8: La Antolefinina presenta una estructura secundaria no convencional.	49
VIII. Análisis de resultados y conclusiones.	51
IX. Aportaciones y perspectivas futuras de trabajo.	59
X. Bibliografía.	60
Anexo	65

ÍNDICE DE FIGURAS

Número y título de la figura:	Página
Figura 1: La información genética se encuentra organizada en los cromosomas.	3
Figura 2: Dibujo esquemático que muestra las diferencias entre la síntesis de ADNc y el procesamiento normal del ADN.	6
Figura 3: Secuenciación nucleotídica con ayuda de un secuenciador automático de un capilar.	7
Figura 4: Rastreo inmunológico de un gen de interés.	9
Figura 5: Micrografía electrónica y dibujo esquemático a nivel de un poro nuclear.	15
Figura 6: Dibujo esquemático que muestra el ciclo de importación y exportación hacia el núcleo para una molécula con una señal NLS o NES.	17
Figura 7: Diagrama estructural de la secuencia consenso para la α -amidación.	23
Figura 8: Mapa circular del vector pBluescript@SK.	29
Figura 9: Análisis electroforético del inserto de ADN en gel de agarosa.	37
Figura 10: Registro en la base de datos del GenBank de la secuencia primaria de nucleótidos del ARNm y de la secuencia de aminoácidos de la proteína codificada por el único marco de lectura abierto de traducción encontrado en este ARNm.	40
Figura 11: Dibujo esquemático del ARNm clonado.	39
Figura 12: Representación esquemática en formato del GenBank del estudio de alineamiento estructural entre el ARNm clonado con secuencias de ácidos nucleicos reportados en esta base de datos.	42

Figura 13: Posibles sitios funcionales y de procesamiento postraduccionales encontrados con ayuda del programa PROSITE.	45
Figura 14: Alineamiento de la secuencia primaria de las proteínas homólogas a la Antolefinina.	48
Figura 15: Predicción de la estructura secundaria.	50
Figura 16: Actividad moduladora de la caja CAAT	52
Figura 17: Propuesta 1. Se trata de una proteína funcional de 246 aminoácidos capaz de translocarse al interior y exportarse al exterior del núcleo de células del sistema nervioso de ratón.	54
Figura 18: Propuesta 2. Generación de tres péptidos bioactivos (A, B y C) de acuerdo a la presencia de sitios consenso para procesamiento proteolítico.	56
Figura 19: Propuesta 3. Se trata de una pre-proteína pero se obtienen solo dos péptidos bioactivos (A y B) de acuerdo a la presencia de sitios consenso para procesamiento proteolítico.	57
Figura 20: Propuesta 4. Se generaron dos péptidos bioactivos	57

ÍNDICE DE TABLAS

Número y título de tabla:	Página
Tabla 1. Vectores de clonación: moléculas transportadoras de ADN.	5
Tabla 2: Secuencias consenso buscadas con el programa PROSITE para predecir la localización de una proteína.	12
Tabla 3: Principales modificaciones postraduccionales para las cuales se busca secuencias consenso con el programa PROSITE.	12
Tabla 4: Programas para identificar secuencias consenso que ayudan en la predicción de funcionalidad en una proteína nueva.	13
Tabla 5: Actividades moduladoras de EM-1 y EM-2.	14
Tabla 6: Proteínas involucradas en el transporte de moléculas al interior y exterior del núcleo.	16
Tabla 7: Secuencias señal para transporte núcleo-citoplasma.	18
Tabla 8: Ejemplos de secuencias NES	19
Tabla 9: Tipos de Proteína Covertasas (PC's) y sus características	22
Tabla 10: Enzimas que se pueden emplear en el sitio de clonación múltiple del vector pBluescript®SK.	28
Tabla 11: Características del vector de clonación pBluescript®SK.	30
Tabla 12: Carga de la muestra en el gel para correr electroforesis.	33
Tabla 13: Preparación de la reacción para marcaje con nucleótidos fluorescentes.	33
Tabla 14: Protocolo de secuenciación.	33
Tabla 15: Resumen de las características estructurales de la secuencia del ARNm completo codificado por el ADNc clonado.	41
Tabla 16: Resumen de las secuencias estructuralmente homólogas al ARNm clonado.	42
Tabla 17: Proteínas homólogas a la Antolefinina.	46
Tabla 18: Predicción de parámetros fisicoquímicos con el programa ProtParam.	49

I.-RESUMEN

Nuestro grupo de trabajo generó anticuerpos policlonales con la capacidad de identificar material inmunorreactivo a las Endomorfinas en material proteico de alta masa molecular.

Teniendo como antecedente la generación de estos anticuerpos para reconocimiento potencial de las Endomorfinas, el objetivo de mi tesis se concentró en la implementación y aplicación de una metodología inmunomolecular para identificar, aislar y caracterizar la estructura del o los genes que codifican a el o los precursores de tales moléculas.

Sin embargo, a partir de la técnica de rastreo inmunológico diseñada y aplicada como parte inicial del trabajo experimental de esta tesis se logro un hallazgo no anticipado. Éste fue el aislamiento y caracterización molecular de una clona de ADN que codifica un ARN mensajero con un marco de lectura abierto de traducción de una proteína denominada por nuestro grupo de trabajo Antolefinina. Debido a este hallazgo, el interés de trabajo experimental de esta tesis se centró entonces en el análisis estructural y funcional de esta proteína a través del empleo de programas computacionales especializados para predecir funciones biológicas de biomoléculas a través de la identificación de firmas estructurales de secuencias de aminoácidos con función identificada en otras proteínas clonadas y reportadas en bases de datos.

De esta manera, se encontró una señal para localización en núcleo (secuencia NLS), regiones ricas en prolinas y leucinas (secuencia NES), sitios potenciales para fosforilación, amidación y procesamiento proteolítico relacionado con la presencia de pares de aminoácidos básicos. A partir de un análisis detallado de las firmas estructurales encontradas se establecieron cuatro propuestas para predecir función biológica de la Antolefinina. Se llegó a la conclusión de que la Antolefinina tiene altas probabilidades de ser una proteína funcional completa capaz de translocarse al interior y exportarse al exterior del núcleo de células del sistema nervioso de roedor ya que de acuerdo a la bibliografía revisada se ha encontrado que para algunas señales de

localización para núcleo (NLS) es importante la presencia adicional regiones ricas en prolina, lo cual se observa en la secuencia primaria de aminoácidos de la Antolefinina.

Finalmente, el análisis efectuado permitió obtener información necesaria para el diseño futuro de estudios dirigidos a la caracterización de funciones biológicas que permitirán obtener nuevos conocimientos acerca de la biología del sistema nervioso de mamífero.

II. INTRODUCCIÓN

ANTECEDENTES GENERALES

a) La información genética y su organización.

Se llama genoma a la totalidad del material genético de un organismo. El genoma puede estar constituido de una o más moléculas de ADN organizadas como cromosomas cuyo número esta bien definido para cada especie (Klug W. et al 1999). (ver figura 1)

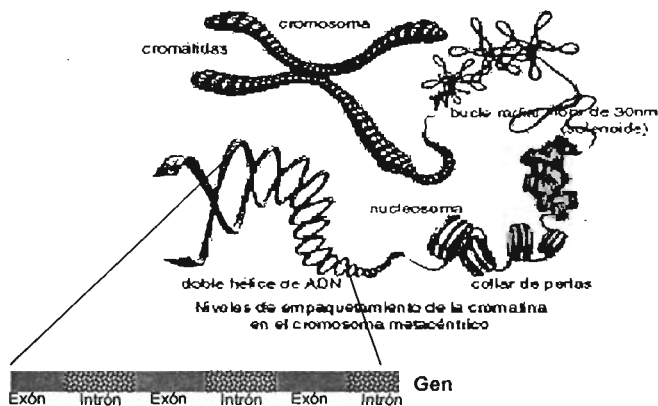


Figura 1: La información genética se encuentra organizada en los cromosomas

Un cromosoma es una molécula de ADN muy larga que transporta una serie de genes. Los genes contienen la información necesaria para sintetizar todas las proteínas requeridas por un organismo. Estas proteínas determinan entre otras cosas el aspecto del organismo, su metabolismo, sus mecanismos de defensa e incluso su comportamiento. Los genes están constituidos por segmentos de ADN que se transcriben y se traducen en proteínas (los exones) y de porciones de ADN entre las regiones codificantes que se transcriben pero no aparecen en el ARNm maduro (los intrones). (Klug W. et al 1999)

El genoma humano está constituido aproximadamente de 20 000 a 25 000 genes funcionales y sin embargo, actualmente aún con el ambicioso Proyecto del Genoma Humano que tiene como uno de sus objetivos obtener un conocimiento básico de la función de cada uno de tales genes, todavía existe una gran cantidad de genes sin caracterizar (Human Genome Project Information). Por lo tanto resulta de interés

involucrarse dentro del campo de la clonación de nuevos genes, y de manera particular de los que se encuentran involucrados en la regulación de funciones fisiológicas aún desconocidas.

Otro de los aspectos interesantes del Proyecto del Genoma Humano es el estudio de otros genomas como son el de la bacteria *Escherichia coli*, la mosca de fruta y el del ratón de laboratorio. La información derivada de la caracterización de los genes presentes en los genomas de dichos organismos se puede utilizar para obtener nuevos conocimientos acerca de la biología humana dado que sus secuencias de ADN se encuentran relacionadas con la del humano. Asimismo permite el empleo potencial de estos organismos como modelos experimentales. (Human Genome Project Information).

b) Estrategias experimentales empleadas usualmente en la caracterización de genes de interés.

Actualmente, existen diferentes estrategias que ayudan en la investigación de nuevos genes: la tecnología del ADN recombinante, métodos de análisis de las secuencias clonadas, así como técnicas para caracterizar los productos de éstos.

El término ADN recombinante hace referencia a la creación de nuevas combinaciones de segmentos o de moléculas de ADN que no se encuentran juntas de manera natural (Harvey et al 2000). El procedimiento base consiste en la extracción y corte del ADN en fragmentos que contienen desde uno hasta varios genes para ser insertados individualmente en moléculas de ADN capaces de replicarse autónomamente, en la tabla 1 se describen algunos ejemplos de vectores de clonación. (Griffiths et al 1999).

El éxito del método de clonación se debe a que los vectores de clonación entran en células hospederas individuales. Una célula individual transformada con un vector recombinante se divide en una colonia que contiene millones de células, cada una de ellas conteniendo el mismo vector recombinante, esta población es llamada entonces clona de ADN. De esta forma la clonación permite la amplificación y recuperación de un segmento específico de ADN proveniente de una muestra compleja como es un genoma (Alberts et al 2002).

Tabla 1. Vectores de clonación: moléculas transportadoras de ADN

Vector de clonación	Tamaño del inserto de ADN	Características
Plásmidos	5-10 kb	Moléculas de origen bacteriano conformadas por ADN circular de doble cadena que no pertenecen al cromosoma de la bacteria, sino que son adicionales a él y se replican de manera independiente. Para poder utilizarlos en ingeniería genética se han modificado de manera que contengan un número limitado de sitios de restricción y genes de resistencia a antibióticos específicos. Un ejemplo de un plásmido sencillo es el pBR322 que sólo posee dos genes de resistencia, tet^R y amp^R , cada uno de los cuales tiene un sitio único para ser utilizado como blanco de restricción.
Vectores virales	Hasta 15 kb	El o los genes de interés son incorporados al genoma del virus. Por ejemplo, Lambda es un bacteriófago, donde el tercio central de su cromosoma puede reemplazarse por ADN exógeno sin afectar la capacidad del fago para infectar células y formar placas líticas. Otro caso son el grupo de bacteriófagos filamentosos M13 (incluyendo los fagos fd y f1) cuyo genoma consiste de ADN circular de cadena sencilla. Las moléculas recombinantes pueden introducirse en células hospederas bacterianas por transfección.
Cósmidos.	Hasta 50 kb	Son vectores híbridos de fago lambda y plásmido. Los cósmidos contienen la secuencia COS del fago lambda necesaria para el empaquetamiento del ADN dentro de la cubierta proteica del fago y secuencias plasmídicas de replicación y de genes de resistencia a antibióticos que permiten identificar las células hospederas que los contienen. Un ejemplo específico de este grupo es el vector Uni-ZAP®XR que combina la alta eficiencia de construcción de bibliotecas del fago lambda y la conveniencia de selección del plásmido pBluescript. Acomoda insertos de hasta 10 Kb de longitud y puede ser rastreado utilizando sondas de ADN o sondas marcadas con anticuerpo.

Puesto que cada segmento de ADN clonado es relativamente pequeño deben construirse muchos clones diferentes para incluir todas las porciones del genoma de un organismo. Una colección de clonas es llamada biblioteca de ADN. Hay diferentes tipos de bibliotecas, clasificadas de acuerdo a la fuente de ADN empleada. Así, se tienen las bibliotecas genómicas y de ADNc (ADN complementario). Las bibliotecas genómicas contienen grandes fragmentos cromosómicos, todos los genes están representados en su forma nativa incluyendo intrones y secuencias reguladoras.

Por otra parte, la expresión de un gen puede variar en células distintas y en las diferentes etapas de desarrollo del organismo de interés, por lo que el material del cual se parte para comenzar a preparar bibliotecas de ADNc es ARNm proveniente de

tejidos específicos o de una etapa en particular de desarrollo (Griffiths et al 1999). (Ver figura 2)

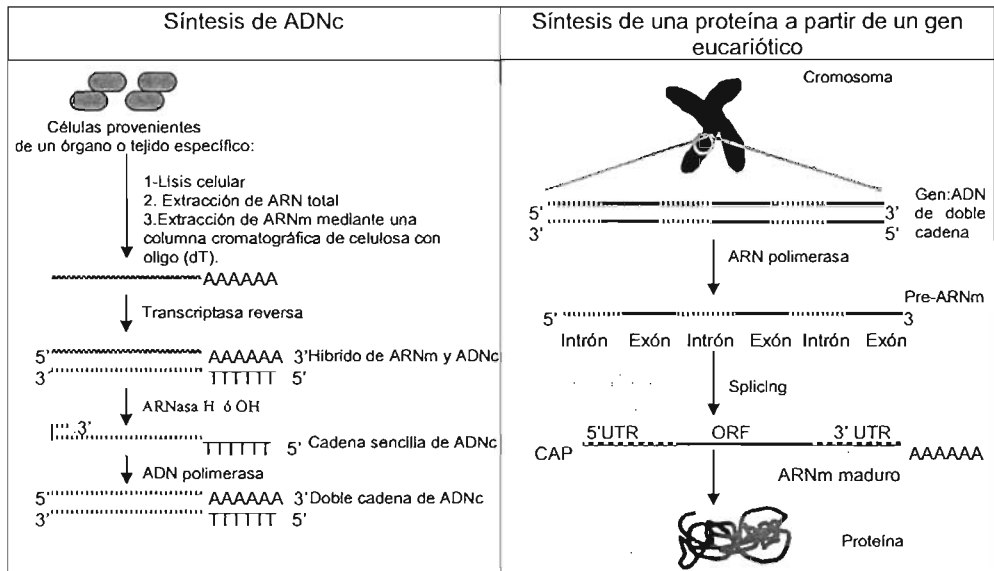


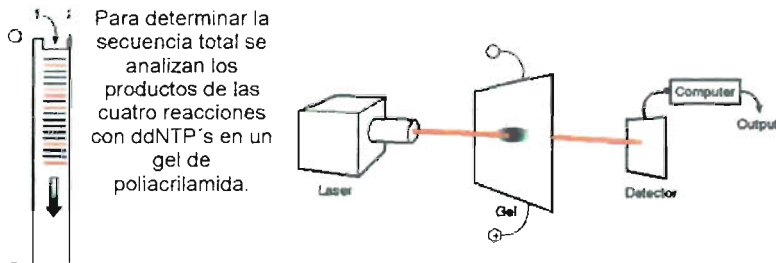
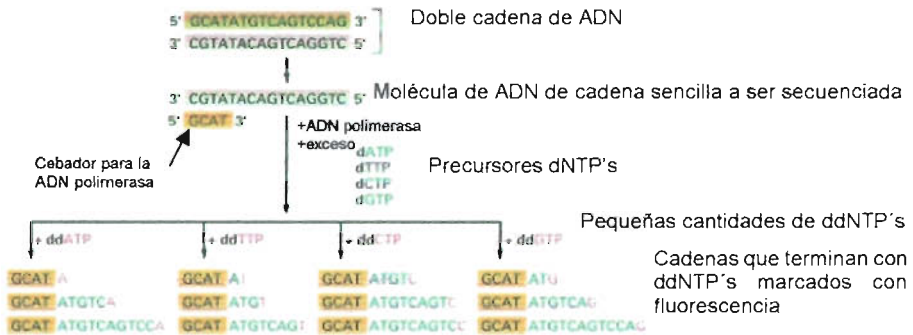
Figura 2: Dibujo esquemático que muestra las diferencias entre la síntesis de ADNc y el procesamiento normal del ADN.

El ADNc es un ADN sintético obtenido a partir de un ARNm maduro con la ayuda de la enzima transcriptasa reversa, aislada originalmente de un retrovirus que la utiliza para formar un híbrido ADN-ARN cuando se encuentra replicando su ARN genómico. Dado que, el ADNc está hecho a partir de ARNm éste se encuentra desprovisto de intrones. Esto es lo que hace posible que el ADNc proveniente de eucariotes pueda ser traducido en una proteína funcional dentro de una bacteria.

Por otra parte, el ADN genómico en un proceso denominado corte y empalme (splicing, en inglés), se eliminan los intrones del transcrito primario o pre-ARNm y se unen los exones para formar la secuencia que da origen a un polipéptido funcional. Otros eventos de procesamiento involucran la modificación de los extremos 5' y 3' (estructura de capucha-CAP en el extremo 5' y residuos de adenilato en el extremo 3'). Otro aspecto importante de los ARNm maduros eucarióticos es la presencia de regiones flanqueantes no codificantes 5' corriente arriba y 3' corriente abajo (5'UTR y 3'UTR, por sus siglas en inglés Untranslated region) que desempeñan una función moduladora de la transcripción.

Por otra parte, para la investigación de nuevos genes el análisis de la secuencia clonada constituye otra herramienta primordial. El método de secuenciación nucleotídica más utilizado fue desarrollado por Fred Sanger. Este método está basado en la síntesis de ADN en presencia de ddNTP's (2,3' didesoxirribonucleósidos trifosfato), los cuales difieren de los dNTP's (2'-desoxirribonucleósidos trifosfato) normales en la falta de un grupo hidroxilo en la posición 3': Los respectivos ddNTPs

pueden ser agregados a una cadena creciente, pero cuando son incorporados terminan la síntesis porque les falta el grupo 3' hidroxilo necesario para formar el enlace trifosfato con el nucleótido siguiente. Se maneja una concentración menor de ddNTP's comparada con la de sus análogos dNTP's, de modo que la terminación de la cadena concluye eventualmente en diferentes posiciones (Harvey et al 2000). En la figura 3 se muestra el caso en que se utiliza un método de secuenciación automático.



La secuencia obtenida se presenta como un electroferograma donde cada nucleótido tiene un color determinado.

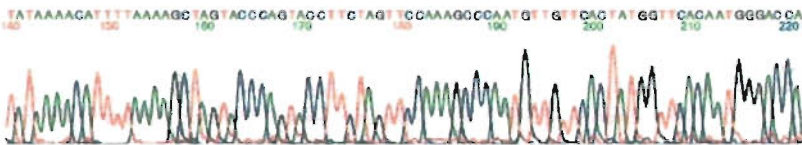


Figura 3: Secuenciación nucleotídica con ayuda de un secuenciador automático de un capilar.

El ADN a ser secuenciado es utilizado como molde para síntesis *in vitro* con la ayuda de la ADN polimerasa. La rara incorporación del ddNTP's por la ADN polimerasa bloquea el crecimiento posterior de la molécula de ADN ya que no puede formar un enlace fosfodiéster con el siguiente dNTP. Posteriormente cada colorante con el cual fue marcado el respectivo ddNTP emite luz a una longitud de onda diferente cuando es excitado por un láser, entonces los cuatro colores pueden ser detectados y distinguidos en un capilar, fragmentos que difieren en tamaño por un único nucleótido pueden ser separados en el gel de poliacrilamida (ABI PRISM DNA Sequencing 1995 The Perkin-Elmer Corporation).

El método de secuenciación nucleotídica basado en la síntesis de ADN en presencia de ddNTP's se efectúa con ayuda de la reacción en cadena de la polimerasa (PCR); otra herramienta clave en la investigación *in vitro* de nuevos genes. La técnica de PCR es un procedimiento efectivo para amplificar grandes cantidades de secuencias específicas de ADN. Los requerimientos básicos para efectuar una reacción de PCR son (1) dos oligonucleótidos sintéticos como cebadores (~20 nucleótidos de longitud) que son complementarios a regiones de las cadenas opuestas que flanquean la secuencia de ADN de interés, (2) una secuencia de ADN de una longitud de 100 a ~35 000 pares de bases como blanco, (3) la enzima ADN polimerasa termoestable (*Taq* polimerasa, que proviene de la bacteria *Thermus aquaticus*) y (4) los cuatro desoxiribonucleótidos.

Un proceso típico de PCR incluye los siguientes pasos:

- 1.Desnaturalización de una cadena molde de ADN: se logra aumentando la temperatura hasta 95°C.
- 2.Renaturalización: la temperatura se disminuye hasta 55°C para permitir que los cebadores (oligoprimers) se unan por complementaridad de sus bases nitrogenadas a las de a la cadena de ADN desnaturalizada.
- 3.Síntesis: la temperatura se aumenta hasta 75°C, que es la óptima para que funcione la *Taq* polimerasa que es la enzima encargada de efectuar la amplificación del fragmento de ADN de interés.

Ciclos repetidos de síntesis y desnaturalización dan como resultado un incremento exponencial del número de segmentos de ADN replicados. La técnica de PCR es muy sensible y puede detectar secuencias de interés que están presentes en la muestra en nanogramos (Glick B.R. et al 1998).

Sin embargo, para la implementación de una estrategia experimental que de forma específica y dirigida conduzca hacia la identificación y aislamiento de genes aún no caracterizados por la tecnología del ADN recombinante se requiere de técnicas y metodologías de abordaje experimental auxiliares.

c) Técnicas auxiliares para la caracterización de nuevos genes y sus productos.

La técnica de rastreo inmunológico fundamentada en el uso de anticuerpos específicos, representa una metodología experimental de alta especificidad y sensibilidad (Girjes et al. 1993, Burns et al. 1987) que detecta los productos proteicos de los genes de interés y permite evaluar de manera cuantitativa y cualitativa los niveles de expresión de tales productos (Cheng et al. 1996). En la figura 4 se describe en que consiste esta técnica.

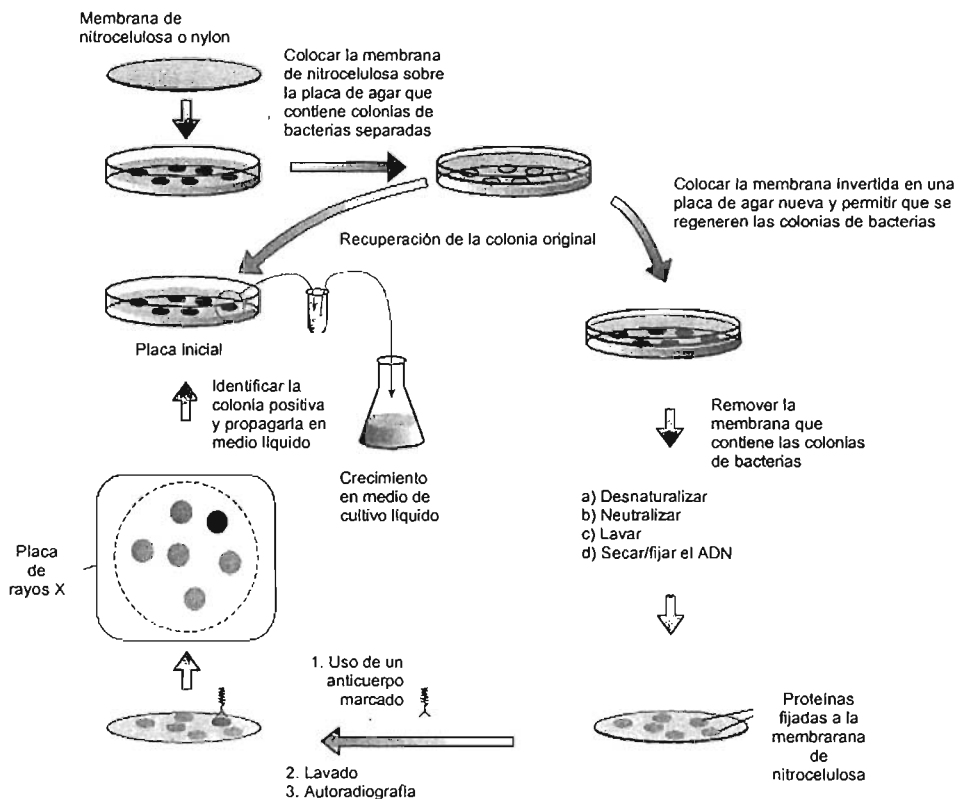


Figura 4: Rastreo inmunológico de un gen de interés

Se puede rastrear una biblioteca de expresión con la ayuda de anticuerpos capaces de identificar un producto proteico. Las bacterias que contienen los genes son colocadas en filtros de forma individual y pueden ser manipuladas para que expongan los productos peptídicos expresados para ser identificados mediante el empleo de anticuerpos específicos (Strachan et al 1999). Las proteínas son identificadas con un anticuerpo marcado (también se puede emplear un primer anticuerpo que identifica la proteína de interés y un segundo anticuerpo marcado contra éste). Una vez identificada la clona de interés, se puede aislar, propagar en medio de cultivo líquido y purificar el vector que contiene el gen investigado.

Si no se cuenta con información concerniente a la secuencia nucleotídica del gen en cuestión la disponibilidad de la secuencia primaria de un producto proteico purificado que pueda ser utilizado como reactivo para generar anticuerpos es suficiente. Nos podemos beneficiar de la expresión de proteínas eucarióticas en células bacterianas para manipular a éstas como a un sistema para rastrear genes aun no identificados y caracterizados al nivel molecular.

La técnica de rastreo inmunológico se ha empleado de manera exitosa en una gran variedad de modelos experimentales de clonación de genes. La clonación del gen TnAV-CP que da origen a la proteína estructural de mayor importancia en el ascovirus *Trichoplusia ni-2^a* es un ejemplo de lo anterior. Para lograrlo, se realizó el rastreo inmunológico de una biblioteca de ADNc con la ayuda de anticuerpos específicos contra proteínas totales de dicho virus. (Zhao et al 2003) La clonación de la enzima acetiltransferasa también se logró a partir del rastreo inmunológico de una biblioteca de expresión de ADNc. Las clonas que poseían proteínas con una actividad enzimática capaz de expresarse en las bacterias hospederas se detectaron con un antisuero con especificidad inmunológica para distinguir entre las enzimas propias de la bacteria y la enzima de interés (Eberwine et al 1987). De igual manera, la clonación, expresión y caracterización del gen que codifica la proteína ribosomal S15 de *Taenia solium*, se logró con el uso de anticuerpos que identificaron de manera específica material proteico recombinante expresado en extractos tisulares de este céstodo (Jimenez L, et al 2004). Otro ejemplo adicional del uso de la técnica de rastreo inmunológico es la clonación y caracterización del gen que codifica un receptor de unión a inmunoglobulina en la superficie celular de miembros de la familia *Tripanosomatidae*. La identificación de tal receptor no había sido lograda, por lo que para clonar el gen que codificara para la proteína representativa de este receptor, se utilizó el rastreo con ayuda de antisueros de una biblioteca de ADNc construida a partir de *Leishmania major*. Se logró el aislamiento de un gen nombrado Lmspl que da origen de manera potencial a la proteína representativa del receptor buscado (Campos-Neto A. et al 2003).

Por otro lado, un ejemplo de una aplicación distinta a lo mencionado anteriormente, es el protocolo experimental implementado para la identificación de nuevos fármacos y vacunas potenciales contra *Onchocerca volvulus* causante de

cegueras infecciosas en el ser humano. Se partió de la generación de bibliotecas de expresión de ADNc construidas a partir de *O. volvulus* en diferentes etapas críticas para el desarrollo de la infección. El rastreo inmunológico para la detección de genes candidato para la terapéutica contra oncocerciasis se realizó con la ayuda de sueros provenientes de individuos identificados como inmunes a esta enfermedad. El resultado obtenido fue el aislamiento de ocho genes con un amplio potencial para ser empleados como blanco para la generación de vacunas y nuevos fármacos (Lizotte-Waniewski M. et al 2000).

Como se puede observar, la técnica de rastreo inmunológico combinada con el uso de bibliotecas de ADNc se ha empleado satisfactoriamente para aislar genes de interés. Asimismo, el uso simultáneo de estas herramientas permite aislar de manera adicional genes nuevos que codifiquen proteínas con una naturaleza similar a la de interés.

Es importante notar que cuando se están aislando genes nuevos, una herramienta adicional para obtener información preliminar acerca de las identidades estructurales y funcionales de la nueva secuencia nucleotídica y de la secuencia proteica codificada, esta constituida por las bases de datos especializadas en efectuar estudios de homología que identifican estructuras que ya se encuentran caracterizadas a nivel molecular y funcional. Las bases de datos utilizadas con mayor frecuencia son FASTA (Introducida en 1988 por W. Pearson y D. Lipman) y BLAST (Basic Local Alignment Search Tool, herramienta del GenBank fundamentada en una rigurosa teoría estadística).

d) Programas especializados de computadora para localizar secuencias consenso para las cuales ya hay funciones identificadas.

Se sabe que los distintos tipos de modificaciones postraduccionales son importantes para alcanzar el estado funcional específico de las proteínas. De manera que cuando se esta ante un gen nuevo que codifica para un proteína de función desconocida, lo que se busca son motivos o firmas estructurales de secuencias consenso (sequence motif o structural functional signatures, respectivamente, por su nomenclatura en el inglés) constituidas por secuencias parciales de residuos contiguos

de aminoácidos que se conservan y están asociados con alguna función particular dentro de la estructura de una proteína dada. La localización de dichas secuencias puede ayudar a predecir el tipo de estructura, la ubicación celular, la(s) función(es) y la clasificación dentro de una familia o superfamilia de miembros moleculares de una proteína clonada. PROSITE es una base de datos especializada y usada con confiabilidad para detectar dichas secuencias. (Ver tabla 2 y 3).

Tabla 2. Secuencias consenso buscadas por el programa PROSITE para predecir la localización de una proteína

Destino de la proteína	Localización y características de la secuencia consenso
Retículo endoplasmático y proteínas secretadas	20 o más aminoácidos muy hidrofóbicos en el extremo N-terminal (por ejemplo la insulina humana, que tiene como péptido señal 24 aminoácidos altamente hidrofóbicos NH ₂ MALWMRLPLLALLALWGPDPAAA)
Mitocondria	α -hélice en el extremo N-terminal con residuos cargados positivamente en una cara e hidrofóbicos en la otra (un ejemplo es la enzima aldehído deshidrogenasa de la mitocondria humana NH ₂ MLRAAARPGPRLGRLL)
Núcleo	Secuencias internas de aminoácidos básicos acompañadas de manera común de residuos de prolina (por ejemplo la secuencia del antígeno SV40 del virus del VIH TPPKKKRKV)
Lisosoma	Adición de residuos de Manosa 6-fosfato

Tabla 3: Principales modificaciones postraduccionales para las cuales se busca secuencias consenso con el programa PROSITE

Tipo de modificación (grupo adicionado)	Aminoácido blanco	Comentario
Fosforilación (PO ₄ ⁻)	Tirosina, serina, treonina	Llevado a cabo por kinasas específicas, puede ser reversible por las fosfatasas
Metilación (CH ₃)	Lisina	Llevado a cabo por metilasas
Hidroxilación	Prolina, lisina, ácido aspártico	Hidroxiprolina e hidroxilisina son particularmente comunes en los colágenos
Acetilación (CH ₃ CO)	Lisina	Llevado a cabo por una acetilasa
Carboxilación (COOH)	Glutamato	Llevado a cabo por γ -carboxilasa
N-glicosilación (carbohidrato complejo)	Asparagina, usualmente en la secuencia Asn-X-Ser/Thr	Se efectúa en el retículo endoplasmático, X es cualquier aminoácido diferente a prolina
O-glicosilación (carbohidrato complejo)	Serina, treonina, hidroxilisina	Se efectúa en el complejo de Golgi, es menos común que la N-glicosilación
GPI (Glicolípido)	Aspartato en el carbono terminal	Sirve para anclar proteínas en la membrana celular
Miristoilación	Glicina en el amino terminal	Sirve para anclar proteínas en la membrana celular
Ácido graso C ₁₄)		

Palmitoilación (Ácido graso C ₁₆)	Cisteína para formar la unión S-palmitoil	Sirve para anclar proteínas en la membrana celular
Farnesilación (Grupo prenilo C ₁₅)	Cisteína en el carbono terminal	Sirve para anclar proteínas en la membrana celular
α -Amidación (NH ₂)	Glicina que provee el grupo amida	Llevado a cabo por la enzima PAM (por sus siglas en inglés Peptidylglycine α -amidating monooxygenase)

Además de PROSITE se dispone de otros programas que buscan secuencias consenso que dan información adicional de una secuencia polipeptídica nueva (ver tabla 4).

Tabla 4: Programas para identificar secuencias consenso que ayudan en la predicción de funcionalidad en una proteína nueva

Tipo de predicción	Utilidad	Programa
Péptido señal	Presente en polipéptidos que actúan como moléculas precursoras, sirve como una señal de reconocimiento para transporte a través de las membranas celulares, esta señal es posteriormente escindida y degradada.	SignalIP PSORT
Segmentos transmembranales	Para establecer perfiles de hidrofobicidad de las cadenas polipeptídicas y en algunos casos establecer la topología de la proteína en la membrana celular.	TMHMM TopPred2 PhDhtm PSORT DAS TmPred HMMTOP TMAP Memsat
Ubicación de la proteína	Puede tratarse de una proteína citosólica, destinada a un organelo en particular o para excreción.	PSORT
Dominios "coiled coil"	Las proteínas no globulares son generalmente α -superhélices formadas por series de repeticiones de aminoácidos hidrofóbicos, usualmente leucinas.	COILS Parcoil Multicoil
Estructura secundaria	La predicción por sí misma de la estructura secundaria puede dar una indicación de función, ya que la estructura de una proteína se conserva más que la secuencia primaria de aminoácidos.	PHD PREDATOR JPred

Como se puede deducir, el análisis estructural para identificar la presencia de firmas o secuencias motivo estructurales en una molécula proteica nueva, puede proveer al investigador de información preliminar importante sobre la predicción de las relaciones estructura-actividad biológica de esta molécula, aún en la ausencia de

homólogos estructurales ya clonados con funciones identificadas. La identificación de motivos estructurales de secuencias consenso es un prerrequisito que ayuda a predecir funciones potenciales o anticipadas de moléculas nuevas clonadas y da la pauta para diseñar los bioensayos pertinentes para confirmarlas en experimentación ulterior (Koonin E.V et al 2004).

ANTECEDENTES PARTICULARES

La técnica de rastreo inmunológico antes mencionada, fue diseñada y aplicada como parte del trabajo experimental inicial desarrollado en esta tesis, y estaba dirigida hacia el aislamiento de una clona de ADNc que codificara el gen que expresa la proteína precursora de las Endomorfinas (EM-1, EM-2). La Endorfina-1 (EM-1) y Endorfina-2 (EM-2), aisladas de extractos de tejido nervioso cerebral del bovino y humano, son dos tetrapéptidos α -amidados que exhiben la más alta afinidad y selectividad de unión por el receptor opioide μ (Zadina et al. 1997) de todos los neuropéptidos endógenos conocidos a la fecha actual. Su capacidad para modular la nocicepción y antinocicepción en el roedor, entre otras actividades (ver tabla 5), les confieren un potencial papel farmacológico en la clínica del dolor (Soignier et al. 2000, Wilson et al. 2000).

Tabla 5: Actividades modulatorias de la EM-1 y EM-2

Endorfina	Actividad
EM-1 y EM-2	<ul style="list-style-type: none"> • Potente actividad antinociceptiva en una gran variedad de modelos animales de dolor agudo, neuropático y alodínico (Wilson et al 2000, Sakurada 2001, Sakurada et al 2002) • Vasodilatación de capilares sanguíneos mediada por un mecanismo dependiente de óxido nítrico endotelial (Champion et al. 1999) • Actividad vasodepresora (Champion et al. 1998) • Actividad cardiovascular y respiratoria en roedor (Czapla et al. 2000) • Regulación de la fagocitosis, quimiotaxis y producción de ión superóxido en células de microglía (Azuma et al. 2001) • Papel importante en desórdenes de movimientos hipercinéticos a través de la inducción sostenida de la activación de receptores opioides (Mehta et al. 2001).

Sin embargo, a lo largo del desarrollo experimental de este objetivo inicial de trabajo de la tesis, se logró un hallazgo experimental no anticipado. Éste fue el aislamiento y caracterización molecular de una clona de ADNc que codifica un ARN mensajero

(ARNm) con un marco de lectura abierto de traducción de una proteína nueva denominada por nuestro grupo de trabajo Antolefinina. Debido a este hallazgo, el interés de trabajo experimental se centró entonces en el análisis estructural y funcional de esta proteína a través del empleo de programas computacionales especializados para predecir funciones biológicas de biomoléculas a través de la identificación de la presencia de firmas estructurales de secuencias de aminoácidos con función identificada en otras proteínas clonadas. De hecho, en esta sección de antecedentes (ver siguiente inciso), nos parece justificado hacer una revisión de algunas de estas firmas estructurales de aminoácidos, las cuales consideramos de importancia para llevar cabo nuestro estudio predictivo funcional de la proteína Antolefinina, que es expresada por este nuevo gen clonado en mi proyecto de tesis.

e) Poros nucleares y señales para importación y exportación de proteínas al núcleo.

La compartimentalización es una característica que ofrece a la célula la oportunidad de regulación. Así, la importación y exportación de proteínas y ácidos nucleicos entre el núcleo y el citoplasma es un proceso estrictamente controlado. Éste se efectúa con la ayuda de proteínas acarreadoras que interaccionan tanto con la carga eléctrica de la molécula a transportar como con el complejo de poro nuclear (NPC, por su nomenclatura en el inglés) (Macara I.G, 2001). Ver figura 5.

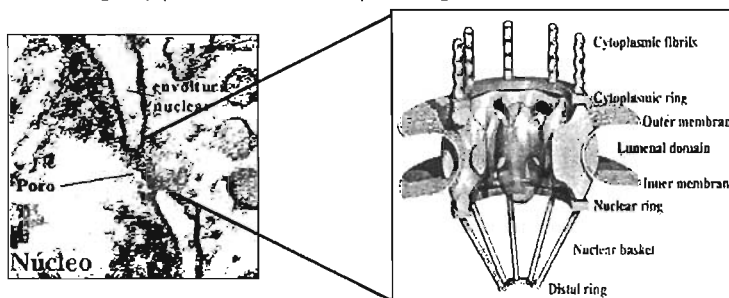


Figura 5: Micrografía electrónica y dibujo esquemático de un poro nuclear

La envoltura nuclear consiste de dos membranas (interna y externa, inner and outer membrane, por su nomenclatura en inglés) que periódicamente se unen definiendo poros, los cuales se rodean de proteínas que conforman el complejo de poro nuclear encargado de regular selectivamente la entrada y salida del núcleo. Este complejo restringe la libre difusión de partículas y proteínas de diámetro superior a 9 nm (tamaño correspondiente a proteínas de 40-60 kD). Sin embargo complejos de hasta 25 nm son transportados eficientemente siempre que lleven adicionados una señal de exportación o importación nuclear. (Hipertextos del área de la Biología).

La mayor parte de las proteínas encargadas de transportar la carga eléctrica pertenecen a una familia denominada importina β /karioferina β (Ver tabla 6). Una característica común a todas las karioferinas (es así como se refiere de manera general a los miembros de esta familia) es que se unen a las nucleoporinas, las proteínas que conforman el complejo de poro nuclear. Otra característica importante es que las karioferinas pueden formar un complejo con la enzima RanGTPasa (Ran es el miembro más abundante de la superfamilia de GTPasas Ras) y de hecho es a partir de tal interacción que se regula la unión de la carga eléctrica. Al igual que otros miembros de la superfamilia Ras, Ran funciona como un swich molecular y sufre un cambio conformacional entre los estados GDP-GTP (Macara I.G, 2001). Ver figura 6.

Tabla 6: Proteínas involucradas en el transporte de moléculas al interior y exterior del núcleo

Tipo de acarreador	Función
IMPORTINAS	
Importina	Importa la Importina α proteínas ribosomales, las proteínas virales HIV Rev, HIV Tat y HIV Rex
Transportina I, kap β 2	Importa proteínas de unión al ARNm y proteínas ribosomales
Transportina SR	Importa proteínas de unión al ARNm
Hmtr10	Desconocida
Importina 7	Importa proteínas ribosomales
Importina II	Importa a UbcM2
Importina 5, Kap β 3	Importa proteínas ribosomales
Importina α	Funciona como un adaptador para importina β
EXPORTINAS	
Exportina 5	Exporta proteínas fosforiladas
Crm 1	Exporta acarreadores para proteínas con una señal NES rica en leucinas
Cas	Exporta Importina α
Exportina-t	Exporta ARNt
Exportina 4	Exporta eIF5A
OTROS TRANSPORTADORES	
Tap/Mex 67	Transporte de ARNm y ARNr
Calreticulina	Promueve la exportación de moléculas que contienen una señal NLS

El transporte mediado por importinas y exportinas depende de una señal o firma estructural de aminoácidos en la molécula a ser transportada (Macara I.G, 2001). Ver tabla 7.

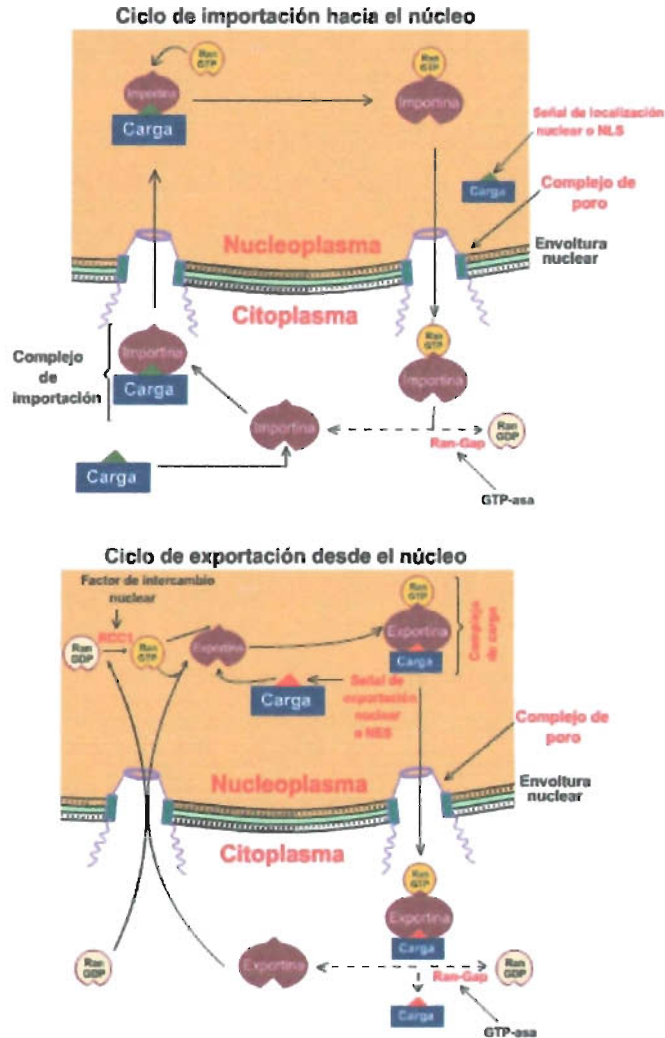


Figura 6: Dibujo esquemático que muestra el ciclo de importación y exportación hacia el núcleo para una molécula con una señal NLS o NES

El poro se puede dilatar hasta 26 nm cuando recibe la señal apropiada (secuencias ricas en lisina, arginina y prolina o aminoácidos hidrofóbicos presentes en las proteínas). La molécula Ran-GTP es la que impone direccionalidad al transporte. Se encuentra presente en grandes concentraciones solo en el núcleo donde disocia el complejo de importación constituido por la carga más importina y estabiliza en complejo de exportación. La transformación cíclica de la Ran es regulada por dos proteínas: RCC1 y la Ran-Gap que generan un gradiente Ran-GTP/Ran-GDP a ambos lados de la membrana nuclear.

Tabla 7: Secuencias señal para transporte núcleo-citoplasma.

Señal	Secuencia consenso propuesta
Secuencia NLS (Nuclear Localization Signal) monopartita	BBBBP(BBBX), BBB(H/P)
Secuencia NLS bipartita	BBXXXXXXXXXXBBBXX
NLS Viral	RXXRRXRBR
Secuencia NES (Nuclear Export Signal) rica en leucinas	LX ₂₋₃ (F/I/L/N/M)X _{1,2,3} LX(I/V/L)
Secuencia NES inusual	IXXXIXLXT, WXKIXLXP

B=Residuo básico (K o R)

X=Cualquier residuo

(/)=Residuos permitidos en esa posición

Las señales o firmas estructurales NLS y NES se encuentran caracterizadas principalmente en los virus, ya que la replicación de su material genético depende de exportar las proteínas necesarias al núcleo de la célula hospedera. La primera secuencia señal identificada para exportación nuclear fue identificada en el antígeno T del virus del simio (SV40). Esta secuencia **PKKKKK**, y la que se encuentra en la nucleoplasmina (una proteína involucrada en el ensamble de la cromatina), **KRPAATKKAGQAKKKKLD**, se consideran los prototipo para las señales NLS monopartita y bipartita (Macara I.G, 2001). Por lo tanto la señal bipartita se caracteriza por (Dingwall C. et al 1991) :

- (1) Dos aminoácidos básicos consecutivos (Arginina o Lisina).
- (2) Una región espaciadora de aproximadamente 10 aminoácidos.
- (3) Por lo menos tres residuos de aminoácidos básicos (Arginina o Lisina) en las cinco posiciones después de la región espaciadora.

Así, la estructura primaria de aminoácidos de esta secuencia consenso bipartita sería [R/K]XXXXXXXXXX[H/R/K], donde X es cualquier aminoácido.

En el caso de señales para exportación de proteínas (NES), la secuencia prototipo esta constituida por segmentos pequeños hidrofóbicos (alrededor de 10 aminoácidos) ricos en residuos de leucina (Gasca et al 2002). La estructura primaria de aminoácidos de esta secuencia consenso de exportación de proteínas sería LX₁₋₃LX₂₋₃LXL, donde X es cualquier aminoácido. Ver tabla 8.

Tabla 8: Ejemplos de secuencias NES

Proteína	Secuencia
MAPKK (Mitogen Activated Protein Kinase Kinase)	NLVDLQKKLEELELD
Rev (Human Immunodeficiency virus Rev))	PVPLQLPPLERLTLD
PKI (Protein Kinase Inhibitor)	ELALKLAGLDIN

f) Importancia de las proteínas que contienen NLS y NES.

Este tipo de señales o firmas estructurales se encuentran presentes en proteínas con una importante actividad biológica. Por ejemplo, en el caso de Diacilglicerol kinasa que se encuentra involucrada en la transducción de señales metabotrópicas de células eucarióticas, Se comprobó que esta enzima encargada de regular los niveles citoplásmicos de diacilglicerol, posee una NLS que provoca que su localización subcelular dependa del tipo de célula y el estado de crecimiento de la misma. Esta enzima, participa en importantes funciones neuronales (Hozumi Y. et al 2003). Otro ejemplo, esta constituido por la SPHK2 (enzima encargada de catalizar la fosforilación de la esfingosina, un péptido bioactivo que regula diversos procesos biológicos), la cual posee una NLS que le confiere la propiedad de localizarse en el núcleo induciendo la inhibición de la síntesis de ADN en varios tipos celulares (Igarashi N. et al).

Por lo que respecta a la presencia de NES, podemos citar a la Antizima-1, que es una proteína inhibidora de la ornitina descarboxilasa, una enzima clave en la síntesis de poliaminas. La Antizima-1 contiene dos NES, una en el extremo N-terminal y otra en el extremo C-terminal, las cuales le confieren la propiedad de regular otras moléculas en el núcleo como son transductores de señal y proteínas reguladoras del ciclo celular en el núcleo (Murai N. et al 2003). En el caso de SOX9, una proteína humana que contiene una secuencia NES localizada entre dos NLS's, esta proteína se encuentra involucrada en la determinación del fenotipo sexual de las gónadas, ya que solo en las gónadas masculinas se localiza en el núcleo (Gasca S. et al 2002).

Como se puede deducir las señales NLS y NES por si mismas desempeñan un papel muy importante en las proteínas donde se encuentran ya que determinan su función.

Por otra parte, en la investigación dentro del campo de la terapia génica (por ejemplo en el cáncer), uno de los métodos más eficientes para la transfección de genes

en células neoplásicas, es el basado en los vectores virales (retrovirus). Sin embargo, se tienen problemas de seguridad debido a las repuestas inflamatorias, inmunes y transformación maligna debido a la mutagénesis provocada por la inserción de estos agentes virales. Una alternativa más segura a este procedimiento, esta constituida por el uso de vectores no virales ("polyplexes", "lipoplexes"), aunque con estos procedimientos se observa bajos niveles de expresión génica *in vivo*. La transferencia ineficiente de ADN del citoplasma hacia del núcleo se ha identificado como una causa principal para lo anterior. Como ya se mencionó, las proteínas celulares y virales entran eficientemente al núcleo gracias a la presencia de NLS's constituidas por intervalos de aminoácidos que se unen a los receptores intracelulares (como son la importinas) para el transporte, facilitando de esta manera el transporte a través del poro nuclear. Experimentos recientes revelan que las NLS's conservan su actividad cuando se conjugan con proteínas no nucleares. La aplicación de péptidos-NLS para la transferencia de genes no virales ha sido investigada ampliamente y parece tener un gran potencial para mejorar la introducción de ADN al núcleo (Bremner K.H. et al 2004). De hecho, uno de los primeros transportes efectuados exitosamente lo constituye el de la proteína Tat- β -Galactosidasa que tiene una masa molecular de aproximadamente 120 kDa. Esto se logró al crear un híbrido entre la proteína de interés y la señal de internalización rica en aminoácidos básicos. Esta metodología de unir proteínas de gran tamaño con los péptidos que contienen las señales apropiadas para el transporte al interior del núcleo da lugar también a una nueva herramienta para incorporar moléculas no peptídicas como son las sondas fluorescentes. (Schwarse S. R. et al 1999)

Otro aspecto importante de la caracterización de NLS y NES es su uso en el estudio de rutas involucradas en la exportación de ARNm (Gallouzi I-E. et al 2001) y en la caracterización de las rutas de proteínas de interés. Un ejemplo, es el uso de la señal del virus HIV Tat (48 YGRKKRRQRRRPPQ 60) para detectar secuencias peptídicas capaces de unirse con alta afinidad a cGPK (en el inglés, cGMP-deependent protein kinases type α y β), ya que el aislamiento de tales secuencias permitirá estudiar la ruta de señalamiento relacionada con este tipo de proteínas kinasas implicadas en una gran variedad de respuestas celulares como son la relajación del músculo liso o la sinapsis neuronal (Dostmann W. R. G. et al 2000).

En este contexto, es fácil deducir que las señales para localización en el núcleo (NLS) y las de exportación (NES) constituyen una herramienta experimental importante, no solo para propósitos terapéuticos, sino que también para la entrega alternativa de genes (sientan las bases para definir los requerimientos para el diseño y construcción de vectores con NLS) y para comprender mecanismos celulares como son la transducción de señales y la transcripción genética.

g) Recorte de las cadenas polipeptídicas.

Una amplia variedad de moléculas (principalmente hormonas y neuropéptidos), son sintetizadas como grandes precursores proteicos que posteriormente son editados enzimáticamente. Los precursores proteicos de péptidos y polipéptidos biológicamente activos pueden ser subdivididos por lo menos en cuatro tipos diferentes, de acuerdo al sitio donde se lleva a cabo el corte enzimático de escisión (Seidah NG et al. 1998).

- a. Precursor tipo I, secuencia consenso X-X-R-X-(K/R)-R↓-X-X (donde X es cualquier aminoácido y la flecha indica el sitio de corte en el aminoácido básico): esta clase de precursor incluye pro-proteínas sintetizadas en células secretorias constitutivas, como son los factores de crecimiento, neurotrofinas, ciertos receptores, toxinas bacterianas y glicoproteínas de superficie virales.
- b. Precursor tipo II, secuencia consenso X-X-X-X-(K/R)-(K/R)↓-X-X (donde X es cualquier aminoácido y la flecha indica el sitio de corte en el aminoácido básico): normalmente son pro-proteínas que contienen la secuencia consenso en el carbono terminal. Estos incluyen a la mayoría de los precursores de las hormonas polipéptidicas y sustancias peptídicas bioactivas procesadas.
- c. Precursor tipo III, secuencia consenso B-X-B-X-B-X-R↓-X-X (donde X es cualquier aminoácido, B un aminoácido básico y la flecha indica el sitio de corte en el aminoácido básico): esta clase de precursores incluye a los de las hormonas polipéptidicas así como factores de crecimiento.
- d. Precursor tipo IV, secuencia consenso B-X-B-X-B-X-B-(K/R)↓-X-(R/K) (donde X es cualquier aminoácido, B un aminoácido básico y la flecha indica el sitio de corte en el aminoácido básico): en este grupo se incluyen algunos precursores hormonales, hormonas pancreáticas y algunos factores de crecimiento.

Asimismo, las enzimas procesadoras especializadas en cortes proteolíticos pertenecen a una familia de serina proteinasa del tipo subtilisina/kexina. Hasta la fecha se conocen siete tipos de estas convertasas: PC1/PC3, PC2, furina, PACE4, PC4, PC5 y PC7. Para obtener más información acerca de los péptidos generados a partir de una proteína precursora es importante considerar los siguientes aspectos funcionales sobre la familia de las convertasas (Seidah NG et al. 1998):

Tabla 9: Tipos de Proteína-Convertasas (PC's) y sus características

Convertasa	Tipo de convertasa	Tipo de precursor	Ejemplo de Precursor escindido
PC1	Clase II: se expresa principalmente en células endocrinas y neurales (tejido periférico y SNC). Se localiza en gránulos secretorios y TGN.	Hormonas polipeptídicas	POMC ProInsulina Humana ProDinorfina ProGlucagón ProMCH ProEncefalina ProRenina Humana
PC2	Clase II: se expresa principalmente en células endocrinas y neurales (tejido periférico y SNC). Se localiza en gránulos secretorios.	Enzimas Hormonas polipeptídicas	POMC ProInsulina Humana ProDinorfina proTHR ProGlucagón Proneurotensina Proencefalina
PC5-A	Clase III: se expresa tanto en células endocrinas como no endocrinas (corteza adrenal, endoteliales y células sertoli y sistema digestivo)	Hormonas polipeptídicas receptores	MIS (sustancia inhibidora proMulleriana) R-PTP μ

Como se observa en la tabla 9, PC1 y PC2 son dos convertasas responsables de las escisiones proteolíticas efectuadas en las rutas de biosíntesis peptídica reguladas de las células endocrinas y neuronales. Su expresión restringida asegura que sólo se lleve a cabo el procesamiento de secuencias en el tipo de células apropiadas. (Korner J. et al 1991). Es pertinente mencionar que análisis específicos de Northern-blot en tejidos que expresan PC2 y PC1/PC3 revelaron que su expresión está altamente restringida a tejidos neuroendócrinos, incluyendo los islotes de Langerhans, la médula pituitaria adrenal y varias regiones del cerebro (Steiner et al. 1992). Asimismo, la expresión de PC2 y PC1/PC3 en varios fenotipos de células endocrinas y neuronales, muchas veces da lugar a una variedad de mezclas de productos peptídicos que se

derivan del mismo precursor con actividades antagónicas o divergentes. (An Zhou et al 1999).

j) La α -amidación.

Los péptidos biológicamente activos como son las hormonas que actúan a grandes distancias en el cuerpo o los neurotransmisores son sintetizados a partir de grandes precursores proteicos inactivos a los que se les ha efectuado una gran cantidad de modificaciones post y co-traduccionales (adición de péptido señal, adición de cadenas de O y N-oligosacáridos y presencia de un par de aminoácidos básicos). Tales péptidos tienen la característica de tener α -amidado el último aminoácido del extremo carboxilo terminal (Cuttitta F. 1993). Esta modificación covalente es única para los péptidos con una función reguladora, y de hecho, ha sido utilizada como parámetro para descubrir y aislar nuevas moléculas con una actividad biológica potencial (Cuttitta F. 1993). La α -amidación de un polipéptido sucede como último paso de una serie de modificaciones postraduccionales efectuadas en el extremo C-terminal (Zastrow M. et al 1986). La enzima PAM (por sus siglas en inglés Peptidylglycine α -Amidating Monooxygenase) es la encargada de efectuar esta modificación covalente (Eipper B.A. et al. 1993). La secuencia consenso para que se lleve a cabo la α -amidación es la que se muestra en la figura 7.

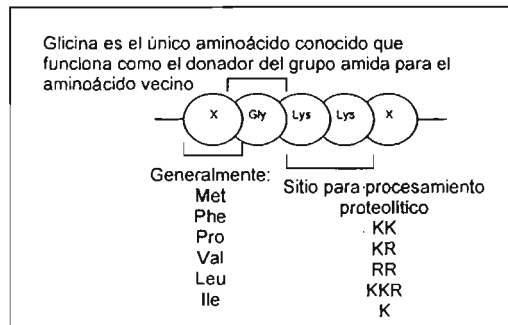


Figura 7: Diagrama estructural de la secuencia consenso para la α -amidación

La importancia de la α -amidación reside en que esta modificación confiere resistencia a la acción degradativa por el extremo carboxilo terminal de la carboxipeptidasa-E, aumentando la vida media biológica y la afinidad del polipéptido amidado por su receptor, traduciéndose así esto en un efecto biológico más potente y de larga duración (Cuttitta F. 1993).

III.-PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN DEL TRABAJO EXPERIMENTAL.

Como ya se mencionó con anterioridad, existe un número significativo de genes que codifican nuevas sustancias de tipo proteico y peptídico que esperan ser caracterizadas a nivel estructural y funcional en los mamíferos, incluyendo el humano. Durante el trabajo experimental inicial de rastreo inmunológico dirigido a la clonación del precursor de las endomorfina, se aisló en forma no anticipada, como una falsa inmunopositiva, una clona con un ADNc que codifica una molécula proteica no caracterizada aún en el sistema nervioso de mamífero. Estudios de homología estructural efectuados en la base de datos del GenBank y con el programa de análisis del EMBL-EBI se confirmó que esta proteína, denominada por nuestro grupo de trabajo "Antolefinina", no presenta identidad estructural con ninguna otra molécula ya clonada y con funciones identificadas. En este contexto, debido a este hallazgo experimental no anticipado y a la ausencia de identificación funcional de esta molécula, mi trabajo de tesis fue replanteado y dirigido hacia estudios estructurales de esta nueva proteína para identificar la presencia en ella de secuencias consenso o firmas estructurales, las cuáles, al estar codificadas en proteínas y/o polipéptidos ya clonados con funciones caracterizadas, nos pondría en posibilidad de proponer funciones fisiológicas de esta nueva proteína dentro del sistema nervioso de mamífero. Es claro deducir que se necesita implementar una estrategia analítica que permita iniciar en forma sistemática y dirigida la caracterización funcional de la Antolefinina. Esta estrategia analítica requiere de programas y bases de datos especializadas en la identificación de secuencias consenso para las cuales ya hay funciones establecidas. Anticipamos que la información de estructura-función derivada de tales análisis, servirá de base para estudios futuros con bioensayos específicos, llevados a cabo por nuestro y otros grupos de trabajo experimental, dirigidos hacia la caracterización funcional extensiva de esta nueva proteína.

IV.-HIPÓTESIS DEL TRABAJO EXPERIMENTAL.

La diversidad funcional de las proteínas está estrechamente relacionada con su estructura, de manera que estas moléculas codifican secuencias parciales de aminoácidos que son blanco de cambios químicos estructurales a nivel post-traducciona, efectuados a través de reacciones enzimáticas. Muchas de estas modificaciones están implicadas en la regulación de la actividad de las proteínas. Es por esta razón, que el análisis de la estructura primaria de una proteína representa un punto de partida importante para predecir sus funciones. Esta información inicial es de importancia crítica para el investigador, ya que da lugar al diseño y aplicación de bioensayos *in vitro* e *in vivo* específicos para caracterizar tales funciones.

En este contexto, dado que se clonó un ARNm aún no caracterizado en el sistema nervioso de roedor y que codifica una proteína nueva que no presenta homólogos estructurales con funciones identificadas, si se procede a la identificación de secuencias peptídicas consenso o firmas estructurales específicas con ayuda de programas disponibles (PROSITE, por ejemplo), entonces esta información no solo servirá de base para proponer la función de esta proteína llamada Antolefinina, sino que también para plantear estudios futuros dirigidos a su caracterización.

V.-OBJETIVOS DEL TRABAJO EXPERIMENTAL.

General:

El objetivo central de mi proyecto de tesis estuvo enfocado hacia la implementación de una metodología experimental para aislar y caracterizar la estructura del gen que da origen a la proteína precursora de los péptidos opioides llamados Endomorfina-1-2. Sin embargo, debido a que en el curso de estos estudios, se aisló en forma no anticipada, una clona de ADNc que codifica para una molécula proteica que no es el precursor de estos péptidos y que no tiene identidad estructural con proteínas ya identificadas, decidimos replantear el objetivo global hacia el análisis estructural detallado de esta nueva proteína, con el fin de proponer sus posibles roles funcionales a través de la identificación de la presencia en ella de secuencias parciales de aminoácidos con funciones identificadas.

Particulares:

- Identificación de material proteico inmunoreactivo a las endomorfina expresado por una biblioteca de ADNc del cerebro total de ratón empleando anticuerpos contra estos péptidos generados por nuestro equipo de trabajo.
- Aislamiento de las clonas que contienen el vector plasmídico con el o los ADNc's que codifican para las proteínas identificadas por los anticuerpos.
- Purificación del vector plasmídico con los ADNc's identificados y aislados.
- Caracterización del tamaño molecular del los ADNc's clonados mediante un análisis electroforético en geles de agarosa.
- Caracterización de la secuencia nucleotídica de los ADNc's clonados con ayuda del método de síntesis de ADN en presencia de ddNTP's marcados con fluorescencia a través de un secuenciador automático de ADN de un capilar.
- Análisis estructural de la secuencia nucleotídica del ADNc clonado obtenida para determinar si se trata de un ARNm completo. Identificación de posibles marcos de lectura abiertos de traducción y de cajas

nucleotídicas reguladoras en las regiones 5' y 3' no codificantes del ARNm aislado.

- Análisis estructural comparativo para la búsqueda de secuencias nucleotídicas con identidad estructural a nuestro ARNm clonado con ayuda de la base de datos estructural del GenBank.
- Análisis estructural comparativo para la búsqueda de secuencias de aminoácidos con identidad estructural a la proteína deducida (llamada Antolefinina) por el marco de lectura abierto de traducción de nuestro ARNm clonado en las bases de datos estructurales del GenBank y el programa de análisis del EMBL-EBI.
- Análisis de la secuencia primaria de aminoácidos codificada por la Antolefinina a través del uso de programas computacionales de predicción de función proteica (por ejemplo PredictProtein, PROSITE), con el objetivo de determinar la presencia de secuencias parciales de aminoácidos con funciones específicas.

VI. -MATERIALES Y MÉTODOS

a) Rastreo inmunológico de la biblioteca de ADNc de cerebro total de ratón.

1. Descripción de la biblioteca de ADNc

Se utilizó una biblioteca comercial (Stratagene # Catálogo 937319) de expresión de ADNc del cerebro total de ratón adulto (*Mus musculus*, BALB/c), la cual tiene aproximadamente un 80% de ADNc's clonados orientacionalmente (5'—3'). El tamaño molecular promedio de los insertos de ADNc's de esta biblioteca es de 1.8 Kb en el sistema plasmídico pBlueScript SK (+/-) subclonado en el sistema del vector del fago lambda Uni-ZAP-XR (ver tabla 10 y 11 y figura 8). Esta construcción de sistema vector de clonación está diseñada para infectar y transformar ADNc en sistemas celulares bacterianos como es la cepa XL1-Blue MRF' de la bacteria *E.coli*.

Tabla 10: Enzimas que se pueden emplear en el sitio de clonación múltiple del vector pBluescript® SK.

Enzima	No. de sitios de corte	Posición nucleotídica
<i>Kpn I</i>	1	653
<i>Apa I</i>	1	659
<i>EcoO109 I</i>	1	659
<i>Dra II</i>	1	659
<i>Xho I</i>	1	668
<i>Hinc II</i>	1	674
<i>Acc I</i>	1	674
<i>Sal I</i>	1	674
<i>Bsp106 I</i>	1	683
<i>Cla I</i>	1	683
<i>Hind III</i>	1	689
<i>EcoR V</i>	1	695
<i>EcoR I</i>	1	701
<i>Pst I</i>	1	707
<i>Sma I</i>	1	713
<i>BamH I</i>	1	719
<i>Spe I</i>	1	725
<i>Xba I</i>	1	731
<i>Not I</i>	1	737
<i>Eag I</i>	1	738
<i>BstX I</i>	1	744
<i>Sac II</i>	1	749
<i>Sac I</i>	1	755

Tabla 11: Características del vector de clonación pBluescript® SK

Característica	Posición nucleotídica
Origen de replicación f1	24-330
Secuencia codificante del fragmento α de la β -Galactosidasa ($lacZ'$)	463-816
Sitio del promotor T7 para inicio de la transcripción	643
Sitio de clonación múltiple	653-760
Sitio del promotor T3 para inicio de la transcripción	774
Promotor lac	847-938
Origen de replicación pUC	1158-1825
Marco de lectura abierto para la resistencia a ampicilina	1976-2833

2. Titulación y amplificación de la biblioteca

La biblioteca fue titulada y amplificada de acuerdo a un protocolo estándar recomendado por el proveedor (Uni-ZAP® Premade Library Instruction Manual, 2003, Stratagene). Para la formación de placas líticas, se mezcló una alícuota de la biblioteca equivalente a 5×10^4 pfu (unidades formadoras de placas líticas) /placa de Petri de 15 mm de diámetro y 600 μ L de células hospederas *E. coli* de la cepa XL1-BLue MRF' ajustadas a una concentración de 0.5 OD₆₀₀ (Densidad Óptica medida a una longitud de onda de 600 nm). Se incubaron las bacterias y fagos durante 15 minutos a 37 °C para permitir la adhesión e infección de los fagos a estas células. Se mezcló 6.5 mL de agar suave NZY con la mezcla del fago y las bacterias, se esparció en placas de agar NZY y se agitó cuidadosamente la placa para permitir que las células se distribuyeran homogéneamente. Se dejó reposar la placa durante 10 minutos hasta solidificación del agar suave NZY. Se incubaron las placas invertidas a 40°C por 4 horas para permitir la aparición de las placas líticas. Una vez transcurrido este tiempo las placas se enfriaron durante 2 horas a 4°C para evitar que el agar se adhiriera posteriormente a las membranas de nitrocelulosa.

3. Ensayo de inmunodot-blot

Para eliminar problemas que afectaran la sensibilidad del ensayo, como son la presencia de falsos positivos y aumento de la señal de fondo debidos a la reactividad cruzada de los anticuerpos policlonales con algunas de las proteínas de *E. coli* y el fago, se realizó lo siguiente: el estuche con el que se trabajó para el rastreo inmunológico (picoBlue™, Stratagene No. cat.200371) posee un lisado del fago y *E.*

coli, el cual se utilizó para unirlo primero a las membranas de nitrocelulosa embebidas previamente en una solución 10mM (milimolar) de IPTG (Isopropil-1-tio-β-D-galactopiranosido), y enseguida se incubaron éstas con la solución de anticuerpo primario, esto para adsorber los anticuerpos que presentaran una reacción cruzada con las proteínas de la bacteria o del fago.

A la caja de Petri con cultivo bacteriano lisado se le aplicó la membrana de nitrocelulosa tratada como se describió anteriormente. Se permitió que las membranas se humedecieran, evitando el depósito de burbujas. Se incubaron las cajas con agar durante 4 horas a 37°C. Este procedimiento se realizó por duplicado para cada placa de cultivo. Se identificó perfectamente la orientación de la membrana de nitrocelulosa con respecto a la caja con agar. Al cabo del tiempo de incubación se removieron cuidadosamente las membranas con las placas líticas transferidas con ayuda de pinzas y se lavaron por 10 minutos a temperatura ambiente en una solución de PBS 1X (Solución amortiguadora de fosfatos, por sus siglas en inglés Phosphate-Buffered Saline) para eliminar restos del medio de cultivo. Una vez perfectamente lavadas las membranas de nitrocelulosa se incubaron por 12 horas a 4°C con los anticuerpos dirigidos contra la proteína precursora de las Endomorfina, diluidos 1:20 en PBS/ 0.3% gelatina de teleostato (SIGMA)/ 0.3% Tween 20 (SIGMA). Después de esta incubación, las membranas fueron lavadas con la misma solución pero sin anticuerpos para eliminar el exceso de antisuero unido de manera inespecífica a las membranas.

4. Revelado de las membranas

Finalmente las membranas se incubaron en una solución de anticuerpo anti-IgG (Amersham) de conejo marcado radiactivamente con Iodo¹²⁵ (dilución 1:250) durante 2 horas a temperatura ambiente. Las clonas positivas fueron identificadas exponiendo las membranas a películas X-O-Mat (Kodak) por 7 días y revelando con D-19 Developer (Kodak).

b) Purificación del plásmido que contiene el ADNc que da origen a la proteína identificada con los anticuerpos.

Para la purificación del plásmido, se empleó el estuche QIAprep®Spin Miniprep Kit (de la marca QIAGEN # de catálogo 27106). El protocolo usado para este objetivo experimental fue el recomendado por el distribuidor. El procedimiento de este estuche esta basado en la lisis alcalina de las células bacterianas seguido por la absorción de ADN sobre una membrana de sílica-gel en presencia de altas concentraciones de sales. Se siguen tres pasos básicos:

- 1) Preparación y limpieza del lisado de bacterias. Los componentes de la bacteria son desnaturalizados y precipitados y removidos por filtración.
- 2) Adsorción del ADN sobre la membrana QIAprep de sílica-gel. Las características de esta membrana permiten que solo sea retenido ADN de doble cadena, mientras que el ARN, proteínas y otros metabolitos no son retenidos.
- 3) Lavado y elución del ADN plasmídico.

El rendimiento de este estuche es de 5-10 μg de ADN plasmídico/ 5mL de cultivo de bacteria crecido en medio LB (Luria-Bertani) durante toda la noche.

c) Tratamiento con las enzimas de restricción *Xba I* y *Xho I*.

Para la realización de los ensayos de restricción se consideró que una unidad de enzima requiere 1 μg de ADN, y dos horas de incubación a 37°C.

Las enzimas utilizadas *XbaI* (No. catálogo 15226-012) y *XhoI* (No. catálogo 15231-012) fueron de la marca GIBCOBRL® 2000U (10U/ μL) al igual que el marcador de peso molecular (1 kb DNA Ladder 1.0 μg / μL No. catálogo 15615-016).

La determinación de la cantidad de ADN presente en la muestra se realizó mediante medición espectrofotométrica a una longitud de onda de 260nm, la concentración del plásmido fue 0.87 μg / μL .

d) Análisis electroforético del inserto de ADNc en gel de agarosa.

Se preparó un gel de agarosa al 1% con 2.5% de bromuro de etidio. En cada pozo del gel se cargaron 1.74 μg (2.0 μL) de ADN (ver tabla 12).

Tabla 12: Carga de la muestra en el gel para correr la electroforesis

	Pozo 1	Pozo 2	Pozo 3	Pozo 4	Pozo 5
	MW	Plásmido con inserto	Plásmido+XbaI	Plásmido+XhoI	Plásmido+XbaI+XhoI
ADN	0.5 μ L	2.0 μ L	5.0 μ L	5.0 μ L	5.0 μ L
Sol. de carga de muestra (SB 5X)	2.0 μ L	2.0 μ L	2.0 μ L	2.0 μ L	2.0 μ L
Agua mQ	7.5 μ L	6.0 μ L	3.0 μ L	3.0 μ L	3.0 μ L
Volumen total	10 μ L	10 μ L	10 μ L	10 μ L	10 μ L

MW: Marcador de peso molecular
Agua mQ: Agua desionizada

La electroforesis se desarrolló en solución amortiguadora TAE 0.5X con 0.5 μ g/ μ L de bromuro de etidio.

e) Caracterización estructural del ADNc clonado a través de secuenciación nucleotídica automatizada con la ayuda del método de marcaje de nucleótido terminal fluorescente por PCR:

Se empleo el estuche ABI PRISM® BigDye™ (Terminator Cycle Sequencing Ready Reaction Kit Versión 2.0) y un secuenciador automático de un capilar ABI PRISM 310 Genetic Analyzer de Applied Biosystems. El protocolo seguido fue el recomendado por el proveedor y se describe en las tablas 13 y 14.

Tabla 13: Preparación de la reacción para marcaje con nucleótidos fluorescentes

Reactivo	Concentraciones	Cantidades / volúmenes
Molde	200-500 ng (nanogramos) ADN de doble cadena	1.5 μ L
Cebador	3.2 pmol	2.0 μ L
Agua mQ	-	c.b.p. 20 μ L

Tabla 14: Protocolo de secuenciación

Paso	Acción
1	Colocar los tubos para PCR en el termociclador (Eppendorf Mastercycler gradient)
2	Repetir lo siguiente por 25 ciclos: <ul style="list-style-type: none"> • Rampa térmica rápida a 96°C por 10 segundos • Rampa térmica rápida a 50°C por 5 segundos • Rampa térmica rápida a 60°C por 4 minutos
3	<ul style="list-style-type: none"> • Rampa térmica rápida a 4°C, mantenerla hasta el momento de la purificación.

Se emplearon como cebadores el oligonucleótido sentido complementario al promotor T3 de 18 nucleótidos de longitud (5'-TAACCCTCACTAAAGGGA-3') (Promoter Sequencing Primer, marca GIBCO BRL®, 5µg/mL No.catálogo 18417-014), y el oligonucleótido anti-sentido complementario al promotor T7 de 16 nucleótidos de longitud (5'-TAATACGACTCACTATAGGG-3') (Promoter Sequencing Primer, marca GIBCO BRL®, 5µg/mL No.catálogo 18716-016).

Para purificar los productos de extensión de la PCR se usaron las columnas Centri-Sep™ Spin Columns (marca Princeton Separations No.catálogo CS-901) y el protocolo seguido fue el sugerido por el proveedor. Estas columnas se emplean para una purificación rápida y eficiente de moléculas grandes (proteínas, ácidos nucleicos, carbohidratos complejos, etc), eliminándose las moléculas pequeñas como son nucleótidos, sales de la solución amortiguadora, etc.

El diseño de la columna esta basado en un gel deshidratado que tiene la capacidad de retener el exceso de ddNTP's marcados con fluorescencia que no se emplearon en la reacción, dejando en solución solo la cadena nucleotídica marcada fluorescentemente por la extensión de la PCR. El gel remueve con una eficiencia superior al 98% los dNTP's marcados sin reaccionar.

El gel de la columna es hidratado con agua desionizada. El exceso de agua es removido por centrifugación. La muestra es colocada en la columna y recuperada por centrifugación a 14 000 x g durante 1 minuto.

f) Búsqueda de secuencias estructuralmente homólogas a la secuencia nucleotídica del ARNm clonado con ayuda del programa BLAST (NCBI) de la base de datos del GenBank.

El estudio de homología estructural nucleotídica de nuestro ARNm clonado se realizó con el programa BLAST en su modalidad de comparación de nucleótido-nucleótido (Altschul et al 1990) del NCBI (por sus siglas en inglés National Center for Biotechnology Information). Este programa se encargó de comparar secuencia primaria nucleótido-nucleótido de un ARNm contra secuencias de ARNms clonadas y reportadas en la base de datos del GenBank.

g) Búsqueda de secuencias primarias de aminoácidos de péptidos y proteínas homólogas a la secuencia de aminoácidos codificada por el ARNm clonado.

El estudio de homología estructural de la proteína deducida (Antolefinina) de nuestro ARNm clonado en mi proyecto se realizó en la base de datos del EMBI-EBI (European Bioinformatics Institute). Este programa se encargó de comparar aminoácido-aminoácido la secuencia proporcionada de la Antolefinina contra la base de datos de secuencias peptídicas y proteínas clonadas y reportadas en esta base de datos para organismos eucarióticos.

h) Análisis estructural de la secuencia primaria de aminoácidos obtenida para la identificación de la presencia de secuencias parciales de aminoácidos (secuencias consenso) con función identificada en otras proteínas.

Para el análisis estructural de la secuencia primaria de aminoácidos de la Antolefinina codificada por el ARNm clonado en mi proyecto se emplearon los siguientes programas computacionales disponibles en-línea a través del Internet:

■ PROSITE (Bairoch A. et al, 1997)

PROSITE es una base de datos para localizar patrones y perfiles estructurales y funcionales dentro de la secuencia de aminoácidos de péptidos y proteínas.

■ ProParam

Este programa se utilizó para predecir los parámetros fisicoquímicos de la Antolefinina a partir de la secuencia primaria de aminoácidos de ésta. (Gill et al (1989). Anal. Biochem. 182:319-326(1989).

■ TMHMM

TMHMM es un servidor en-línea por internet para llevar a cabo estudios de predicción de tipos de hélices transmembranales en una proteína (Sonnhammer E.L.L. et al 1998. Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology, p 175-182).

■ PHD (Rost B et al, 1996 Methods in Enzymology, 266: 525-539)

PHD es un conjunto de programas encargados de predecir la estructura secundaria de una proteína, (v.g., la Antolefinina) a partir de alineamientos

múltiples con las secuencias de aminoácidos de proteínas disponibles en esta base de datos.

■ PROF (Rost B. 2004, Mol. Biol, Artículo en revisión para ser publicado)

PROF es un conjunto de programas encargados de predecir la estructura secundaria de una proteína a partir de alineamientos múltiples con las secuencias de aminoácidos de proteínas disponibles en esta base de datos.

■ Predictor of NON-Regular Secondary Structure

Esta herramienta es utilizada para predecir estructura secundaria no regular en una proteína. (B Rost and J Liu (2003) The PredictProtein Server. Nucleic Acids Research 31(13): 3300-3304).

■ GLOBE prediction of globularity (Rost B. 1998, artículo no publicado)

Este programa se encarga de predecir si la proteína tendrá una conformación globular.

Todos los programas anteriores se encuentran disponibles en el servidor de consulta por internet en-línea del PredictProtein (B. Rost and J.Lui, 2003).

VIII. Resultados

Resultado 1: Obtención de una clona inmunopositiva a partir del rastreo inmunológico efectuado.

Como resultado de la metodología experimental implementada se llegó a la identificación de una clona inmunopositiva de ADNc. Ésta última, fue sometida a un proceso de purificación para tener en solución únicamente al plásmido con el inserto de ADNc subclonado. Una primera aproximación para conseguir información acerca del tamaño del inserto de ADNc subclonado en el vector de clonación aislado fue un análisis electroforético en geles de agarosa.

Resultado 2: Obtención de un inserto de ADNc menor a 1kb de tamaño molecular purificado de la clona inmunopositiva en el ensayo.

El ADNc purificado del plásmido de la clona inmunopositiva aislada en mi trabajo de tesis fue sometido a la caracterización de su tamaño molecular. La figura 9 muestra la identificación del tamaño molecular del inserto de ADNc escindido con las enzimas de restricción *Xba*I y *Xho*I en un análisis electroforético en un gel de agarosa al 1% y tinción con bromuro de etidio.

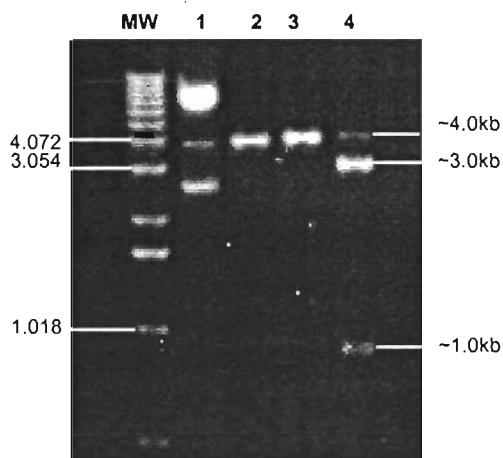


Figura 9: Análisis electroforético del inserto de ADNc en gel de agarosa

- MW: Marcador de peso molecular
- 1: Plásmido con el inserto de ADNc
- 2: Plásmido tratado con la enzima *Xba*I
- 3: Plásmido tratado con la enzima *Xho*I
- 4: Plásmido tratado con la enzima *Xba*I y *Xho*I

En el carril 1 de este gel se observan los tres productos de ADN resueltos por el gel que corresponden al tamaño molecular del plásmido con el inserto de ADNc sin ningún tratamiento previo de restricción enzimática. De la parte superior a la inferior, del carril, la primera de estas bandas de ADN pertenece a la forma estructural de plásmido desplegado con un tamaño molecular superior a las 4.0 Kb (kilobases), la segunda banda es representativa de la forma estructural de plásmido enrollada (≈ 4 Kb) y la tercera, con mayor movilidad electroforética, es la forma estructural del plásmido superenrollado (menor de 3 Kb). Las bandas de ADN obtenidas en los carriles 2 y 3 son el resultado del tratamiento del plásmido con el inserto de ADNc restringido respectivamente con la enzima *XbaI* o *XhoI*. De este resultado se deduce que el plásmido tiene dos sitios únicos de corte, uno para cada enzima, dado que se observa una única banda por carril de aproximadamente 4.0 kb. Esta masa molecular corresponde al plásmido enrollado no restringido enzimáticamente del carril 1.

En contraste, en el último carril del gel (No. 4), se pueden apreciar tres bandas de ADN, la primera de 4.0 Kb representa material plasmidico que no fue digerido por las enzimas utilizadas, la segunda banda de ADN, con un tamaño molecular aproximado de 3.0 kb representa a el plásmido vacío y linearizado, ya que la masa molecular corresponde con la reportada por el fabricante de la biblioteca (Stratagene). Finalmente la última banda de ADN, con una masa molecular menor a 1.0 Kb corresponde al inserto de ADNc clonado que fue escindido por el tratamiento simultáneo con ambas enzimas.

Con esta información preliminar sobre el tamaño molecular aproximado del inserto de ADNc clonado (1Kb, aproximadamente), se procedió a caracterizar la estructura primaria de ambas cadenas de la secuencia nucleotídica total de esta molécula con la técnica de síntesis de ADN en presencia de ddNTP's marcados con fluorescencia, ya que la metodología de secuenciación estandarizada en nuestro laboratorio (BigDye Terminator Cycle Sequencing Ready Reaction V. 2.0) era capaz de proporcionar la secuenciación de segmentos nucleotídicos con confiabilidad y resolución arriba de los 600 nucleotidos en longitud. Así, llevamos a cabo la secuenciación bi-direccional del inserto de ADNc a partir de sus extremos 5' y 3', a través de PCR con oligonucleótidos cebadores complementarios a los sitios promotores T3 y T7, respectivamente.

Resultado 3: Caracterización de la secuencia primaria de nucleótidos del ADNc clonado.

La secuenciación nucleotídica completa del ADNc clonado dio lugar a la identificación de un ADNc de 821 pares de bases de longitud que fue registrado en la base de datos del GenBank con el registro de acceso GI: 37992801. El registro original se muestra en la figura 10.

La secuencia nucleotídica del ADNc clonado codifica un ARNm que presenta hechos estructurales de un ARNm completo: un codón de inicio (ATG) en la posición 22, otro de término (TGA) en la posición 761 y un marco de lectura abierto de 736 nucleótidos (ORF, por sus siglas en inglés Open Reading Frame). Adicionalmente, se tienen los segmentos nucleotídicos no codificantes 5'-UTR y 3'-UTR (por sus siglas en inglés Untranslated Region), los cuales son de 21 nucleótidos y 58 nucleótidos respectivamente. El segmento 3'-UTR cuenta con la secuencia de poli(A) característica de los ARNm eucarióticos maduros y por otra parte la señal consenso de poliadenilación en la posición 784 de la región 3'-UTR. En la región 5'-UTR, esta codificada una secuencia nucleotídica CAAT, la cual es representativa de una secuencia consenso regulada por factores de transcripción (ver tabla 15 y figura 11).

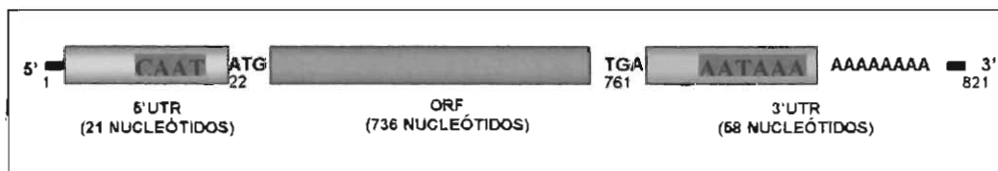


Figura 11: Dibujo esquemático del ARNm clonado

El ARNm codificado por el ADNc clonado presenta características estructurales de un ARNm completo, es decir, regiones 5'- y 3'-UTR, un solo codón preferencial de inicio y otro de término de traducción, una sola región codificante para proteína (ORF), una región nucleotídica consenso de poliadenilación y una cola en el extremo 3' de poli-A encargada de proporcionar estabilidad a este ARNm.

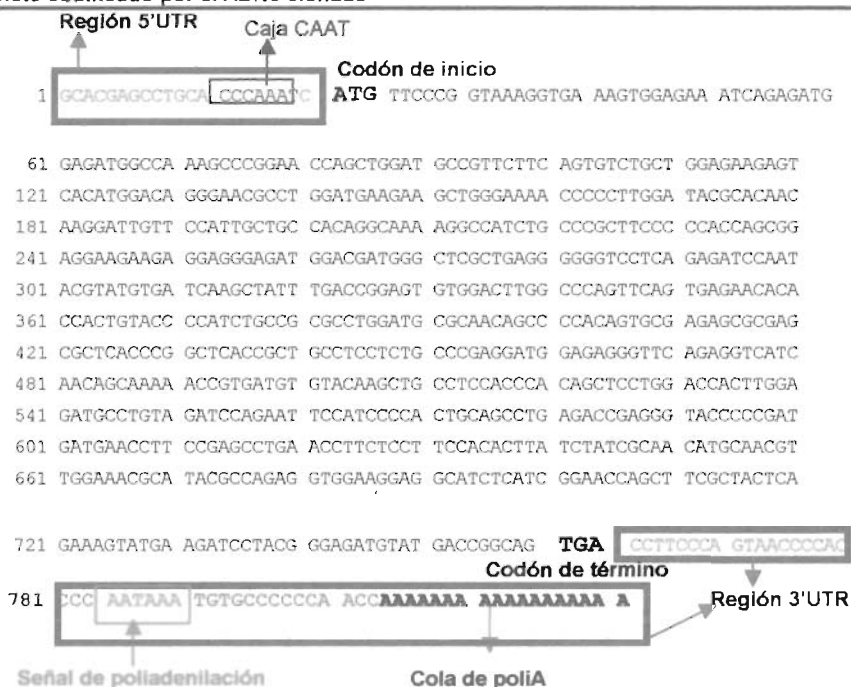
```

1: AY428767. Mus musculus stra...[gi:37992801]
LOCUS      AY428767                821 bp    mRNA    linear    ROD 01-NOV-2003
DEFINITION Mus musculus strain BALB/c tissue-type brain antolefinine mRNA,
complete cds.
ACCESSION  AY428767
VERSION    AY428767.1    GI:37992801
KEYWORDS   .
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 821)
AUTHORS    Anton,B., Leff,P., Matus,M., Gonzaga,R., Medina,R., Calva,J.C.,
            Acevedo,R., Martinez,C., Retana,I., Saavedra,R., Arias,A.,
            Zavala,E., Pavon,L. and Alagon,A.
TITLE      Cloning of a novel 821 bp cDNA from a whole mouse brain with an ORF
            encoding a 246 aa protein referred to as Antolefinine
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 821)
AUTHORS    Anton,B., Leff,P., Matus,M., Gonzaga,R., Medina,R., Calva,J.C.,
            Acevedo,R., Martinez,C., Retana,I., Saavedra,R., Arias,A.,
            Zavala,E., Pavon,L. and Alagon,A.
TITLE      Direct Submission
JOURNAL    Submitted (07-OCT-2003) Laboratory of Molecular Neurobiology and
            Addictive Neurochemistry, National Institute of Psychiatry, Calzada
            Mexico-Xochimilco # 101, Mexico City, D.F. 14370, Mexico
FEATURES   Location/Qualifiers
            source                1..821
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="BALB/c"
            /db_xref="taxon:10090"
            /chromosome="5"
            /map="LOC231020; between LOC231019 and LOC330032"
            /tissue_type="brain"
            /clone_lib="Uni-Zap XR, Stratagene Cat. # 937319"
            CDS                    22..762
            /note="unknown neurobiological function"
            /codon_start=1
            /product="antolefinine"
            /protein_id="AAR06606.1"
            /db_xref="GI:37992802"
            /translation="MPPVKVKVEKSEMEMAKARNQLDAVLQCLLEKSHMDRERLDEEA
            GKTFLDTHNKDCSIAATGKRPSARFPHQRRKKRREMDDGLAEGGPQRSNTYVIKLFDR
            SVDLQAQFSENTPLYICRAWMRNSPTVREERERSPGSPLPLPEDGEGSEVINSKNRDV
            YKLPPTAPGPLGDACRSRIPSPFLQPETEGETPDDEPSEPEPSPSTLIYRNMQRWKRIR
            QRWKEASHRNQLRYSESMKILREMYDRQ"
ORIGIN     1 gcacgagcct gcacccaaat catgttcccc gtaaaggatga aagtggagaa atcagagatg
           61 gagatggcca aagcccgcaa ccagctggat gccgttcttc agtgtctgct ggagaagagt
           121 cacatggaca ggaacgcctc ggatgaagaa gctgggaaaa cccctctgga tacgcacaac
           181 aaggattgtt ccatgtctgc cacaggcaaa aggccatctg cccgctcccc ccaccagcgg
           241 aggaagaaga ggaggagatg ggacgatggg ctgctgagg ggggtcctca gagatccaat
           301 acgtatgtga tcaagetatt tgaccggagt gtggacttgg cccagttcag tgagaacaca
           361 ccactgtacc ccactctgcc cgcttggatg cgcaacagcc ccacagtgcg agagcgcgag
           421 cgctcaccgg gctcaccgct gcctctctct cccgaggatg gagagggttc agaggtcatc
           481 aacagcaaaa accgtgatgt gtacaagctg cctccaccca cagctcctgg accacttgya
           541 gatgcctcta gatccagaat tccatcccca ctgcagcctg agaccggagg tacccccgat
           601 gatgaacctt ccgagcctga accttctcct tccacactta tctatcgcaa catgcaacgt
           661 tggaaacgca tacgccagag gtggaaggag gcattctcat ggaaccagct tcgctactca
           721 gaaagtatga agatctctac ggagatgtat gaccggcagt gaccttccca gtaacccca
           781 cccaataaat gtgcccccca accaaaaaaa aaaaaaaaaa a

```

Figura 10: Registro en la base de datos del GenBank de la secuencia primaria de nucleótidos del ARNm y de la secuencia de aminoácidos de la proteína codificada por el único marco de lectura abierto de traducción encontrado en este ARNm.

Tabla 15: Resumen de las características estructurales de la secuencia del ARNm completo codificado por el ADNc clonado



Para obtener información adicional acerca de la identidad estructural de la secuencia nucleotídica de ARNm clonada y registrada en la base de datos del GenBank se procedió a una búsqueda de secuencias nucleotídicas homólogas en esta misma base de datos.

Resultado 4: Resultado de la búsqueda de secuencias nucleotídicas homólogas a la secuencia del ARNm clonado.

En la figura 12 (ver abajo), se presenta en forma esquemática el resultado de la búsqueda en la base de datos del GenBank del alineamiento estructural del ARNm clonado con secuencias homólogas de ácidos nucleicos ya clonados y reportados en esta base de datos. Los resultados de estos estudios, nos mostraron que al menos diez secuencias nucleotídicas, reportadas después que la de nosotros, presentan

segmentos nucleotídicos de diferente longitud con una alta homología estructural contra casi toda la secuencia nucleotídica de nuestro ARNm. Estas secuencias, de diferente longitud de nucleótidos, tienen un perfil de alta identidad estructural de segmentos de sus secuencias con nuestra secuencia completa de ARNm ($\approx 85-99\%$, ver tabla 16).

Clave de color para las marcas de alineamiento

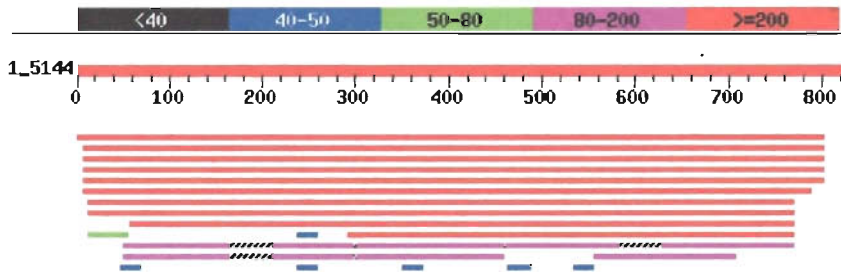


Figura 12: Representación esquemática en formato del GenBank del estudio de alineamiento estructural entre el ARNm clonado con secuencias de ácidos nucleícos reportados en esta base de datos.

La longitud de la secuencia de nuestro ARNm está representada por la línea roja gruesa en la parte superior de la figura, y esta subdividida por números (en negro) que representan el número progresivo de residuos de nucleótidos en una orientación 5'-3' de la izquierda a la derecha. Los índices numéricos de homología estructural está definida arbitrariamente por el programa por una clave de cinco colores diferentes que dividen el intervalo de marcas en cinco grupos. El rojo indica una homología muy alta y el negro muy baja. Los colores restantes indican índices de homolías intermedias.

Tabla 16: Resumen de las secuencias estructuralmente homólogas al ARNm clonado.

Nota: La descripción se da en el mismo orden, de arriba hacia abajo, en que se presentan las líneas en color rojo en la figura 12. Estas líneas son representativas de segmentos de ADN que tienen muy alta homología estructural con nuestra clona de referencia (la primera línea roja gruesa con números de la figura).

Descripción	%de Identidad	No. de bases comparadas	Comentarios
1. ADNc de páncreas de ratón macho de 10 días de nacido Fecha de registro: 03 abril 2004	99%	795	La secuencia nucleotídica del ADNc de páncreas difiere por 8 nucleótidos adicionales presentes en la región 5'UTR, el resto de la secuencia presenta 100% de identidad.
2. ADN de ratón Gen 1810054G18 Fecha de registro: 21 diciembre 2003	99%	795	Esta secuencia nucleotídica proviene de ADN genómico y tiene un marco de lectura abierto que presenta el 100% de identidad con la secuencia nucleotídica obtenida.
3. ADN proveniente del cromosoma 5 de ratón Clona RP23-212F6 Fecha de registro: 01 enero 2004	95%	795	Corresponde a una porción de una secuencia genómica con marco de lectura abierto no identificado.

4. ADN de rata (<i>Rattus norvegicus</i>) Fecha de registro: 23 octubre 2003	94%	781	Este registro se deriva de una secuencia genómica para la cual se aplicó el método de predicción de genes GNOMON. La secuencia nucleotídica sólo difiere por 8 nucleótidos adicionales presentes en la región 5'UTR, el resto de la secuencia presenta 100% de identidad.
5. ADNc de humano (<i>Homo sapiens</i>) Proteína F25965 Fecha de registro: 12 noviembre 2003	87%	758	El ARNm se extrajo de retinoblastoma de ojo. Este gen codifica una proteína que difiere en 12 aminoácidos con respecto a la proteína presentada. Esta secuencia cuenta con todas las características estructurales de un ARNm completo.
6. ADNc de humano Fecha de registro: 01 abril 2004	87%	712	Secuencia parcial de ARNm que no presenta marco de lectura abierto.
7. ADNc de humano Proteína MSTP064 Fecha de registro: 01 agosto 2003	85%	477	La identidad estructural observada con nuestro ARNm se localiza principalmente al nivel de la región codificante.
8. ADN humano proveniente del cromosoma 19. Fecha de registro: 16 octubre 1997	88%	664	Se trata de ADN genómico. La identidad estructural al nivel nucleótido-nucleótido aparece en diferentes porciones debido a la presencia de intrones: 53-298, 304-583 y 631-771
9. ADNc de humano Fecha de registro: 18 febrero 2000	88%	362	La secuencia estructural de este ADNc con nuestra secuencia nucleotídica del ARNm clonado en nuestro proyecto se localiza en diferentes segmentos: 53-166 y 211-460. No se presenta marco de lectura abierto.
10. ADN de ratón proveniente del cromosoma X Clona RP23-434M18 Fecha de registro: 24 agosto 2003	82%	149	Corresponde a un segmento de ADN genómico secuenciado, la cual no presenta marco de lectura abierto.

El ADNc que codifica el ARNm completo clonado mostró una identidad estructural superior al 80%, en casi toda su longitud, con segmentos nucleotídicos de 10 secuencias de ADN reportadas en el GenBank. La primera de ellas, pertenece a un ADNc clonado de páncreas de ratón, el cual no codifica para una cola de poli(A) característica de los ARNm maduros. En esta forma, al no saber la orientación correcta de este ADNc, su marco de lectura abierto identificado fue reportado para traducir una proteína de carácter hipotético, cuya secuencia primaria es idéntica a la codificada por el ARNm clonado. Una situación similar se presenta con las secuencias 2, 3, 4, 8 y 10, ya que se trata de secuencias nucleotídicas de ADN genómico, que aún contiene intrones, y por lo cual las regiones codificantes deducidas se derivan de programas de predicción (por ejemplo GNOMON) o no se llegan a presentar. Para las secuencias

nucleotídicas 6 y 9, aún cuando fueron determinadas de ADNc's, no codifican para ARNm's completos, por lo que en ningún caso se reportó región codificante identificada. La secuencia nucleotídica 5, corresponde a un ADNc clonado del humano y codifica a un ARNm completo con una región codificante que presenta una alta identidad estructural ($\geq 95\%$) con la secuencia de ARNm clonada. Por último, la secuencia nucleotídica 7 que es también de un ADNc clonado del humano, da origen a un producto proteico estructuralmente diferente al codificado por el ARNm aislado en este trabajo experimental.

De acuerdo a todo lo descrito anteriormente, el ADNc clonado representa a un ARNm completo que no había sido aislado ni caracterizado estructuralmente con anterioridad. Además, tiene un ortólogo del humano, por lo que se deduce que se expresa en otras especies de mamíferos. En este contexto, el siguiente trabajo experimental de mi tesis fue enfocado hacia estudios de la caracterización estructural y de predicción funcional de la proteína que codificaba este nuevo ARNm clonado. Como se observa en el registro original de la bases de datos pública del GenBank de la figura 10, esta nueva proteína se llamó Antolefinina.

Resultado 5: Análisis estructural de la secuencia primaria de aminoácidos de la Antolefinina.

Como se mencionó en el inciso anterior la secuencia primaria de la proteína llamada por nuestro grupo de trabajo Antolefinina fue registrada en la base de datos del GenBank (Figura 10). Asimismo, en la figura 13 se observa el análisis predictivo de la secuencia peptídica de la Antolefinina con el programa PROSITE.

Si bien en la secuencia de la Antolefinina no aparece el tetrapéptido correspondiente a la EM-1 o la EM-2, se observan características estructurales interesantes que por orden de importancia se describen a continuación:

- Fragmento localizado en la posición 71 para localización en el núcleo (NLS), una región rica en prolina hacia el extremo C-terminal y región rica en leucinas en el extremo N-terminal que constituye una señal para exportación del núcleo (NES). Como se mencionó en la introducción estas tres características corresponden a proteínas destinadas a núcleo.

- Sitio consenso en la posición 61 para α -amidarse. Sabemos que la α -amidación es sinónimo de actividad biológica de la molécula.
- Sitios para fosforilación. La fosforilación representa una manera de regular la actividad biológica de una proteína por lo que la presencia de esta modificación es también un indicativo fuerte de actividad biológica.
- Otras propiedades relevantes son la presencia de tres sitios potenciales para el procesamiento proteolítico con ayuda de enzimas especializadas que reconocen y escinden donde se encuentran pares de aminoácidos básicos.

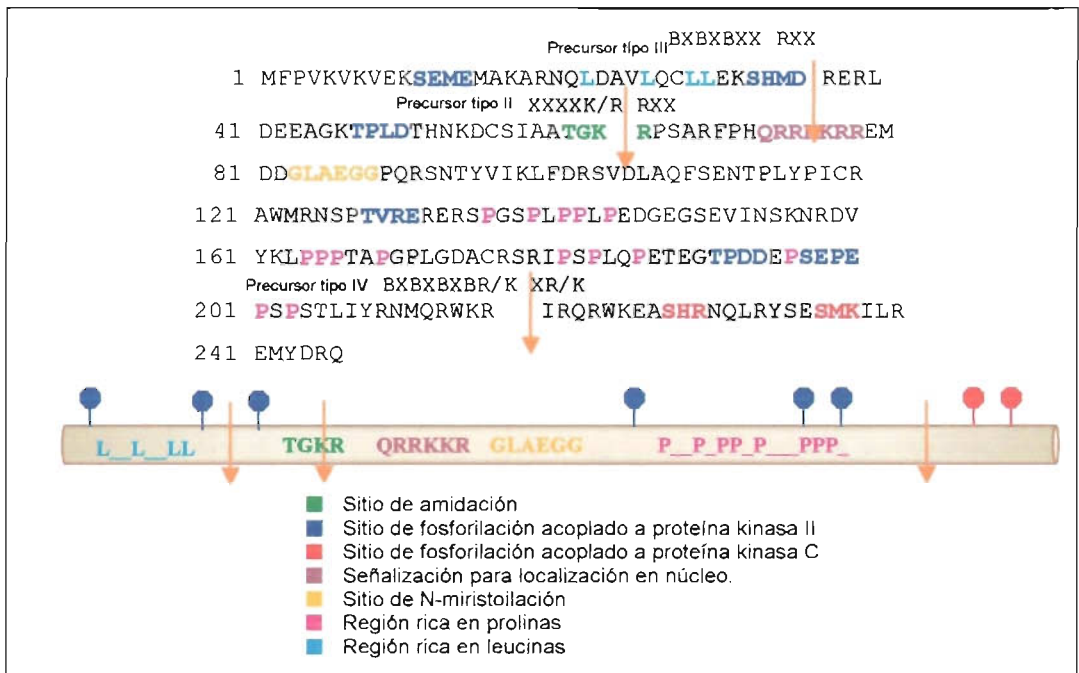


Figura 13: Posibles sitios funcionales y de procesamiento postraduccional encontrados con ayuda del programa PROSITE.

En la secuencia se muestran sitios potenciales de corte proteolítico (flechas color naranja), de fosforilación (Kishimoto A., et al 1985, Pinna L.A. 1990), de amidación, de N-miristoilación, secuencias ricas en prolinas y leucinas así como una señal para localización en el núcleo.

Esta serie de características resultan interesantes desde el punto de vista funcional, en la discusión de resultados se analiza cada uno de estos sitios consenso para tratar de establecer la actividad de la Antolefinina en la fisiología de las células nerviosas.

La idea de tener una proteína cuya secuencia nucleotídica aún no se ha caracterizado y con posibles sitios funcionales muy interesantes, nos hace preguntarnos ahora si se trata de una proteína exclusiva del sistema nervioso o si se encuentra en otros tejidos. Por otra parte es interesante conocer también si comparte identidad a nivel estructural con otras proteínas ya caracterizadas.

De esta manera se procedió a realizar una segunda búsqueda de estructuras homólogas pero ahora a nivel de la estructura primaria de aminoácidos.

Resultado 6: Resultado de la búsqueda de secuencias peptídicas homólogas a la Antolefinina.

De acuerdo a lo observado en la tabla 17 y figura 14 las proteínas que comparten una homología superior al 30% con la Antolefinina son proteínas hipotéticas aún no caracterizadas. No obstante, las proteínas que si se encuentran descritas como son la 8, 9 y 10 (que pertenecen a la familia SPRP) tienen la particularidad de compartir con la estructura primaria de la Antolefinina solo la región rica en prolinas, por lo cual se podría intuir que nuestra proteína puede estar relacionada a la familia antes mencionada. La secuencia de señalización para ser transportada a núcleo y las múltiples regiones con pares de aminoácidos básicos para procesamiento proteolítico solo se conservan en las secuencias de aminoácidos derivadas de páncreas, hígado, ojo y aorta.

Tabla 17: Proteínas homólogas a la Antolefinina

No.	Homología (%)	Tejido y especie	Observaciones
1	100	Páncreas Mus musculus (ratón)	Proteína hipotética (corresponde a la mencionada en la tabla 12, secuencia No.1).
2	100	Hígado Mus musculus (ratón)	Proteína derivada de un ADNc, todavía no caracterizada
3	95	Ojo Homo sapiens (humano)	Proteína derivada de un ADNc, todavía no caracterizada
4	91	Aorta Homo sapiens (humano)	Proteína derivada de un ADNc, todavía no caracterizada
5	35	Ventrículo Homo sapiens (humano)	Proteína derivada de un ADNc, todavía no caracterizada
6	34	Timo Mus musculus (ratón)	Proteína derivada de un ADNc, todavía no caracterizada
7	34	Cerebro Homo sapiens (humano)	Proteína derivada de un ADNc, todavía no caracterizada

8	25	Traquea Sus scrofa (cerdo)	Clasificación: Proteína estructural Proteína regulada por la vitamina A en células epiteliales de la traquea, ésta es inducida en los tumores pulmonares. No se detecta en tejidos pulmonares sanos, solo en los tejidos tumorales. Familia: pertenece a la familia SPRP (por sus siglas en inglés Small Proline-Rich Proteins).
9	25	Epitelio traqueobranquial Macaca mulata (primate)	Clasificación: Proteína estructural Función: proteína ligada a la envoltura de los keratinocitos. Primero aparece en el citosol, pero posteriormente se convierte en una proteína ligada a la membrana, lo que da como resultado la formación de una envoltura insoluble por debajo de la membrana plasmática. Inducción: Durante la diferenciación escamosa de los keratinocitos Familia: pertenece a la familia SPRP
10	24	Piel Homo sapiens (humano)	Clasificación: Proteína estructural Función: proteína ligada a la envoltura de los keratinocitos. Primero aparece en el citosol, pero posteriormente se convierte en una proteína ligada a la membrana, lo que da como resultado la formación de una envoltura insoluble por debajo de la membrana plasmática. Inducción: Durante la diferenciación escamosa de los keratinocitos Familia: pertenece a la familia SPRR

Con esta segunda búsqueda de secuencias homólogas se confirmó que la Antolefinina es una proteína no caracterizada en tejido nervioso o algún otro órgano. Por lo tanto, el último paso para obtener información adicional sobre la Antolefinina fue la predicción de parámetros fisicoquímicos y estructura secundaria.

```

1 - - - - - 101
SECUENCIA MFPPVKVKEKSEMEMAQARNQLDAVLQ LLEKSHMDRERLDEEAGKTPLDTHNKD SIAATGKRPSARFPFHQRKRRKREDDGLAEGGPQRSNTYVIKLF
1 PANCREAS: MFPPVKVKEKSEMEMAQARNQLDAVLQ LLEKSHMDRERLDEEAGKTPLDTHNKD SIAATGKRPSARFPFHQRKRRKREDDGLAEGGPQRSNTYVIKLF
2 HIGADO: -----MAQARNQLDAVLQ LLEKSHMDRERLDEEAGKTPLDTHNKD SIAATGKRPSARFPFHQRKRRKREDDGLAEGGPQRSNTYVIKLF
3 OJO: MFPPVKVKEKSELEMAQARNQLDAVLQ LLEKSHMDRERLDEEAGKTPSDTHNKD SIAATGKRPSARFPFHQRKRRKREDDGLAEGGPQRSNTYVIKLF
4 AORTA: -----
5 VENTRICULO: -----
6 TIMO: -----
7 CEREBRO: -----
8 TRAQUEA: -----
9 EPITELIO TRAQUEOBRANQUIAL: -----
10 PIEL (H.SAPIENS): -----

202 - - - - - 202
SECUENCIA RSVDLAQFSENTPLYPICRAMRNSPTVREERERSPGSLPPLPEDGEGSEVINSGNRDVYKLPPTAPGPLGDA CSRIPSPQLQFETEGTTPDDEPSEPEPS
1 PANCREAS: RSVDLAQFSENTPLYPICRAMRNSPTVREERERSPGSLPPLPEDGEGSEVINSGNRDVYKLPPTAPGPLGDA CSRIPSPQLQFETEGTTPDDEPSEPEPS
2 HIGADO: RSVDLAQFSENTPLYPICRAMRNSPTVREERERSPGSLPPLPEDGEGSEVINSGNRDVYKLPPTAPGPLGDA CSRIPSPQLQFETEGTTPDDEPSEPEPS
3 OJO: RSVDLAQFSENTPLYPICRAMRNSPTVREERECSPSSPLPPLPEDEEGSEVTNSKSRDVYKLPPTPPGPPGDA CSRIPSPQLQFEMQGTTPDDEPSEPEPS
4 AORTA: -----MNSPSVREERECSPSSPLPPLPEDEEGSEVTNSKSRDVYKLPPTPPGPPGDA CSRIPSPQLQFEMQGTTPDDEPSEPEPS
5 VENTRICULO: -----PKLEKVKKPAVPEPPPPKPVVEEVEVPTVKRE -RKIPEPTKVFEIKPAIPLPAPEPKKFEAEKTIKPPVVEPEPT
6 TIMO: -----ETSPGSP -PPLSLPSGPGQP-ETSPGSPCPPLPSGPGQPETSPGSPCPPLSLPSP-GPQFEGSP PPSLSPFPQFEASGSPDSSPQQPEPH
7 CEREBRO: -----ETSPGSP -PPLSLPSGPGQP-ETSPGSPCPPLPSGPGQPETSPGSPCPPLSLPSP-GPQFEGSP PPSLSPFPQFEASGSPDSSPQQPEPH
8 TRAQUEA: -----QQQKQPCTP-PPQFQQQQVKQPCQPPQVEVKTKEPCHP-KVPEP-QPKVPEPCQPKVPEVPEPQPKVPEPC
9 EPITELIO TRAQUEOBRANQUIAL: -----QQQKQPCTP-PPQFQQQQVKQPCQPPQVEVKTKEPCLP-KVPEP-HPKVPEPCQPKV---PEPHKVPPEPC
10 PIEL (H.SAPIENS): -----QQQKQPCTP-PPQFQQQQVKQPCQPPQVEVKTKEPCQK-KVPEP-HPKVPEPCQPKI---PEPQPKVPEPC

203 - - - - - 246
SECUENCIA PSTLIYRNMQRWKRIRQRWKEASHRNQLRYSESMCILREMYDRQ
1 PANCREAS: PSTLIYRNMQRWKRIRQRWKEASHRNQLRYSESMCILREMYDRQ
2 HIGADO: PSTLIYRNMQRWKRIRQRWKEASHRNQLRYSESMCILREMYDRQ
3 OJO: PSTLIYRNMQRWKRIRQRWKEASHRNQLRYSESMCILREMYERQ
4 AORTA: PSTLIYRNMQRWKRIRQRWKEASHRNQLRYSESMCILREMYERQ
5 VENTRICULO: P-----
6 TIMO: PGSL-----
7 CEREBRO: PGSL-----
8 TRAQUEA: PSPVI-----
9 EPITELIO TRAQUEOBRANQUIAL: PSTV-----
10 PIEL (H.SAPIENS): PSTV-----

```

Figura 14: Alineamiento de la secuencia primaria de las proteínas homólogas a la Antolefinina

Resultado 7: Predicción de los parámetros fisicoquímicos de la Antolefina

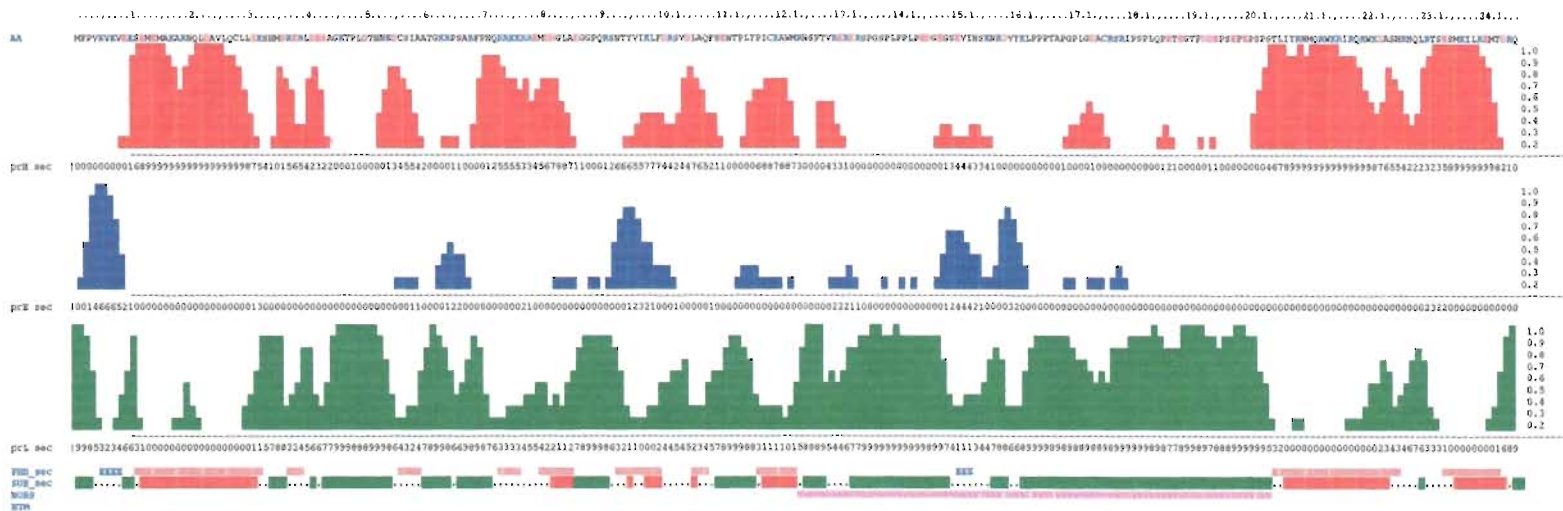
La predicción de parámetros fisicoquímicos tiene su utilidad sobre todo a nivel de estudios futuros para aislar y purificar a la nueva proteína.

Tabla 18: Predicción de los parámetros fisicoquímicos con el programa ProtParam:

Peso molecular estimado: 28378.1					
Composición:					
Ala (A)	13	5.3%	Ile (I)	8	3.3%
Arg (R)	28	11.4%	Leu (L)	18	7.3%
Asn (N)	9	3.7%	Lys (K)	16	6.5%
Asp (D)	15	6.1%	Met (M)	9	3.7%
Cys (C)	4	1.6%	Phe (F)	4	1.6%
Gln (Q)	10	4.1%	Pro (P)	26	10.6%
Glu (E)	23	9.3%	Ser (S)	20	8.1%
Gly (G)	11	4.5%	Thr (T)	10	4.1%
His (H)	4	1.6%	Trp (W)	3	1.2%
Tyr (Y)	6	2.4%	Val (V)	9	3.7%
Número total de residuos cargados negativamente: (Asp + Glu): 38					
Número total de residuos cargados positivamente (Arg + Lys): 44					

Resultado 8: La Antolefinina presenta una estructura secundaria no convencional

Los programas GLOBE prediction of globularity, Ambivalent Sequence Predictor, PROF predictions y Predictor of NON-Regular Secondary Structure dan resultados consistentes entre ellos, es decir todos muestran que la Antolefinina posee una estructura secundaria no globular constituida principalmente de α -hélices y asas (ver figura 15). Se predice que esta proteína no tendrá una conformación globular.



Donde:

AA: secuencia de aminoácidos

■ Conformación de hélice α

■ Conformación de hoja β

■ Conformación de asa

PrH: Probabilidad de tener una conformación de hélice (9= alta. 0=baja)

PrE: Probabilidad de tener una conformación de hoja (9= alta. 0=baja)

PrL: Probabilidad de tener una conformación de asa (9= alta. 0=baja)

PHD_sec: Predicción de estructura secundaria (Profile network prediction HeiDelberg)

H= hélice

E= hoja

"..."= asa

SUB_sec: Predicción de estructura secundaria para la cual se tiene una exactitud superior al 82%

NORS: Estructura secundaria no común (N= región NORS)

HTM: Hélices transmembranales (m= hélice transmembranal, espacio en blanco= no se tiene)

Figura 15: Predicción de estructura secundaria

VII. Análisis de resultados y conclusiones.

El objetivo central de mi proyecto de tesis estuvo dirigido hacia la implementación de una metodología experimental para aislar una clona de ADN que codificara el gen que expresa a la proteína precursora de las Endomorfina, sin embargo a lo largo del desarrollo experimental de este objetivo inicial se aisló de manera no anticipada una clona de ADNc que codifica un ARNm con un marco de lectura abierto de traducción de una proteína denominada por nuestro grupo de trabajo Antolefinina. Debido a este hallazgo el abordaje experimental de mi tesis se replanteo para caracterizar el producto de dicho gen del sistema nervioso de mamífero.

En primer lugar, a partir de un análisis electroforético (resultado 2) se dedujo que el tamaño aproximado del inserto de ADNc era menor de 1000 pares de bases, confirmando esto posteriormente con la determinación de la secuencia nucleotídica (resultado3).

Es importante mencionar que se tomaron diversos criterios para considerar relevante a la secuencia nucleotídica obtenida. En primer lugar se confirmó que el inserto de ADNc se hubiera secuenciado de manera total, es decir que las secuencias obtenidas con los cebadores T3 y T7 para los extremos 5' y 3' respectivamente fueran complementarias en algún sitio y que además se tuviera la cola de poli(A) característica de los ARNm maduros:

```

T3 → 5' GCACGAGCCTGCACCCAAATCATGTTCCCGGTAAAGGTGAAAGTGGAGAAATCAG
AGATGGAGATGGCCAAAGCCCAGGACCCAGCTGGATGCCGTTCTTCAGTGTCTGCTGGAGAA
GAGTCACATGGACAGGGAACGCCTGGATGAAGAAGCTGGGAAAACCCCTGGATACGCAC
AACAAAGGATGTTCCATTTGCTGCCACAGGCAAAAGGCCATCTGCCCGCTTCCCCACCAGC
GGAGGAAGAAGAGGAGGGAGATGGACGATGGGCTCGCTGAGGGGGG TCCTCAGAGATCC
Complementarias ← T7 → TCCTCAGAGATCCAA
TACGTATGTGATCAAGCTATTTGACCCGAGTGTGGACTTGGCCCAGTTCAGTGAGAACACA
CCACTGTACCCCATCTGCCGCGCCTGGATGCGCAACAGCCCACAGTGGAGAGCGCGAGC
GCTCACCCGGCTCACCGCTGCCTCCTCTGCCCGAGGATGGAGAGGGTTCAGAGGTCATCAA
CAGCAAAAACCCGTGATGTGTACAAGCTGCCTCCACCCACAGCTCCTGGACCACTTGGAGAT
GCCTGTAGATCCAGAATTCCATCCCCACTGCAGCCTGAGACCGAGGGTACCCCCGATGATG
AACCTTCGAGCCTGAACCTTCTCCTTCCACACTTATCTATCGCAACATGCAACGTTGGAA
ACGCATACGCCAGAGGTGGAAGGAGGCATCTCATCGGAACCAGCTTCGCTACTCAGAAAGT
ATGAAGATCCTACGGGAGATGATGACCGGCAGTGACCTTCCAGTAACCCACCCCAATA
AATGTGCCCCCAACC.AAAAAAAAAAAAAAAAAAAAAA 3'

```

Asimismo, se consideró importante que la secuencia se encontrara orientada en sentido 5'→3', lo cual se corroboró con la presencia de la cola de poli(A) dispuesta

específicamente en el extremo 3'. Otro punto analizado fue la presencia de elementos de regulación ubicados en las regiones 5' y 3'UTR. Se sabe que la expresión de un gen puede ser regulada a varios niveles, incluyendo la transcripción, el procesamiento alternativo del ARNm (splicing en inglés), estabilidad del ARNm y eventos postraduccionales (Strachan T. et al. 1999). En el caso del control del inicio de la traducción del ARNm, éste se encuentra determinado principalmente por las propiedades estructurales en la región 5' UTR. Dentro de este contexto, se realizó la búsqueda de elementos de regulación en el extremo 5' (resultado 3) y se encontró a la caja CAAT que representa un promotor proximal. Lo relevante de esta caja es que funge como un modulador de la transcripción basal de otros promotores ubicados cerca del sitio donde inicia la transcripción, y además opera como una secuencia intensificadora (enhancer sequence en inglés) de acuerdo al mecanismo mostrado en la figura 16. (Strachan T. et al. 1999)

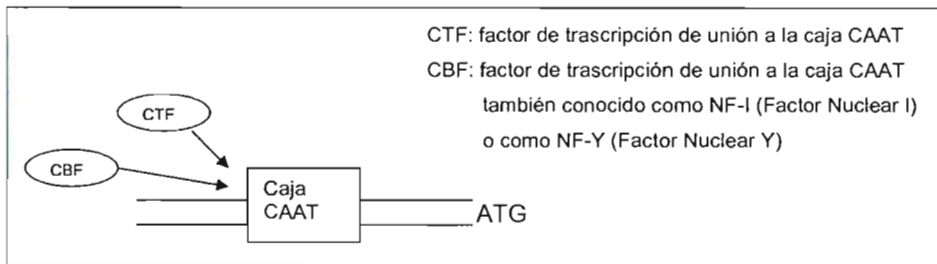


Figura 16: Actividad moduladora de la caja CAAT

A esta secuencia se asocian proteínas (CTF y CBF) que asisten en la unión de la ARN polimerasa II a su promotor lo cual da como resultado la optimización de las velocidades de transcripción.

La velocidad de síntesis del ARNm no es el único factor determinante de los niveles de un ARNm dentro de la célula sino también la vida media de éste último. Así, la cola de poli(A) presente en la región 3'UTR protege la cadena de ARNm de la degradación efectuada por las exonucleasas, se encarga de facilitar el transporte de las moléculas de ARNm hacia el citoplasma, de estabilizarlas en éste, así como de facilitar la traducción ya que permite la unión de elementos de reconocimiento en el ARNm que son rastreados por la maquinaria ribosomal. Otro elemento importante asociado a la cola de poli(A) es la secuencia AAUAAA encargada de establecer el sitio de corte en el

extremo 3' para la gran mayoría de los transcritos de la ADN polimerasa II. El corte ocurre en un lugar específico localizado entre 15 y 30 nucleótidos corriente abajo del elemento AAUAAA. Una vez que la ruptura ha ocurrido corriente abajo de este elemento se adicionan los residuos de adenilato con ayuda de la enzima poli(A) polimerasa para formar la cola de poli(A). Los dispositivos analizados anteriormente corresponden a elementos reguladores cis-acting (por su significado en inglés, que actúan en el mismo sitio donde se encuentran) ya que su función esta limitada a ser ejercida sobre la cadena doble de ADN en la cual residen.

Todo lo mencionado reafirma que el ARNm aislado es funcional, por lo que se puede establecer que la proteína que se deriva del marco de lectura abierto no es hipotética, sino que es sintetizada en el roedor dado que la existencia de un ARNm es sinónimo de que la proteína se expresa.

A este nivel experimental, mi interés se centró entonces en esta proteína que de acuerdo a los resultados de homología a nivel de nucleótidos y de aminoácidos (resultados 4, 5 y 6), definitivamente se trata de un ARNm completo que no había sido aislado en el roedor y cuya proteína no ha sido caracterizada. Información adicional que se puede obtener de los resultados anteriores es que la Antolefinina no solo esta presente en el cerebro del roedor sino también en otros órganos: páncreas, hígado, ojo, aorta y tiene homología con proteínas de otros tejidos: ventrículo, timo, traquea, epitelio traqueobranquial y piel.

A fin de lograr proponer una función para la Antolefinina las deducciones obtenidas en el resultado 5 se analizan dando origen a la siguiente serie de propuestas.

- **Propuesta 1:**

La propuesta 1 corresponde al caso más simple, es decir, cuando toda la proteína es funcional sin sufrir ningún procesamiento enzimático ligado a PC's y solo son funcionales la señal bipartita para transporte a núcleo (NLS), la señal para transporte fuera del núcleo (NES) y la región rica en prolinas, ya que como se pudo observar en las búsquedas de homología (figura 14), la región rica en prolinas tiende a ser conservada, por lo que hay una fuerte probabilidad de que ésta sea funcional. Ejemplos de actividad relacionada con una región rica en prolinas es el de la proteína

Hox-4.2 en la cual dicha región (ubicada en el extremo amino terminal) es un dominio de activación transcripcional fuerte (Rambaldi I. et al. 1994), o el del factor hepático nuclear 4 (HNF4) donde la región rica en prolinas localizada hacia el extremo carboxilo terminal se encarga de reprimir la actividad del dominio 2 de activación funcional (AF2) (Iyemere V.P. et al. 1998), y finalmente el ejemplo de las enzimas Kinasa MEK-1 y MEK-2 donde la región rica en prolinas (ubicada en medio de su secuencia) regula la función de éstas y es además necesaria para la asociación de importantes factores de activación (Catling A.D et al. 1995). Ver figura 17.

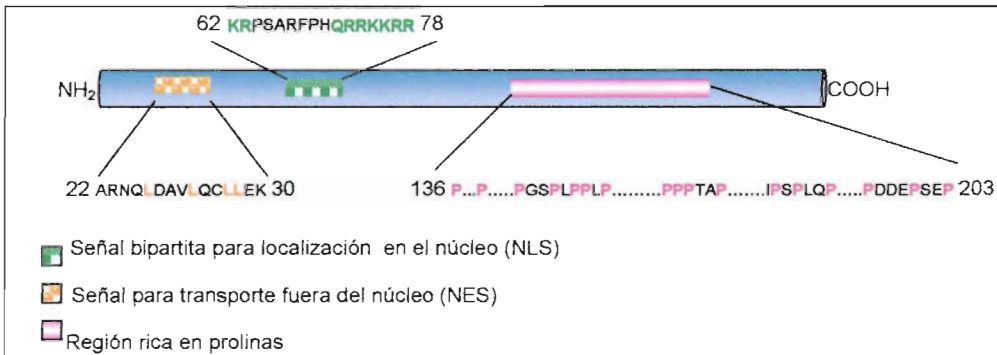


Figura 17. Propuesta 1: Se trata de una proteína funcional de 246 aminoácidos capaz de translocarse al interior y exportarse al exterior del núcleo de células del sistema nervioso del ratón.

Retomando lo de la secuencia de señalización para localización en núcleo, se sabe que la exportación de una proteína hacia el núcleo es extremadamente selectiva, por lo que el procesamiento de proteínas nucleares debe incluir la presencia de señales estructurales que especifiquen la acumulación selectiva en el núcleo (Dingwall C et al 1986, García-Bustos et al 1991). Como ya se mencionó estudios en algunas proteínas nucleares tales como el antígeno T del virus SV40, han indicado que la parte de la secuencia requerida para la translocación a núcleo esta relacionada con la presencia de residuos de aminoácidos básicos (Gómez-Márquez et al 1988), pero el parámetro más confiable para establecer que se trata de una proteína que será exportada a núcleo es la señal denominada bipartita. Esta última esta estructurada por un primer segmento que contiene dos aminoácidos básicos juntos (Arginina o Lisina) y una segunda región

a una distancia de diez aminoácidos aproximadamente conformada por lo menos de tres aminoácidos básicos consecutivos. Adicional a lo mencionado, se ha encontrado que para algunas señales para transporte a núcleo además de estar constituidas por repeticiones de aminoácidos básicos se tienen regiones ricas en prolina, las cuales se encargan de conformar estructuras extendidas que son importantes para la conformación de señales de localización para núcleo (NLS) (Yuko et al. 1998). En el caso de la Antolefinina estas dos condiciones (señal bipartita y la región rica en prolinas) están presentes, lo cual sugiere entonces que si se podría tratar de una proteína que se va a núcleo.

Por otra parte, sabemos que una de las formas en que la célula incrementa la diversidad de sus productos reside en modificaciones postraduccionales (como es la fosforilación y la amidación) y en particular, en su habilidad para generar múltiples péptidos y proteínas vía la proteólisis sitio-específica de una estructura precursora única, la Antolefinina cuenta con tres sitios potenciales para este último tipo de procesamiento proteolítico.

- **Propuesta 2:**

En la propuesta 2 se considera que se trata de una proproteína de la cual el primer péptido derivado está α -amidado y se genera al eliminarse la proregión de la proteína. Se predijo la α -amidación en el residuo 61 porque el precursor de hormonas y otros péptidos activos que se encuentran amidados en el C-terminal siempre está seguido por un residuo de glicina el cual provee el grupo amida y de manera frecuente por al menos dos residuos básicos consecutivos (Arg o Lys) que funcionan generalmente como precursores de sitios activos de corte. A pesar de que todos los aminoácidos se pueden encontrar amidados, los residuos hidrofóbicos neutros tales como Val o Phe son buenos sustratos, mientras que residuos cargados como Asp o Arg son menos reactivos (Kreil G. 1984, Smyth D.G, et al 1987). Esta posibilidad se considero también de alta relevancia para la Antolefinina porque sabemos que la amidación se encuentra relacionada con la resistencia a la carboxipeptidasa lo cual se traduce en aumento de la estabilidad y vida media biológica del péptido amidado una vez que es escindido de su molécula precursora, lo cual da como resultado un aumento

en la afinidad del péptido amidado por su receptor, un efecto biológico más potente y expresión correcta de la actividad biológica (Cuttita F.1993).

Por todo lo mencionado anteriormente parece estimable la posibilidad de que se genere un péptido amidado ya que es insignia de actividad biológica.

Asimismo en esta propuesta se sigue considerando la posibilidad de un péptido que es exportado a núcleo debido a la presencia de una señal monopartita y un tercer péptido (C) que podría tener una función neural desconocida.

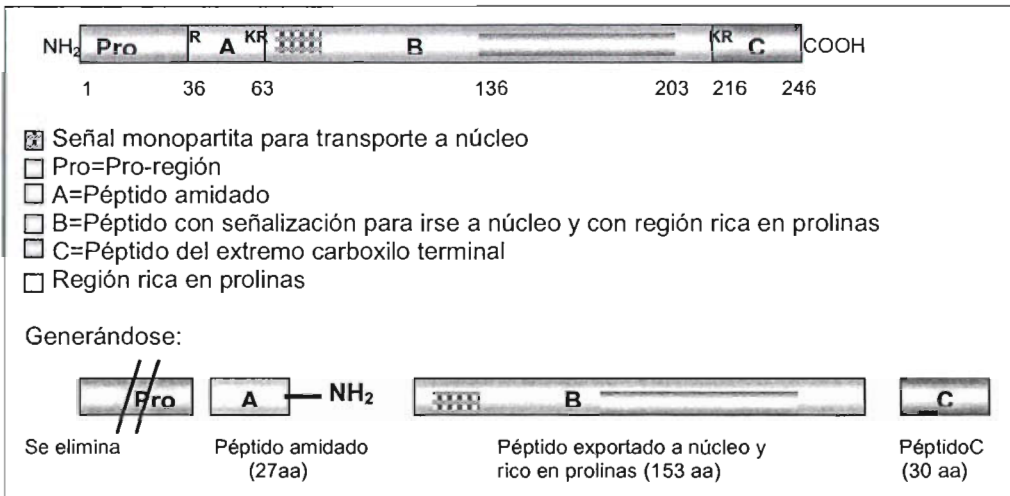


Figura 18. Propuesta 2: Generación de tres péptidos bioactivos (A, B y C) de acuerdo a la presencia de sitios consenso para procesamiento proteolítico.

• **Propuesta 3:**

La propuesta 3 solo difiere con respecto a la anterior en que el péptido central es degradado, lo cual sucede en otras proproteínas, generándose de esta forma solo dos péptidos bioactivos: uno amidado de 27 aminoácidos y otro de 30 aminoácidos.

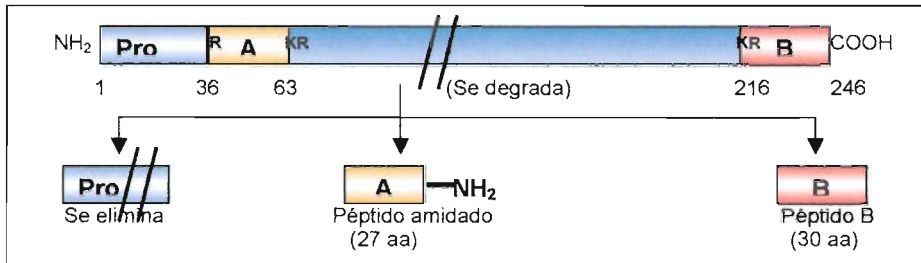


Figura 19. Propuesta 3: Se trata de una pre-proteína pero se obtienen solo dos péptidos bioactivos (A y B) de acuerdo a la presencia de sitios consenso para procesamiento proteolítico.

- **Propuesta 4:**

Finalmente en la propuesta 4 se establece que solo se crean dos péptidos bioactivos: el que se va a núcleo con la señal bipartita y el de 30 aminoácidos.

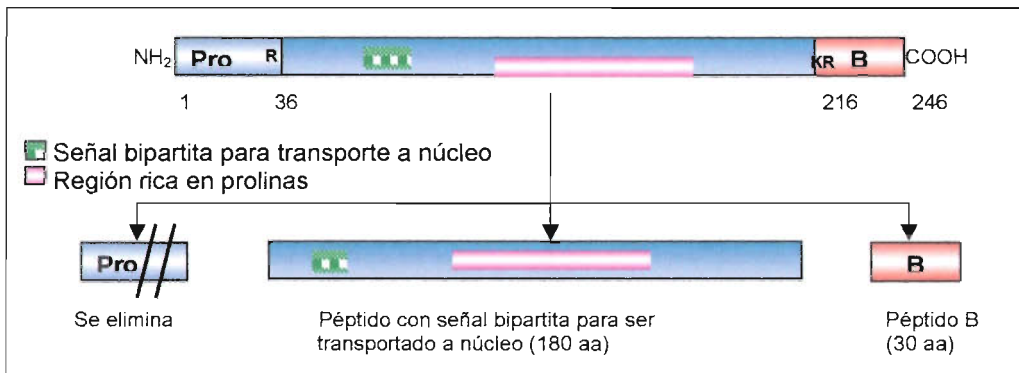


Figura 20. Propuesta 4: Se generan dos péptidos bioactivos

Para las últimas tres propuestas se podría establecer que los niveles y la duración de las bioactividades específicas de los péptidos que se generen a partir del precursor inactivo estarían limitadas a la proteólisis de este último. Asimismo, si se trata de una preproteína, los péptidos derivados de su procesamiento pueden tener su órgano blanco lejos de donde fueron sintetizadas, esto último podría explicar porque se encontraron secuencias homólogas en una gran variedad de tejidos.

En último lugar, otra forma de conocer información sobre una proteína desconocida es mediante el análisis de su estructura secundaria para localizar dominios

que tienden a ser conservados entre familias de proteínas. Así, en la figura 15 se muestran los resultados de los análisis de predicción de estructura secundaria efectuados con diferentes programas, los cuales dan resultados consistentes entre ellos. De acuerdo a esto, se puede considerar que la Antolefinina sería una proteína no globular donde solo los primeros treinta aminoácidos tienen sin lugar a duda una conformación de hélice, el resto de la estructura parece no estar bien definida (presencia de asas) o tener una conformación secundaria no común (regiones NORS), lo cual no nos permite establecer la presencia de dominios conservados.

Finalmente, con todo lo desarrollado se pudo comprender que el abordaje de una proteína nueva se puede hacer desde diferentes puntos de partida. Sin duda, el más simple es partir de la comparación contra lo que ya se encuentra reportado para ver si la proteína pertenece a una superfamilia ya caracterizada, lo cual se puede hacer con la base de datos del GenBank (comparación de una secuencia de nucleótidos) o la del Instituto Europeo de Bioinformática (EBI) (comparación de una secuencia de aminoácidos). Una segunda aproximación es sin duda la búsqueda de patrones importantes en la estructura primaria de la proteína, ya que son éstos los que nos permiten intuir sobre funcionalidad, que fue lo que sucedió con las propuestas suministradas.

En conclusión, el ADNc identificado corresponde a un gen que da origen a una proteína en espera de ser caracterizada experimentalmente para identificar y confirmar su(s) función(es), al menos en el sistema nervioso del roedor. De acuerdo a las propuestas presentadas, la 1 es la que contiene un mayor número de elementos que sugieren que la funcionalidad de la Antolefinina será de una proteína que tiene movilidad en el núcleo al ser importada y exportada de éste.

IX. Aportaciones y perspectivas de trabajo.

Con este trabajo experimental de tesis se logró clonar un ARNm que da origen a una proteína no caracterizada en el mamífero. De acuerdo a un análisis estructural cuidadoso se establecieron cuatro propuestas, cada una de las cuales abre la posibilidad a una serie de experimentos necesarios para poder establecer el papel funcional de esta proteína dentro de la fisiología de distintos órganos.

Adicionalmente, se generó una estrategia metodológica para obtener información acerca de una proteína totalmente nueva, la cual sin lugar a duda puede aplicarse en futuros experimentos donde el objetivo central sea caracterizar nuevas proteínas. En particular, nuestro grupo de trabajo desarrolla estudios para la caracterización de nuevos sistemas peptídicos, por lo que lo mencionado con anterioridad es una herramienta de apoyo básica para esto.

ESTA TESIS NO SE
DE LA BIBLIOTECA

VI. BIBLIOGRAFÍA

ABI PRISM DNA Sequencing (1995) The Perkin-Elmer Corporation.

ALBERTS B., BRAY D., LEWIS J., RAFF M., ROBERTS K.; WATSON J. 2002 3th edition. Molecular Biology of the Cell. U.S.A: Garland Publishing.

AN Z., GENE W., ZHOU X., STEINER D. 1999. Proteolytic processing in the secretory pathway. *J. Biol. Chem.* **274**: 20745-20748.

AZUMA Y., OHURA K., WANG PL., SHINOHARA M. 2001. Endomorphins 1 and 2 modulate chemotaxis, phagocytosis and superoxide anion production by microglia. *J. Neuroimmunol.* **3;119(1)**:51-56.

BAIROCH A., BUCHER P., HOFMANN K. 1997. *Nucleic Acids Res.* **25**:217-221. PROSITE <http://www.expasy.ch/prosite> Version: 99.07.

BREMNER K., SEYMOUR L., LOGAN A., READ M. 2004. Factors influencing the ability of nuclear localization sequence peptides to enhance nonviral gene delivery. *Bioconjugate Chem.* **15**: 152-161.

BURNS D., CROWL R. 1987. An immunological assay for chloramphenicol acetyltransferase. *Anal. Biochem.* **1;162(2)**: 399-404.

CAMPOS-NETO A., SUFFIA I., CAVASSANI A., JEN S., GREESON K0., OVENDALE P., SILVA J., REED S., SKEIKY Y. 2003. Cloning and characterization of a gene encoding an immunoglobulin-binding receptor of the cell surface of some members of the family *Trypanosomatidae*. *Infect. Immun.* **71(9)**: 5065-5076.

CATLING A., SCHAEFFER J., REUTER C., REDDY G., WEBWR M. 1995. A proline-rich sequence unique to MEK1 and MEK2 is required for raf binding and regulates Mek function. *Mol. Cell. Biol.* **15(10)**: 5214-5225.

CUTTITA F. 1993. Peptide amidation: signature of bioactivity. *Anat. Rec.* **236**: 87-93.

CZAPLA M., GOZAL D., ALEA O., BECKERMAN R., ZADINA J. 2000. Differential cardiorespiratory effects of endomorphin 1, endomorphin 2, DAMGO, and morphine. *Am. J. Respir. Crit. Care Med.* **162(3 Pt 1)**:994-999.

CHAMPION H., BIVALACQUA T., ZADINA J., KASTIN A., KADOWITS P. 1999. Vasodilator responses to the Endomorphin Peptides, but not nociceptin/OFQ, are mediated by nitric oxide release. *Ann. N. Y. Acad Sci.* **897**:165-172.

CHAMPION H., ZADINA JE., KASTIN A., KADOWITZ P. 1998. Endomorphin 1 and 2 have vasodepressor activity in the anesthetized mouse. *Peptides.* **19(5)**: 925-929.

DINGWALL C. 1986. Protein import into cell nucleus. *Annu. Rev. Cell Biol.* **2**:367-390.

DINGWALL C., LASKEY R. 1991. Nuclear targeting sequences -- a concensus? *Trends Biochem. Sci.* **16**:478-481.

DOSTMANN W., TAYLOR M., NICKL C., BRAYDEN J., FRANK R., TEGGE W. 2000. Highly specific, membrane-permeant peptide blockers of GMP-dependent protein kinase Ia inhibit NO-induced cerebral dilation. *PNAS* **97(26)**: 14772-14777.

EBERWINE J., BARCHAS J., HEWLETT W., EVANS C. 1987. Isolation of enzyme cDNA clones by enzyme immunodetection assay: Isolation of a peptide acetyltransferase. *Proc.Natl.Acad.Sci. USA.* **84**: 1449-1453.

EIPPER B., GREEN C., CAMPBELL T., STOFFERS D., KEUTMAN H., MANIS R., OUAFIK L. 1992. *Protein Sci.* **2**: 489-497.

EMBL-EBI (European Bioinformatics Institute). <http://www.ebi.ac.uk/>.

GALLOUZI I., STEITZ J. 2001. Delineation of mRNA export pathways by the use of cell-permeable peptides. *Science.* **294**: 1895-1901.

GASCA S., CAÑIZARES J., PASCAL DE S., MÉJEAU C., POULAT F., BERTA P., BOIZET-BONHOURE B. 2002. A nuclear export signal within the high mobility group domain regulates the nucleocytoplasmic translocation of SOX9 during sexual determination. *PNAS.* **99(17)**: 11199-11204.

GARCÍA-BUSTOS. 1991. *Biochim. Biophys. Acta.* **1071**:83-101.

GILL S., VON HIPPEL P. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**:319-326.

GIRJES A., ELLIS W., LAVIN M., CARRICK F. 1993. Immuno-dot blot as a rapid diagnostic method for detection of chlamydial infection in koalas (*Phasolarctos cinereus*). *Vet. Rec.* **133(6)**: 136-141.

GLICK.B., PASTERNAK J. 1998 2nd edition. Molecular Biotechnology. Principles and Applications of Recombinant DNA. EUA: ASM Press.

GÓMEZ-MÁRQUEZ J., SEGADE F. 1988. Prothymosin alpha is a nuclear protein. *FEBS Lett.* **226**:217-219.

GRIFFITHS A., GELBART W.; MILLER J.; LEWONTIN R. 1999. Modern Genetic Analysis. U.S.A: W. H. Freeman & Co.

HARVEY L., BERK A., ZIPURSKY S., MATSUDAIRA P., BALTIMORE D., DARNELL J. 2000 4th edition. Molecular Cell Biology. U.S.A: W. H. Freeman & Co.

HIPERTEXTOS DEL ÁREA DE LA BIOLOGÍA. Universidad Nacional del Nordeste. 1998-2003. <http://fai.une.edu.ar/biología>.

HOZUMI Y., ITO T., NAKAMO T., NAKAGAWA T., AOYAGI M., KOMDO H., GOTO K. 2003. Nuclear localization of diacylglycerol kinase ζ in neurons. *Eur. J. Neurosci.* **18**: 1448-1457.

HUMAN GENOME PROJECT INFORMATION.

http://www.ornl.gov/sci/techresources/Human_Genome/faq/faqs1.shtml

IGARASHI N., OKADA T., HAYASHI S., FUJITA T., JEHANGEE S., NAKAMURA S. 2003. Sphingosine kinase 2 is a nuclear protein and inhibits DNA synthesis. *J. Biol. Chem.* **278**(47): 46832-46839.

IYEMERE V., DAVIS N., BRWNLEE G. 1998. The activation function 2 domain of hepatic nuclear factor 4 is regulated by a short C-terminal proline rich repressor domain. *Nucleic Acid Res.* **26**(9): 2098-2104.

JIMENEZ L., BECERRA A., LANDA A. 2004. Cloning, expression and partial characterization of a gene encoding the S15a ribosomal protein of *Taenium solium*. *Parasitol. Res.*

KISHIMOTO A., NISHIYAMA K., NAKANISHI H., URATSUJI Y., NOMURA H., TAKEYAMA Y., NISHIZUKA Y. 1985. Studies on the phosphorylation of myelin basic protein kinase C and adenosine 3', 5' monophosphate-dependent protein kinase. *J. Biol. Chem.* **260**:12492-12499.

KLUG W., CUMMINGS M. 1999, 5ª edición. Conceptos de Genética. España: Prentice Hall.

KOONIN E., GALPERIN M. 2003. SEQUENSE-EVOLUTION-FUNCTION. Computational Approaches in Comparative Genomics. USA: Kluwer Academic Publishers.

KORNER J., CHUN J., O'BRYAN L., AXEL R. 1991. Prohormone processing in *Xenopus* oocytes: characterization of cleavage signals and cleavage enzymes. *Proc. Natl. Acad. Sci. USA* **88**: 11393-11397.

KREIL G. 1984. Occurrence, detection, and biosynthesis of carboxy-terminal amides. *Meth. Enzymol.* **106**:218-223(1984).

LIZOTTE-WANIEWSKI M., TAWE W., GUILIANO D., LU W., LIU J., WILLIAMS S., LUSTIGMAN S. 2000. Identification of potential vaccine and drug target candidates by expressed sequence tag analysis and immunoscreening of *Onchocerca volvulus* larval cDNA libraries. *Infect. Immun.* **68**(6): 3491-3501

MACARA I. 2001. Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.* **65** (4): 570-594.

MEHTA A., BOT G., REISINE T., CHESSELET M. 2001. Endomorphin-1: induction of motor behavior and lack of receptor desensitization. *J. Neurosci.* **15;21(12)**:4436-4442.

Molecular Cloning a Laboratory Manual (2000 3a edición).

MURAI N., MURAKAMI Y., MATSUFUJI S. 2003. Identification of Nuclear Export Signal in Antizyme-1. *J. Biol. Chem.* **278(45)**: 44791-44798.

PINNA L.A. 1990. Casein Kinase 2: an eminence rise in cellular regulation? *Biochim. Biophys. Acta* **1054**:267-284.

RAMBALDI I., KOVÁCS E., FEATHERSTONE M. 1994. A proline-rich transcriptional activation domain in murine HOXd-4 (HOX-4.2). *Nucleic Acid Res.* **22**: 376-382.

ROST B., LIU J. 2003. The PredictProtein Server. *Nucleic Acids Res.* **31(13)**: 3300-3304. <http://cubic.bioc.columbia.edu/> Version: 1.99.08.

ROST B. 1996. PHD: predicting one-dimensional proteins structure by profile-based neural networks. *Methods Enzymol.* **266**:525-539. HTML. <http://cubic.bioc.columbia.edu/>.

ROST B. 1998. Unpublished. GLOBE <http://cubic.bioc.columbia.edu/> Version: 1.98.05.

SAKURADA S., HAYASHI T., YUHKI M., ORITO T., ZADINA J., KASTIN A., FUJIMURA T., MURAYAMA K., SAKURADA C., SAKURADA T., NARITA M., SUZUKI T., TAN-NO K., TSENG L. 2001. Differential antinociceptive effects induced by intrathecally administered endomorphin-1 and endomorphin-2 in the mouse. *Eur. J. Pharmacol.* **21;427(3)**:203-210

SAKURADA S., ZADINA J., KASTIN A., KATSUYAMA S., FUJIMURA T., MURAYAMA K., YUKI M., UEDA H., SAKURADA T. 1999. Differential involvement of mu-opioid receptor subtypes in endomorphin-1- and -2-induced antinociception. *Eur. J. Pharmacol.* **7; 372(1)**:25-30.

SCHWARSE S., HO A., VOCERO-AKBANI A., DOWDY S., 1999. In vivo protein transduction: delivery of biologically active protein into the mouse. *Science.* **285**: 1569-1573.

SEIDAH N., DAY R., MARCINKIEWIEZ M. 1998. Precursor convertases: an evolutionary ancient, cell-specific, combinatorial mechanism yielding diverse bioactive peptides and proteins. *Ann. N. Y. Acad. Sci.* **15;839**: 9-24.

SMYTH D., BRADBURY A. 1987. Biosynthesis of the C-terminal amide in peptide hormones. *Biosci. Rep.* **7**:907-916.

SOIGNIER R., VACCARINO A., BRENNAN A., KASTIN A., ZADINA J. 2000. Analgesic effects of endomorphin-1 and endomorphin-2 in the formalin test in mice. *Life Sci.* **14;67(8)**:907-912.

SONNHAMMER E., GUNNAR H., ANDERS K. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology.* 175-182.

STEINER D., SMEEKEENS S., OHAGI S., CHAN S. 1992. The new enzymology of precursor processing endoproteases. *J. Biol. Chem.* **267**: 23435-23438.

STRACHAN T., READ A. 1999, 2nd edition. Human Molecular Genetics 2. England: BIOS Scientific Publishers Ltd.

Uni-ZAP[®]XR Premade Library. Instruction Manual 2003 Stratagene.

WILSON A., SOIGNIER R., ZADINA J., KASTIN A., NORES W., OLSON R., OLSON G. 2000. Dissociation of analgesic and rewarding effects of endomorphin-1 in rats. *Peptides.* **21(12)**: 1871-1874.

YUKO S., KOBAYASHI T., KODA T., IKUTA K., KAKINUMA M., KISHI M., 1998. Two-proline rich nuclear localization signals in the amino- and carboxyl-terminal regions of the borna disease virus phosphoprotein. *J. Virol.* **72(12)**: 9755-9762.

ZADINA J., KASTIN A., HACKLER L. 1997. A potent and selective endogenous agonist for the mu-opiate receptor. *Nature.* **386**: 499-592.

ZHAO K., CUI L. 2003. Molecular characterization of the major virion protein gene from the *Trichoplusia ni* ascovirus. *Virus Genes.* **27(1)**: 93-102.

ZASTROW M., TRITTON T., CASTLE J. 1986. Exocrine secretion granules contain peptoid amidation activity. *Proc. Natl. Acad. Sci. USA* **83**: 3297-3301.

ANEXO

<p>Caldo NZY</p> <p>5g NaCl</p> <p>2g de MgSO₄.7H₂O</p> <p>5g de extracto de levadura</p> <p>10g de amina NZ</p> <p>Adicionar agua desionizada hasta obtener un volumen final de 1L. Ajustar pH a 7.5 con NaOH. Esterilizar con autoclave</p>
<p>Agar NZY</p> <p>5g NaCl</p> <p>2g de MgSO₄.7H₂O</p> <p>5g de extracto de levadura</p> <p>10g de amina NZ</p> <p>15g de agar</p> <p>Adicionar agua desionizada hasta obtener un volumen final de 1L. Ajustar pH a 7.5 con NaOH. Esterilizar con autoclave. Colocar en cajas petri (~25 mL/100-mm placa)</p>
<p>Agar suave NZY</p> <p>Preparar 1L de caldo NZY, adicionar 0.7% (m/v) de agarosa. Esterilizar por autoclave.</p>
<p>Agar LB (para 1L)</p> <p>10 g NaCl</p> <p>10 g triptofano</p> <p>5 g extracto de levadura</p> <p>20 g de agar</p> <p>Ajustar a pH 7.0 con NaOH 5N, se adiciona agua desionizada para llevar a un volumen final de 1L. Esterilizar con autoclave. Colocar en cajas petri (~25 mL/ 100mm placa)</p>
<p>Solución IPTG</p> <p>Disolver 2g de IPTG en 8 mL de agua destilada. Ajustar el volumen de la solución a 10 mL con agua destilada y esterilizar por filtración a través de un filtro con un poro de 0.22 micras. Almacenar la solución en alícuotas de 1 mL a -20°C.</p>
<p>Solución amortiguadora PBS 1X</p> <p>Disolver 8g de NaCl, 0.2g KCl, 1.44g Na₂HPO₄, y 0.24g de KH₂PO₄ en 800 mL de agua</p>

destilada. Ajustar el pH a 7.4 con HCl. Adicionar agua destilada para completar 1 litro. Repartir la solución en alícuotas apropiadas y esterilizar por autoclave durante 20 minutos en ciclo para líquidos. Almacenar a temperatura ambiente

Solución Amortiguadora SB (Sample Buffer) 5X:

0.125 % Azul de Bromofenol
25 % Glicerol

Solución Amortiguadora TAE (Tris-acetate/EDTA)

Solución 50X:
242g Tris
57.1 mL de ácido acético glacia
100 mL 0.5M EDTA. (pH 8.0)

Solución de EDTA (0.5M pH 8.0)

Adicionar 186.1g de EDTA·2H₂O a 800 mL de H₂O. Agitar vigorosamente con ayuda de un agitador magnético. Ajustar el pH a 8.0 con ayuda de NaOH. Repartir la solución en alícuotas apropiadas y esterilizar por autoclave.