

03088 4
21

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

INSTITUTO DE BIOTECNOLOGIA, DEPARTAMENTO DE BIOESTRUCTURA Y RECONOCIMIENTO MOLECULAR, UNAM, CUERNAVACA, MORELOS



EXPRESION Y CARACTERIZACION DE LOS PRODUCTOS DEL GENE CNGT II DEL ALACRAN CENTRUROIDES NOXIUS HOFFMANN Y DE VARIANTES GENERADAS POR MUTAGENESIS SITIO ESPECIFICA

T E S I S

QUE PARA OBTENER EL GRADO DE
D O C T O R
P R E S E N T A
LA M. EN C. MA. CONSUELO GARCIA RODRIGUEZ

Biotecnología

CUERNAVACA, MORELOS 1997

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México



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.

INSTITUTO DE BIOTECNOLOGIA, DEPARTAMENTO DE BIOESTRUCTURA Y
RECONOCIMIENTO MOLECULAR. UNAM. CUERNAVACA, MORELOS.

TESIS:

**Expresión y caracterización de los productos del gene
Cngt II del alacrán *Centruroides noxius* Hoffmann y de
variantes generadas por mutagénesis sitio específica.**

COMITE TUTORIAL:

DR. BALTAZAR BECERRIL L.
DR. LOURIVAL D. POSSANI P.
DR. ARTURO LIEVANO M.
DR. XAVIER SOBERON M.

ALUMNO:

M. en C. MA. CONSUELO GARCIA RODRIGUEZ

CUERNAVACA, MORELOS, AGOSTO DE 1997.

Esta tesis fue realizada en el laboratorio del Dr. Lourival Domingos Possani Postay, bajo la tutoría principal del Dr. Baltazar Becerril Luján y contó con la ayuda financiera de los siguientes organismos:

-Howard Hughes Medical Institute, donativo N-s75191-527104 y 75197-527107.

-CONACyT, 4734-N9406.

-DGAPA-UNAM, IN211996.

-Beca Crédito para estudios de Doctorado. CONACyT No. de Registro 85802.

INDICE

CONTENIDO	PAGINA
PRESENTACION	1
RESUMEN	2
SUMMARY	2
ABREVIATURA	3
I INTRODUCCION	4
II ANTECEDENTES	13
III HIPOTESIS	20
IV OBJETIVOS	20
V METODOLOGIA	21
VI DATOS NO PUBLICADOS	32
VII RESULTADOS PUBLICADOS	47
-García <i>et al.</i> 1997. <i>Comp. Biochem. Physiol.</i> 116b (3): 315-322	
VIII DISCUSION Y CONCLUSIONES	48
IX PERSPECTIVAS	63
X ANEXO	64
-Becerril <i>et al.</i> 1995. <i>J. Toxicol.-Toxin Reviews</i> , 14 (3): 339-357	
-Selisko <i>et al.</i> 1996. <i>Eur. J. Biochem.</i> 242 : 235-242	
XI BIBLIOGRAFIA	65

PRESENTACION

Esta tesis está enfocada al estudio de la relación estructura-función de las toxinas del veneno del alacrán *Centruroides noxius* Hoffmann, en particular a las toxinas que reconocen canales de sodio dependientes de voltaje.

La tesis se encuentra dividida en seis partes: la primera es una breve introducción a las toxinas y su interacción con canales iónicos; en la segunda parte se hace una descripción de los antecedentes que sirvieron de base para formular la hipótesis de trabajo; en la tercera parte se presenta un trabajo publicado con los resultados de esta tesis; en la cuarta parte se describen las metodologías y datos que no han sido publicados; en la quinta parte se hace una discusión de los resultados y por último un anexo que comprende dos trabajos en los cuales participo como coautora.

RESUMEN

En este trabajo se reporta la expresión de la toxina Cn 5 del alacrán *Centruroides noxius* Hoffmann la cual reconoce canales de sodio. Inicialmente se hace una caracterización de la toxina nativa lo cual nos permite catalogarla como una toxina contra artrópodos que muestra una mayor toxicidad hacia crustáceos. Se hacen comparaciones de secuencia primaria entre varias toxinas de alacrán, tanto de toxinas contra artrópodos como de toxinas contra mamíferos, con el fin de buscar los sitios o regiones de la toxina que pudieran estar participando en la especificidad y/o toxicidad de la molécula. La expresión se lleva a cabo probando diferentes estrategias de expresión en *Escherichia coli*. También se reporta el plegamiento *in vitro* de la toxina recombinante. Se discute el problema del plegamiento de proteínas en relación a este trabajo. Así mismo se inicia la construcción de toxinas mutantes que nos permitan confirmar los resultados de las comparaciones.

SUMMARY

We are presenting a work on expression of toxin Cn 5 from the scorpion *Centruroides noxius* Hoffmann. This protein affects sodium channels. First we characterize the native toxin which shows that it presents toxic activity against crustaceans. In order to identify the motifs that participate in the specificity or toxicity we compared the primary sequence of toxins against mammals and toxins against arthropods. Three different expression strategies in *Escherichia coli* and the *in vitro* refolding of the recombinant toxin are reported. We discuss the protein folding problem for the interest of this work. Finally, the initiation of the construction of toxin mutants is reported, this will enable us to test the results of the sequence comparison.

ABREVIATURAS

β-ME-	beta-mercaptoetanol
CNBr-	bromuro de cianógeno
DTT-	ditiotreitól
EGTA-	bis (β-aminoéster) etilenglicol
ELISA-	ensayo inmunoabsorbente unido a enzima (siglas en inglés)
Gnd HCl-	cloruro de guanidina
GSH-	glutación reducido
GSSG-	glutación oxidado
HPLC-	cromatografía líquida de alta resolución (siglas en inglés)
IPTG-	isopropiltiogalactopiranosido
mbp-	proteína acarreadora de maltosa (siglas en inglés)
NMR-	resonancia magnética nuclear (siglas en inglés)
PBS-	solución salina amortiguadora (siglas en inglés)
PCR-	reacción en cadena de la polimerasa (siglas en inglés)
PAGE-	electroforesis en gel de poliacrilamida- sodio dodecil sulfato (siglas en inglés)
SDS-	sodio dodecil sulfato

AaH	= <i>Androctonus australis</i>
Be	= <i>Buthus eupeus</i>
BmK	= <i>Buthus martensii</i> Karsch
Bot	= <i>Buthus occitanus tunetanus</i>
Cll	= <i>Centruroides limpidus limpidus</i>
CnH	= <i>Centruroides noxius</i> Hoffmann
CsE	= <i>Centruroides sculpturatus</i> Ewing
Css	= <i>Centruroides suffusus suffusus</i>
Lqh	= <i>Leiurus quinquestriatus hebraeus</i>
Lqq	= <i>Leiurus quinquestriatus quinquestriatus</i>
Pi	= <i>Pandinus imperator</i>
Ts	= <i>Tityus serrulatus</i>

INTRODUCCION

EL VENENO DE ALACRAN

El veneno de alacrán es, en su estado natural, un líquido blanquecino que presenta una gran estabilidad el cual está constituido por proteínas, sales, mucoproteínas, aminoácidos libres, lípidos, nucleótidos y pequeños péptidos. El veneno liofilizado y almacenado a 4°C conserva sus propiedades tóxicas por tiempo indefinido. Este material liofilizado se reconstituye en agua y mediante centrifugación se puede separar la fracción tóxica que es soluble, de los restos celulares insolubles (Stahnke 1978). En general el veneno carece de actividad enzimática y cuando la presenta, ésta, es muy pobre. La única enzima que se ha reportado es la hialuronidasa (Zlotkin 1978, Possani 1983). Los venenos de alacranes contienen una familia de proteínas homólogas, las cuales son responsables de la actividad neurotóxica que produce la picadura del alacrán (Babin 1974, Possani 1984). Esta actividad tóxica varía de acuerdo a la especie de alacrán, desde malestares locales en el área de la picadura hasta envenenamiento severo con manifestaciones sistémicas que pueden ocasionar la muerte (Keegan 1980).

TOXINAS DEL VENENO

Las neurotoxinas del veneno de alacrán presentan una alta afinidad hacia canales iónicos (Rochat 1979, Carbone 1982, Granier 1989, Blaustein 1991) en membranas de células excitables. Cada especie de alacrán contiene una serie de toxinas propias, entre las cuales existen diferencias en cuanto a secuencia primaria y un amplio rango en su grado de toxicidad y especificidad hacia los diferentes grupos animales. Las neurotoxinas son pequeños péptidos de una sola cadena, generalmente básicos y altamente estables debido a la presencia de varios puentes disulfuro. Dependiendo de la longitud de la cadena, que puede ser de 30 a 70 aminoácidos, las toxinas se dividen en toxinas de cadena corta y toxinas de cadena larga.

TOXINAS DE CADENA CORTA

El primer grupo llamado toxinas de cadena corta constan de 31 a 39 aminoácidos y presentan tres o cuatro puentes disulfuro (Possani, 1984). La mayoría de las toxinas de este grupo son estabilizadas mediante tres

puentes disulfuro (Carbone 1982, Granier 1989), aunque recientemente se han reportado toxinas de cadena corta con 4 puentes disulfuro (Olamendi-Portugal 1996). Estas toxinas bloquean la conductancia iónica de los canales de potasio, sin modificar los mecanismos de compuerta. Un ejemplo de toxinas bloqueadoras de canales de potasio es la Noxiustoxina (Ntx) del alacrán mexicano *Centruroides noxius* Hoffmann (Carbone 1982, Possani 1982).

Existe un segundo grupo de toxinas de cadena corta las cuales presentan actividad contra insectos; son moléculas estabilizadas por cuatro puentes disulfuro y presentan reconocimiento hacia diferentes receptores. Una de estas toxinas, llamada Clorotoxina (Cltx), es un bloqueador específico del canal de cloro de baja conductancia (Lippens 1995). Sin embargo, esta actividad sólo se presenta cuando la toxina es aplicada en el lado intracelular del canal, por lo que es poco probable que dicho canal sea el blanco natural de este tipo de toxinas.

Otro ejemplo de toxinas de este grupo es la toxina I_5A del alacrán *Buthus eupeus*, la cual reconoce al receptor de glutamato (Arseniev, 1984).

TOXINAS DE CADENA LARGA

Las toxinas de cadena larga poseen de 60 a 70 aminoácidos y están estabilizadas por cuatro puentes disulfuro. Este grupo de toxinas presenta reconocimiento hacia canales de sodio. Se han descrito varias toxinas que presentan especificidad hacia mamíferos, insectos o crustáceos. Sin embargo, generalmente la selectividad hacia una especie no es absoluta. Se han reportado varias toxinas que muestran actividad hacia varios grupos filogenéticos.

A.- TOXINAS CONTRA MAMIFEROS

Las toxinas que reconocen al canal de sodio en mamíferos se subdividen, de acuerdo a su mecanismo de acción en dos tipos α y β (Couraud 1982, Meves 1984, Meves 1987, Caterall 1991). Las toxinas tipo α se unen al sitio 3 del canal de sodio sensible a voltaje y retardan la inactivación, lo cual produce una prolongación del potencial de acción. Además, la unión es dependiente del potencial de membrana por lo que su afinidad decrece con la depolarización. La toxina mejor caracterizada del grupo es la toxina II del alacrán Norte Africano *Androctonus australis* Hector (AaH II). Las toxinas tipo β se unen, de forma independiente del voltaje, al sitio 4 del canal e inducen la aparición de una corriente anormal de sodio, debida a un

cambio transitorio en la dependencia de la activación al voltaje. Un ejemplo de toxinas tipo β es la toxina II del alacrán Mexicano *Centruroides suffusus suffusus* (Css II). Ambos tipos de toxinas, α y β , modifican la dependencia de la inactivación al voltaje, haciéndola incompleta (Meves 1987).

En general se considera que el mecanismo de acción de las toxinas tipo α es característico del veneno de los alacranes del viejo mundo mientras que el de las toxinas tipo β es característico de alacranes del nuevo mundo. Sin embargo, para alacranes de los géneros *Centruroides* y *Tityus* (alacranes americanos) se han reportado toxinas con mecanismo de acción tipo α (Meves 1984, Kirsch 1989). Aunque, se debe mencionar que estas toxinas tipo α americanas no desplazan ni a las toxinas tipo α ni a las tipo β , en ensayos de unión a sinaptosomas (Jover 1980, Wheeler 1983) y que sólo se han descrito dos sitios de unión al canal de sodio para las toxinas de alacrán (Tabla I).

B.- TOXINAS CONTRA INSECTOS

La clasificación de toxinas que reconocen canales de sodio de insectos, es más compleja. Actualmente se pueden definir tres grupos los cuales no abarcan a todas las toxinas contra insectos reportadas hasta el momento. Los tres grupos comprenden toxinas altamente tóxicas que en general pertenecen a alacranes del viejo mundo (Zlotkin 1985). El primer grupo está conformado por toxinas α contra insectos, esto es porque al igual que las toxinas tipo α de mamífero, prolongan el potencial de acción evocado mediante una inhibición de la inactivación del canal. Este grupo ha sido descrito con la toxina α contra insectos del alacrán *Leiurus quinquestratus hebraeus* (Lqh α IT, Eitan 1990). El segundo grupo corresponde a las toxinas depresoras, las cuales producen en el insecto una parálisis flácida progresiva mediante un bloqueo del potencial de acción debido a la depolarización de la membrana y supresión de la corriente de sodio. Un ejemplo de una toxina depresante es la toxina Lqh IT2, del alacrán israelita *Leiurus quinquestratus hebraeus*. El tercer grupo corresponde a las toxinas excitatorias las cuales producen una rápida parálisis espasmódica y se unen al canal de forma independiente del voltaje. La toxina AaHIT, de *A. australis* Hector es un claro ejemplo de dicho grupo. El mecanismo de interacción de las toxinas excitatorias es

altamente similar al de las toxinas tipo β de mamíferos. También existe cierta similitud entre el mecanismo de acción de las toxinas depresoras y las excitatorias pues en ambas se ha reportado (Zlotkin 1991) una fase de contracción excitatoria que en las depresoras es transitoria mientras que en las excitatorias se presenta de forma sostenida (Zlotkin, 1991). Por último, se puede mencionar un cuarto grupo de toxinas contra insectos, o artrópodos en general, que no pertenecen a ninguno de los tres grupos descritos y que son, principalmente, toxinas de alacranes americanos que presentan baja toxicidad y baja especificidad hacia los modelos biológicos en los que han sido ensayadas. En este grupo se encuentran toxinas como las variantes 1-6 del alacrán Norte Americano *C. sculpturatus* Ewing (CsEv1-CsEv6) y las toxinas de alacranes Mexicanos: Cn1, Cn5 y Cn10 de *C. noxius* y la toxina Cll1 de *C. limpidae limpidae* (Meves 1984, Vázquez 1995, Lebreton 1994, Selisko 1996, García 1997).

Otro aspecto que hace más compleja la clasificación de las toxinas de alacrán es la existencia de toxinas que presentan características típicas de dos o más tipos de toxinas. Por ejemplo, algunas toxinas tipo β del alacrán brasileño *Tityus serrulatus* presentan alta toxicidad, tanto contra mamíferos como contra insectos (Martin 1985, De Lima 1986). Una de estas toxinas Ts γ , ha sido probada en ensayos de unión a sinaptosomas y se encontró que es capaz de desplazar tanto a toxinas contra insectos, como toxinas contra mamíferos (De Lima 1986). Otro caso similar es el de la toxina 4 contra insectos del alacrán *Androctonus australis* Hector (AaH IT4). Esta toxina depresante es capaz de unirse a ambos sitios en el canal de sodio en sinaptosomas de mamíferos (Loret 1991). En las toxinas del alacrán norte americano *Centruroides sculpturatus* Ewing, se presentan toxinas caracterizadas estructuralmente como tipo β , y sin embargo, en el nódulo de Ranvier su efecto electrofisiológico es tipo α : CsEV, CsEv3, etc. (Meves, 1984).

CANALES DE SODIO

Los canales iónicos dependientes del voltaje son los responsables de generar el potencial de acción en células excitables y participan en procesos reguladores en células no excitables (Caterall 1991). La característica más importante de estos canales es la capacidad de controlar su conductancia iónica, en una escala temporal de milisegundos, mediante cambios en el potencial de membrana. Así mismo, presentan un

mecanismo de compuerta dependiente del voltaje para el que se han descrito dos actividades: la activación, que controla la velocidad de apertura y la inactivación, que controla la velocidad de cerrado del canal. La conductancia de los canales activados por voltaje es extremadamente rápida y altamente selectiva. Entre los canales activados por voltaje se encuentran los de sodio, para los que se han descrito 5 sitios receptores de neurotoxinas, cada uno con diferentes efectos en la función del canal (Tabla 1, tomada de Caterall, 1991).

Tabla 1. Drogas y toxinas que se unen a los sitios receptores 1-5 de neurotoxinas y algunos sitios adicionales no identificados en el canal de Na⁺.

1	Tetrodotoxina Saxitoxina μ -Conotoxinas	Inhibición de la conductancia iónica.
2	Veratridina Aconitina Grayanotoxina Batracotoxina	Activación persistente.
3	Toxinas α de alacrán Toxinas de anémone marina	Inhibición de la inactivación; aumenta la activación persistente.
4	Toxinas β de alacrán	Modifica la dependencia de voltaje de la activación.
5	Brevetoxinas Ciguatoxinas	Disparos repetitivos; modifican la dependencia del voltaje de la activación.
•	Insecticidas	Disparos repetitivos; activación persistente.
	Toxina de <i>Gonolopora</i> Toxina de <i>Conus striatus</i>	Inhibe la inactivación.
	Anestésicos locales Antiarrítmicos Anticonvulsivos	Inhiben la frecuencia y la dependencia del voltaje.

• Sitios no identificados.

ESTRUCTURA DE LAS TOXINAS

La homología entre todas las toxinas de alacrán se manifiesta claramente cuando se hace un análisis de las secuencias primarias. Como se muestra en la Tabla 2, al hacer alineamientos de secuencias primarias en base a la posición de las cisteínas, es posible observar posiciones o segmentos conservados así como la conservación del patrón de puentes disulfuro

(Meves 1984, Possani 1985, Bontems 1991, Becerril 1995).

En cuanto a su estructura tridimensional, se ha obtenido suficiente información con base en estudios de difracción circular, resonancia magnética nuclear y cristalografía de rayos X, lo cual permite establecer la conservación de un marco estructural entre todas las toxinas del veneno de alacrán (Fontecilla-Camps 1980, Arseniev 1984, Pashkov 1988, Bontems 1991, Lebreton 1994, Dauplais 1995, Jablonsky 1995, Lippens 1995, Li 1996, Delepierre 1997). Los elementos principales de la estructura secundaria son: una hélice α de dos y media vueltas y una lamina β plegada antiparalela formada por dos o tres segmentos de la cadena (Fig. 1). La hélice α se encuentra conectada, mediante dos puentes disulfuro, a la lámina β formando el dominio hélice α estabilizada por cisteínas (CSH). Este dominio se ha propuesto como un motivo estructural común de péptidos neurotóxicos que reconocen canales iónicos (Kobayashi, 1991). Cerca de esta región se localiza un tercer puente disulfuro que une dos asas. Estos tres puentes disulfuro se encuentran conservados en todas las toxinas de alacrán. El cuarto puente disulfuro puede modificar su posición, como en el caso de las toxinas excitatorias o no estar presente como en algunas toxinas de cadena corta (Tabla 2).

Estos elementos conforman un núcleo compacto y son superponibles entre todas las toxinas de las cuales se ha determinado la estructura, tanto entre toxinas de cadena corta (Arseniev 1984, Bontems 1991, Duplais 1995, Lippens 1995), como toxinas de cadena larga, sin importar el grado de toxicidad (Fontecilla-Camps 1980, Jablonsky 1995, Hong-Min 1996) o la especificidad (Pashkov 1988, Loret 1991, Lebreton 1994, Delepierre 1997). Varias asas salen de este núcleo, dos de ellas las asas B y J. Estas asas representan elementos estructurales que permiten diferenciar entre toxinas tipo α y β . El asa B es de mayor longitud en las toxinas tipo α mientras que el asa J es mayor en toxinas tipo β (Meves 1984).

Las regiones más flexibles de la estructura se encuentran en los extremos N-terminal y C-terminal de la molécula. Es de esperarse que las diferencias funcionales deben residir en los segmentos variables ya sea en las asas y/o en las regiones terminales, tanto en el tipo de aminoácidos como en la longitud de la cadena.

IMPORTANCIA DEL ESTUDIO DE LAS TOXINAS

El estímulo inicial para el estudio de su veneno fue el aspecto médico de

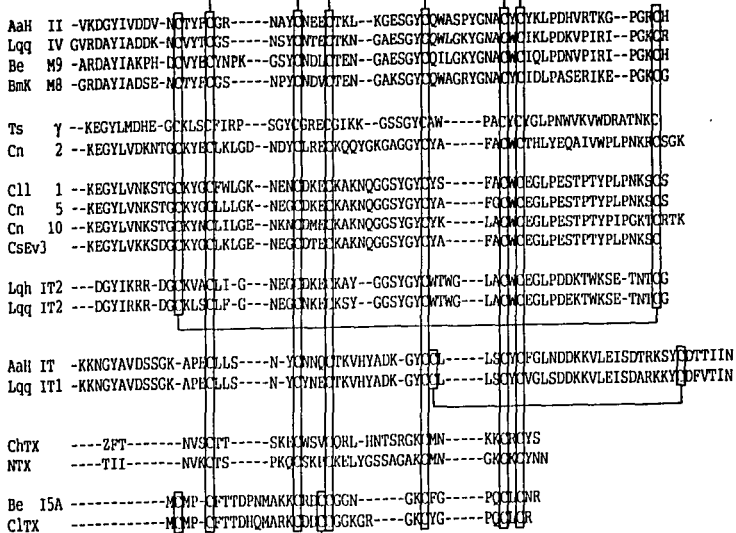


Tabla 2 Alineamiento de secuencias primarias representativas de los diferentes grupos de toxinas de alacrán. Las secuencias han sido alineadas en base a las cisteínas con objeto de mostrar la conservación de tres de los puentes disulfuro. Este arreglo de puentes disulfuro se mantiene en todas las toxinas de alacrán. El cuarto puente puede o no estar presente y tener una posición diferente en algunas toxinas, como en las depresantes contra insectos.

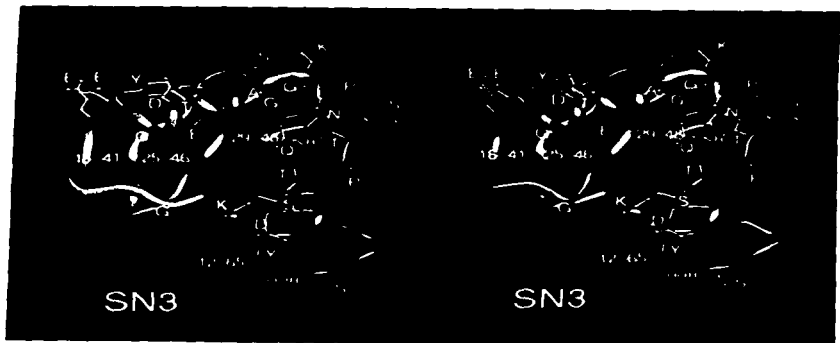


Fig. 1 A: Representación esquemática de la estructura tridimensional de la toxina CsEv3. Muestra los principales motivos estructurales que se conservan en todas las toxinas de alacrán: tres hebras, que comprenden de los aminoácidos: 1 a 4, 37 a 41 y 46 a 50, formando una lámina β -plegada y una hélice α de aproximadamente 2.5 vueltas (aminoácidos del 20 al 32). Dos puentes disulfuro, entre los residuos 25 y 46 y los residuos 29 y 48, que unen ambos elementos formando el motivo CSH (hélice α estabilizada por cisteínas) y un tercer puente cercano a este núcleo, entre los residuos 16 y 41. Uniendo los extremos amino y carboxilo se encuentra el cuarto puente formado entre el aminoácido 12 y el 65.

la picadura de alacrán. Por lo tanto las toxinas que afectan mamíferos son las mejor caracterizadas. Actualmente el alacranismo (envenenamiento producido por la picadura del alacrán) aún representa un serio problema de salud en varias regiones del mundo. Afecta grandes núcleos de población, tanto en el medio rural como en el urbano, y da lugar a altas tasas anuales de morbilidad. Se estima que en México por lo menos 200,000 personas son picadas anualmente por estos arácnidos, ocasionando entre 700 a 800 muertes (Possani 1996).

Aunque el interés médico por el estudio del veneno de alacrán sigue siendo válido han surgido nuevos motivos para esta investigación. Al ser péptidos neuroactivos son frecuentemente usados en el estudio de procesos fisiológicos donde están involucradas membranas excitables (Kirsch 1989, Caterall 1991). En el estudio de la estructura y función de los canales iónicos así como en su aislamiento, las toxinas de alacrán han jugado un papel muy importante (De Lima 1986, Blaustein 1991, Chul-Seung 1992). El número y diversidad de canales iónicos expresados en las neuronas es muy grande y para evaluar esta diversidad molecular son necesarias sondas específicas que permitan diferenciar entre los subtipos. Debido a la actividad altamente específica, de las neurotoxinas, sobre sus blancos moleculares representan herramientas efectivas para caracterizar y diferenciar el creciente número de canales y receptores del sistema nervioso. Asimismo, en el campo de los insecticidas algunas de las toxinas de alacranes pueden ser de gran utilidad debido a su actividad insecticida y a su alta especificidad. Con la participación conjunta de la biología molecular y la toxicología podemos esperar que se desarrollen nuevas moléculas las cuales podrán ser diseñadas específicamente para reconocer con alta afinidad a un receptor determinado (Possani 1990, Adams 1994).

ALACRANES MEXICANOS DE IMPORTANCIA MEDICA

Es importante mencionar que de las aproximadamente 800 especies de alacranes descritas, menos de 50 representan algún peligro para el hombre (Keegan, 1980). Estas especies se encuentran agrupadas en la familia Buthidae a la cual pertenece el género *Centruroides* y cuyo centro de radiación es México. En nuestro país se han reportado 28 especies de este género, 8 de las cuales se consideran peligrosas al humano: *C. elegans*, *C. infamatus infamatus*, *C. limpidus tecomanus*, *C. limpidus limpidus*, *C. sculpturatus* Ewing, *C. noxius* Hoffmann., *C. suffusus suffusus* y *C. pallidiceps*.

ANTECEDENTES

TOXINA CsEv3

En 1974 en un trabajo publicado por Babin *et al.* se describieron las secuencias de aminoácidos de tres toxinas del alacrán *C. sculpturatus* Ewing (Babin, 1974). Denominadas variantes 1, 2 y 3, los tres péptidos fueron fácilmente purificados debido a su gran abundancia en el veneno. Las tres secuencias resultaron altamente homólogas y en los ensayos de toxicidad mostraron una actividad muy baja contra mamíferos. Estas fueron secuenciadas sólo como apoyo para determinar las secuencias de otras toxinas poco abundantes, pero de alta toxicidad. Las clasificaron como toxinas contra insectos pues en los bioensayos fueron capaces de producir una parálisis total en grillos, con duración de dos a tres horas. Se determinaron las dosis efectivas medias (ED_{50}), las cuales fueron de 11, 21.5 y 255 mg/kg de peso para las variantes 1, 2 y 3, respectivamente. En 1980 se logró determinar la estructura tridimensional de la variante 3 (CsEv3) utilizando cristalografía de rayos X, con una resolución de 3.0 Å. Esta fue la primera toxina de alacrán cuya estructura tridimensional se determinó (Fig.1) y la primera vez que se propuso un modelo general para la estructura de las toxinas de alacrán (Fontecilla-Camps 1980). En 1983 el mismo grupo publicó un refinamiento de la estructura obtenida con una resolución de 1.8 Å (Almassay, 1983). En este trabajo comprobaron que la estructura propuesta es correcta y reportaron la configuración en *cis* de la prolina 59.

GENE *Cngt II*

Como trabajo de maestría realicé una búsqueda en un banco de ADNc de *C. noxius* H., cuyo objetivo era buscar el gene que codificara para una toxina semejante a la CsEv3 (Becerril 1993). Consideramos importante para el estudio de la relación estructura-función trabajar con una toxina cuya estructura tridimensional fuera conocida. Se utilizó como sonda una clona truncada que codifica para un péptido altamente similar al extremo carboxilo terminal de la CsEv3. Se aislaron varias clonas dos de las cuales corresponden a genes que codifican para dos proteínas diferentes muy parecidas a las variantes de *C. sculpturatus* E. A estos genes se les llamó *Cngt II* y *Cngt III*. Ambas clonas están completas e incluyen un péptido

señal, región codificadora para la toxina, codón de terminación, una región 3' no codificadora donde se encuentra la señal de poliadenilación y la secuencia poli A (Fig. 2).

FLEXIBILIDAD DEL EXTREMO CARBOXILO DE ALGUNAS TOXINAS DE ALACRAN

En un estudio de dicroísmo circular presentado por Loret *et al.*, (Loret, 1990), se propone que la ausencia de prolina en el extremo carboxilo de una toxina es determinante para la flexibilidad de esa región. En este trabajo se relaciona la flexibilidad molecular con la especificidad de las toxinas, proponiendo una explicación a la capacidad que tiene la toxina Ts γ , del alacrán brasileño *Tityus serrulatus*, para afectar tanto insectos como mamíferos. Los autores se refieren a la toxina γ como toxina VII sin embargo, dicha toxina fue reportada originalmente como toxina γ (Possani, 1977) por lo tanto seguiré esta nomenclatura durante todo el texto. Concluyen que esta flexibilidad le permite al extremo carboxilo adoptar las diferentes conformaciones necesarias para reconocer los dos tipos de canales. Esta interpretación se basa en el análisis de los espectros de dicroísmo circular de tres toxinas, las cuales representan tres actividades farmacológicas diferentes. La toxina C5s II que afecta mamíferos, AaH IT2 que afecta insectos y Ts γ que afecta ambos. Los espectros de cada una de las toxinas en agua fueron comparados con aquellos obtenidos después de hacer un cambio gradual del solvente, a TFA (2,2,2-trifluoroetano) que promueve la formación de hélices α . Hacen una comparación y la única molécula capaz de modificar su conformación es la toxina Ts γ .

Posteriormente, el mismo grupo realizó otro estudio de dicroísmo circular con una toxina contra canales de sodio de insectos, AaH IT4 (Loret, 1991). Esta molécula presenta características de diversos grupos de toxinas de alacrán, a pesar de tener una baja homología de secuencia primaria con el resto de ellas. Presenta toxicidad tanto hacia insectos como a mamíferos; en estudios de pegado al canal de sodio de mamíferos es capaz de desplazar toxinas tipo α y tipo β . En experimentos inmunológicos encuentran que la toxina AaH IT4, que es una toxina de un alacrán del norte de África, presenta reactividad cruzada sólo con C5s II que es una toxina tipo β de un alacrán Norte Americano. También encuentran que esta toxina no contiene prolina, que es un aminoácido común en las toxinas de alacrán. Al igual que en el caso de Ts γ , el espectro de dicroísmo circular muestra

```

-19          -10
M  N  S  L  L  M  I  T  A  C  L  F  L  I  G  T
ATG AAC TCG TTG TTG ATG ATC ACT GCT TGT TTG TTC CTG ATC GGA ACA
-----péptido señal-----

      +1          10
V  W  A  K  E  G  Y  L  V  N  K  S  T  G  C  K
GTG TGG GCA AAA GAA GGT TAT CTG GTA AAC AAG AGC ACG GGC TGC AAA
-----

          20
Y  G  C  L  L  L  G  K  N  E  G  C  D  K  E  C
TAC GGT TGC CTG CTA TTG GGG AAA AAC GAA GGC TGC GAT AAG GAA TGC

      30          40
K  A  K  N  Q  G  G  S  Y  G  Y  C  Y  A  F  G
AAA GCG AAG AAC CAA GGA GGT AGT TAC GGC TAT TGC TAC GCT TTT GGG

          50          60
C  W  C  E  G  L  F  E  S  T  P  T  Y  P  L  P
TGC TGG TGC GAA GGT TTG CCC GAA AGT ACA CCG ACT TAT CCC CTT CCT

          68  Fin
N  K  S  C  S  K  K  *
AAT AAA TCT TGC AGC AAA AAA TAA

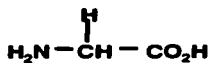
```

Fig 2 Secuencia nucleotídica correspondiente a la clona *Cngt II* obtenida a partir de ADNc. Esta secuencia codifica para la toxina Cn 5 del alacrán *Centruroides noxius* Hoffmann.

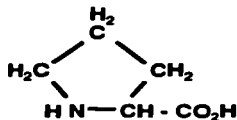
que la región carboxilo presenta una flexibilidad que le daría la capacidad de modificar su conformación. Por lo tanto, en este trabajo también relacionan la ausencia de prolina con una mayor flexibilidad estructural de la región carboxilo de la molécula y con la capacidad de toxinas de afectar tanto canales de sodio de mamíferos como de insectos.

EFEECTO DE LAS PROLINAS EN LA ESTRUCTURA DE PROTEINAS

La relación entre flexibilidad y la ausencia de prolina se puede explicar al observar la estructura de la prolina. Presenta un grupo amino secundario (realmente es un iminoácido), el cual se mantiene en una conformación estructural rígida debido a que el último átomo de la cadena lateral se une al primer átomo de la cadena principal formando un anillo. Esto reduce considerablemente la flexibilidad estructural cuando en la cadena se encuentra una prolina. En el caso de la glicina sucede lo contrario, dado que presenta una cadena lateral con la estructura más simple: un átomo de hidrógeno. Por lo tanto, el impedimento estérico es mínimo y esto le proporciona una mayor flexibilidad a la estructura donde se encuentra la glicina (Branden 1991).



Glicina



Prolina

Debido a la conformación especial de la prolina presenta un efecto sobre el enlace peptídico que precede a una prolina. Existen 2 conformaciones posibles del enlace peptídico, una en la que los carbonos α (C^α) mantienen una posición *trans* y otra en que su posición es *cis*. La forma *trans* se favorece energéticamente debido a que existe una menor repulsión entre los átomos que no participan en el enlace, incluyendo la cadena lateral de cada aminoácido. En general la forma *trans* se presenta en las proteínas en relación 1000:1 sobre la *cis*. Esta regla se rompe cuando el residuo siguiente ($i+1$) es una prolina, en este caso la relación *trans/cis* es de solo 4:1. Esto es posible gracias a que la configuración cíclica de la cadena

lateral de la prolina reduce la repulsión entre los átomos y la estabilidad intrínseca del isómero *cis* es comparable a la del isómero *trans* (Creighton, 1993)

ESTUDIO DE LA RELACION ESTRUCTURA-FUNCION DE LAS TOXINAS

Para poder analizar la relación estructura-función de las toxinas es posible hacer uso de la tecnología de ADN recombinante. Mediante mutaciones dirigidas se puede estudiar la participación de regiones o sitios específicos en la función. Adicionalmente esta tecnología puede aplicarse en la optimización de toxinas ya conocidas o para introducir nuevas funciones a estas moléculas.

En este tipo de análisis debemos considerar los siguientes aspectos (Adams 1994):

a) Que ya se conocen múltiples isoformas naturales que nos permiten hacer comparaciones de las secuencias primarias y estructuras tridimensionales, para tratar de determinar cuáles son las posiciones o regiones asociadas con cierta función.

b) Que existen posiciones altamente conservadas en todas las toxinas de alacrán, las cuales pueden tener la función de mantener el marco estructural, como son las cisteínas y la glicina 39 (o posición correspondiente). Estas posiciones deben respetarse en los ensayos de mutagénesis.

c) Que la afinidad hacia un receptor no está siempre en relación directa con el efecto funcional de la toxina. En la mayoría de los estudios con neurotoxinas se asume que cuando una toxina se une a un receptor de igual forma lo afecta funcionalmente y esto no siempre es cierto. Por ejemplo, en ensayos de unión a membranas de cerebro de rata *in vitro*, la toxina AaH IT4 presenta una afinidad similar con AaH IT2, mientras que en ensayos de toxicidad *in vivo* fue 50 veces menos tóxica (Loret 1991).

EXPRESION DE TOXINAS RECOMBINANTES

Hasta el momento existen varios estudios de expresión de toxinas decadena larga: la expresión de estas toxinas se ha intentado en varios sistemas, tanto procariones como eucariotes, desde *E. coli* hasta células en cultivo, por ejemplo células COS-7. También se ha ensayado en baculovirus que se considera un buen modelo gracias a su potencial en el control de insectos y su probada eficiencia como vector de expresión heteróloga. El otro modelo de expresión que ha sido ensayado es el de

levadura. A continuación describo varios trabajos de expresión de toxinas de cadena larga, en los diferentes sistemas.

a) En 1989, Bougis *et al.*, ensayaron la expresión de la toxina contra mamíferos, AaH II. Transfectaron células COS-7 con un plásmido que lleva el ADNc de esta toxina. Lograron una expresión transitoria que les permitió confirmar la expresión de una toxina activa. Sin embargo, mencionan que la cantidad de toxina recombinante fue tan pequeña que no les permitió la caracterización de la toxina recombinante. Enfatizaron en sus conclusiones que es necesario analizar otros sistemas de expresión que les permita obtener mayor cantidad de toxina.

b) Dee *et al.*, 1990, presentaron un trabajo de expresión de la toxina AaIT en células de fibroblastos murinos (NIH/3T3). Prepararon un gene sintético de la toxina precedido por el péptido señal de la interleucina-2 de humano, con el fin de dirigir la secreción de la toxina al medio de cultivo. Lograron probar la actividad tóxica de la proteína recombinante, pero la cantidad de toxina producida no fue suficiente para realizar otros ensayos.

c) En 1990 se publicó un trabajo de Stewart *et al.* en el cual mencionan que construyeron un baculovirus recombinante derivado del virus de la poliédrosis nuclear de *Autographa californica* y la toxina AaH II. Su objetivo fue reducir el tiempo que tarda el virus en matar al insecto hospedero (el virus silvestre requiere de 4 a 5 días para matar al hospedero). Primero ensayaron la expresión de su construcción en células de *Spodoptera frugiperda* y encontraron que sí expresaban una toxina activa. En el paso siguiente alimentaron larvas de *Trichoplusia ni* con poliédra y observaron que las patologías asociadas con la infección del virus recombinante y el silvestre eran distintas. En el caso del recombinante se presentan los síntomas típicos de envenenamiento por la toxina. Lograron reducir el tiempo de vida de las larvas, después de ingerir el virus recombinante, a 50 % del correspondiente al del virus silvestre. Esto es importante pues se traduce en menor daño a los cultivos que se busca proteger. La producción de toxina se estimó en aproximadamente 400 ng/ml.

d) Resultados similares fueron obtenidos por Maeda *et al.*, 1991, empleando otro baculovirus y la misma toxina. Sintetizaron el gene para AaIT y lo insertaron en un vector de transferencia para el virus de la poliédrosis nuclear del gusano de seda *Bombyx mori*. El gene quedó bajo el promotor de la poliédrina y con el péptido señal de la bombyxina. Inyectaron larvas de *B. mori* con el virus para determinar el efecto

insecticida. Observaron una reducción de 40% en el tiempo necesario para matar al hospedero. Se observó un cambio en la actividad motriz de los organismos, aproximadamente 40 hr después de la inyección, que les impide comer. Estos dos efectos se suman para que haya un daño menor a los cultivos. La producción de toxina recombinante fue de 5 µg/ml de hemolinfa.

e) Martin-Eauclaire *et al.*, (1994) realizaron un trabajo donde expresaron la toxina AaH IT1 en la levadura *Saccharomyces cerevisiae*. En un plásmido de alto número de copias (pMA 91) que contiene un promotor fuerte (PGK1), insertaron la secuencia que codifica para la toxina, generando dos construcciones: una con el péptido señal propio de la toxina y la otra con un péptido señal de la levadura. Aunque obtuvieron una toxina recombinante activa, la producción en el caso de la construcción con el péptido señal de la toxina fue muy baja (0.7 a 4 µg /L). Para la construcción, donde usan el péptido señal de la levadura, la producción fue aún más baja. Los autores analizaron el problema de la expresión de toxinas de cadena larga y comentan que el principal problema es la baja expresión, pues aún en el caso del baculovirus que tiene mayor producción, la purificación de la proteína a partir de la hemolinfa limita mucho la cantidad final que se obtiene. Ellos creen que los bajos niveles de producción que se obtienen pueden ser debido principalmente a tres causas:

- 1) Una tasa baja de síntesis de la proteína. Una causa podría ser el diferente uso de codones.
- 2) Baja estabilidad de la proteína. Principalmente debido a digestión proteolítica por ser una proteína extraña al organismo.
- 3) Toxicidad de la proteína a las células del cultivo. Esto podría suceder si la toxina reconoce canales por el lado intracelular, como en el caso de la cloro toxina (Lippens 1995).

HIPOTESIS

Loret *et al.* (1990; 1991) sugieren que la ausencia de prolina del carboxilo terminal les proporciona la flexibilidad que les permite a las toxinas Ts γ y AaH IT4 competir por el sitio de unión al canal tanto de las toxinas α y β de mamíferos como por los sitios de las toxinas que reconocen al canal de sodio en insectos. Por lo tanto nuestra hipótesis de trabajo es: la prolina por ser un iminoácido introduce una constricción que impide la estructuración del extremo carboxilo. Al sustituirla por glicina o por otros aminoácidos con mayor grado de movilidad que las prolina, se favorecería la flexibilidad necesaria para unirse a ambos tipos de canales de sodio (mamífero e insecto).

OBJETIVOS

OBJETIVO GENERAL

El objetivo general de este trabajo es participar en el estudio de la relación estructura-función de las toxinas de alacrán, utilizando como modelo la toxina contra crustáceos Cn 5.

OBJETIVOS ESPECIFICOS

- Purificación de la toxina nativa
- Caracterización de la toxina Cn 5:
 - a) determinación de la secuencia primaria
 - b) determinación de las posiciones de los puentes disulfuro
 - c) determinación de la toxicidad en acociles (LD₅₀)
 - d) determinación del sitio de unión a sinaptosomas
- Clonación y expresión del gene *Cngt II*
 - a) expresión como péptido de fusión
 - b) expresión directa utilizando el péptido señal propio

- c) expresión directa utilizando un péptido señal bacteriano
- Purificación de los productos de expresión
- Plegamiento *in vitro* de los productos de expresión
- Caracterización de los productos de expresión. Además de los puntos propuestos para la caracterización de la toxina nativa, se probará el reconocimiento por los anticuerpos policlonales obtenidos contra la toxina nativa.
- Análisis y comparaciones de secuencias primarias para, determinar las mutaciones necesarias que nos permitan analizar la flexibilidad del extremo carboxilo de la proteína Cn 5
- Clonación y expresión de genes mutantes

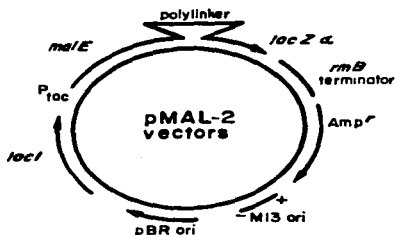
METODOLOGIA

En esta sección se presentan únicamente las metodologías que no han sido descritas en las publicaciones que acompañan este trabajo (ver trabajos anexos).

EDICION Y AMPLIFICACION DEL GENE

La edición del gene se hizo mediante PCR (Vent Polimerasa) y dependiendo del plásmido donde se clonó se diseñaron diferentes oligos.

a) pMal-C2 (BioLabs New England): plásmido para expresión citoplásmica, como péptido de fusión unido a la proteína mbp (proteína acarreadora de maltosa). Se utilizaron los siguientes oligos: 2-Mal-N, el cual elimina el péptido señal e introduce una metionina que permite liberar la toxina mediante corte con bromuro de cianógeno (CNBr), incluye los codones para los 7 primeros aminoácidos del extremo amino. 1-Mal-C, contiene los codones para los últimos 7 aminoácidos, con este oligo se eliminan las dos lisinas finales, y un codón de terminación. Es un oligo reverso por lo que



pMAL-c2,-p2 polylinker

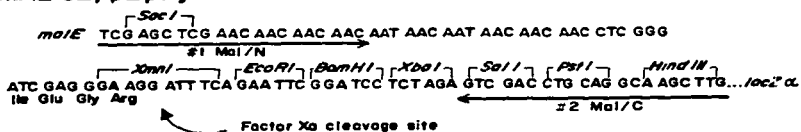


Fig 3. pMAL-C2: vector de expresión citoplásmica como péptido de fusión unido a la proteína acarreadora de maltosa (mbp). La construcción permite insertar la secuencia de interés unida al gene *malE* que codifica para *mbp*, el cual se encuentra bajo un promotor fuerte (*tac*). Presenta un sitio múltiple de clonación. La selección de colonias se hace por color ya que cuenta con *lacZ* α . El péptido de fusión queda unido mediante un espaciador poli asparagina y además, presenta un sitio de corte para el factor Xa, que permite la liberación de la proteína de interés.

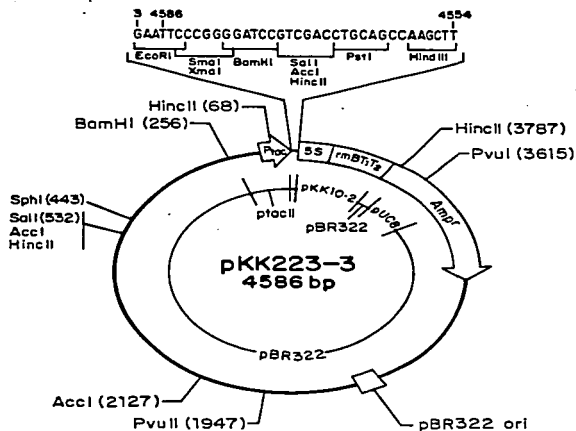
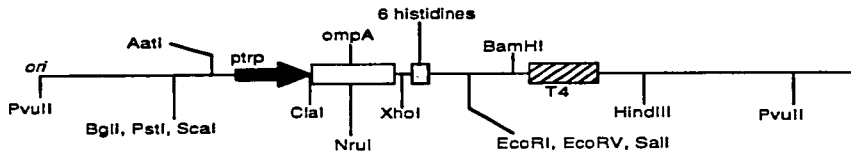


Fig 4. pKK223-3: vector de expresión directa por lo que las proteínas pueden ser dirigidas fuera del citoplasma mediante una señal adecuada. Contiene al promotor tac el cual es inducido mediante adición de IPTG al medio de cultivo. La construcción tiene un sitio múltiple de clonación, así como un terminador ribosomal (*rrnB*) que impide una lectura corrida. Presenta un sitio de unión al ribosoma que puede ser utilizado si el codón de iniciación del inserto se encuentra entre cinco y nueve bases del sitio de reconocimiento de la enzima *Eco* RI.



pPACIB.1 (ca. 2.8 kb)

Secuencia señal de *omp A*, secuencia del oligopéptido y sitios de clonación.

Met Lis Lis Tre Ala Ile Ala Ile Ala Val Ala Leu Ala Gli Fen
 ATG AAA AAG ACA GCT ATC GCG ATT GCA GTG GCA CTG GCT GGT TTC

Ala Tre Val Ala Gln Ala Ser Ser His His His His His His Gli
 Gct ACC GTA GCG CAG GCTTCGACC CAC CAC CAC CAC CAC CAC GGG
 xho I

Gli
GGGAATTC GATATC GTCGAC GCATCC
 EcoRI EcoRV Sall BamHI

Fig 5. pPACIB.1: vector de expresión periplásmica. La construcción queda bajo el promotor de triptofano el cual es inducible mediante adición de ácido β -indol-acrílico al medio de cultivo. Cuenta con el péptido señal de *omp A*, un oligopéptido de histidina, y el terminador del fago T4. La proteína se expresa fusionada a un dominio de afinidad a metales, lo cual facilita la purificación mediante cromatografía de afinidad.

CLONACION

En los diagramas correspondientes a los tres vehículos de clonación se pueden observar los sitios de clonación en cada caso (Figuras 3, 4 y 5). En los tres casos la transformación se hizo en la cepa *Escherichia coli* DH5 α . Para probar la presencia del inserto se llevó a cabo la purificación y la digestión de los plásmidos. Para comprobar que las construcciones estuvieran correctas se hizo la secuencia nucleotídica de cada construcción.

- a) pMal-C2; la clonación se hizo en el sitio de *Xmn* I.
- b) pKK223-3; la clonación se hizo en el sitio de *Sma* I.
- c) pPACIB.1; la clonación se hizo en el sitio de *Eco* RV.

EXPRESION Y PURIFICACION

a) pMal-C2/Cngt II: la expresión se llevó a cabo en la cepa de *E. coli* CMK, en medio Luria adicionado con glucosa (2g/l) y ampicilina (200 μ g/ml), a 37°C. La inducción se hizo mediante IPTG, 0.5 mM, cuando el cultivo alcanzó una A₆₀₀=0.5. Se tomaron muestras antes y después de la inducción para determinar el nivel de expresión en gel SDS-PAGE. La cosecha se realizó aproximadamente 3.5 h después de la inducción. La purificación del péptido de fusión se llevo a cabo siguiendo las instrucciones descritas por el proveedor mediante cromatografía de afinidad, utilizando una resina de agarosa-amilosa a la cual se une el péptido acarreador. Para eluir se utilizó el mismo amortiguador que para la corrida (fosfatos, pH=7.2, 10mM, NaCl 0.5M, azida de sodio 1mM, β -mercapto etanol 10mM, EGTA 1mM) más 10mM de maltosa. Para quitar la maltosa se pasa por una columna (62.0 cm x 1.0 cm) de Sephadex^R G-75 con un amortiguador de corrida de 20 mM NH₄ Ac, pH=4.7. El corte con CNBr se hizo según el protocolo descrito por Smith *et al.* (1996). Brevemente, se solubilizó la proteína en 70% de ácido fórmico y cloruro de guanidina (Gnd HCl) 6M, se mantuvo la reacción por 16-20 hr a 20-25°C en un medio libre de oxígeno (burbujeando N₂ por 15 min) y se agregó el CNBr en una relación molar de 100 veces por cada metionina presente en la proteína. El corte con factor Xa se realizó siguiendo el protocolo del proveedor, el cual consiste en solubilizar la proteína en el amortiguador de corte de la enzima (Tris-Cl, pH=7.2, 20 mM, NaCl 100mM, CaCl₂ 2mM, azida de sodio 1mM) y agregar la enzima en una relación 1:200 peso/peso respecto a la proteína de fusión. Se deja a temperatura ambiente por 24 hr.

b) pKK-223-3/Cngt II, se ensayó la expresión en 5 diferentes cepas de *E. coli*; BL21 DE3, Ec W3110, Ec BMH71.18, EcMM294 y Ec LE392. Los cultivos se hicieron en medio Luria con ampicilina (200µg/ml) a 30°C, la inducción se hizo 3.5 hr. después de iniciado el cultivo, con 2.5 mM de IPTG. Se tomaron muestras antes y después de la inducción, de 2 a 24 horas, para probar la expresión en electroforesis en gel desnaturalizante de poliacrilamida (SDS-PAGE) y mediante ensayo inmunoabsorbente unido a enzima (ELISA).

c) pPACIB.1/Cngt II, expresión en las cepas de *E. coli*; BL21, Ec W3110, Ec BMH71.18, Ec MM294 y Ec LE392. Al igual que para el caso anterior las bacterias fueron crecidas en medio Luria adicionado con ampicilina (200µg/ml) a 30°C; 2 hr después se indujo la expresión agregando 50 µg/ml de ácido β-indolacrilico. Se cosechó 3.5 hr después de la inducción y la expresión se demostró mediante SDS-PAGE y por ELISA. La purificación de la proteína se llevó a cabo siguiendo las instrucciones del proveedor mediante cromatografía de afinidad en una resina de agarosa-níquel. Se hizo en condiciones desnaturalizantes. La elución se llevó a cabo modificando el pH del amortiguador de columna (8M urea, 0.1 M fosfato de sodio, 0.01 M Tris-HCl, pH= 8.0, 6.3, 5.9 y 4.5).

ENSAYOS DE PLEGAMIENTO *in vitro*:

Los protocolos de plegamiento se presentan separados de acuerdo a la proteína que se utilizó para el ensayo. En todos los casos la reacción de plegamiento se detuvo acidificando la mezcla con ácido acético hasta pH= 4.5. Para analizar el resultado de cada reacción se hizo una separación de las diferentes fracciones mediante cromatografía líquida de alta presión (HPLC) de una o varias muestras. En algunos ensayos se realizó ELISA y SDS-PAGE de las diferentes fracciones obtenidas de HPLC.

1º Péptido de fusión sin digerir:

a) Diálisis de la proteína (1.0 mg/ml) con el siguiente amortiguador: 50 mM Tris-HCl (pH=7.0), 50 mM NaCl, 5mM 2-βME, en bolsas con tamaño de exclusión del poro de 12 - 14,000 Da. Se mantuvo a 4°C por 48 hrs y se hizo un cambio del amortiguador a las 24 hrs.

b) Dilución de la proteína (4mg/ml) 1:10 en 20 mM de acetato de amonio, pH= 8.0. Se mantuvo en agitación y se tomaron muestras a las 44, 90 y 140 hrs.

c) Diálisis del péptido (1.0 mg/ml) contra H₂O a 4°C. Se hicieron cuatro cambios del amortiguador para diálisis, cada doce horas y en cada cambio se modificó el pH (7.5, 7.0, 6.5 y 6.0). El tamaño de exclusión del poro de las bolsas de diálisis utilizadas fue de 12- 14,000 Da.

2º Péptido de fusión digerido (Factor Xa o CNBr):

En todos los siguientes ensayos se desnaturalizó y redujo la proteína en 6M Gnd HCl, 50 mM DTT y 100 mM Tris-HCl. A 37°C en condiciones libres de oxígeno, por 12 hrs (Jaenicke, 1989). Para eliminar el Gnd HCl se pasó por una columna de Sephadex G15 con 20 mM de NH₄ Ac, pH=4.7. En el caso de los ensayos de plegamiento por dilución se aplicó la mezcla directamente sin eliminar Gnd HCl por goteo a la solución de plegamiento en agitación.

a) Diálisis de la proteína (2mg/ml), digerida con CNBr, contra concentraciones decrecientes de Gnd HCl en PBS más 0.5 mM de 2-βME. Los cambios de amortiguador fueron 7 y se hicieron cada 15 min, utilizando 5M, 4M, 3M, 2M, 1M, 0M y 0M de Gnd HCl a temperatura ambiente y en agitación. El tamaño de exclusión del poro en las bolsas de diálisis fue de 3,000 Da. Este protocolo fue tomado de Hattori *et al*, (1993).

b) Plegamiento por dilución, 1:10, utilizando un amortiguador sin Gnd HCl, a temperatura ambiente en agitación, se tomaron muestras a las 12, 24 y 48 hr (Hattori, 1993). Posteriormente las muestras se dializaron contra agua por 4 hrs con 4 cambios.

c) Se utilizó un protocolo descrito por Huth *et al*, (1993); dilución de la proteína (2mg/ml), cortada con factor Xa, 1:15 en el siguiente amortiguador: 50 mM Tris-HCl (pH=7.4), 10 mM cisteamina/cistamina (en 10mM HCl), se dejó en agitación a temperatura ambiente y en un medio libre de oxígeno. Se tomaron muestras a las 24, 48 y 72 h.

d) Este ensayo es igual al anterior pero con la proteína digerida con factor Xa. El ensayo se hizo a temperatura ambiente y a 4°C.

e) Es similar al ensayo c pero se ensayaron 2 amortiguadores diferentes: 50 mM Tris-HCl (pH=8.7) y solución salina de fosfatos (PBS).

f) El péptido digerido con factor Xa se resuspendió (0.2 mg/ml) en el siguiente amortiguador: 50 mM fosfato de sodio (pH=7.8), 0.2M NaCl, 5mM glutation reducido (GSH) y 0.5 mM glutation oxidado (GSSG), en agitación y a temperatura ambiente por 2 h.

g) Filtración molecular del péptido de fusión digerido con factor Xa, en Sephadex^R G-15 y para la corrida se usó un amortiguador de ac. acético 10%. Se liofilizaron todas las fracciones proteicas y se resuspendió en ácido acético 10%.

h) El péptido de fusión digerido con factor Xa, sin liofilizar después de la columna de Sephadex^R G-15, se dializó en el siguiente amortiguador: 0.1M fosfato de sodio (pH=8.0), 0.1M NaCl, 0.02M guanidina, 1mM GSH y 1mM GSSG. Se mantuvo a temperatura ambiente y en membranas con un límite de exclusión aproximado de 1,200 Da.

i) El péptido de fusión digerido con factor Xa se dializó con el siguiente amortiguador: 100 mM fosfato de potasio (pH=8.0), 1 mM GSSG, 0.02M guanidina, por 24 hr, a temperatura ambiente utilizando bolsas con un límite de exclusión del poro de 12-14,000 Da. Se colectó el amortiguador de diálisis y se dializó contra: 1mM fosfato de potasio (pH=8.0), en bolsas con un tamaño de exclusión del poro de 3,000 Da a temperatura ambiente, por 2 hr y con cambios cada 30 min. Al terminar se liofilizó y resuspendió en 1 ml de ac. acético 10%.

3° Toxina Cn 5 nativa:

En todos los ensayos con la proteína nativa se desnaturalizó y se redujo igual que la anterior. Para quitar el Gnd HCl se pasó por una columna de Sephadex^R G-15 con 10% de ac. acético. Las fracciones correspondientes a la toxina se liofilizaron para ser utilizadas en los ensayos posteriores.

a) La proteína liofilizada (20 μg) se resuspendió en ácido acético 10% ($\text{pH}=3.2$) y se pasó por HPLC.

b) La proteína desnaturalizada y reducida (50 μg) se pasó por Sephadex^R G-15 con 20 mM de NH_4Ac , $\text{pH}=4.7$.

c) Se dializó la proteína (35 $\mu\text{g}/\text{ml}$) en el siguiente amortiguador: 0.1M fosfato de sodio ($\text{pH}=8.0$), 0.1M NaCl, 1 mM GSH, 1 mM GSSG, en bolsas con un límite de exclusión de 3,500 Da, a temperatura ambiente, por 48 horas, con un cambio a las 24 horas.

d) Se dejó la proteína (50 μg) a 4°C , 1 h, en una solución de pre-plegamiento: 0.1M fosfato de sodio ($\text{pH}=7.8$), 0.1M NaCl, 1mM GSH y 1mM GSSG. El plegamiento se hizo por diálisis en la misma solución pero sin el par redox, usando bolsas con un tamaño de exclusión del poro de 1,200 Da a temperatura ambiente por cuatro horas, con un cambio del amortiguador de diálisis a las dos horas.

e) Igual que el anterior pero la diálisis se dejó toda la noche.

f) La proteína (50 μg) desnaturalizada y reducida se agregó por goteo al amortiguador: 50mM fosfato de sodio ($\text{pH}=7.8$), 0.2M NaCl, 5mM GSH y 0.5 mM GSSG hasta alcanzar una dilución de 1:80. Se dejó en agitación por dos horas a temperatura ambiente.

g) Igual que el ensayo anterior pero usando los dos componentes del par redox en una concentración de 5mM.

4º Toxina recombinante libre (His-Cn 5r):

a) A 50 μg de la toxina recombinante desnaturalizada y reducida se le hizo un ensayo de plegamiento por dilución. Se llevó a cabo mediante el protocolo g utilizado para la toxina nativa.

HPLC

En todos los ensayos de plegamiento el protocolo de HPLC fue el mismo. Se usó una columna C_{18} de fase reversa. La elución se hizo mediante un

gradiente linear de amortiguador A (0.12% de ácido trifluoroacético en agua) a 60% de amortiguador B (0.1% de ácido trifluoroacético en acetónitrilo). La elución fue por 60 minutos con un flujo de 1 ml/ min. Para cada ensayo de plegamiento que se pasó por HPLC se hizo una corrida control con la toxina nativa.

MUTAGENESIS

Los cambios realizados al gene silvestre se hicieron mediante amplificación por PCR, utilizando oligos mutagénicos que se diseñaron específicamente para introducir los cambios programados en los sitios exactos. En este caso se diseñaron sólo los oligos mutagénicos, ya que para hacer la reamplificación completa, después de amplificar con el oligo mutante, se pueden utilizar los oligos que se diseñaron para la clonación del gene silvestre. El primer oligo: mut-1, es un oligo para amplificar a partir del extremo 3', modifica los codones para las 4 prolinas (posiciones 52, 56, 59 y 61), sustituyéndolas por alaninas. Este oligo comprende desde el codón para el aminoácido numero 48 hasta los dos codones de terminación y presenta la delección correspondiente a las dos lisinas finales del péptido maduro (Fig. 2). Se usó como ADN templado la construcción del gene *Cngt II* en pPACIB.1. Para amplificar por el extremo 5' utilizamos el oligo PACIB/Dir.

El segundo oligo: MUT-3 introduce los codones para sustituir las prolinas 56 y 59 por tirosina y valina, respectivamente. Este oligo comprende desde el codón para el aminoácido 47 hasta el codón para el aminoácido 62, por lo que fue necesaria una segunda amplificación para tener el gene completo. El ADN templado que se usó es la construcción del gene silvestre en pPACIB.1. Para amplificar a partir del extremo 5' se utilizó el oligo PACIB1/Dir y para la reamplificación se usó, además del primer producto de PCR, el oligo PACIB1/Rev.

OLIGO mut-1

5' CTA TTA GCT GCA AGA TTT ATT **AGC AAG AGC** ATA AGT **AGC** TGT ACT
3' TTC **AGC** CAA ACC TTC GCAOLIGO mut-3

5' A TTT ATT AGG AAG GGG ATA AGT **AAC** TGT ACT TTC **GTA** CAA ACC TTC
3' GCA CCA

Los productos de PCR que se obtuvieron fueron clonados y secuenciados de la misma forma que se describe para el gene silvestre.

RESULTADOS

En este apartado se describen únicamente resultados no publicados. Los resultados restantes son descritos en los dos artículos del anexo y, principalmente, en el artículo incluido en la tesis.

EDICION DEL GENE

Con el objeto de clonar el gene *Cngt* II en los tres vectores de expresión que se probaron, este se amplificó mediante PCR utilizando los diferentes oligos que generan las modificaciones descritas en la metodología. La Fig. 6 muestra que para cada caso se obtuvo el producto de PCR del tamaño esperado, de acuerdo a los oligos empleados. Utilizando los oligos para hacer la clonación en pMal-C2 esperábamos un fragmento de 204 pb. Para la clonación en pKK223-3 el tamaño del fragmento fue de 264 pb y un fragmento de 218 pb para la clonación en pPACIB.1.

CLONACION

La construcción en pMal-C2 se generó mediante la clonación del producto de PCR en un sitio de extremos rasurados, por lo que a varias de las clones positivas se les hizo secuencia de nucleótidos para encontrar aquellas que tuvieran la orientación y fase de lectura correcta. Para las construcciones en pKK223-3 y pPACIB.1 se diseñaron oligos que permitían hacer una clonación dirigida. Sin embargo, debido a problemas con la doble digestión del inserto se optó posteriormente por la clonación en sitios de extremos rasurados con los productos de PCR sin digerir. Una vez que se comprobó por secuencia nucleotídica que la orientación de estos insertos era correcta se digirieron con la enzima *Eco* RI (figuras 4 y 5) y se volvieron a ligar sobre si mismas con el fin de dejar en fase ambas construcciones. Finalmente, mediante secuencia nucleotídica, se comprobó la fase de lectura de ambos insertos.

EXPRESION Y PURIFICACION

a) Construcción en pMal-C2. Para detectar la expresión se prepararon

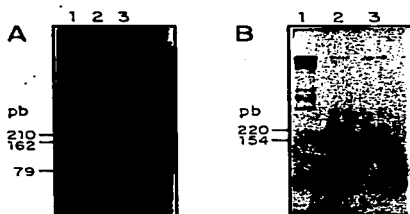


Fig 6. Geles de agarosa 1% que muestran el tamaño de los tres productos de amplificación por PCR del gene *Cngt II*. A: carril 1, marcador de peso molecular; carril 2 y 3, fragmentos de 201 pb para la construcción en pMAL. B: carril 1, marcador de peso molecular; carril 2, fragmento de 267 pb para la construcción en pKK223-3; carril 3, fragmento de 224 pb para la construcción en pPAC1B.1.

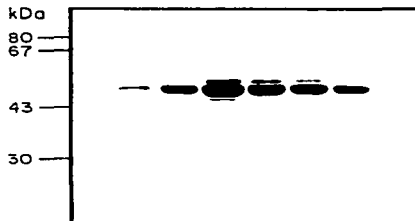


Fig 7. En un gel de poliacrilamida 10% se muestran las fracciones, 2 a 8, de la elución de la columna de afinidad de agarosa-maltosa. El tamaño esperado de la proteína de fusión es de aprox. 49.5 kDa (42.0 kDa de la proteína mbp y 7.5 kDa de la toxina Cn5).

cultivos de 5-25 ml a partir de varias clonas positivas. Se comprobó, mediante electroforesis en SDS-PAGE, la expresión de una proteína con el peso molecular aproximado al esperado (42,000 Da de la proteína acarreadora más 7,200, aproximadamente, de la toxina recombinante). Posteriormente se prepararon cultivos de 0.5 l cada uno para llevar a cabo la inducción y purificación de la proteína de fusión mediante cromatografía de afinidad. Se obtuvieron aproximadamente 2.5 g de células (peso seco) por cultivo. Para la purificación, se pasó el extracto celular sonificado y centrifugado por la columna de afinidad y se obtuvieron 56.25 mg de proteína de fusión total ó 112.5 mg/l. Esto equivale aproximadamente al 20% del total de la proteína celular. Como se observa en la Fig. 7 la pureza de la proteína de fusión es alta.

Al hacer la digestión con el factor Xa de la proteína purificada, se obtuvieron dos bandas que corresponden a los tamaños esperados de la toxina recombinante y de la proteína acarreadora. Finalmente, se hizo una inmunodetección para comprobar la identidad de la proteína recombinante. Como se observa en la Fig. 8, los anticuerpos policlonales contra la toxina nativa reconocen la proteína de fusión y al péptido pequeño que resulta de la digestión pero no a la proteína acarreadora (mbp). Se determinó la secuencia de aminoácidos de la banda pequeña producto de la digestión. Los primeros 20 aminoácidos corresponden al extremo amino de la toxina Cn 5 recombinante. Esta toxina recombinante presenta además de la secuencia del amino terminal de la toxina nativa la metionina adicionada.

b) Construcción en pKK223-3. Se hicieron ensayos de expresión en varias cepas y debido a la baja expresión en este sistema sólo fue posible detectarla mediante ELISA. Se hicieron varios esfuerzos por tener una mejor producción de la toxina recombinante pero no fue posible obtenerlo por lo que se abandonó este sistema.

c) Construcción en pPACIB.1. Se hicieron análisis de expresión en varias cepas y se demostró mediante ELISA que en las cepas Ec W3110, Ec BMH 71.18 y Ec MM 294 se lograron los mejores rendimientos (Fig. 9). Para la purificación de la proteína recombinante, se preparó un cultivo de 0.5 l en el cual se obtuvieron, 2.9 g de células (peso húmedo). Se probaron diferentes condiciones de expresión y se encontró que las mejores condiciones para la producción de la toxina recombinante eran a 30°C, con 50 µg/ml de ácido β-indol acrílico y en medio Luria. El extracto celular

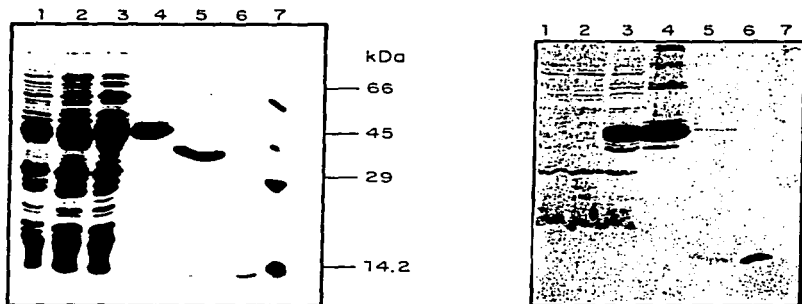


Fig 8. Expresión de la proteína de fusión mbp-Cn 5. A: gel de poliacrilamida 10% y B: Western blot. Carril 1: cepa sin transformar, carril 2: cepa transformada sin inducir, carril 3: cepa transformada e inducida, carril 4: proteína de fusión después de la purificación, carril 5: proteína de fusión purificada y digerida con el factor Xa, carril 6: toxina Cn 5, carril 7: marcador de peso molecular. Las dos bandas que aparecen después de la digestión corresponden a los tamaños esperados (42 y 7.5 kDa).

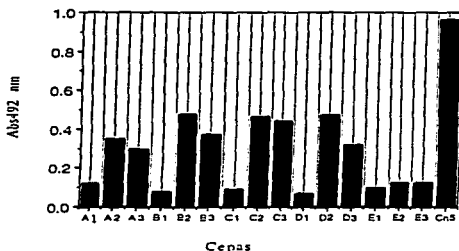


Fig 9. Ensayo por ELISA de expresión de la construcción en pPACIB.1, utilizando diferentes cepas. A: BL21, B: EcW3110, C: EcBMH71.18, D: EcMM294, E: EcLE392. 1: antes de la inducción, 2 y 3: dos y 6 horas después de la inducción, respectivamente.

desnaturalizado se pasó por la columna de afinidad y después de eluir se obtuvieron cerca de 3 mg de proteína por lo que el rendimiento de la proteína recombinante corresponde aproximadamente a 2.5 % de la proteína total. Mediante gel SDS-PAGE se demostró la presencia de dos bandas protéicas, una de las cuales presenta un peso molecular similar a la toxina nativa (Fig. 10). La banda mayor corresponde a una proteína bacteriana que co-purifica siempre que se utiliza esta columna (Wülfing, 1994). También se demostró por ELISA que los anticuerpos policlonales contra la toxina nativa son capaces de reconocer las fracciones de la elución. En ensayos de Western blot, se demostró que estos anticuerpos reconocen únicamente a la banda de peso molecular similar al de la toxina nativa (Fig. 11).

Una vez purificada la proteína se hizo un corte con bromuro de cianógeno y para recuperar la toxina recombinante libre se utilizó HPLC. A este material se le hizo secuencia de aminoácidos y se comprobó que los primeros 25 aminoácidos corresponden a la secuencia exacta de la toxina nativa. Por lo que podemos decir que la toxina recombinante presenta un extremo amino idéntico al de la toxina nativa.

PLEGAMIENTO *in vitro*

1º. Péptido de Fusión sin Digerir (ver Tabla 3).

Para tratar de plegar la toxina recombinante antes de separarla de la proteína acarreadora (mbp) se utilizaron dos métodos: dilución y diálisis modificando varios parámetros como tiempo de plegado, amortiguador y pH. En todos estos ensayos los resultados fueron negativos, no conseguimos recuperar la toxina recombinante una vez que se hizo la digestión (CNBr o Factor Xa). Como ejemplo, en la Fig. 12 podemos ver un cromatograma de HPLC que se obtuvo después de aplicar el protocolo a y digerir el péptido de fusión con factor Xa. No aparece ningún pico que tenga un tiempo de retención similar al de la toxina nativa. Se ven una serie de pequeños picos irregulares, cercanos a los 38 min, los cuales mediante ELISA demostramos que corresponden a la toxina recombinante. Este resultado nos indica que la toxina no está presentando el plegamiento correcto y se asume que esa serie de pequeños picos correspondan a diferentes estados de plegamiento y/o agregación de la toxina recombinante.

2º. Péptido de Fusión Digerido (ver Tabla 3).

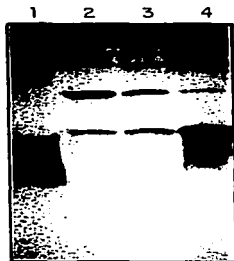


Fig 10. Electroforesis en gel de poliacrilamida 12.5% de las fracciones obtenidas durante la purificación de la proteína recombinante (His-Cn5r). La elución mediante cambios de pH produjo diferentes fracciones en las que se ven dos bandas, una de las cuales presenta una migración cercana a la nativa. Carril 1: toxina Cn 5, carriles 2, 3 y 4: fracciones I, II y III, respectivamente.

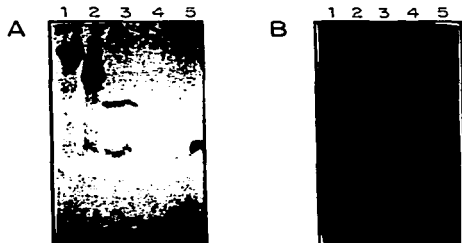


Fig 11. Electroforesis en gel de poliacrilamida 10% y membrana de inmunotransferencia (Western); carril 1: proteína BSA (Albúmina de suero bovino), carril 2: proteína de fusión (mbp-Cn5 recombinante), carril 3: proteína eluida de la columna de agarosa-níquel, carril 4: igual que tres, diferente fracción, carril 5: toxina Cn 5 nativa. El reconocimiento de los anticuerpos es únicamente hacia la banda pequeña y al péptido de fusión, lo cual indica que la proteína expresada corresponde a la toxina recombinante.

PROTEINA	METODO	AMORTIGUADOR	pH	TEMP.	TIEMPO	AGENTE REDOX	No. DE FIGURA
mbp-Cn 5 1mg/ml	Diálisis	50mM Tris-HCl 50mM NaCl	7.0	40C	48 h	5mM B-ME	12
mbp-Cn 5/Xa 2mg/ml	Dilución 1:15	50mM Tris-HCl	7.4	T. A.	24, 48 y 72 h	10mM Cisteam/Cistam	14
Cn 5 50 ug	Cromato- grafía	20mM NH4 Ac.	4.7	T. A.			17
Cn 5 35 ug/ml	Diálisis	0.1 M Na-fosfato 0.1 M NaCl	8.0	T. A.	48 h	1mM GSH 1mM GSSG	18
Cn 5 50 ug	Dilución 1:80	50mM Na-fosfato 0.2 M Na Cl	7.8	T. A.	2 h	5mM GSH 0.5mM GSSG	19
Cn 5 50 ug	Dilución 1:80	50mM Na-fosfato 0.2 M NaCl	7.8	T. A.	2 h	5mM GSH 5mM GSSG	20
His-Cn 5 50 ug	Dilución 1:80	50mM Na-fosfato 0.2 M NaCl	7.8	T. A.	2 h	5mM GSH 5mM GSSG	21

Tabla 3 La tabla muestra un resumen de los protocolos de plegamiento *in vitro* correspondientes a cada uno de los cromatogramas de HPLC, que se muestran en este trabajo. Presenta las variables correspondientes a los siguientes ensayos: Fig. 12 ensayo a de la proteína de fusión sin digerir, Fig. 14 ensayo c de la proteína de fusión digerida con factor Xa, Figuras 17, 18, 19 y 20 ensayos b, c, f y g, respectivamente, de la toxina nativa y Fig. 21 ensayo a correspondiente a la toxina recombinante His-Cn 5. En todas las figuras el asterisco (*) indica el pico que presenta el tiempo de retención similar a la nativa.

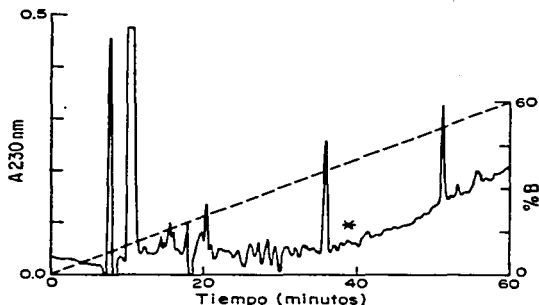


Fig 12. Cromatograma de HPLC correspondiente al ensayo *a de* plegamiento *in vitro* utilizando el péptido de fusión (mbp-Cn 5) sin digerir, la digestión se hizo después del plegamiento. Se recupera un pico con el tiempo de retención del péptido sin digerir (51.15 min), el resto del material se ha perdido por precipitación.

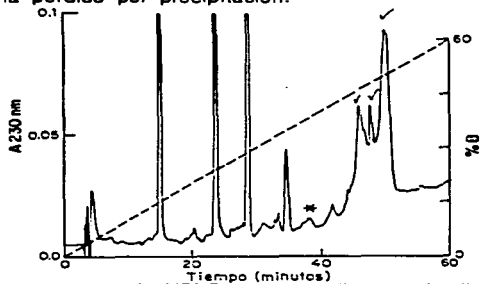


Fig 13. Cromatograma de HPLC correspondiente a la digestión mediante CNBr del péptido de fusión. La presencia de varios picos (✓), los cuales son reconocidos por los anticuerpos policlonales contra Cn 5, nos indica que la digestión fue parcial y que se están generando diversos péptidos debido a la presencia de metioninas en la proteína mbp.

Para este caso también se probaron diferentes parámetros y en el caso del corte con CNBr se observó que no hubo plegamiento. Se hizo ELISA de las diferentes fracciones obtenidas después del corte y se demostró que todas son reconocidas por los anticuerpos contra la toxina nativa por lo que pensamos que la digestión fue muy parcial. Esto se puede observar en el cromatograma de la Fig. 13. Para evitar el problema de la digestión parcial decidimos hacer solamente cortes con factor Xa, pues mediante geles SDS-PAGE habíamos observado que esta digestión es casi total. Con los ensayos de plegamiento de la proteína digerida con el Factor Xa se obtuvieron fracciones que presentaban el tiempo de retención de la nativa. Los ensayos que dieron resultados positivos, los más exitosos fueron aquellos en que se hizo el plegamiento mediante dilución utilizando un par redox. Como ejemplo, podemos observar los cromatogramas del ensayo *c*, en el cual ya se observa un pico que eluye en el tiempo de retención de la toxina nativa (Fig. 14). En el ensayo *c* se realizó un ELISA a las fracciones obtenidas en los diferentes tiempos del proceso de plegamiento. Se observó que el reconocimiento de los anticuerpos aumentaba con el tiempo (Fig.15). En cambio, en los ensayos que se hicieron mediante diálisis, no se observó plegamiento. En el caso particular del ensayo *i* se formó un precipitado en la bolsa. Se comprobó que en el precipitado se encuentra la mayor parte de la toxina recombinante y parte de la proteína mbp como se puede ver en el gel SDS-PAGE de la Fig. 16. Este fenómeno de agregación ya se había observado en otros ensayos de plegamiento en los que se usó la proteína de fusión digerida con el Factor Xa.

3°. Toxina Nativa (ver Tabla 3).

En todos los ensayos de plegamiento con la toxina nativa se obtuvo una fracción con el tiempo de retención de la proteína nativa. En algunos ensayos el porcentaje fue menor como, en el caso de los ensayos *a* y *b* que fueron por filtración molecular. Se observaron varias conformaciones o diferentes plegamientos que se manifiestan por la presencia de varios picos bien definidos en el cromatograma (Fig. 17). En la base de los picos se puede ver el material que seguramente presenta fases de plegamiento muy diferentes. El método de diálisis permitió una buena recuperación del material, como se muestra en la Figura 18 para el caso del ensayo *c*. En este cromatograma se ve un pico bien definido que presenta el tiempo de retención correspondiente a la toxina nativa. Además, la cantidad de material recuperado nos está indicando que el proceso de plegamiento fue

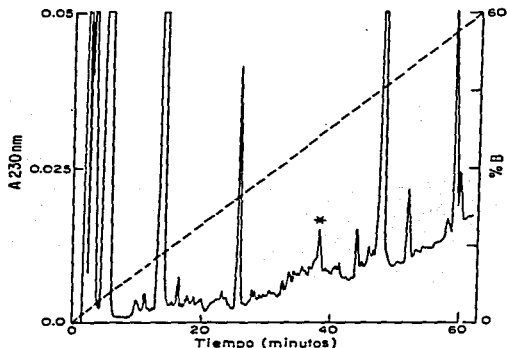


Fig 14. Cromatograma de HPLC correspondiente a un ensayo de plegamiento mediante dilución. Se utilizó el péptido desnaturalizado, reducido y digerido con factor Xa (ensayo c). Se observa la presencia de un pequeño pico que presenta el tiempo de retención de la toxina nativa (*), el cual es reconocido por los anticuerpos contra la toxina Cn 5.

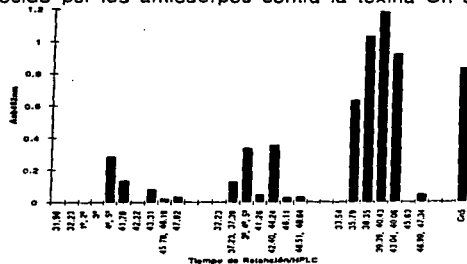


Fig 15. ELISA de diferentes alícuotas del ensayo de plegamiento c. Podemos observar un aumento en el reconocimiento de los anticuerpos al transcurrir el tiempo del ensayo. Los tres conjuntos de datos corresponden a los tiempos de plegamiento: 24, 48 y 72 h

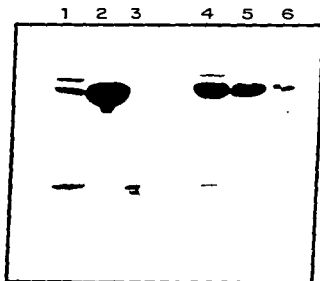


Fig 16. Gel de poliacrilamida 12.5% correspondiente al ensayo *i* de plegamiento de la toxina digerida con factor Xa. Carril 1: precipitado (desnaturalizado y reducido), carril 2: proteína mbp, carril 3: toxina Cn 5, carril 4: alícuota del ensayo reducido, carril 5: sobrenadante reducido, carril 6: precipitado (sin reducir y desnaturalizar). Podemos observar que la mayor parte de la toxina recombinante se encuentra en el precipitado, mientras que en el sobrenadante se encuentra la proteína acarreadora.

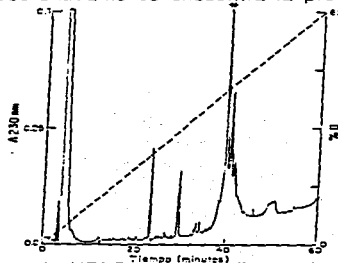


Fig 17. Cromatograma de HPLC correspondiente al ensayo de plegamiento *b* de la proteína nativa. Se muestra la gran recuperación del material, aún cuando no todo está plegado. Suponemos que los tres picos corresponden a tres isoformas estructurales de la proteína plegada, mientras que la base seguramente representa material soluble no plegado.

favorecido sobre el proceso de agregación. También podemos asumir que no todas las vías de plegamiento fueron productivas, pues la base del pico presenta muchas irregularidades las cuales reflejan diferentes conformaciones de la molécula. En los ensayos de plegamiento anteriores, utilizando la toxina recombinante, especialmente los que se hicieron mediante diálisis, la pérdida del material por agregación fue casi del 100%. Por esto al no haber proteína soluble, el tamaño de los picos en el cromatograma es apenas detectable (Fig. 12). En general podemos decir que con la toxina nativa se aumentó la cantidad de toxina plegada aunque no toda con la configuración correcta pues, aún que es proteína soluble, no eluye como un solo pico bien definido. Se observó un incremento en el plegamiento cuando se usa el par redox GSH/GSSG y parece ser mejor cuando la relación de este par es 1:1 que 1:10. Esto se mostró en los ensayos *f* y *g* que se presentan en los cromatogramas de la Fig 19 y Fig 20, respectivamente.

4° Toxina Recombinante Libre (ver Tabla 3).

Con la toxina recombinante (His-Cn5), ya liberada del oligopéptido de histidinas, se hicieron los ensayos de plegamiento similares a los de la toxina nativa. En la Fig. 21 se observa el cromatograma correspondiente a un ensayo por dilución en el que se utiliza el par redox en relación 1:1 (protocolo *g* de la toxina nativa). Aún cuando el pico presenta el tiempo de retención correspondiente a la nativa, la poca definición indica una alta heterogeneidad molecular. En pruebas de toxicidad en cochinillas, este material presenta actividad tóxica letal aunque con mayor dosis que la toxina nativa.

MUTAGENESIS

Hasta el momento se cuenta con la clona correspondiente a la mutante 1 en la cual se sustituyeron las 4 prolinas por alaninas. Mediante secuencia de nucleótidos se comprobó que la clona está en fase, que tiene las cuatro mutaciones y que no se encuentran las 2 lisinas finales. Antes de hacer la construcción se analizó (Dra. Barbara Selisko) la estructura preliminar obtenida por NMR de la toxina nativa (datos sin publicar). Utilizando el programa INSIGHT II de Biosym Technologies, San Diego en una estación de trabajo SILICON GRAPHICS IRIS 4D/35. Se midieron los ángulos ϕ y ψ que presentan las prolinas en la estructura y se encontró que corresponden a valores permitidos para alaninas (Cantor, 1980). Adicionalmente, se

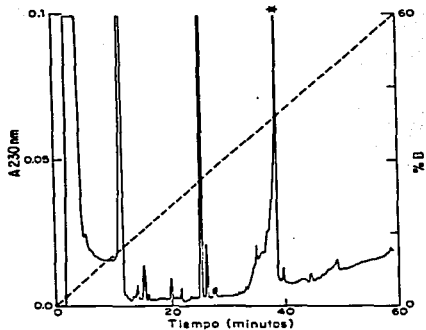


Fig 18. Cromatograma de HPLC correspondiente al ensayo de plegamiento *c* de la proteína nativa. En este ensayo de diálisis podemos observar un pico bien definido, que presenta el tiempo de retención de la toxina nativa.

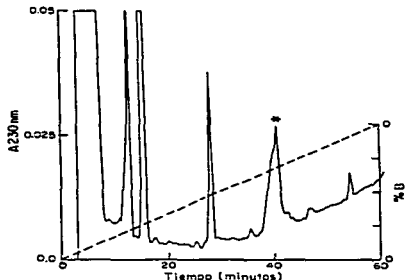


Fig 19. Cromatograma de HPLC correspondiente al ensayo *f* de plegamiento de la toxina nativa. En este ensayo por dilución, se utilizó un par redox con una relación uno a diez del oxidante sobre el reductor. Se observa un pico mayoritario aunque no totalmente definido.

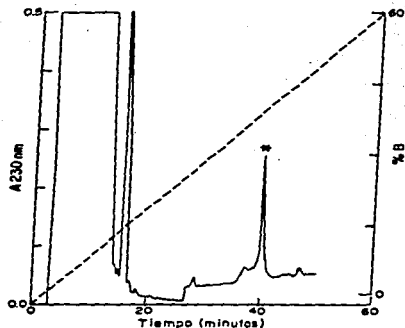


Fig 20. Cromatograma de HPLC correspondiente al ensayo de plegamiento *g* de la proteína nativa. Ensayo similar al *f* sólo que la relación del par redox es uno a uno. Se observa un pico mayoritario muy bien definido.

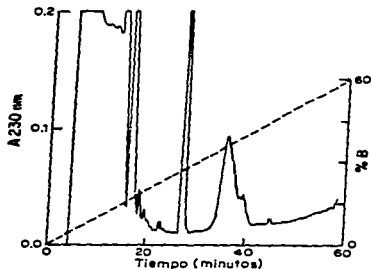


Fig 21. Cromatograma de HPLC correspondiente al ensayo de plegamiento *a* de la proteína recombinante His-Cn5r. Como se puede ver existe una alta recuperación del material lo cual indica que la proteína está soluble y no precipitada. El pico mayoritario esta muy cerca del tiempo de retención de la toxina nativa y el material recuperado presenta actividad tóxica.

modificaron las prolinas a alaninas usando el programa BIOPOLYMER de Biosym y se hizo una dinámica del modelo mediante el programa DISCOVER de Biosym (20 ps a una temperatura de 300 K). La estructura final muestra cambios mínimos en las posiciones C α de la región C terminal, residuos 50 a 65, de hasta 1.29 Å. Este resultado muestra que el modelo presenta una forma energéticamente estable. Este análisis nos permite asumir que las sustituciones no son un factor que impida una correcta estructuración de la toxina mutante. Sin embargo, la situación real del plegamiento de la mutante va a diferir por la presencia de la Ala 59 en el lugar de la Pro 59, que presenta un enlace peptídico en *cis*. La forma *trans* en un residuo como alanina estará favorecida 1000:1 sobre la *cis* por lo que es poco probable que se forme un enlace *cis*, similar a la nativa. Sólo en el caso en que la situación entre las cadenas laterales cerca de la posición 59, presentando un enlace *cis* antes de la Ala 59, sea tan energéticamente favorable que permita su formación.

RESULTADOS PUBLICADOS

-García C., Becerril B., Selisko B., Delepierre M. and Possani L. D. (1997). Isolation, characterization and comparison of a novel crustacean toxin with a mammalian toxin from the venom of the scorpion *Centruroides noxius* Hoffmann. *Comp. Biochem. Physiol.* 116B(3): 315-322.



Isolation, Characterization and Comparison of a Novel Crustacean Toxin with a Mammalian Toxin from the Venom of the Scorpion *Centruroides noxius* Hoffmann

Consuelo García,¹ Baltazar Becerril,¹ Barbara Selisko,¹ Muriel Delepiere² and Lourival D. Possani^{1*}

¹DEPARTMENT OF MOLECULAR RECOGNITION AND STRUCTURAL BIOLOGY, INSTITUTE OF BIOTECHNOLOGY, NATIONAL AUTONOMOUS UNIVERSITY OF MEXICO, AVENIDA UNIVERSIDAD, 2001, APARTADO POSTAL 510-3, CUERNAVACA, MORELOS 62271 MEXICO, AND ²LABORATORY OF NUCLEAR MAGNETIC RESONANCE (CMRS URA 1159) PASTEUR INSTITUTE, 28 AVE DU DR. ROUV, 75015 PARIS, FRANCE

ABSTRACT. A novel crustacean-specific toxin, Cn5, containing 66 amino acid residues was isolated from the venom of the scorpion *Centruroides noxius* Hoffmann. It is stabilized by four disulfide bridges, formed between Cys12-Cys65, Cys16-Cys41, Cys25-Cys46 and Cys29-Cys43. Toxicity tests revealed that Cn5 is a toxin that affects arthropods but not mammals. However, at high concentrations, Cn5 does displace the mammalian-specific toxin Cn2 from rat brain synaptosomes. The concentration of Cn5 that produces half-maximal inhibition (IC₅₀) was estimated to be 100 μ M. Sequence comparison of Cn5 with toxin Cn2, a mammal-specific toxin from the same scorpion, showed the presence of two sequence stretches, at positions 30 to 38 and 49 to 58, where the majority of the differences are concentrated. On the three-dimensional structure of Cn5 it is demonstrated that these two sequence stretches form a continuous surface region near the site thought to bind to the sodium channel. We assume that this region might be implicated in determining species specificity. *COMP. BIOCHEM. PHYSIOL.* 116B,3:315-322, 1997. © 1997 Elsevier Science Inc.

KEY WORDS: Amino acid sequence, *Centruroides noxius*, crustacean-specific toxin, scorpion toxins, sodium channel, species specificity, three-dimensional modelling

INTRODUCTION

Scorpion toxins have been shown to affect ion permeability of excitable cells (1). Four different families of peptides have been described, which specifically recognize membrane-bound proteins, Na⁺-channels (1,2,3), K⁺-channels (4,5), Cl⁻-channels (6) and Ca²⁺-channels (7). Despite the fact that the primary structures of these proteins can be quite different (8), their three-dimensional structure is extremely similar. Except for the Ca²⁺-channels specific scorpion toxins, whose structure is still unknown (7), all these toxins were shown to contain a core structure, formed by an α -helix and a three-stranded β -sheet that are connected by two disulfide bridges formed between two conserved sequence stretches, Cys-X-X-X-Cys and Cys-X-Cys, where X is a variable amino acid (9-11).

ALLIENI requests for Lourival D. Possani, Institute of Biotechnology, UNAM, Avenida Universidad, 2001, Apartado Postal 510-3, Cuernavaca, 62271 MEXICO, Tel (+5273) 1711269, Fax (+5273) 172385. E-mail: possani@ibitec.unam.mx

Abbreviations: Cn, *Centruroides noxius* Hoffmann; CM-Cellulose, carboxymethyl-cellulose; RC-toxin, toxin reduced and carboxymethylated; ¹H-NMR, proton nuclear magnetic resonance.

Received 25 June 1996, accepted 6 September 1996.

All known Na⁺-channel specific scorpion toxins are composed of 61-70 amino acid residues, stabilized by four disulfide bridges (3,12-16). According to toxicity tests, binding experiments or electrophysiological experiments, they can be divided into toxins that are specifically toxic to mammals or arthropods. The term "specific," however, does not mean "exclusively effective" but refers to low toxicity or affinity in one group and/or higher toxicity or affinity in the other. There are many reports of toxins that show effects towards both mammals and arthropods (17-20).

Mammalian-specific toxins can be grouped into α -toxins binding to site 3 of mammal Na⁺-channels, slowing down their inactivation process, and β -toxins binding to site 4 of the channel, affecting the activation of the channel (21,22). Displacement experiments on excitatory tissues showed that the two binding sites are not related (21,23). Comparing structural features, α - and β -toxins vary in the site of two loop regions, the J-loop and the D-loop (24).

Arthropod-specific toxins have been tested so far mainly in insects (13,17-20), e.g. blowfly larvae, crickets, cockroaches, and some of them in crustaceans (13,17,25), e.g. crayfishes and cochinillas. Insect α -toxins show the classic α -toxin effect on insect sodium channels (17,26). This type

of toxins are believed to bind to receptor site 3 (26). Toxin T α 7 (also called T α VII) from the New-World scorpion *Tityus serrulatus* is a β -toxin that shows toxic effects in insects and in mammals (13,27). β -Toxins like T α 7 are thought to bind to the receptor site 4 on insect sodium channels (26). An aspect which distinguishes toxins against insect sodium channels from their vertebrate counterpart is that the former can be further divided into excitatory and depressant toxins, causing a contractory or flaccid paralysis, respectively (28). Both toxin types displace each other (28-30), but their receptor sites are not identical (31).

The question why scorpion toxin molecules with their similar three-dimensional fold show marked differences in their functional characteristics, i.e., specific affinity to sodium channels of a particular animal phyla, still remains unclear. Rather small differences, i.e., single amino acid modifications within the critical binding region of the toxin to its receptor, seem to be one possible explanation. Thus far, there has been little information published concerning the identification of certain regions or residues that might determine the specificity of scorpion toxins.

In the present communication three aspects of a novel crustacean-specific toxin, Cn5, from the scorpion *Centruroides noxius*, are presented: 1) the isolation, the determination of the primary structure and the position of the disulfide bridges, 2) its species specificity and ability to compete with the mammal-specific toxins for binding to brain synaptosomes and 3) a sequence comparison of Cn5 with mammal-specific toxins of the genus *Centruroides* resulting in the identification of a continuous region on the surface of the toxin molecule, which might be responsible for their specificity.

MATERIALS AND METHODS

Source of Venom

Venom from scorpions of the species *Centruroides noxius* Hoffmann was obtained by electrical stimulation of the anesthetized animals. The venom was resuspended in water and centrifuged for 15 min at 10,000 \times g. Subsequently, it was freeze dried and kept at -20°C .

Chemical and Reagents

Only analytical grade chemicals were used throughout this work. Reagents for sequencing were all obtained from Milligen/Bioscience, Division of Millipore (Burlington, MA). Acetonitrile and trifluoroacetic acid were from Pierce (Rockford, IL). Trypsin and chymotrypsin were sequence grade enzymes from Boehringer Mannheim (Mannheim, Germany). Frotase VS from *Staphylococcus aureus* was kindly provided by Dr. Brian Martin (NIMH, Bethesda, MD). Sephadex G-50 (medium-grade) was purchased from Pharmacia Fine Chemicals (Lppsala, Sweden), carboxymethyl-cellulose (CM-32) from Whatman (Clifton, NJ).

Reagents for reduction and alkylation were obtained from Sigma (St. Louis, MO). Deionized, double distilled water was used throughout.

Separation Procedures

Soluble venom from the scorpion *Centruroides noxius* Hoffmann was separated by chromatographic procedures as earlier described (3,32). Briefly, the first purification step consisted in a gel filtration step using a Sephadex G-50 column. For further purification the material was applied onto a carboxymethyl cellulose (CM-1) Cellulose column, equilibrated and developed in the presence of 20 mM ammonium acetate buffer, pH 4.7, with a continuous salt gradient from 0 to 0.5 M NaCl. The third and last purification step was conducted on a Waters (Millipore Co., Milford, MA, USA) high performance liquid chromatography (HPLC) system using a C_{18} reverse phase column (Vydac, Hesperia, CA). For details refer to corresponding figure legends. The homogeneity of the resulting components were assessed by three independent means: HPLC separation on a C_{18} reverse phase microbore column (2 mm I.D.) using the Waters system (Millipore Co., Milford, MA, U.S.A.), polyacrylamide gel electrophoresis in acidic conditions (33) or in presence of sodium dodecyl sulfate (34) and microsequencing as described below.

Bio-assays

Lethality tests were performed in three distinct species of animals, following approved protocols by the Animal Care Commission of the Institute of Biotechnology. Adult albino mice (strain CD1, average body weight 20 g) were used to test the various chromatographic fractions in the purified toxin. A given amount of protein dissolved in 200 μl phosphate-buffered saline (PBS) was injected intraperitoneally. Bioassays with insects and crustaceans were performed using crickets (*Acheta* spp.) and freshwater crayfishes (*Cambarellus montezumae* spp.), respectively, where up to 150 μg toxin dissolved in 20 μl water were injected. As control experiments test animals were injected with pure PBS or water.

Primary Structure Determination of Cn5

The amino acid sequence of toxin Cn5 was obtained by direct sequencing the native toxin, the reduced and carboxymethylated toxin (RC-toxin) as well as fragments generated by enzymatic cleavage of RC-toxin. The reduction of the toxin with dithiothreitol, the alkylation with iodoacetic acid and the enzymatic cleavage with trypsin, chymotrypsin and protease VS were performed as described (35,36). The peptides produced by enzymatic hydrolysis were separated by HPLC as described (37). Microsequence determination was carried out with a Milligen ProSequencer model 6600

(Division of Millipore Co, Milford, MA, U.S.A.) using protocols with the peptide either adsorbed or covalently attached to Sepulon[®] membranes (Aryl amine or DITC) as previously described (35-37).

Amino acid analysis was performed on samples hydrolyzed in 6 N HCl with 0.5% phenol at 110°C in evacuated, sealed tubes as described (38).

Determination of Disulfide Bridges

Aliquots (usually 100 µg protein) of native Cn5 were digested with trypsin, chymotrypsin and endopeptidase V8, or a combination of these enzymes. The corresponding peptides were separated by HPLC and sequenced using the protocol described above.

Toxin Binding to Rat Brain Synaptosomes

The purification of rat brain synaptosomes (fraction P3) and binding-displacement assays were performed as described (39). Briefly, rat brain membranes were added to a buffered solution (140 mM NaCl, 5 mM KCl, 2.5 mM CaCl₂, 1.5 mM MgSO₄, 25 mM Tris-HCl, pH 7.4, 0.1% BSA) of ¹²⁵I-labeled Cn 2 with or without unlabeled toxin Cn2 or Cn5. After 60 min at room temperature the reaction was stopped by adding 5 ml of cold binding medium. The membranes were immediately collected on glass fiber filters, washed, dried and counted in a gamma counter. All values are an average of at least triple experiments.

RESULTS

Purification Amino Acid Sequence and Determination of Disulfide Bridges

PURIFICATION AND AMINO ACID SEQUENCE. The initial separation of the venom on CM-cellulose resulted in three fractions, from which number II was toxic. It was further separated by ion-exchange chromatography on CM-cellulose giving 14 sub-fractions, as described in detail in (S,32). Fraction II-5 (the fifth component from the separation on CM-cellulose) was not toxic to mice, but to insects and crustaceans. We purified this fraction further by reverse phase HPLC (Fig. 1). In order to guarantee its purity, the fifth component of the chromatogram given in Fig. 1 was reappplied to an HPLC C₁₈ reverse phase column (Fig. 1, inset). The homogeneity of the obtained peptide named Cn5 was confirmed by gel electrophoresis (data not shown) and by direct sequencing the native peptide. A single sequence was obtained for the N-terminal peptide: KEGYLVNKS⁺TG⁺NYGNLLG⁺ONE (where X is a blank space corresponding to cysteine), demonstrating the purity of the peptide. The toxin was reduced, carboxymethylated and completely sequenced. In Fig. 2 we show the results of the overlapping sequences determined. The RC-toxin gave an unequivocal

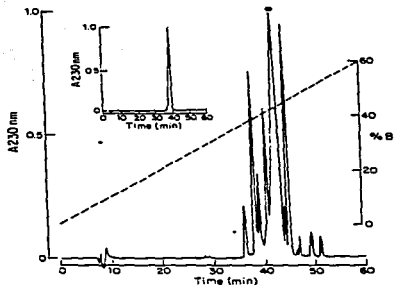


FIG. 1. HPLC separation of toxin Cn5. Sub-fraction II-5 (500 µg) from the second purification step on CM-cellulose (see Material and Methods) was applied to a semi-preparative C₁₈ reverse-phase column (25 cm × 1 cm I.D., particle size 10 µm, Vydac, Hysperia, U.S.A.) and separated during 60 min using a linear gradient from solution A (0.12% trifluoroacetic acid in water) to 60% solution B (acetonitrile in presence of 0.1% trifluoroacetic acid). Component 5, labeled with a star, corresponds to Cn5. The inset shows the profile of the re-chromatography of Cn5 on an HPLC C₁₈ reverse phase column using the linear gradient given above.

direct amino acid sequence up to residue at position 51. A tryptic fragment (eluted at 53 min in a 90 min gradient—data not shown) of the RC-toxin resulted in an overlapping sequence defining residues 31 to 52. An aliquot of the same peptide was digested with protease V8. The sub-peptide (T1V8), eluting at 29.20 min, gave the amino acid sequence from residues at position 53 to 65. The last amino acid residue was identified by amino acid analysis of the full length toxin and of peptide T1V8. The composition obtained demonstrated the presence of all the residues already sequenced



FIG. 2. Amino acid sequence of toxin Cn5. Automatic Edman degradation (sequence underlined with ⁺) provided the first 51 positions of the primary sequence. Peptide 1 (T1) obtained by tryptic hydrolysis of Cn5 permitted the identification of residues 31 to 52. Sequencing the sub-peptide generated by protease V8 digestion of peptide T1 (-T1V8-) identified residues 53 to 65. The last residue was identified by amino acid analysis (AA).

TABLE 1. Amino acid sequence of heteropolymeric peptides corresponding to the disulfide bridges of Cn5

Elution HPLC (min)	Amino acid sequence heteropolymer	Corresponding disulfide bridge																				
28.75	<table border="0" style="margin-left: 20px;"> <tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr> <tr><td>Gly</td><td>Tyr</td><td>Cys</td><td>Tyr</td><td>Ala</td></tr> <tr><td>Gly</td><td>Xxx</td><td>Leu</td><td></td><td></td></tr> </table>	1	2	3	4	5	Gly	Tyr	Cys	Tyr	Ala	Gly	Xxx	Leu			Cys16-Cys41					
1	2	3	4	5																		
Gly	Tyr	Cys	Tyr	Ala																		
Gly	Xxx	Leu																				
31.90	<table border="0" style="margin-left: 20px;"> <tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr> <tr><td>Asn</td><td>Glu</td><td>Glu</td><td>Cys</td><td>Asp</td><td>Leu</td></tr> <tr><td>Ala</td><td>Phe</td><td>Gly</td><td>Cys</td><td></td><td></td></tr> </table>	1	2	3	4	5	6	Asn	Glu	Glu	Cys	Asp	Leu	Ala	Phe	Gly	Cys			Cys23-Cys46		
1	2	3	4	5	6																	
Asn	Glu	Glu	Cys	Asp	Leu																	
Ala	Phe	Gly	Cys																			
36.14	<table border="0" style="margin-left: 20px;"> <tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr> <tr><td>Ser</td><td>Tyr</td><td>Gly</td><td>Cys</td><td></td></tr> <tr><td>Ser</td><td>Xxx</td><td></td><td></td><td></td></tr> <tr><td>Asn</td><td>Gln</td><td>Gly</td><td>Gln</td><td>(Tyr)</td></tr> </table>	1	2	3	4	5	Ser	Tyr	Gly	Cys		Ser	Xxx				Asn	Gln	Gly	Gln	(Tyr)	Cys12-Cys65 Cys16-Cys41
1	2	3	4	5																		
Ser	Tyr	Gly	Cys																			
Ser	Xxx																					
Asn	Gln	Gly	Gln	(Tyr)																		
36.62	<table border="0" style="margin-left: 20px;"> <tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr> <tr><td>Glu</td><td>Cys</td><td>Leu</td><td></td><td></td></tr> <tr><td>Xxx</td><td>Glu</td><td>Gly</td><td>Leu</td><td>Phe</td></tr> </table>	1	2	3	4	5	Glu	Cys	Leu			Xxx	Glu	Gly	Leu	Phe	Cys29-Cys45					
1	2	3	4	5																		
Glu	Cys	Leu																				
Xxx	Glu	Gly	Leu	Phe																		

The given amino acid sequences were obtained by microsequencing peptide purified by HPLC after the cleavage of Cn5 with trypsin and chymotrypsin. The numbers on top of the amino acid mean positions from the sequence Xxx means a blank position corresponding to cysteine, while CysX refers to the position of a cysteine (a component that absorbs at 205 nm π - π corresponds to this amino acid). Tyrosine between parentheses was assumed based on the amino acid sequence and enzymatic cleavage specificity.

plus one additional serine. We concluded that the last position, amino acid 66, was occupied by a serine. These findings were confirmed by the amino acid sequence deduced from the nucleotide sequence of the cDNA clone CngrII encoding toxin Cn5 (40).

DETERMINATION OF THE DISULFIDE BRIDGES. Several aliquots of native Cn5 were hydrolyzed with trypsin followed by an additional cleavage with chymotrypsin. The resulting peptides were separated by reverse phase HPLC and sequenced. Over forty different fractions were sequenced (not shown), some of which gave two amino acids per step, indicating the presence of two peptides connected by a disulfide bridge (dimeric peptide). Table 1 shows four sequences that allowed us to propose how the disulfide bridges are arranged in the native toxin. The dimer eluting at 28.75 min gave sequences consistent with the disulfide bridge Cys16-Cys41. Similarly, the peptide-pair eluting at 31.90 min corresponded to peptide fragments around the disulfide bridge Cys23-Cys46. The oligomer eluting at 36.14 gave three sequences from which the disulfide bridge Cys12-Cys65 was identified, since the remaining peptide of the trimeric sequence corresponded to the pair of Cys16 and Cys41, already identified. Finally, the peptide eluting at 36.62 min gave a dimeric sequence compatible with the disulfide bridge Cys29-Cys45.

Species Specificity

Lethality tests in mice showed that Cn5 is not toxic up to a quantity of 370 μ g per 20 g average body weight. In contrast, in arthropods the injection of Cn5 yielded toxic effects as excitability, paralysis and death. In both crickets

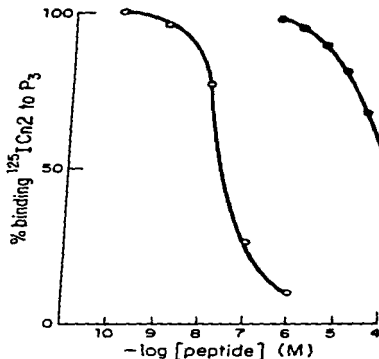


FIG. 3. Displacement of 125 I-Cn2 binding to rat brain synaptosome membranes by Cn5. Rat brain synaptosome membranes (90 μ g/200 μ l per assay) were incubated with 300 pM of 125 I-Cn2 at room temperature for 1 hr, in the presence of increasing concentration of native Cn2 (○) or Cn5 (●), then filtered and counted as described in Materials and Methods. One hundred percent binding was obtained in absence of Cn5. All values are means of triplicates.

and crayfishes, the onset of toxicity lies at doses of less than 20 μ g per g body weight. Nevertheless, applying 20 μ g in crickets we observed a fast recuperation (after 30 min) whereas the toxic effect in crayfishes was stronger and longer lasting (more than 6 hr). The onset of lethality (at least one test animal dead of intoxication) in crickets was detected at 150 μ g per g body weight whereas in crayfishes at considerably lower doses, at 30 μ g per g body weight. In crayfishes a preliminary LD₅₀ was determined using three groups of six test animals applying doses of 25, 30 and 35 μ g per g body weight (results not shown). From this data a LD₅₀ of 28.5 μ g per g body weight was estimated.

Binding and Displacement Assays

In order to obtain additional information on the specificity of toxin Cn5, its binding properties to rat brain synaptosomes were studied using the mammal-specific toxin Cn2 as a comparison. Figure 3 shows the results of binding and

displacement experiments accomplished with 125 I-Cn2 and Cn5. Toxin Cn5 displaced labeled Cn2 with a difference of four orders of magnitude with respect to the displacement by non-labeled Cn2. The 50% inhibition constants IC_{50} were determined to be 16 nM and 100 μ M for Cn2 and Cn5, respectively.

DISCUSSION

In this communication we report the amino acid sequence of an unknown toxin of *Centruroides noxius* Hoffmann which we named Cn5. The amino acid sequence corresponds to a mature peptide derived from a precursor of which the cDNA was reported earlier (40). The determination of the full amino acid sequence by peptide chemistry resulted to be important, because the comparison with the amino acid sequence deduced from the cDNA shows that the toxin is synthesized *in vivo* as a pre-pro-peptide. The precursor includes a signal peptide of 19 amino acids and two additional basic residues in the C-terminal. The processing of the C-terminal was reported earlier for other scorpion toxins from Old-World as well as from New-World scorpions (40-43). The C-terminal of the mature peptide can include a free or an amidated carboxyl group. Amidation takes place when in the precursor peptide a glycine follows the last residue of the resulting mature peptide. The glycine is processed leaving the preceding residue amidated (44). Since the cDNA of Cn5 does not contain a triplet coding for a glycine after the last residue, Ser66, of the mature peptide, we assume that Ser66 is not amidated.

With 66 amino acids Cn5 belongs to the group of long-chain toxins specific for sodium channels. A sequence comparison of toxin Cn5 with other sodium-channel-affecting toxins of the genus *Centruroides*, shown in Fig. 4, reveals that Cn5 shares higher similarity with a consensus sequence of toxins specific for arthropods (panel A) than with a consensus sequence of toxins against mammals (panel B). Cn5 shares typical characteristics of arthropod-specific toxins like the presence of four prolines at positions 52, 56, 59 and 61 in the C-terminal. Moreover, the C-terminal is less hydrophobic in its character than its counterpart in mammalian-specific toxins.

The position of the disulfide bridges, Cys12-Cys65, Cys16-Cys41, Cys25-Cys46 and Cys29-Cys48, corresponds to the typical pattern found in the majority of sodium-channel-binding toxins with the exception of arthropod-specific toxins from old-world scorpions (16). Having a long J-loop (residues 16 to 24) and a short B-loop (residues 42 to 45), Cn5 belongs structurally to the group of β -toxins.

The biological assays support the above observation of Cn5 being an arthropod-specific toxin. It did not show any symptoms of toxicity towards mammals using up to 370 μ g/20 g average body weight mouse. In this respect, it is important to note, that a scorpion of the species *Centruroides noxius* can provide up to 50 μ g venom per single electrically

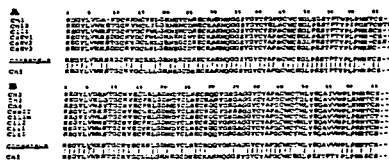


FIG. 4. Sequence comparison of arthropod-specific (A) and mammal-specific (B) toxins of the genus *Centruroides*. Gaps were introduced to maximize similarities. Below each toxin family the consensus sequence is shown which represents the sequence with the most representative amino acid. X stands for a variable residue with no obvious preference. Cn1 (43), Cn2 and Cn3 (43-46), Cn4 (42) and Cn10 (B. Saliako, C. Garcia, B. Becerril, M. Desleppere, L.D. Possani, *in prep.*) are from *Centruroides noxius*, Cn11 from *Centruroides suffusus suffusus* (see review B), CnE1, CnE2 and CnE3 are from *Centruroides sculpturatus* Ewing (34), Cn11 (13), Cn11m (53) and Cn12 (39) from *Centruroides limpidus limpidus*, Cn1 from *Centruroides infamatus infamatus* (39) and Cn11 from *Centruroides limpidus neomatus* (16). The vertical bars indicate identity of the respective residues.

stimulated extraction (8). Thus, the maximum doses we applied per mouse is approximately 10 times higher than the amount of venom produced by a single scorpion in a given time. In contrast to the absence of toxic effects in mice, Cn5 shows toxic effects applying 20 μ g/g average body weight of arthropods (crickets and crayfishes). Therefore, Cn5 can be considered a toxin with high specificity towards arthropods. The comparison of its activity in insects versus crustaceans shows that it is more effective in crustaceans.

In order to investigate the extent and the molecular base of the species specificity of Cn5 further, we compared it with mammal-specific toxins of the genus *Centruroides*, especially with toxin Cn2 of *Centruroides noxius* (45). Cn2 has been characterized, functionally as a β -toxin (39). Therefore, it is expected to bind to site 4 at vertebrate sodium channels. Lethality tests, conducted with Cn2 in mice, showed an LD_{50} of 0.4 μ g/20 g mouse (45) whereas in crustaceans it did not show any toxic effect up to 25 μ g/g body weight (S. Sereno and L.D. Possani, unpublished results).

The displacement of 125 I-Cn2 by Cn5 and Cn2 itself showed that, against our expectation, Cn5 was able to displace Cn2 from its binding sites on rat brain synaptosomes. The high amount of Cn5, necessary to displace Cn2, reflects the differences in species specificity of these toxins. Both toxins belong structurally to the β -toxin class. This indicates that the displacement shown might be due to competition for the same site. Since Cn2 binds to site 4 of the channel, we expect Cn5 to bind to site 4 as well. The low

affinity of Cn5 for the mammalian Na^+ -channel might be the cause of the absence of a toxic response observed in mammals even at high doses. Wheeler et al. (23) using three arthropod-specific toxins: C5Ev1, C5Ev2 and C5Ev3, from *Centruroides sculpturatus* Ewing, in binding and displacement experiments found no competition among the arthropod-specific toxins and those of two mammal-specific toxins: *Androctonus australis* toxin II (an α -toxin), and *Centruroides suffusus suffusus* II (a typical β -toxin). This apparent discrepancy to the results of our study is very likely due to the limited concentration range used by Wheeler et al. (23). While they used concentrations up to 1 μM , we tested concentrations up to 122 μM , which was the estimated IC_{50} . Recently, Gordon et al. (22) proposed the existence of a functional macro-receptor site 3 at mammalian and insect sodium channels, where homologous α -toxins bind to various overlapping receptor sites, causing the physiological effect of α -toxins but with different efficacies. It might well be that something similar is occurring in the case of β -toxins, i.e., that Cn5 binds to a macro-receptor site 4 on the mammal sodium channel, but not exactly to the same site as Cn2. In this case the displacement could be caused by an allosteric effect or by steric hindrance of the two toxins binding to overlapping sites.

The sequence comparison of toxin Cn5 with a consensus sequence of mammal-specific toxins including Cn2 is shown in Fig. 4 (panel B) reveals that the amino acid differences are concentrated mainly in two sequence stretches, between residues 32 to 38 and 49 to 58. We assume that these two stretches are involved in the definition of the sequence specificity. In order to localize the position of residues 32-38 and 49-58 within the three-dimensional structure, we used the structure of Cn5 determined by NMR spectroscopy (F. Lebreton, C. Garcia, L. D. Possani, M. Delepiere, in preparation). A solid-sphere representation of the model is shown in Fig. 5. The side facing the viewer presents the β -sheet structure. Following the proposal of Fontecilla-Camp et al. (12) we will call this side the "face" of the molecule. The α -helix, in turn, is situated on the top-back of the model. Figure 5 demonstrates that the sequence stretches 32-38 and 49-58 (shown in black) form a continuous surface region at the top and left side of the molecule. In terms of secondary structure, residues 32-38 comprise the turn (residues 30-31) between the α -helix and the second β -strand and the first residues of this β -strand. Residues 49-58 are involved in the third (middle) β -strand (residues 49 and 50) and the beginning of the C-terminal (residues 51-58) which does not adopt a particular secondary structure. Lorer et al. (46), after analyzing the secondary structure elements of three toxins that show differences in their specificity, also proposed the C-terminal as being involved in the determination of the species specificity. In this respect, it is interesting to note, that Cn5 as other arthropod-specific toxins (see Fig. 4) contains two prolines at positions 52 and 56. They are expected to confer a certain rigidity to the

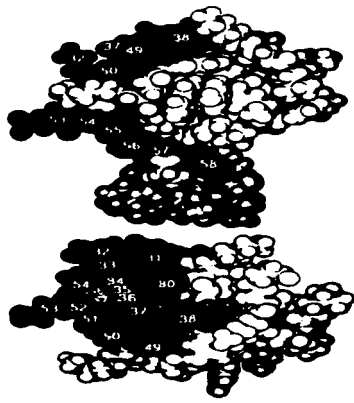


FIG. 5. Region of variable residues (30-38, 49-58, colored in dark grey) identified by comparison of arthropod-specific toxin Cn5 and mammal-specific toxin Cn2. A space-filled presentation of the NMR-structure of Cn5 (F. Lebreton, C. Garcia, L.D. Possani, M. Delepiere, in prep.) is shown presenting the "face" of the molecule in the upper figure and the "top" in the lower figure.

backbone. In contrast, in Cn2 these residues are occupied by Tyr and Leu, respectively.

The possible involvement of the identified surface region in the definition of the species specificity is further supported by its being situated at the side of the toxin that is thought to bind to the channel. For α as well as for β -toxins exists evidence that the flat surface of the toxin, facing the viewer in Figure 5, contains the channel-binding site (47-52). Thereby, special importance is given to the aromatic cluster, formed by residues 4, 42, 42', 47 and 58 in mammal-specific toxins from New-World scorpions. In arthropod-specific toxins the aromatic cluster is extended by Tyr35. The surface region, that we propose here to be specificity-determining, is situated in the vicinity of the aromatic cluster. In fact, residues 38 and 58 are part of it. The maximum distance between residues 32-38 and 49-58, and

the corresponding nearest residue of the aromatic cluster in Cn5 is 14.6 Å. Considering that the dimensions of the binding site of Charybdotoxin, a K⁺-channel-binding toxin of similar structure, are 12 × 15 Å (53), it is reasonable to suppose that all these residues can be situated within or in the immediate vicinity of the binding site and therefore be of influence. However, the distance given is the direct distance between the Cα atoms, it does not consider the "effective" distance described by the surface of the molecule. The visual inspection of the three-dimensional structure provides more information in that respect. It revealed that residues 31, 33 and 34 are oriented rather to the back of the molecule whereas residues 31 to 36 to the left side. The question remains if these particular residues could still influence the binding specificity of the toxins.

In conclusion, our comparative studies indicate that the surface region comprising residues 30-38 and 49-58 might be responsible for the species specificity of a toxin. These residues present therefore interesting candidates for mutation analyses of Cn5. As especially promising candidates we propose Pro52 and Pro56 as well as the residues that are part of residues 38 and 58) or in close vicinity to the aromatic cluster (residues 30, 32, 35, 36, 37, 49, 50, 57).

This work was supported partially by grants from the Howard Hughes Medical Institute (75491-82724) and the European Commission (CIIH-CT 94-22451), L.D.P. from CONACYT-Mexico (IN-4734) and DGAPA-UNAM (IN-225893) to B.B. and a scholarship from the Alexander von Humboldt Foundation, Germany to B.S. We would like to thank Fernando Jimenez, Timoteo Olmeda, Fredy Gironas and Cristian Ballarín for their technical assistance.

References

- Catterall, W. Neurotoxins that act on voltage-sensitive sodium channels in excitable membranes. *Annu. Rev. Pharmacol. Toxicol.* 33: 1-34 (1993).
- Rischar, H.; Bernini, P.; Couraud, F. Scorpion toxins: chemistry and mode of action. In Ceccarelli, B.; Clementi, F. (Eds.), *Advances in Cytopharmacology*, Vol. 38. New York: Raven Press, 1979: 323-334.
- Latzunki, M.; Frelin, G.; Burhanini, J.; Lambert, A.; Meiri, H.; Paron, D.; Romey, G.; Schmid, A.; Schwett, H.; Vigne, P.; Vuyssier, H.P.M. Polypeptide toxins as tools to study voltage-sensitive sodium channels. *Ann. N. Y. Acad. Sci.* 478: 254-272 (1986).
- Carbone, E.; Wänke, E.; Prestipino, G.; Posani, L.D.; Muehleck, A. Selective blockage of voltage-dependent K⁺ channels by a novel scorpion toxin. *Nature* 296: 82-81 (1982).
- Miller, C.; Macdonald, E.; Latorre, R.; Phillips M. Charybdotoxin, a protein inhibitor of single Ca²⁺-activated K⁺ channels from mammalian skeletal muscle. *Nature* 313: 316-318 (1985).
- Urbán, J. A.; Moczy, J. E.; Strichartz, G.R. Purification and characterization of chlorotoxin, a chloride channel ligand from the venom of the scorpion *Centruroides*. *J. Biol. Chem.* 264: C101-C109 (1989).
- Valdivia, H.H.; Kirby, M.S.; Lederer, W.J.; Cossentino, R. Scorpion toxin-treated against the acetylcholine receptor (C12-3)-receptor channel of skeletal and cardiac muscle. *Proc. Natl. Acad. Sci. USA* 89: 12185-12189 (1992).
- Posani, L.D. Structure of scorpion toxins. In *Handbook of Natural Toxins*, Vol. 2. New York: Marcel Dekker, Inc. 1984: 31-38.
- Kobayashi, Y.; Takashima, H.; Tamaki, H.; Kogokua, Y.; Lambert, P.; Kusuda, H.; Chino, N.; Watanabe, T.; Ni; Kimura, T.; Mutsaers, S. M.; Lederer, W. The Coxsone-Stäubliin molecule: A Common Structural Motif of Ion-Channel Blocking Neurotoxic Peptides. *Biopolymers* 31: 1213-1220 (1991).
- Ménez, A.; Bontems, F.; Roumestand, C.; Gilquin, B.; Toma, F. Structural basis for functional diversity of animal toxins. *Proc. Roy. Soc. Edinburg* 99B: 3-13 (1992).
- Bontems, F.; Roumestand, C.; Gilquin, B.; Ménez, A.; Toma, F. Refined structure of Charybdotoxin: conformation of scorpion toxins and insect defensins. *Science* 254: 1521-1523 (1991).
- Fonfellella-Camps, J.C.; Almazo, R.J.; Suddarth, F.L.; Watt, D.D.; Buzsáki, G.E. Three-dimensional structure of a protein from scorpion venom: A new structural class of neurotoxins. *Proc. Natl. Acad. Sci. USA* 77: 6496-6502 (1980).
- Leffert, F.; Delepiere, M.; Ramirez, J.; Baldern, C.; Pavesi, L. D. Primary and NMR three-dimensional structure determination of a novel crustacean toxin from the venom of the scorpion *Centruroides limpulus limpulus* Karsch. *Biochem. J.* 311: 1135-1149 (1994).
- Lee, W.; Moore, C.H.; Watt, D.D.; Rama Krishna, N. Solution structure of the variant-3 neurotoxin from *Centruroides limpulus* Ewing. *Eur. J. Biochem.* 215: 89-95 (1994).
- Huwoet, D.; Habersetter, R.; Schatz, G.; Astner, J.-P.; Fonfellella-Camps J.C. Crystal structure of toxin II from the scorpion *Androctonus australis* Hector refined at 1.3 Å resolution. *J. Mol. Biol.* 218: 88-131 (1994).
- Darbin, H.; Weber, C.; Braun, W. Two-dimensional 1H Nuclear Magnetic Resonance study of AaH II, an antineurotoxin from the scorpion *Androctonus australis* Hector. Sequence resonance assignments and folding of the polypeptide chain. *Biochem. J.* 183: 1843-1941.
- Eitan, M.; Fowler, E.; Herrmann, R.; Duvall, A.; Feltham, M.; Zlotkin, E. A scorpion venom neurotoxin paralytic to insects that affects sodium current inactivation, purification, primary structure and mode of action. *Biochem. J.* 29: 941-947 (1992).
- De Lima, M.E.; Martin, M.F.; Ding, C.R.; Rischat, H. Tityus serrulatus toxin VII bears pharmacological properties of both β-bungarotoxin and insect toxin from scorpion venoms. *Biochem. Biophys. Research Commun.* 139: 236-322 (1986).
- Lorez, E.P.; Martin-Eauclaire, M.-F.; Mansuelle, P.; Sampieri, F.; Girard, C.; Rischat H. An antineurotoxin purified from the scorpion *Androctonus australis* Hector also acts on the α- and β-sites of the mammalian sodium channel: sequence and circular dichroism study. *Biochem. J.* 278: 643-649 (1991).
- Corlano, B.; Martin-Eauclaire, M.; Geisler, S.; Kopyev, E.C.; Khalifa, R.B.; Feltham, M.; Rischat, H. Scorpion toxins affecting sodium current inactivation bind to sodium homologous receptor sites on rat brain and insect sodium channels. *J. Biol. Chem.* 271: 8244-8245 (1996).
- Jover, E.; Couraud, F.; Rischat, H. Two types of scorpion neurotoxins characterized by their binding to two separate receptor sites on rat brain α-scorpion toxins. *Biochem. Biophys. Res. Commun.* 64: 187-184 (1982).
- Catterall, W. A. Structure and function of voltage-gated sodium and calcium channels. *Current Opinon Neurobiol.* 11: 3-9 (1991).
- Wheeler, K.P.; Watt, D.D.; Latzunki, M. Classification of Na channel receptors specific for various scorpion toxins. *Pluggers Arch.* 37: 167-175 (1987).
- Meses, H.; Simard, J.M.; Watt, D.D. Biochemical and electrophysiological characteristics of toxins isolated from the venom

- of the scorpion *Centruroides sculpturatus*. *J. Physiol., Paris* 79: 155-191(1984).
25. Algrain, A.C.; Guzman, H.S.; Martin, B.M.; Ramirez, A.N.; Carbone, E.; Possani, L.D. Isolation and characterization of two toxins from the Mexican scorpion *Centruroides impudicus* Karsch. *Comp. Biochem. Physiol.* 59B:153-161; 1988.
 26. Cestié, S.; Khalifa, R.B.; Pelhate, M.; Rochat, H.; Gordon, D. α -Scorpion toxins binding on rat brain and insect sodium channels reveal divergent allosteric modulations by brevevenin and veratridine. *J. Biol. Chem.* 270:15153-15161; 1995.
 27. Yatani, A.; Kirsch, G.E.; Possani, L.D.; Brown, M. Effects of New World scorpion toxins on single-channel and whole cell cardiac sodium currents. *Am. J. Physiol.* 254:H443-H451; 1988.
 28. Zlotkin, E.; Kadouri, D.; Gordon, D.; Pelhate, M.; Martin, M.F.; Rochat, H. An excitatory and a depressant insect toxin from scorpion venom both affect sodium conductance and possess a common binding site. *Arch. Biochem. Biophys.* 242: 877-887; 1995.
 29. Gordon, D.; Jover, E.; Couraud, F.; Zlotkin, E. The binding of the insect-selective neurotoxin (A1T) from scorpion venom to locust synaptosomal membranes. *Biochem. Biophys. Acta* 778:349-355; 1984.
 30. Zlotkin, E.; Gurevitz, M.; Fowler, E.; Adams M.E. Depressant insect selective neurotoxins from scorpions: venom chemistry, action, and gene cloning. *Arch. Insect Biochem. Physiol.* 22: 55-73; 1993.
 31. Gordon, D.; Moskowitz, H.; Eitan, M.; Warner, C.; Carterall, W.A.; Zlotkin, E. Localization of receptor sites for insect-selective toxins on sodium channels by site-directed antibodies. *Biochem. J.* 31:7622-7625; 1992.
 32. Possani, L.D.; Dent, M.A.R.; Martin, B.M.; Maelicke, A.; Svendsen, I. The amino terminal sequence of several toxins from the venom of the Mexican scorpion *Centruroides noxius* Hoffmann. *Carlberg Res. Commun.* 46:207-214; 1981.
 33. Reifield, R.A.; Lewis, V.J.; Williams, D.E. Disk electrophoresis of basic proteins and peptides on polyacrylamide gels. *Nature* 195:281-285; 1979.
 34. Lammli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685; 1970.
 35. Possani, L.D.; Martin, B.M.; Svendsen, I.; Rede G.S.; Erickson, B.W. Scorpion toxins from *Centruroides noxius* and *Tityus serrulatus*. *Biochem. J.* 229:739-750; 1985.
 36. Martin, B.M.; Carbone, E.; Yatani, A.; Brown, A.M.; Ramirez, A.N.; Gurrula, G.B.; Possani, L.D. Amino acid sequence and physicochemical characterization of toxins from the venom of the scorpion *Centruroides impudicus* Karsch. *Toxicon* 26:785-794; 1988.
 37. Valdivia, H.H.; Martin, B.M.; Ramirez, A.N.; Fletcher, P.L.; Possani, L.D. Isolation and pharmacological characterization of four novel Na⁺-channel-blocking toxins from the scorpion *Centruroides noxius* Hoffmann. *J. Biochem. Med.* 116:1353-1391; 1994.
 38. Possani, L.D.; Martin, B.M.; Fletcher, M.D.; Fletcher, P.L. Discharge effect on pancreatic exocrine secretion produced by toxins purified from *Tityus serrulatus* scorpion venom. *J. Biol. Chem.* 266:3175-3181; 1991.
 39. Dehesa-Dávila, M.; Ramirez, A.N.; Zamudio, F.Z.; Gurrula-Briones, G.; Lévano, A.; Darson, A.; Possani, L.D. Structural and functional comparison of toxins from the venom of the scorpion *Centruroides impudicus* Karsch, *Centruroides impudicus* Karsch and *Centruroides noxius*. 113B:331-339; 1996.
 40. Becerril, B.; Vázquez, A.; García, C.; Concha, M.; Bolívar, F.; Possani, L.D. Cloning and characterization of cDNA that code for Na⁺-channel-blocking toxins of the scorpion *Centruroides noxius* Hoffmann. *Gene* 125:165-171; 1993.
 41. Bougin, F.E.; Rochat, H.; Smith, L.A. Precursors of *Anandrodium australe* Hector. *J. Biol. Chem.* 264:19259-19265; 1989.
 42. Vázquez, A.; Becerril, B.; Martin, B.M.; Zamudio, F.; Bolívar, F.; Possani, L.D. Primary structure determination and cloning of the cDNA encoding toxin 4 of the scorpion *Centruroides noxius* Hoffmann. *FEBS Lett.* 320:43-46; 1993.
 43. Vázquez, A.; Tapia, J.V.; Elmson, W.E.; Martin, B.M.; LeFevre, F.; Desjardet, M.; Possani, L.D.; Becerril, B. Cloning and characterization of the cDNAs encoding Na⁺-channel-specific toxins 1 and 2 of the scorpion *Centruroides noxius* Hoffmann. *Toxicon* 33:1163-1172; 1995.
 44. Creighton, T.E. *Proteins: Structures and Molecular Properties*, 2nd edition. New York: W. H. Freeman; 1993:90.
 45. Zamudio, F.; Saavedra, R.; Martin, B.M.; Gurrula-Briones, G.; Heron, P.; Possani, L.D. Amino acid sequence and immunological characterization with monoclonal antibodies of two toxins from the venom of the scorpion *Centruroides noxius* Hoffmann. *Eur. J. Biochem.* 224:281-292; 1992.
 46. Lorez, E.P.; Sampieri, F.; Roussel, A.; Granier, C.; Rochat, H. Conformational flexibility of a scorpion toxin active on mammals and insects: a circular dichroism study. *Protein Sci.* 1:64-172; 1992.
 47. Fontegilletto-Camps, J.C.; Almsay, R.J.; Suddath, F.L.; Bugz C.E. The three-dimensional structure of scorpion neurotoxins. *Toxicon* 22:1-7; 1982.
 48. Kharrat, R.; Darbon, H.; Rochat, H.; Granier, C. Structure/activity relationships of scorpion α -toxins. *Eur. J. Biochem.* 181:351-392; 1989.
 49. Kharrat, R.; Darbon, H.; Granier, C.; Rochat H. Structure-activity relationships of scorpion neurotoxins: contributions of arginine residues. *Toxicon* 28:329-323; 1990.
 50. El-Aveli, M.; Darbon, H.; Bahroun, E.; Vargas, C.; Rochat, H. Differential effects of defined chemical modifications on antigenic and pharmacological activities of scorpion alpha and beta-toxins. *Eur. J. Biochem.* 155:289-294; 1986.
 51. Jover, E.; Bahro, J.; Couraud, F. Binding of β -toxin: a physicochemical study. *Biochem. J.* 231:147-152; 1984.
 52. Darbon, H.; Angelides, K.J. Structural mapping of the voltage-dependent sodium channel. *J. Biol. Chem.* 259:6674-6684; 1984.
 53. Stämpfli, P.; Kukulakova-Parkensky, L.; Miller, C. Intimations of K⁺ channel structure from a complete functional map of the molecular surface of charybdotoxin. *Biochem. J.* 333:443-452; 1994.
 54. Babin, D.R.; Watt, D.D.; Gow, S.M.; Mlejnek, R.V. Amino acid sequences of neurotoxic protein variants from the venom of *Centruroides sculpturatus* Ewing. *Arch. Biochem. Biophys.* 168:694-726; 1974.
 55. Ramirez, A.N.; Martin, B.M.; Gurrula, G.B.; Possani, L.D. Isolation and characterization of a novel toxin from the venom of the scorpion *Centruroides impudicus* Karsch. *Toxicon* 32:479-492; 1994.

DISCUSION Y CONCLUSIONES

Con el objeto de realizar un estudio sobre la relación estructura-función de las toxinas de alacrán, iniciamos la caracterización de la toxina Cn 5 del alacrán *Centruroides noxius* Hoffmann. Algunos de los elementos que nos ofrecía esta toxina como modelo en este análisis fueron los siguientes: a) Se conocía la estructura tridimensional de una toxina muy semejante (Fontecilla-Camps, 1980), con la cual presenta una identidad de 88%. b) Contábamos con el gene *Cngt* II, que codifica para Cn 5 (Becerril 1993). c) La toxina presenta algunas características que posiblemente nos permitirían, mediante mutaciones puntuales, analizar la participación de la región carboxilo en la función.

Iniciamos el trabajo con la purificación de la toxina nativa para así poder caracterizarla. Esto con el objetivo de conocer el funcionamiento de la toxina y para tener parámetros que permitan identificar y comparar los productos de la expresión, tanto de la toxina recombinante como de las mutantes.

SECUENCIA PRIMARIA DE LA TOXINA Cn 5.

Los primeros datos que obtuvimos fueron respecto a la secuencia primaria, la cual corresponde con la secuencia de aminoácidos que se deriva del gene *Cngt* II. Este gene codifica para una pre-proteína de 87 aminoácidos, la proteína madura que se encuentra en el veneno ha sido procesada y únicamente presenta 66 aminoácidos. Este tipo de procesamiento es similar al de otros péptidos neuroactivos y ya ha sido reportado para toxinas de alacrán (Bougis 1988, Vázquez 1993, Becerril 1993). Para las hormonas se han descrito varios tipos de estas modificaciones postraduccionales. En las toxinas de alacrán se llevan a cabo tres de estas modificaciones: la primera consiste en la eliminación del péptido señal que se lleva a cabo mediante una peptidasa localizada en la membrana del retículo endoplásmico rugoso (Karp, 1990). La segunda consiste en la remoción de los aminoácidos básicos que se encuentran al final del extremo carboxilo; por lo general, en las toxinas se presentan uno o dos aminoácidos básicos. Para este corte proteolítico se han descrito diferentes carboxipeptidasas, las cuales pueden remover desde uno hasta cuatro residuos básicos (Creighton 1993). Estos dos pasos de maduración se han descrito para todas las toxinas de alacrán cuya secuencia nucleotídica se conoce. La tercera modificación se presenta sólo en

algunas toxinas y consiste en la adición de un grupo amida al extremo carboxilo de la proteína. En este proceso participan dos enzimas y es necesaria la presencia de una glicina, la cual al ser procesada actúa como donador del grupo amida (Creighton, 1993). Es por esto que en los ARN mensajeros que codifican para las toxinas se encuentra la información para el péptido señal, la secuencia que codifica para la proteína y los codones para uno o dos aminoácidos básicos, los cuales en ocasiones son precedidos por una glicina (Fig.2). Esta glicina participa en la amidación de la proteína, lo cual se piensa que afecta la estabilidad biológica e incluso que en algunos casos es indispensable para la actividad biológica. Todavía no se conoce el significado de este procesamiento y se ha reportado una toxina contra canales de potasio en la que la toxina recombinante no requiere la amidación para su actividad, aún cuando la toxina nativa si esta amidada (Martínez, 1996). En la secuencia de Cn 5 no se encontró esta glicina por lo que no esperábamos un péptido amidado.

Se determinó la posición de los 4 puentes disulfuros, la cual corresponde a las posiciones típicas en la mayoría de las toxinas de alacrán. Esto era de esperarse, por que hasta el momento todas las toxinas de cadena larga de alacranes americanos han presentado el patrón consenso de los puentes disulfuro. Las únicas toxinas con un puente en posición atípica que se conocen son las toxinas excitatorias contra insectos de alacranes del Viejo Mundo (Tabla 2).

ACTIVIDAD BIOLÓGICA Y SU RELACION CON LA ESTRUCTURA.

Para hacer un análisis de la relación estructura-función es necesario definir esta función. Por lo general se utilizan, para determinar esta función alguno de los siguientes tipos de ensayo. Ensayos de toxicidad, utilizando diferentes modelos biológicos lo cual nos permite determinar especificidad y grado de toxicidad de la molécula (determinación de LD₅₀). Ensayos de unión, utilizando la toxina marcada radioactivamente para determinar afinidad y especificidad hacia ciertos sitios. Diferentes ensayos fisiológicos, en los cuales se puede medir el efecto funcional de la toxina sobre su receptor. A pesar de la información que estos ensayos nos dan presentan algunas limitantes que deben considerarse al analizar la función de estas moléculas. Primero, el modelo biológico que se utiliza para determinar la actividad y especificidad de estas neurotoxinas el cual en muchas ocasiones es un organismo taxonómicamente muy alejado de su blanco natural. Por ejemplo en el estudio de toxinas de caracoles marinos

(predadores de peces), arañas (predadores de insectos) y de alacranes (predadores de artrópodos) se han utilizado membranas de mamíferos para determinar la actividad y especificidad de sus toxinas. Segundo, en los ensayos de unión es común asumir que sí existe una alta afinidad por el receptor existe un efecto funcional. En la revisión sobre neurotoxinas de Adams y Olivera (1994) mencionan varios ejemplos donde se muestra que no siempre existe una relación directa entre afinidad por un receptor y efecto funcional. Los autores mencionan como un hecho biológico fundamental que todas las toxinas naturales han evolucionado para funcionar en un contexto biológico específico. Que mientras la toxina cumpla los requerimientos del organismo que la produce, todas aquellas características que no afecten la eficiencia de la toxina pueden, en principio, variar libremente. Por esto pensamos que para hacer un análisis completo de la función de una toxina e intentar relacionarla con su estructura es necesario basarse en los resultados de los ensayos mencionados considerando siempre contexto biológico donde actúa dicha molécula.

Los ensayos de toxicidad con la toxina Cn 5 mostraron que ésta actúa sobre insectos y crustáceos pero no sobre mamíferos (Tabla 1, artículo 2 del anexo). Se observó mayor toxicidad hacia crustáceos por lo que se determinó la dosis letal media en dos organismos pertenecientes a este grupo. En *Armadillidium vulgare* (cochinillas), en el cual presento una LD₅₀ un poco mayor que en *Cambarellus montezumae* spp. (acociles). La cochinilla representa un modelo mas conveniente para este trabajo ya que su pequeño tamaño nos permite utilizar cantidades menores de toxina, en comparación a los acociles. Esto es una ventaja importante sobre todo para las toxinas recombinantes. Además, se hicieron ensayos de desplazamiento de la toxina ¹²⁵I-Cn 2, que es específica a mamíferos, con la Cn 5 en sinaptosomas de cerebro de rata. En contra de lo esperado se observó que la toxina Cn 5 es capaz de desplazar a la Cn 2 de mamíferos (Fig. 3 del artículo 3). Estructuralmente, las dos toxinas se pueden clasificar como toxinas tipo β (Meves, 1984) por lo que suponemos que ambas están reconociendo el mismo sitio de unión en el canal. Para la toxina Cn 2 se ha demostrado que en un canal de sodio dependiente del voltaje su actividad corresponde a una toxina tipo β (Dehesa-Dávila, 1996). Es de esperarse que la toxina Cn 2, por ser una toxina tipo β , se una al sitio 4 del canal de sodio dependiente de voltaje y que la toxina Cn 5 este uniéndose al mismo sitio. Sin embargo, con los experimentos hechos

no podemos determinar el tipo de inhibición que existe y así identificar el sitio de unión de la toxina Cn 5. La alta concentración de toxina Cn 5 necesaria para desplazar a la toxina Cn 2 podría estar relacionada con la ausencia de actividad tóxica en mamíferos. En los ensayos de actividad se demostró que la toxina Cn 5, aún en altas concentraciones, no presenta actividad tóxica en mamíferos (ratón) y sí en artrópodos. Otra posibilidad es que la ausencia de efectos que se observa en mamíferos sea debido a que el canal al cual se une en cerebro sea particular de ese tejido y no se encuentre en las regiones corporales a las cuales tiene acceso la toxina una vez que es inyectada en músculo. Por último esto puede ser un indicio de que, aún cuando la toxina se une al canal, no tiene ningún efecto funcional sobre el mismo.

En el análisis de la relación estructura-función de la toxina Cn 5 hicimos dos comparaciones de secuencias primarias de toxinas de cadena larga de alacranes del género *Centruroides*. La primera comparación la hicimos con toxinas contra artrópodos y en la segunda toxinas contra mamíferos. En la comparación del grupo de toxinas con especificidad hacia artrópodos observamos que los residuos que varían con mayor frecuencia se presentan hacia el lado derecho de la molécula, en este caso dichas posiciones no presentan una continuidad clara (Fig 4, panel A, del artículo). Se pueden observar regiones donde se presentan con mayor frecuencia estos cambios como el inicio de la hélice α (aminoácidos 17 a 21) o el asa entre la segunda y tercera hebras β (aminoácidos 42 a 45). Diferencias entre toxinas de un mismo grupo podrían indicarnos sitios o regiones posiblemente involucrados en el grado de toxicidad. En la comparación con toxinas contra mamíferos observamos que la variación de aminoácidos se agrupa principalmente en dos regiones de la molécula, del residuo 30 al 38 y del 49 al 58. Pensamos que estas regiones podrían estar involucradas en la especificidad. Al analizar sus posiciones en la estructura tridimensional (Fig. 5, del artículo) se observó que los cambios se presentan dentro de una región continua en el lado izquierdo de las caras frontal y superior de la molécula. Estos datos apoyan la hipótesis de la participación de las prolínas, principalmente las 52 y 56, en la especificidad de las toxinas. Tomando en cuenta que en la secuencia consenso de toxinas contra mamíferos se presentan tirosina y valina, en las posiciones 52 y 56 respectivamente, se considera importante probar una mutante de Cn 5 con tirosina y valina en dichas posiciones. Con estos resultados podemos diseñar algunas mutantes que nos permitan analizar la

relación estructura-función. Para evaluar el efecto de estas mutaciones será necesario, además de los ensayos de toxicidad y unión a membranas, contar con un ensayo que nos permita medir el efecto directo al receptor.

EXPRESION.

Para la expresión del gene *Cngt* II decidimos ensayar en un sistema procarionte: *Escherichia coli*. Las razones de elegir este sistema fueron dos. Primero, los datos que encontramos en la literatura nos indicaban que la expresión de toxinas de cadena larga recombinantes ha sido problemática: sin importar el sistema empleado la producción de toxina ha sido muy pobre. Se podría pensar que por ser una proteína eucariote el sistema ideal sería uno eucariote; sin embargo, esto no ha sido así. En los sistemas eucariotes que se han probado la producción de toxina activa ha sido muy baja (Bougis 1989, Dee 1990, Maeda 1991, Martin-Eauclaire 1994). Segundo, la expresión de genes heterólogos en bacteria es hasta el momento el sistema más sencillo y económico para la producción de proteínas con fines de investigación o comercialización (Georgiou, 1996). Las desventajas que con frecuencia se presentan al expresar en *E. coli* son las siguientes: 1) que la proteína no presente la estructura nativa, 2) que sea reconocida por la maquinaria proteolítica de la célula y; 3) que en este sistema no se pueden realizar modificaciones postraduccionales de sistemas eucariotes. Para este proyecto las modificaciones postraduccionales, no representan un problema pues para la toxina Cn 5 ya se conocen cuáles son estas modificaciones y se pueden obtener los mismos resultados con la edición del gene mediante PCR. El unico paso de maduración de las toxinas de alacrán que requiere un sistema eucariote es la amidación que para esta toxina no se presenta y por lo tanto no requerimos de un sistema eucariote. Las otras dos desventajas no se pueden predecir y es necesario experimentar con cada proteína que se planea expresar. Además, cada día surgen nuevas estrategias que permiten minimizar estas desventajas y así obtener una proteína activa de cultivos en *E. coli* (Georgiou 1996, Makrides 1996). Por ejemplo, para obtener proteínas cuya estructuración requiere de la formación de puentes disulfuro, se ha optado por dirigir la proteína expresada al espacio periplásmico. En este compartimiento existe un ambiente oxidante, se encuentran proteínas que participan en el plegamiento y la concentración de proteínas es mucho menor que en citoplasma (Georgiou, 1996). Todas estas condiciones favorecen el plegamiento de proteínas con puentes

disulfuro. Para el transporte de proteínas a través de la membrana normalmente es necesario la presencia de una región de la secuencia que actúe como señal de secreción. Este péptido señal consiste en una región de la secuencia, la cual no se encuentra presente en la proteína madura y que sirve para iniciar el transporte a través de las membranas citoplásmica en procariontes o la del retículo endoplásmico en eucariotes (Von Heijne, 1986). Ya se han descrito las secuencias para varios péptidos señal de *E. coli*, lo cual permite la construcción de los vectores necesarios para la expresión en periplasma.

En el citoplasma el ambiente reductor favorece la agregación de las proteínas que presentan puentes disulfuro (Georgiou 1996, Makrides 1996). Estas proteínas, al no presentar la estructura nativa, exponen superficies hidrofóbicas que favorecen la agregación intermolecular. Esta agregación de moléculas es la base para la formación de cuerpos de inclusión que son agregados de alta densidad de la proteína expresada.

Otra estrategia común para la expresión citoplásmica es expresar la proteína de interés unida a una proteína bacteriana (Makrides, 1996). De esta forma la proteína queda protegida de la degradación por proteasas. Adicionalmente en muchos casos la proteína acarreadora favorece la solubilidad y facilita la purificación de la proteína de interés. Es común diseñar un sistema de separación mediante cromatografía de afinidad para la proteína acarreadora.

Para este proyecto utilizamos tres vectores, los cuales nos ofrecen diferentes características para expresar proteínas heterólogas en bacterias. Decidimos ensayar diferentes estrategias de expresión, esperando encontrar una que nos diera la cantidad de proteína adecuada para desarrollar el proyecto. Esto podría ser producción de proteína activa o inactiva. En el caso de proteína inactiva, ésta podría ser recuperada mediante un proceso de plegamiento *in vitro*.

La primer construcción nos permitió expresar a Cn 5 en citoplasma como péptido de fusión, unida a una proteína acarreadora que es propia de la bacteria (mbp). Esta fusión ayudó a evitar que la proteína fuera reconocida por las proteasas de la célula. Con este sistema no esperábamos, por estar en citoplasma, que la toxina presentara la estructura nativa ya que presenta 4 puentes disulfuro. Además, la proteína acarreadora podría estar interfiriendo en el plegamiento. Sin embargo, esperábamos tener una cantidad adecuada como para permitirnos establecer un protocolo de plegamiento *in vitro*. En este sistema, además del sitio de corte del factor

Xa que nos permite liberar la toxina del péptido acarreador, se introdujo un codón para metionina lo cual nos permite hacer un corte con CNBr y así obtener una toxina recombinante que presente el extremo amino de la toxina nativa. Con este sistema obtuvimos un alto rendimiento (aproximadamente 20% de la proteína total) lo cual sugiere que el reconocimiento por proteasas fue muy poco gracias a la presencia de la proteína acarreadora. También este producto se mantuvo en forma soluble gracias a la proteína acarreadora, pues aún cuando la concentración fue muy alta no se presentaron problemas de agregación de la proteína. El corte con factor Xa nos dio únicamente dos péptidos del tamaño esperado. Sin embargo, una separación eficiente de estos productos no fue posible debido a la formación de polímeros de la toxina. El corte con bromuro de cianógeno resultó poco conveniente debido a la presencia de ocho metioninas en la proteína acarreadora.

La segunda construcción dirige la proteína hacia el espacio periplásmico, donde es posible la formación de los puentes disulfuro y por lo tanto la correcta estructuración. En esta construcción utilizamos el péptido señal propio de la toxina, el cual presenta todas las regiones necesarias para actuar como señal de exportación. El sitio de corte cumple con los requerimientos necesarios para ser reconocido por una aminopeptidasa (Von Heijne, 1986), por lo que esperábamos que el producto se dirigiera a periplasma y presentara el extremo amino y la estructura de la toxina nativa. Este sistema lo consideramos muy ventajoso, pues en caso de que el péptido señal fuera reconocido nos permitiría obtener una toxina con la estructura nativa, sin necesidad de efectuar algún tipo de corte para liberar el extremo amino o proceso de plegamiento. El principal inconveniente de este sistema fue que la secuencia entera es extraña a la bacteria, por lo que existe mayor posibilidad de que sea sustrato de la maquinaria proteolítica de la célula. Otro inconveniente de este sistema fue no tener un sistema de purificación eficiente. Nuestros resultados indican que la expresión fue muy poco eficiente o el reconocimiento por proteasas muy alto ya que sólo pudimos confirmar la expresión de la toxina mediante ensayos de ELISA. Otra posibilidad es que la proteína esté sufriendo un fenómeno de agregación y por lo tanto no la encontramos en el material soluble. En este caso, sería recomendable recuperar el material agregado, sólo si éste se encuentra en periplasma, pues si está en citoplasma, la toxina estaría unida al péptido señal y no contamos con un protocolo de corte para liberar a la toxina recombinante.

La tercera construcción es también dirigida al espacio periplásmico, pero el péptido señal es propio de la bacteria (Omp A) lo cual asegura que éste sea reconocido por la célula. Además de exportar la proteína a periplasma, el péptido señal tiene la ventaja de ser bacteriano por lo que el reconocimiento por proteasas deberá ser menor. Además, en esta construcción se adiciona a la proteína una cola de histidinas, la cual presenta afinidad a metales y facilita así su aislamiento. Para liberar a la proteína de la región de histidinas se introdujo una metionina que nos permitiera hacer un corte con bromuro de cianógeno y así obtener una toxina recombinante que presente el extremo amino nativo. Otra ventaja de este sistema es que por tener un péptido señal bacteriana el reconocimiento por las proteasas deberá ser menor que en el segundo sistema. Los resultados obtenidos parecen indicar que así fue pues con este sistema logramos obtener un rendimiento de aproximadamente 2% de la proteína total. En este caso se presentó un problema de purificación, pues sólo fue posible hacerlo en condiciones desnaturalizantes. Esto puede indicarnos que la proteína esta formando agregados, ya sea en citoplasma o periplasma, por lo que las histidinas no se encuentran expuestas. Una vez que se desnaturaliza la proteína o el agregado, las histidinas quedan expuestas y es posible aislar el péptido mediante cromatografía de afinidad. Este sistema nos permitió obtener el material suficiente para realizar ensayos de plegamiento *in vitro* e iniciar la caracterización de la toxina recombinante. Consideramos que este sistema fue exitoso ya que tuvo: primero, un nivel de expresión alto y segundo, en éste no se dio un proceso, cuantitativamente importante, de degradación por proteasas de la proteína recombinante. Además, la purificación mediante cromatografía de afinidad es una ventaja muy importante.

En los tres casos logramos expresar la toxina recombinante, con niveles de producción muy diferentes en cada caso, sin embargo, en los tres casos podríamos implementar estrategias que lleven a mejorar el nivel de producción o el nivel de proteína activa.

Para la expresión en citoplasma el nivel de producción se puede elevar considerablemente mediante cultivos de alta densidad. Estos cultivos permiten obtener más de 100g de células/l (peso húmedo). Sin embargo, esta estrategia no nos aumentaría directamente el rendimiento de proteína activa y el rendimiento actual es suficiente para nuestros

requerimientos.

En el caso de la expresión periplásmica se pueden implementar estrategias para obtener la proteína activa tanto a nivel de fermentación o mediante técnicas de ADN recombinante (Georgiou 1996, Mackrides 1996). Para la fermentación se pueden modificar las condiciones de temperatura y composición del medio, a condiciones que favorezcan el plegamiento y reduzcan la agregación de proteínas en periplasma. Bajar la temperatura durante el periodo de crecimiento del cultivo, sobre todo después de la inducción, es una técnica que ha sido ampliamente usada para la obtención de proteínas activas. Otra opción para evitar la agregación y obtener una proteína activa es bajar el nivel de síntesis de proteínas, ya sea utilizando un promotor débil o reduciendo las condiciones de inducción. Una de las estrategias más sencillas para obtener mejores rendimientos de proteínas activas, es modificar el medio de cultivo mediante la adición de reactivos que favorecen directa o indirectamente el plegamiento de proteínas (Georgiou, 1996). Como ejemplo, Blackwell et al., (1991) realizaron un trabajo en el cual logran incrementar la producción de proteína activa cuatrocientas veces. Esto lo llevan a cabo manteniendo el cultivo bajo estrés osmótico, producido por la adición de sorbitol directamente al medio de cultivo. Otros aditivos que favorecen la expresión de proteínas solubles en periplasma incluyen al etanol (induce la expresión de proteínas de choque térmico), tioles y disulfuros de bajo peso molecular (modifican el potencial redox del espacio periplásmico) (Georgiou 1996, Makrides 1996). Para la segunda estrategia las técnicas de ADN recombinante nos permiten modificar el transporte y el plegamiento de las proteínas, pues es posible co-exresar proteínas que participan en estos procesos (Dueñas, 1994). Por ejemplo, algunas proteínas que serían útiles en la expresión de esta toxina son: proteína disulfuro isomerasa, proteína proliil isomerasa y chaperonas del tipo Gro EL (Gething 1992, Saibil 1994). La primera comprende varias enzimas, tanto procariontes como eucariotes, que participan en el plegamiento de proteínas con puentes disulfuro (Freedman 1994, Derman 1993). Estas proteínas catalizan las reacciones de intercambio tiol-disulfuro y, dependiendo del péptido sustrato y del potencial redox, promueven la formación, isomerización o reducción de los disulfuros. Estas proteínas no determinan las vías de plegamiento de las proteínas, más bien facilitan la formación del conjunto correcto de disulfuros mediante un rápido rearrreglo de los puentes disulfuros incorrectos (Fischer, 1990). Las

proteínas que presentan actividad de peptidil prolil *cis-trans* isomerasas catalizan la reacción de isomerización del enlace peptídico que precede a algunas prolinas. Se ha visto que en ausencia de dicha proteína la isomerización constituye una barrera cinética importante que impide la rápida estructuración de la proteína cuando la posición del enlace es en *cis* (Brandts 1975, Texter 1992, Weissman 1995). Otro grupo de proteínas, a las que ahora se conoce con el nombre colectivo de chaperonas, funcionan *in vivo* no como catalizadores de la formación de la estructura sino, en el reconocimiento y estabilización de los intermediarios durante el plegamiento de los péptidos (Gething 1992, Saibil 1994). De esta forma evitan la formación de estructuras aberrantes y dirigen los péptidos hacia vías de plegamiento biológicamente productivas. Las chaperonas incluso se han descrito como cajas o compartimentos que actúan separando a la proteína del ambiente y ofreciendo así un espacio adecuado para la correcta estructuración (Ellis, 1996).

PLEGAMIENTO *in vitro*.

Hasta el momento se tiene un conocimiento "primitivo" sobre el mecanismo mediante el cual las proteínas llegan a su conformación nativa dentro de la célula, pero aún en caso de lograr decifrarlo, es difícil pensar que condiciones idénticas se puedan aplicar para en procesos de plegamiento *in vitro*. Frecuentemente la eficiencia del plegamiento *in vitro* es baja en comparación con el plegamiento *in vivo*, y comunmente se requiere emplear concentraciones de proteína y condiciones físico-químicas muy diferentes a las que ocurren en la célula (Gething, 1992). En general para obtener una proteína biológicamente activa, se requiere que su cadena peptídica presente una conformación tridimensional específica. Esta estructura nativa está determinada por su secuencia primaria; pero únicamente bajo ciertas condiciones físicas del medio. Un aumento en las características caotrópicas del solvente favorece la desnaturalización de la proteína, la cual puede volver a su estado nativo una vez que las condiciones del medio sean adecuadas. Cada proteína presenta características propias de desnaturalización y renaturalización por lo que no existe una técnica de renaturalización común para todas las proteínas. Se considera que las variables más importantes para encontrar las condiciones de plegamiento de una proteína son: concentración, pH, temperatura, fuerza iónica y agentes caotrópicos (Rainer, 1990). Además, para proteínas con puentes disulfuro un factor que influye

sustancialmente tanto en la tasa como en la vía de plegamiento, es el potencial redox de la solución de renaturalización. El agente redox debe presentarse tanto en su forma oxidada como reducida para permitir el rearreglo de los puentes disulfuro incorrectos, los cuales se generan rápidamente al inicio de la reoxidación (Fischer, 1990).

En este trabajo ensayamos varios protocolos de plegamiento de la proteína recombinante en los cuales hicimos modificaciones de uno o varios de estos parámetros. Sin embargo, en la mayoría de los ensayos el porcentaje de renaturalización fue muy bajo. En los ensayos de plegamiento en los cuales utilizamos la toxina expresada como péptido de fusión, consideramos que además de los problemas intrínsecos del plegamiento de la toxina, la presencia del péptido acarreador (mbp) puede ser un problema adicional importante. Es posible que el péptido acarreador influya en el proceso de plegamiento evitando la estructuración y favoreciendo la agregación. Sin embargo, los ensayos con el péptido liberado, con factor Xa o con bromuro de cianógeno, el incremento en la renaturalización no fue significativo. Consideramos que la presencia de otros péptidos en la mezcla de reacción puede estar influenciando de forma importante el proceso de renaturalización, especialmente el corte con bromuro de cianógeno en el cual se obtienen varios péptidos, ya que la proteína acarreadora presenta ocho metioninas.

Al presentarse estos problemas en el plegamiento *in vitro* con las toxinas recombinantes decidimos determinar las condiciones de plegamiento para la toxina nativa y después continuar con la toxina recombinante. Pensamos que era importante determinar estas condiciones para Cn 5; primero, por que así podríamos descartar que es un problema intrínseco de la toxina recombinante y segundo por que era necesario tener dichas condiciones como marco de referencia para los ensayos con la toxina recombinante. Encontramos que, aún para la toxina nativa fue difícil definir las condiciones que nos permitieran obtener un rendimiento alto en el proceso de renaturalización. No obstante, en todos los ensayos de plegamiento utilizando la toxina nativa logramos obtener proteína renaturalizada. Pensamos que el principal factor fue la pureza del péptido que al no estar contaminado con otros péptidos la reacción de plegamiento se favorece. Con la toxina nativa encontramos condiciones que nos permitieron recuperar aproximadamente 30 % del material y este fue el rendimiento más alto que logramos obtener.

Los datos de plegamiento *in vitro* nos indican que para esta proteína la

renaturalización no es un proceso fácilmente reversible. Suponemos que esta baja recuperación se está dando por que existe competencia entre el proceso de plegamiento y la agregación de moléculas desnaturalizadas o parcialmente renaturalizadas. Durante el plegamiento *in vitro* el fenómeno de agregación es considerablemente importante ya que los intermediarios cinéticos presentan una alta tendencia de agregación transitoria debida a que exponen gran parte de sus regiones hidrofóbicas. Fischer & Schmid (1990) explican que este fenómeno se presenta porque no se han encontrado las condiciones adecuadas de la solución de plegamiento o que puede reflejar un "problema genuino" del plegamiento *in vitro*. Mencionan que un problema verdadero de plegamiento *in vitro* se presenta cuando el proceso incluye barreras cinéticas producidas por modificaciones covalentes, como la formación de puentes disulfuro o por la isomerización *cis-trans* del enlace peptídico en el caso de las *cis*-prolinas presentes en la proteína nativa.

La toxina Cn5, además de ocho cisteínas que forman cuatro puentes disulfuro, presenta cuatro prolinas de las cuales tres participan en enlaces *trans* y una, la prolina 59, en *cis*. Estos datos fueron aportados por un estudio preliminar de NMR a la toxina Cn 5 (datos no publicado) y por comparación con los datos estructurales de toxinas similares; cristalografía de rayos-x de la toxina CsEv3 (Almassy, 1983) y NMR de la toxina CII 1 (Lebreton, 1994). Estas tres toxinas homólogas presentan una alta similitud en secuencia primaria.

El proceso de plegamiento para proteínas que no presentan puentes disulfuro puede realizarse en cuestión de milisegundos, en contraste con el plegamiento *in vitro* de toxinas que si los presentan, el cual puede tomar horas e incluso días (Bardwell, 1994). Para el caso de proteínas con puentes disulfuro múltiples el número de combinaciones posibles aumenta considerablemente. Por ejemplo, para la toxina Cn 5 que presenta cuatro puentes disulfuro existen 105 posibles (Jaenick and Rudolph, 1990) y sólo una de ellas representa el arreglo nativo. Además, para la formación completa del juego de puentes disulfuro es necesario que los grupos sulfhidrilos correspondientes se mantengan accesibles. Esto es, que las moléculas parcialmente plegadas y reoxidadas presenten una estructura flexible, que permita el acceso al grupo sulfhidrilo correspondiente o al agente redox. Un ejemplo ilustrativo de este problema es el plegamiento *in vitro* del fragmento Fab de un anticuerpo recombinante (Lilje, 1994). Este

fragmento presenta un plegamiento exageradamente lento y la recuperación de proteína nativa en la mayoría de los casos es menor al 10%. Al adicionar una enzima disulfuro isomerasa (PDI) a la reacción de plegamiento se obtiene un gran incremento en la recuperación de proteína nativa, pero únicamente cuando la PDI se adiciona al inicio del plegamiento. Esto demuestra que la enzima se requiere en los estados tempranos del plegamiento para favorecer las formas productivas de las moléculas desnaturalizadas y de los intermediarios tempranos. La ARNasa T1 es otro ejemplo donde se ha detectado la falta de flexibilidad molecular debido a que la formación prematura de los puentes disulfuro nativos reduce la tasa de plegamiento *in vitro* (Weissman, 1995). La tasa de plegamiento para proteínas varía considerablemente y además una misma proteína puede presentar fases rápida y lenta de renaturalización. La presencia de dichas fases cinéticas en el plegamiento de un proteína se explica en algunos casos, mediante la hipótesis de la prolina propuesta por Brandts et al., en 1975. Su planteamiento asume que la fase lenta es debida al isomerismo *cis-trans* de las prolinas en el estado desnaturalizado. Esto es, que sólo aquella fracción de las moléculas desnaturalizadas que presente el arreglo de isómeros idénticos a la conformación nativa pueden llegar directamente al estado nativo. El resto de las moléculas deberán obtener primero el arreglo correcto de isómeros antes de pasar al estado nativo. Pero no todas las prolinas de un péptido contribuyen a la cinética de plegamiento de la misma forma. Las prolinas, cuyo enlace peptídico presenta la conformación *cis* en el estado nativo tienen un fuerte efecto en la cinética de plegamiento porque la configuración *trans*, no nativa, predomina en la proteína desnaturalizada (Schindler, 1996). Aún en el estado desnaturalizado la tasa de isomerización del enlace peptídico *cis-trans* es, por lo general, más lenta que la tasa de plegamiento (Weissman, 1995). Aunque se han reportado varias proteínas con actividad catalizadora de esta isomerización, la eficiencia de catálisis varía dependiendo de la proteína sustrato. Incluso se ha propuesto que esta reacción pudiera ser usada como mecanismo de regulación, pues es posible que la isomerización de ciertas prolinas prevenga el plegamiento de un péptido hasta que éste encuentre la isomerasa propia, la cual catalizaría esta reacción en el tiempo y espacio adecuado (Kyte, 1995). Para la toxina Cn 5 esto pudiera suceder ya que el veneno es secretado en vesículas (Stahnke, 1978) en las que, además de la toxina, se podrían estar empacando las enzimas del procesamiento y las

proteínas plegadoras. La regulación del procesamiento en gránulos secretorios, donde se empaca la proteína junto con las enzimas necesarias, se ha reportado para otros péptidos neuroactivos (Sossin, 1991). Esto, además de propiciar un espacio adecuado para el plegamiento y procesamiento de la toxina, podría proteger al organismo de la actividad de la proteína madura. Otro aspecto que influye en la tasa de isomerización es el efecto estérico de las cadenas laterales de los aminoácidos vecinos. La isomerización se hace más lenta conforme aumenta el tamaño de la cadena lateral del aminoácido inmediato que precede a la prolina (Brandts, 1975).

La toxina Cn 5 además de ocho cisteínas que forman cuatro puentes disulfuro, tiene cuatro prolina de las cuales tres presentan enlaces *trans* y sólo la prolina 59 presenta enlace *cis*. Lo anterior ha sido propuesto a partir de datos preliminares de NMR a la toxina Cn 5 (Lebreton *et al.*, datos no publicados), así como por los datos de cristalografía de rayos-X de la toxina CsEv3 (Almassy, 1983) y los datos de NMR de la toxina CII 1 (Lebreton, 1994). Estas tres toxinas homólogas presentan una alta similitud en su secuencia primaria. Además, en la toxina Cn 5 el aminoácido que precede a la prolina 59 que presenta la configuración *cis*, es una tirosina. Con esto podemos asumir que la isomerización de la prolina 59 deberá participar de forma muy importante en la cinética de plegamiento. Considerando estas características y basándonos en los datos del plegamiento *in vitro*, podemos decir que la toxina Cn 5 presenta un problema genuino en el proceso de plegamiento. No podemos descartar la posibilidad de que no hemos encontrado las condiciones adecuadas de la solución de renaturalización pues pensamos que para todos los casos puede existir una condición que permita una recuperación mayoritariamente nativa. Sin embargo, para proteínas con verdaderos problemas de plegamiento deberá ser más difícil encontrar dichas condiciones. Pensamos que durante el plegamiento *in vivo* también existen estas barreras cinéticas, sólo que en la célula se presentan una serie de proteínas cuya función es superar esas barreras y favorecer el plegamiento evitando así la agregación de las proteínas.

TRABAJOS RECIENTES

Durante el tiempo en que se desarrolló este trabajo otros grupos trabajaron en la expresión de toxinas de cadena larga. Dos de estos grupos

recientemente lograron obtener toxinas recombinantes funcionales. En ambos casos se utilizó una estrategia de expresión en *E. coli*. En el primer trabajo expresan la toxina Bot XIV como péptido de fusión con dos dominios de la proteína A. de *S. aureus*, los cuales se unen a IgG. El vector de expresión que usaron permite la secreción de la proteína por lo que recuperan proteína recombinante activa directamente del medio de cultivo. Mediante un corte con CNBr obtienen cerca de 1 mg/l de toxina recombinante ya purificada. En el perfil de HPLC obtienen dos picos de los cuales uno, que representa el 33.33% de la proteína pura, muestra actividad similar a la toxina nativa. En el segundo pico (66.66%) detectan un poco de actividad y explican que esto puede deberse a que este pico contiene material parcialmente plegado (Bouhaouala-Zahar, 1996). En el segundo trabajo hacen la expresión de la toxina Lqh α IT, en forma de cuerpos de inclusión y mediante un proceso de plegamiento *in vitro* llegan a lograr obtener 5mg/l de proteína activa, esto es 10% de plegamiento (Zilberberg, 1996). Además, en este trabajo hacen alteraciones a la toxina mediante mutagénesis sitio específica. Hacen tres sustituciones en el extremo carboxilo (Y49I, A50K y N54K), las cuales resultan en una disminución en la toxicidad contra mamíferos de 6.4 veces sin que esto modifique la unión a sinaptosomas de rata. En la actividad tóxica contra insectos no encuentran modificaciones, por lo que creen que estos sitios del extremo carboxilo están participando en la toxicidad hacia mamíferos. Los dos grupos siguen la misma estrategia que nosotros planteamos de expresión en *Escherichia coli*. Obtienen la toxina recombinante activa en cantidades tales que les permiten realizar una serie de estudios. No obstante, en ambos trabajos se observa que el proceso de plegamiento no es ni cercano al 100% (10% en el primer caso y 33% en el segundo). Estos trabajos son hasta el momento los únicos reportes de expresión eficiente de una toxina de cadena larga. Ambos trabajos apoyan la estrategia seguida en nuestro proyecto ya que en los dos hacen la expresión de toxinas de alacrán en *E. coli*. Concluimos que para la expresión de toxinas de cadena larga *E. coli* parece ser la mejor opción para expresión. De igual forma los resultados de estos trabajos nos permiten confirmar que las toxinas de cadena larga presentan verdaderos problemas de plegamiento.

PERSPECTIVAS

1º Continuar con el diseño y expresión de toxinas mutantes que nos permitan participar en el estudio de la relación estructura-función de las toxinas de cadena larga del veneno de alacrán.

2º Buscar sistemas o modelos biológicos para los que la toxina Cn 5 presente una mayor afinidad y así requerir de menores cantidades de toxina para los ensayos.

3º Realizar ensayos electrofisiológicos que nos permitan hacer tanto una caracterización fina de estas proteínas, así como descifrar el mecanismo de acción de las mismas.

4º Continuar con el análisis racional de los datos de estructura primaria, actividad, mecanismo de acción y estructura tridimensional, que están disponibles hasta el momento, de las toxinas que reconocen canales de sodio. Al relacionar estos tres aspectos de las toxinas podremos tener un mejor panorama para compararlas. Con lo anterior intentar inferir los razos estructurales que hacen las diferencias entre estas toxinas.

5º Utilizar a las toxinas de cadena larga como un modelo para estudiar el plegamiento de proteínas. Además del conocimiento básico sobre el plegamiento en general de las proteínas que se puede obtener, para este trabajo es particularmente importante estudiar la participación del plegamiento en la actividad de estas toxinas.

ANEXO

- Becerril B., Corona M., García C., Bolívar F. and Possani L. D. (1995). Cloning of genes encoding scorpion toxins: an interpretative review. *J. Toxicol.-Toxin Reviews*, 14(3): 339-357
- Selisko B., García C., Becerril B., Delepiere M. and Possani L. D. (1996). An insect-specific toxin from *Centruroides noxius* Hoffmann cDNA, primary structure, three dimensional model and electrostatic surface potentials in comparison with other toxin variants. *Eur. J. Biochem.* 242: 235-242

**CLONING OF GENES ENCODING SCORPION TOXINS:
AN INTERPRETATIVE REVIEW**

Baltazar Becerril, Miguel Corona, Consuelo García, Francisco
Bolívar and Lourival D. Possani*

Instituto de Biotecnología, Universidad Nacional Autónoma de
México, Avenida Universidad, 2001, Apartado Postal 510-3,
Cuernavaca, Morelos 62271, México

TABLE OF CONTENTS

ABSTRACT-----	339
INTRODUCTION-----	340
ANALYSIS OF THE SIGNAL PEPTIDE AND 3'NON-CODING SEQUENCES-----	341
CDNA SEQUENCE ANALYSIS-----	343
GENOMIC DNA CLONING-----	345
SYNTHETIC GENE CLONING-----	346
DEDUCED AMINO ACID SEQUENCES AND COMPARISON ANALYSIS-----	348
CONCLUSIONS-----	351
REFERENCES-----	352

ABSTRACT

Scorpion toxins (Stox) are polypeptides intensively investigated because they are excellent models to study protein-structure-function relationships and exquisite tools to access ion channel functions. Recently, techniques of molecular biology have been used to clone the genes encoding toxins specific for Na⁺-channels. Genes from scorpions of the genus *Androctonus*, *Centruroides*, *Ledurus*, *Tityus* and *Bothus* have been cloned and their nucleotide (nt) sequence determined. Data from complementary DNA (cDNA) cloning of stox from the genus before mentioned and genomic cloning of Stox genes from scorpions of the genus *Tityus* and *Androctonus* have

provided information for the deduction of the structure of the transcriptional units encoding toxin genes of these scorpions. It seems that transcription of these genomic regions generates pre messenger RNAs (pre-mRNAs) of approximately 800nt containing an intron of approximately 470nt, located within the region encoding the signal peptide. The processing of these pre-mRNAs originates a mature molecule of approximately 330 nt. The translation of the mRNAs give precursors of approximately 32-89 amino acid (aa) residues, which in some cases are processed at both amino- and carboxyl-terminus ends. At the amino terminus, a 18-21 aa signal peptide is released by a signal peptidase. At the C-terminus, basic residues (Lys or Arg) are eliminated by a carboxypeptidase. In some cases, the C-terminal residues are processed given rise to a final amidated aa. In this review, the information available on this subject is summarized and we also present an interpretative review of the main Stox structure known, with the possible implications for their function on excitable cells.

INTRODUCTION

Scorpion venoms contain an assortment of low molecular weight peptides which are toxic to different organisms including human beings (1-3). Scorpion venoms, so far studied, contain four main classes of toxins: polypeptides of 60-70 aa¹ residues which affect the function of Na⁺-channels of excitable cells (4,5), peptides of 37-39 aa residues blockers of K⁺-channels of excitable cells (6-10) and non excitable cells (11), peptides of 37 aa residues that are specific for chloride channels (12) and finally, low molecular mass proteins of unknown structures that affect the Ca²⁺-release channel of sarcoplasmic reticulum (13).

Among the best studied scorpion toxins (Stox) are those that modify Na⁺-channels and have been classified as α and β (5,14,15). This classification has been proposed based on two distinct physiological effects exerted by Stox on the Na⁺-channels. The α -Stoxs specifically affect axonal conduction by a prolongation of the action potential as a result of a slowing or blockade of the Na⁺-channel inactivation (16-20), while β -Stoxs preferentially alter the activation mechanism of these channels (21-23).

¹ The abbreviations used in this chapter are: a, aromatic residue; aa, amino acid(s); Ah, *Androctonus australis* Hecaton; Am, *Androctonus mauretanicus mauretanicus*; b, basic residue; Bj, *Buthotus judicus*; Bot, *Buthus occitanus tunetanus*; bp, base pair(s); cDNA, DNA complementary to RNA; Cit, *Centruroides limpicus tecomanus*; Cngt, cDNA from CNH cloned into lambda gt11; CNH, *Centruroides nexius* Hoffmann; Cse, *Centruroides sculptus* Ewing; Csf, *Centruroides suffusus suffusus*; h, hydrophobic residue; Lqh, *Leiurus quinquestriatus hebraeus*; Lqq, *Leiurus quinquestriatus quinquestriatus*; nt, nucleotide(s); oligonucleotide(s), oligodeoxyribonucleotide(s); PCR, polymerase chain reaction; pre-mRNA, pre-messenger RNA; Stox, scorpion toxin(s); SP, signal peptide(s); Ts, *Tityus serrulatus*.

Our group has contributed during the last seventeen years to the knowledge of the structure-function relationship of the toxins of the venom from several species of Brazilian and Mexican scorpions. Purification, chemical and physiological characterization of more than 30 different toxins were obtained during this period (2, 24-26). Recently, we have started to use molecular biology strategies for cloning, mutagenization, and heterologous expression of specific Stox-encoding genes (27-29). The first report on the cloning of Stox appeared in 1989 (30). In this work, four cDNA sequences encoding mammal-specific toxins and three cDNAs encoding insect-specific toxins from *Androctonus australis Hector* (AaH), were published. Since then, several other cDNA sequences encoding Stox have been published (27-28, 31-34). The purpose of this review is to present the information available in the literature in this matter, as well as to compare the known structures obtained by cloning and direct nt sequencing, with the known functions. At this moment the knowledge is restricted to the Na⁺-channel specific Stox. The genes that code for other classes of toxins have not been cloned, except for a couple of synthetic genes, corresponding to K⁺-channel blocking peptides (35-36).

ANALYSIS OF THE SIGNAL PEPTIDES AND THE 3' NON-CODING SEQUENCES OF cDNAs ENCODING STOX

A comparison of the signal peptides (SP) encoded by cDNAs of representative reported Stox is shown in Fig. 1. They have been grouped in three distinct categories: SP of 19-20 aa residues of New World Stox (*Centruroides noxius* and *Tityus serrulatus*), SP of 19 aa residues corresponding to Old World Stox (*Androctonus australis* and *Leiurus quinquestratus hebraeus*) and 18 or 21 aa residues SP of Old World Stox (*Androctonus australis* or *Bothus judaicus* and *Leiurus quinquestratus hebraeus*). Consensus sequences are proposed for each group and for the three groups together. These sequences are in agreement with the rules proposed by von Heijne (37) for cleavage sites in eukaryotic SP. For the New World Stox SP, the residues -8 through -11 are mostly hydrophobic or aromatic, probably as part of an hydrophobic core. Worth mentioning is residue -9 which in spite of a different codon in each cDNA (27,33), it is always Leu. It is striking to observe an extraordinary richness of hydrophobic and/or aromatic residues in the region -9 through -19 of group 3 SP. It is also remarkable to find a Val residue in position -3 for all the SP shown in Fig. 1. *Tityus serrulatus* (Ts) VII SP has been included in group 1, but it has several conserved residues with groups 2 and 3. The comparison of 3'-non-coding cDNA sequences is shown in Fig. 2. The nucleotide sequences were aligned and gaps inserted to optimize similarities. This alignment allowed us to detect several well-conserved regions including those located around nt 15, 30, 40, 50 and 75. Particularly conserved are the sequences from nt 35 through 45, and 70 through 80. This last region contains the polyadenylation signal. It is remarkable the similarity between CngtIII and AaHIV in some of these conserved regions. It is not known if the conserved sequences above mentioned might be implicated in mRNA processing. The cDNAs corresponding to CngtIII (encoding toxin Cn5) and CngtIV do not include a typical

		-20					-15				-10					-5			-1
CngtIII	M	N	S	L	L	M	I	X	A	A	L	F	V	L	L	V	V	V	A
CngtIII	M	N	M	L	L	M	I	X	A	A	L	V	L	L	L	V	V	V	A
CngtIV	M	N	M	L	L	L	I	X	A	A	L	V	L	L	L	V	V	V	A
CngtV	M	N	M	S	L	L	M	I	A	A	L	V	L	L	L	V	V	V	A
FvII	M	N	M	S	L	L	M	I	A	A	L	V	L	L	L	V	V	V	A
Cons. 1	M	N	M	S	L	L	M	I	A	A	L	V	L	L	L	V	V	V	A
AeHI	M	N	Y	L	V	M	I	S	L	A	L	L	L	L	L	L	L	L	E
AeHII	M	N	Y	L	V	M	I	S	L	A	L	L	L	L	L	L	L	L	E
AeHIII	M	N	Y	L	V	M	I	S	L	A	L	L	L	L	L	L	L	L	E
EgHIT	M	N	M	L	L	V	M	I	S	L	A	L	L	L	L	L	L	L	E
Cons. 2	M	N	Y	L	V	M	I	S	L	A	L	L	L	L	L	L	L	L	E
AeHIT	M	K	F	L	L	L	F	L	V	V	L	P	I	M	-	-	G	-	V
AeHII	M	K	F	L	L	L	F	L	V	V	L	P	I	M	-	-	G	-	V
AeHIII	M	K	F	L	L	L	F	L	V	V	L	P	I	M	-	-	G	-	V
EgHIT	M	K	L	L	L	L	V	I	S	A	S	M	L	I	-	-	G	-	V
Cons. 3	M	K	F	L	L	L	F	L	V	V	L	P	I	M	-	-	G	-	V
Cons. 1-3	M	N	E	L	L	M	I	X	A	L	L	L	L	L	L	L	L	L	E
	M	K	A	L	V	I	S	A	S	M	L	I	M	G	-	-	G	-	V

Fig. 1. Comparison of signal peptides.

SP of representative toxin-encoding cDNAs from New World and Old World scorpions are compared. Gaps (-) were introduced where necessary to maximize similarities. The aa residues of SP were numbered from -1 to -21 taking BjIT2 as reference. A consensus sequence was derived for each group of SP sequences. A global consensus was deduced from each partial consensus. A second possible global consensus (next most abundant residues), is also shown. For consensus sequences, X means variable residue, a=aromatic residue, b=basic residue and h=hydrophobic residue.

	10	20	30	40	50	60	70	80	90	
CngtIII	TGGCAACGACTTTT	ATATGTCCATAAACA	GAGAATAT	TCGAACGGCTCT	TAATTGCA	GTTAAGT	-----	AAAAATAATGC	T-----	AATAGCATT
CngtIV	TGGCAACGACTTTT	ATATGTCCACCACA	GAAGAATAG	TGAACCGCTCT	TAAITCCAAG	-----	-----	-----	-----	-----
CngtV	TGGCAACGACTTTT	ATATGTTCACCAACA	GAAGAATAT	TGTAACCGCTCT	TAAITGCA	GTTAAAGT	-----	-----	-----	-----
CngtV	TGGCAACGACTTTT	ATATGCCCGCA	CAAGAAATAT	TGAACCGCTCT	TAAITGCA	TAAAGT	-----	AAAAATAAA	-TAT-----	TATAGCTTY
FvII	ATT-TGTTTC	CGTGAATATCCT	TACAAATG	AA-ETG-	TAAT-----	AAAGT-----	TGGCAAAAAATAAA	AAA-	AA-	TGTT-C-
AeHI	ACC-TGTA-	GAGTAAAT	CAGAAAGAA	TGTATC	CTAA-AAAT	ACTGGT-----	-----	AAAAATAAA	-CATA-	AGTA-----AT
LChit	ATC-TGTA-	GAACATAA	CAAGAAGAA	TGTATC	CTAA-GAAT	GTATC-----	-----	AAAAATAAA	-TAT-----	-----
AeHIT	T-TGTA-----	TAT-GAA-GTAT	-----	T-T-GAAT	GTATC-TA-----	-----	AAAAATAAGCACA	-----	TATAT--T	-----
AeHII	ACC-TGTA-	AAGCAAAA	CAACAAGAT	GTATC	CTAA-AAAT	ACTGGT-----	-----	AAAAATAAA	-CATA-	AGTA-----AT
AeHIII	AT--TATA	-GATG-----	GAAT	GTATC	CTAA-GTATC	-----	AAAGT	-----	AAAAATAA	-----TATAACAAAA
EgHIT	CCC-CTAC-	TGGCAT	GAATATGTTGCA	GAAA-AATC	-----	AA-TAT	ACC-----	TG-C	AAAAATAA	-----GTAACITCATGCA-

Fig. 2. Alignment of 3'-non-coding sequences.

The 3'-non-coding regions of the indicated cDNAs were aligned. Gaps for different SpX and cDNAs were introduced where necessary to maximize similarity. Numbers on top of sequences are for reference purposes only. CngtV is encoded by CngtII. Sources of nt sequences are the same as in Fig. 3.

polyadenylation signal but they present a polyA tail (27). A possible interpretation for this discrepancy is that the region between nt 70-80 (which is rich in adenosines) could have hybridized with an oligo(dT) to prime the first strand of cDNA synthesis and what we showed as a polyA in the case of *CngtII* and *CngtIV*, is actually the complementary of such an oligo (dT).

CDNA SEQUENCE ANALYSIS

Fig.3 shows a nt sequence comparison between New World (*CngtII*, *CngtIV* and *TsvII*; 27,33) and Old World (*AaHI*-III and *AaHII*; 30; *SjtI*; 31; *LghoIT*; 32) cDNA nt sequences. This comparison shows some general features about the Stox coding cDNAs. cDNAs encoding *AaHI*-*AaHIII* Stox are closely related. cDNAs that code for *AaHI* and *AaHIII* are also more similar between themselves than with the cDNA encoding *AaHII*. The 5' non-coding regions, SP coding regions and 3' noncoding regions are well-conserved in all these three cDNAs (Fig. 3; also see reference 30). These observations indicate that *AaHI*-*AaHIII* Stox are encoded by genes that might have originated from duplications of an ancestral gene and each duplicate evolved independently. A similar event might have been responsible for *C. noxius* Stox-encoding gene similarities in which cDNAs corresponding to *CngtII*-*CngtIV* (27), seem to have apparently originated also by duplications of an ancestral gene and each duplicate evolved independently. Thus it is clear that Stox are encoded by closely related genes probably originated from an ancestral gene. It is also clear that duplicated genes have been under different selective pressures. For instance, 5' and 3' non-coding sequences are very well-conserved among *C. noxius* or *A. australis* Stox genes but coding regions show different levels of variability. In *Centruroides noxius Hoffmann* (*CnH*) genes, the similarity at the nt level is at least 80% but at the aa level it is 68% (see also Fig.4, below). Part of the explanation for this phenomenon can be found in the existence of regions of high variability at the aa level, maybe, hot spots for mutation (27). Clear examples of this variability are the regions around residue number 10 and the carboxyl terminus. Similar characteristics are present in cDNAs from *AaH*. Another important observation in this comparison is the significant similarity between *LghoIT* and *AaHI*-*AaHIII* cDNAs. At the nt level, *LghoIT* cDNA is more similar to *AaHII* cDNA: a 95% similarity at the SP sequence, a 72% similarity at the mature peptide sequence and a 60% similarity at the 3' non-coding sequence (Figs. 1 and 3). These are clear indications that the cDNAs encoding *AaHII* and *LghoIT* are specified by closely related genes. It is also remarkable the close similarity of the 3' non-coding region of the cDNAs encoding *LghoIT* and *AaHI*. In spite of the noticeable similarity between *LghoIT* and *AaHII* they have different specificities. It has been demonstrated that *AaHII* affects mammals (38) while *LghoIT* is highly toxic to insects and crustaceans but less toxic to mice (20). *LghoIT* (Fig. 4), seems to be processed by means of the hydrolysis of two basic aa residues at its C-terminus (32). Processing at carboxyl-terminus has been observed only for mammalian Stox *AaHI*-*AaHIII* (30). We have reported similar processing for toxin 4 of *C. noxius*, in which it has been demonstrated, by means of mass spectrometry, that the C-terminus of toxin Cn4 is cysteinyl-asparagine-amide (28). Martin-Euclaire et

10 20 30 40 50

AM1111 GTA GGT GAT GGT TAT ATT GTC GAT TCC AAC AAC --- TGT GTA TAC GAT TGT GGT CCA
V N D G V J V D S K N C V Y N C V Y P
AM11 AAA GGT GAC GGT TAT ATT GTC GTA CCC AAT GAC --- TCT GTA TAC CAT TGT GGT CCA
E N D S V I V Y D S K N C V Y N C V Y P
AM111 GTA AAA GAC GGT TAT ATT GTC GAT GAT GTA AAC --- TCC ACA TAC TTT TGT GGT AGA
E N D S V I V Y D S K N C V Y N C V Y P
LMD111 GTA GCT GAC GCT TAT ATT GCC AAG AAT TAT AAT --- TGT GTA TAT GAA TGC TTT GAA
V N D G A F J N K K D B C K V F N
EL112 --- CAT GCT TAT ATA AGA E N Y N B --- GGA TCC AAG GTT TCC ATC ATC
V N D G A F J N K K D B C K V F N
LAV11 --- AAA GAA GGT TAT GTC ATG GAT GAC --- GAA GCT TAC AAA GTT AAT GCT TTT ATC
K E R G U Y L N D N R E E C K L B C F I
KOL11 --- AAA GAA GGT TAT GTC ATG GAT GAC --- GAA GCT TAC AAA GTT AAT GCT TTT ATC
E S G F L V N K S T E C K Y G C L L
KOL11V --- AAG GAC GGT TAT GTC GTC GAT GTA --- AAG GGC TCC AAA AAG TAT TCC TAT AAA
K E N D S V I V Y D S K N C V Y N C V Y P
AM11V AAG AAC AAT GGA TAT GTC GTC GAT AAT --- AGT GCT AAA GCT CCC GAA TCT GTT TTG
K E N D S V I V Y D S K N C V Y N C V Y P
60 70 80 90 100 110

AM1111 CCA --- --- --- --- --- TCC GAC GGT TTA TGT AAG --- AAA AAC GGT GCT AAA GCT
P --- --- --- --- --- TCC GAC GGT TTA TGT AAG --- AAA AAC GGT GCT TCC AGT
AM11 P --- --- --- --- --- TCC GAC GGT TTA TGT AAG --- AAA AAC GGT GCT TCC AGT
AM11V --- --- --- --- --- TCC GAC GGT TTA TGT AAG --- AAA AAC GGT GCT TCC AGT
LMD11V --- --- --- --- --- TCC GAC GGT TTA TGT AAG --- AAA AAC GGT GCT TCC AGT
EL112 --- GGA --- AAT GAA GGT TCC GGT AAA GAA TCC GTT GCT --- --- CAC GGT GGC TCC
D A V Y C N E L C T I K N G A S B
LAV11 --- AGA CCA TCC GCA TAC TCC GGC AGA GAA TCC GCA ATT AAA AAG GGC TCC TCC ---
E P S S S Y C S S E C S I E K D S S
KOL11 TTG GGA AAA AAC GAA GGC TCC GAT AAG GAA TCC AAA GGC AAC CAA GGA GCT AGT
L D K N E S C D P E C R K N D G S G
KOL11V TTG GGA AAA AAC GAT TAT TCC AAT AGC GAA TCC AAA ATG AAA GAC CAA GCA GCT AGT
L G E R D V C N R S E C E N E N R R A D S
AM11T --- --- --- --- --- AGC AAT GAA GAA TCC AAA GAA GAA TCC AAA GAA GAA GAA GAA
S N V C N B E C T K V N Y A D E
120 130 140 150 160 170

AM1111 --- GGC TGT TCC GGT TTT TTA ATT CCA TCC GCA GTT GCT TCC TCC GTC GTA CCG TGG
E S C V C L J C M D W A
AM11 --- GGC TGT TCC GGT TTT TTA ATT CCA TCC GCA GTT GCT TCC TCC GTC AAA GAC TTS
E S C V C L J C M D W A
AM11V --- GGT TAT TCC CAA TCC GCA AGT CAA TAT GCA AAC GCT TGT TAT TCC TAT AAA TTS
S Y C D B W A S P Y N J A C M C Y R L
LMD11V --- GGC TAT TCC CAA TCC GCA AGT CAA TAT GCA AAC GCT TGT TAT TCC TAT AAA TTS
S Y C D B W A S P Y N J A C M C Y R L
EL112 TTT GGT TAC TGT --- --- --- --- --- GGC TCC TCC TCC TCC TCC TCC TCC TCC
J S V C S M A S V S W A C M Y A L
LAV11 --- GGC TAT TCC GGC TCC GGC --- --- --- --- --- GGC TGT TAC TCC TCC GGC GTT
S Y C A M P C Y L A
KOL11 TAC GGC TAT TCC TAC GCT TTT --- --- --- --- --- GGC TCC TCC TCC GAA GGT TTA
V D N W S E S C V C G C Y C A G
KOL11V TAC GGC TAT TCC TAC GCA TTT --- --- --- --- --- GGC TCC TAT TGT GAA GSA TTA
V D N W S E S C V C G C Y C A G
AM11T --- GGA TAT TCC TCC TCA GTT --- --- --- --- --- TCA TGT TAT TCC TCC GCT CTA
S C C L S C F G G
180 190 200 210 220 230 240

AM1111 CCC GAT AAC GTA CCC AGT AAA GAT CCA --- --- --- --- --- TCA TAT AAA --- --- TCC GAT ACC CCA TAA
P D N W P I K D P --- --- --- --- --- TCA GGC AAA --- --- TCC ACC TCC TGA TAA
AM11 CCC GAT AAC GTA CCC AGT AAA GCA --- --- --- --- --- CCA GGA AGA --- --- TCC GAT CCG CCA TAA
P D N W S E S C V C G C Y C A G
LMD11V CCC GAT AAT GTA CCC AGT AGA GTA --- --- --- --- --- CCA GGA AAA --- --- TCC GAT CCG CCA TAA
P D N W S E S C V C G C Y C A G
EL112 CCA GAC GGC GTC AGA TGG AAG --- --- --- --- --- ACC AAT ACA --- --- TCC GGT AAG AAA AAG TAA
P D N W S E S C V C G C Y C A G
LAV11 CCA AAT TCC GTC AAA GTT TCC GAT ASA CCG --- --- --- --- --- TCC GGT AAG AAA AAG TAA
P D N W S E S C V C G C Y C A G
KOL11 CCC GAA AGT ACA CCG AGT TAT CCC GTT --- --- --- --- --- CCA TAT AAA TET --- --- TCC GGT AAG AAA TAA
P K S T P I Y P L --- --- --- --- --- TCA TAT AAA AGA --- --- TCC GGC AAG AAA TAA
S N T M N A S C E K K
KOL11V TCC GAT AGT ACA CCG AGT TCC CCC GTT --- --- --- --- --- CCA TAT AAA AGA --- --- TCC GGC AAG AAA TAA
S N T M N A S C E K K
AM11T AAT GAC GAT AAA AAG GTT TGC GAT TCC GAC ACA AGG AAA AGT TAT TCC ACC ACA TTA ATT AAT TAA
N D D C K V L S D T R E S Y C D Y T I I N W P

al. (33), reported also a similar processing for toxin VII of *Tityus serrulatus*. Bougia et al (30), have proposed a model in which STox specific to mammals would be processed at their C-terminus, contrary to the insect-specific STox which is not processed. Since LqhoIT is an insect-specific STox and it has a C-terminal sequence compatible with the rules for processing (39), we propose that additional elements should be added to the previously proposed model. If the insect-specific STox is excitatory (70 aa residues in length), the C-terminal of the precursor apparently is not processed, but if it is of the LqhoIT type or depressant (Fig. 4, group 9), it may seem to be processed. However, the rules of processing are not still clearly understood in this last group. BJT2, whose precursor has the carboxyl-end -Cys-gly-Arg-Lys-Lys (31), is processed giving the mature carboxyl-end -Cys-Gly. The expected mature carboxyl-end if the rules were consistent with α -amidation (39), would have been cysteine-amide. A similar maturation for the precursor of LqHIT2 has been recently reported (34). Perhaps another rule for the processing of STox would be that when the carboxyl-end of the precursor has three basic residues, in spite of the presence of a Gly residue preceding a basic tripeptide, the mature end is generated after the hydrolysis of the basic triad only. Although some rules for STox precursor processing are clear and in accordance with the rules for processing of neuroactive polypeptides (39), new possibilities are emerging based on the analysis of the cDNAs sequences encoding STox and the determination of the aa sequences of the carboxyl-ends of the corresponding mature STox. These kind of interpretations have been only possible by means of the cloning of toxin genes. The analysis of cDNA sequences and the aa sequences of mature toxins allowed to postulate the existence of new enzyme activities as the carboxypeptidase, necessary for the processing of precursor STox by elimination of three basic residues at the C-terminus as in BJT2 or LqHIT2.

GENOMIC DNA CLONING

To the best of our knowledge, our group has published the only formal report of the cloning of a scorpion gene in which the complete genomic nucleotide sequence is presented (29). The gene encoding toxin gamma from the Brazilian scorpion *Tityus serrulatus* was amplified by PCR from genomic DNA employing synthetic

Fig. 3. Comparison of cDNA nt sequences.

The cDNA sequences from North and South American STox (27,33), were aligned with North African STox cDNAs (30-32), taking Cys codons as reference. Gaps were included where necessary to maximize similarities. From North American STox cDNAs, we have included only CngtIII and CngtIV (27). CngtIII and CngtV were not included because they are very similar to CngtII and CngtIV, respectively. The numbers on top of the sequences were placed for reference purpose; they do not correspond necessarily with the positions in the original publications. Last digits of numerals are aligned with corresponding nt.

Group

		10	20	30	40	50	60	
C#2	REGYLVNKTCKCYECLKLGDDHYCLRECKOQYKAGGGYCFAP----	ACVCTHLYEQAIWVFL--	PNKK--CS					
C#e12	REGYLVNKTCKCYECLKLGDDHYCLRECKOQYKAGGGYCFAP----	ACVCTHLYEQAIWVFL--	PNKK--CM					
C1e1	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	GCWCEGLPEQAVWVFL--	PNKK--CS					
C#3	REGYLVNKTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	GCWCEGLPEQAVWVFL--	PNKK--CR					
C#g12	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	GCWCEGLPEQAVWVFL--	PNKK--CS					
Cons.1	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	GCWCEGLPEQAVWVFL--	PNKK--CS					

		10	20	30	40	50	60	
C#E1	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	ACWCEGLPESTFTVFL--	PNKK--CS					
C#g1211	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	GCWCEGLPESTFTVFL--	PNKK--CF					
C#E2	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	ACWCEGLPESTFTVFL--	PNKK--CS					
C#g12	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	GCWCEGLPESTFTVFL--	PNKK--CS					
C#E3	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	ACWCEGLPESTFTVFL--	PNKK--CS					
Cons.2	REGYLVNKHSTCKCYECLFGLGDDHYCLRECKOQYKAGGGYCFAP----	ACWCEGLPESTFTVFL--	PNKK--CS					

		10	20	30	40	50	60	
C#E1	RDGYLVNK--TCKKTCYKLEGNDFCNRECKMHWGGSTVYCYCF----	GCYCEGLDPSQTWFL--	PNKK--CT					
C#g12	RDGYLVNKK--TCKKTCYKLEGNDFCNRECKMHWGGSTVYCYCF----	GCYCEGLDPSQTWFL--	PNKK--CG					
C#1	RDGYLVNDA--KCKKNCYKLGKNDYCNRECKMHWGGSTVYCYCF----	GCYCEGLDPSQTWFL--	PNKK--CT					
Cons.3	RDGYLVNDS--KCKKNCYKLEGNDFCNRECKMHWGGSTVYCYCF----	GCYCEGLDPSQTWFL--	PNKK--CG					

		10	20	30	40	50	60	
T#V11	KEGYLVNME--GCKLACFP--RPSGYDRECGIF--KSS--GYCANP----	ACYCYGLPHNVAWVDA--	TNKK--CGKK					
F#8	KEGYLVNME--GCKLACFP--RPSGYDRECGIF--KSS--GYCANP----	ACYCYGLPHNVAWVDA--	TNKK--CG					
Cons.4	KEGYLVNME--GCKLACFP--RPSGYDRECGIF--KSS--GYCANP----	ACYCYGLPHNVAWVDA--	TNKK--CG					

		10	20	30	40	50	60	
A#H1	KADGYIYVFN--CVYHCVPF----	CDGLCKKN--GSS--GSCFLFPSGLACWCKXKLPDNYVFKDT--	GRK--CTR					
A#H12	VADGYIYVFN--CVYHCVPF----	CDGLCKKN--GAKS--GSCFLFPSGLACWCKXKLPDNYVFKDP--	STR--CHSR					
Cons.5	XADGYIYVXKN--CVYHCVPF----	CDGLCKKN--GAKS--GSCFLFPSGLACWCKXKLPDNYVFKDK--	SXK--CXXK					

		10	20	30	40	50	60	
Ighe1T	VADYIYAGNYN--CVYECFR----	DAYCNELCTRN--GAS--GYCWVLRKYGNAWCKLDPNVIPIVF--	CR--CHRR					
A#H11	VADGYIYVDDVN--CTYFCGR--	NAYCNECECKL--GES--GYCWASPYGNACVCKLDPNVIPIVF--	GR--CHGR					
A#W1	LKGDIYIYDILN--CTFFCCR--	NAYCNECECKL--GES--GYCWASPYGNACVCKLDPNVIPIVF--	GR--CH					
F#X1	LKGDIYVDRN--CTFFCCR--	NAYCNECECKL--GES--GYCWASPYGNACVCKLDPNVIPIVF--	GR--CH					
IgV1	LKGDIYVDRN--CTFFCCR--	NAYCNECECKL--GES--GYCWASPYGNACVCKLDPNVIPIVF--	GR--CH					
Cons.6	LKGDIYVDRN--CTFFCCR--	NAYCNECECKL--GES--GYCWASPYGNACVCKLDPNVIPIVF--	GR--CHXB					

		10	20	30	40	50	60	
F#E1	G--RDYIAQGFN--CVYEQA--	NSYCNLDLCTR--GAS--GYCDLVRKYGNAWCKLDPNVIPIVF--	GR--CHF					
F#E12	G--RDYIAQGFN--CVYEQA--	NSYCNLDLCTR--GAS--GYCDLVRKYGNAWCKLDPNVIPIVF--	GR--CHF					
IgV1	GVRDAYS--RDNN--CVYEQF--	NSYCNLDLCTR--GAS--GYCDLVRKYGNAWCKLDPNVIPIVF--	GR--CR					
Cons.7	GVRDAYSIAQGFN--CVYEQA--	NSYCNLDLCTR--GAS--GYCDLVRKYGNAWCKLDPNVIPIVF--	GR--CRF					

		10	20	30	40	50	60	70
A#H1T	KKNGYAVD--SSGKAPPELL--	SNYCYNECTR--V--HYADRYGCCL----	SCYCFGLNDKVKLEISDTRKYVCDTTIN					
IgG1T1	KKNGYAVD--SSGKAPPELL--	SNYCYNECTR--V--HYADRYGCCL----	SCYCFGLNDKVKLEISDTRKYVCDTTIN					
Cons.8	KKNGYAVD--SSGKAPPELL--	SNYCYNECTR--V--HYADRYGCCL----	SCYCFGLNDKVKLEISDTRKYVCDTTIN					

		10	20	30	40	50	60	
IgH12	DGYIRKR--DGCKVACL--I--G--NECCDCKEAKY--	GCSGYVC--WT--	MGLACWCEGLDPSKTRWSE--	TH--TCG				
IgG12	DGYIRKR--DGCKVACL--I--G--NECCDCKEAKY--	GCSGYVC--WT--	MGLACWCEGLDPSKTRWSE--	TH--TCG				
A#12	DGYIRKR--DGCKVACL--I--G--NECCDCKEAKY--	GCSGYVC--WT--	MGLACWCEGLDPSKTRWSE--	TH--TCG				
Cons.9	DGYIRKR--DGCKVACL--I--G--NECCDCKEAKY--	GCSGYVC--WT--	MGLACWCEGLDPSKTRWSE--	TH--TCG				

		10	20	30	40	50	60	
C#E1	KKDGYVVDSDN--KPYECLN--	DDYCNLCLER--KAKK--GYCYV--GRVS--	CYV--GCLPDNPSPTK--T--	SGK--CNPA				
A#H124	ENGVLLNKTCKCYKVCVYVINE--E--CCYLC--NKAAGCYGYCYF--	WKLACYCQGRKSE--	LN--MYTKN--	CDL				
Cons.	CKKDDYVDRNGEYCECLGDDHYCLRECKOQYKAGGGYCFAP----	GCWCEGLPEQAVWVFL--	PNKK--CS					

		10	20	30	40	50	60	70
CONS.	CKKDDYVDRNGEYCECLGDDHYCLRECKOQYKAGGGYCFAP----	GCWCEGLPEQAVWVFL--	PNKK--CS					
	EA I M T I I E	---	DG DDL TII-III- S QIF-DE--G Y II ESKITA	Q S- bb				

oligonucleotides designed from the reported cDNA sequence (33). The nt sequence of this gene reveals the presence of an intron of 475 base pairs (bp) which interrupts the region that encodes the signal peptide (near the end) of the precursor toxin. Thus, this gene contains two exons (exon 1 with 55 bp and exon 2 with 260 bp). A comparison of 5' and 3' intron boundaries with the corresponding nt sequences of other introns of genes from other members of the group of arachnids revealed a close similarity within the first six nucleotides located at the 5'-end and the last seven nt of the 3'-end of the introns. When these consensus boundary sequences are compared to other eukaryotic intron boundaries, it appears that they are part of a regular feature of all eukaryotic introns so far reported (29). Similar results with a gene encoding the toxin AaH I' from *Androctonus australis* isolated from a genomic library (40), suggested the presence of an intron of similar size and position but no nt sequence was provided. Recently we have cloned and sequenced the corresponding genes encoding gamma toxin-like from the scorpions *Tityus behniensis* and *Tityus strigulosus* by PCR from genomic DNA (unpublished). The structure of these genes is highly similar to that of the gene from *Tityus serrulatus*. The sequence of the exons is almost identical among them but the sequence of the introns is less conserved. These results suggest that the structure of the genes encoding Stox that affect Na⁺ channels, could be very similar. An intron of about 450 bp located near the end of the region that encodes the signal peptide and two

Fig. 4. Amino acid sequence comparison.

This figure compares aa sequences of the principal representatives of different groups of Stox that affect Na⁺ channels. Toxin sequences were grouped accordingly to their similarities. Gaps (-) were introduced for maximizing similarities. Below each similarity group, a consensus sequence is shown in bold. A global consensus sequence is shown at the bottom of this figure. A second possible consensus (next most abundant residues), is also shown. The regions for J and B loops (42) or corresponding gaps, are indicated below the global consensus sequence by means of rectangular parentheses. The aa sequences deduced from cDNAs encoding several Stox are included: CngtIII-CngtV (CngtIII encoding toxin Cn5) (27), AaHI-AaHIII, AaHIT (30), LqhaIT (32), BjtIT (31) and TsvVII (33) are primary sequences deduced from cDNA sequences. CngtV, deduced primary sequence (toxin Cn4), has been corroborated by direct aa sequence of the mature Stox. The carboxyl-terminus was determined as an asparagine-amide residue (28). The last two or three residues shown in bold at the end of deduced primary sequences from cDNAs, are not present in mature toxins. Direct aa sequences are from Meves et al. (42), Lorez et al. (47), Zamudio et al. (26) and Zlotkin et al. (39). For consensus sequences, X means variable residue, Aaromatic residue, bbasic residue and hhydrophobic residue. The asterisk at the end of AaHI represents a non-sense mutation probably generating a stop codon (TGA) from an Arg codon CGA (see Fig. 4). Conserved structural domains are indicated in which α and β series means α -helix and β -sheet respectively. Pairings of the eight half-cysteines are also indicated. AaIT and CsEV aa sequences conform group "I" (for intermediate) which corresponds to Stox sharing characteristics of both Old World and New World Stox.

exons, one containing the 5' noncoding region plus the main part of the region encoding the signal peptide, and the other exon containing the region encoding the C-terminus of the signal peptide and the mature peptide plus 3' non coding region.

SYNTHETIC GENE CLONING

Complementary DNA cloning of Stox genes has been focused mainly on Na⁺-channel blockers. As far as we know, cDNA cloning of toxins that block K⁺, Cl⁻, or Ca²⁺-channels, has not been reported. Recently, a synthetic gene encoding charybdotoxin (CTX), a Ca²⁺-activated K⁺-channel blocker from *Leiurus quinquestratus* venom, has been cloned and expressed as a fusion protein in *Escherichia coli* (35). Recombinant CTX showed blocking and dissociation kinetics identical to the native venom toxin (35). Specific mutagenesis of recombinant CTX in combination with mechanistic analysis of single Ca²⁺-activated K⁺-channels has shed some light on the important regions of the toxin's molecular surface that interact with the channel pore (41). Another synthetic gene encoding Margatoxin (MgTX) a voltage-activated K⁺-channel blocker from *Centruroides margaritatus* venom, has also been cloned and expressed in similar systems (36). Like ChTX, MgTX blocks the n-type current of human T-lymphocytes (K_v1.3 channel) but compared to ChTX, is 20-fold more potent (36). Our group, using the same system reported for the expression of CTX, has synthesized and expressed the gene encoding Noxiustoxin (7), a K⁺-channel blocker from *Centruroides noxius* venom (unpublished results).

DEDUCED AMINO ACID SEQUENCES DERIVED FROM cDNAs ANALYSIS AND COMPARISON WITH DIFFERENT PRIMARY STOX SEQUENCES

In order to compare the different aa sequences deduced from the cloned *C. noxius* cDNAs (27) with other Stox, we performed a literature survey choosing the principal representative types of different Stox. These aa sequences were grouped according to their similarities, aligning them with respect to Cys residues and introducing gaps (-) where necessary to maximize similarities. The Stox were clustered into nine groups, plus two single sequences (Fig.4). The first three groups correspond to *Centruroides* Stox. Each main group was aligned and a consensus sequence was proposed for each one of them. Some particular differences among groups 1-3 can be observed, especially several deletions of one aa residue. A distinctive feature among these three groups of Stox seems to be the last fifteen residues at the carboxyl-terminus. These alignments impose the existence of three main gaps, located after residues 19, 45 and 60 taking groups 1-3 as reference. The portions of the sequences aligned with these gaps (-) correspond to loops in the respective Stox. The loops around residues 20 and 45 have been called loops J and B respectively (42). We propose consequently, on a comparative basis, that gaps could be called accordingly: gaps J and B, respectively. The third loop (around

position 60) which is characteristic of excitatory insect-specific STox (group 8), is proposed to be called "EITS" (Excitatory Insect Toxin Specific). It is worth mentioning that in group 8 Stox, a B gap is also present. An explanation for the presence of such gaps, might be found in their three-dimensional structure (43). In these STox an atypical disulfide bridge is established between Cys residues 38 and 64. Cys 38 is located just in the vicinity of the region equivalent to loop B of STox from groups 5-7. Residues 63-65 are situated in the space reserved for a B loop. Thus, apparently it seems incompatible to have both a B loop and a larger C-terminus (see Fig. 4 group 8 Stox) when an atypical disulfide bond is formed between Cys residues 38 and 64.

In most cases, groups of STox shown in Fig. 4 have specific pharmacological properties. For instance, group 1 Stox correspond to β -type anti-mammalian toxins of New World scorpions. Based on the similarities of the amino acid sequences, group 2 and 3, would correspond to β -type toxins, however, variants 1-3 from *C. sculpuratus* have been suggested to be α -type toxins based on pharmacological data (42). Yet, they bind to the fourth site of the Na^+ -channels, while the Old World α -toxins, bind to the third site (50). Group 4 correspond to β -type Stox from which TsVII affects mammals and insects (44). Groups 5-7 correspond to α -type anti-mammalian toxins from Old World scorpions except LghoIT which is a new insect toxin, and is not strictly selective for insects. It is highly toxic to crustaceans and it shows a low but measurable toxicity to mice (20). Group 8 comprises excitatory insect-selective toxins, which induce a repetitive firing in the motor nerves due to an increase in peak Na^+ currents and a voltage-dependent slowing of Na^+ current inactivation (45). Finally, group 9 comprises depressive insect-specific toxins which block action potentials by depolarizing axonal membranes and suppressing Na^+ currents (46).

A new insect-specific STox (AaHIT4), has been recently isolated from *Androctonus australis* Hector scorpion venom (47). It is toxic to insects and competes with insect-specific toxins for binding to the Na^+ channel of insects. It also modulates the binding of α -type and β -type mammalian-specific STox to the mammalian Na^+ -channel (47). It is interesting to observe that this toxin (group 1) shares some similarities with New World and Old World toxin sequences. It is striking to observe a 46% similarity at the aa level with toxin Cn5 (encoded by *Cng5II*). It can be surmised that AaHIT4 has structural characteristics of both α - and β -type Stox. In fact it can be observed from Fig. 4 that it might have both loops (B and J) although smaller than the typical ones. A New World Stox sharing sequence similarities with Old World Stox is toxin V of *C. sculpuratus* Ewing. It has a structure intermediary between α and β Stox but it has been demonstrated that it has electrophysiological characteristics of an α toxin (42). It would be interesting to explore the structural characteristics of these particular Stox to study which domains determine their particular properties. Another similar dual specificity has been observed for a β Stox (TsVII), which was found to compete with typical Stox for binding on both insect and mammalian synaptosomes of nervous tissues (44). This dual specificity was attributed to a high molecular flexibility which would enable it to bind to both the mammalian β -type and insect Na^+ -channel receptors (48). What AaHIT4 and TsVII Stox have in common seems to be a molecular flexibility putatively attributed to a lack of Pro residues in their C-terminal regions (47-48). In this regard, it is interesting

to note that among New World STox, there is a variability in the richness of Pro residues in the last 15 aa residues of the C-terminal region. *C. sculpturatus* variants 1-3 and toxins, encoded by *CngtII* and *CngtIII* cDNAs (group 2 STox), are rich in Pro residues, while group 4 (*Tityus serrulatus*) STox have only one Pro residue in this stretch within the C-terminal region. *C. sculpturatus* variants 1-3 (42) and toxin Cn5 (unpublished results), affects crustaceans and insects but not mammals. Toxins of group 1 with a presumably very similar three dimensional structure and a relatively conserved sequence with regard to CSEV3, affect mammals. Also a notorious variability between groups 1 and 2 (Fig. 4) is found within the last fifteen residues of the carboxyl-terminus, in which a variability in Pro richness is also observed. These observations support the proposition made by Loret et al. (47-48), in which one of the elements involved in the specificity of Stox could be the carboxyl-terminus, mainly regarding the presence of prolines. Nevertheless the relationship between Pro richness, molecular flexibility and specificity for these STox still remains to be definitively established. Furthermore, the various species specificity differences found among mammalian, insect and crustacean toxins are not necessarily due exclusively to the Stox structures. It might well be due to the diversity of Na⁺ channels present in distinct tissues and organisms. There is a wide variation of types and sub-types of Na⁺ channel molecules (49-50). Fig.4 also shows that Stox affecting Na⁺ channels, have a polypeptide chain comprising 60-70 aa residues and cross-linked by four disulfide bonds (51). They share similarities in their aa sequences and in their three-dimensional structures as it has been demonstrated by X-ray diffraction (52-53), two-dimensional NMR (43) and modeling with computer graphic techniques (54). From these studies, it is clear that scorpion toxins affecting Na⁺ channels, have a similar three-dimensional structure with a globular core, comprising three strands of antiparallel β -sheet and one stretch of α -helix. The helix motif is linked to the carboxyl-proximal strand of the sheet by two of the four disulfide bonds. The cysteine pair of the α -helix domain establishing this two disulfide bonds is spaced by a tripeptide while the pair of cysteine residues of the carboxyl-proximal strand is spaced by only one aa residue. These structural elements are conserved in all toxins shown in Fig.4. Similar elements have been observed in other toxins that affect ionic channels and in other non related proteins (55-57). A third structurally conserved disulfide bond is established between the cysteine residue located at the middle strand of the β sheet and the second most amino proximal cysteine residue, except for the excitatory insect toxins (group 3), in which there are two contiguous Cys residues in their corresponding middle strand of the β -sheet. One of these establishes a positionally conserved disulfide bond and the other establishes an atypical disulfide bond with cysteine residue at position 64. The three conserved disulfide bonds, participate in the stabilization of a structurally conserved core of scorpion toxins. This structural core is also conserved in Stox affecting K⁺-channels (58). The fourth disulfide bond is established between the most amino and carboxyl-proximal cysteine residues except in group 3 toxins in which it is established between cysteine residues 38 and 64 (43). This difference however, has no substantial effect on the overall conserved folding of this kind of toxins. The regions corresponding to these structural domains are indicated in Fig.4. The inspection of the sequences shown in this figure, permits to suppose that they all adopt an evolutionary conserved folding. A closer analysis indicates that

the different molecules can accommodate insertions deletions and mutations which apparently confer them different specificities. Old World Stox (except group A and A&HT4) have a B loop and a J gap. New World Stox (except CsEV) have the opposite, a B gap and a J loop. Basically, Old World Stox are considered α -Stox and New World Stox are considered β -Stox. However, Cs toxin V and variants 1-3 are pharmacologically α -type but "structurally" β -type. Taking together all these observations, it seems that specificity is located on the loops of these Stox. In α -toxins in which the B loop is relatively conserved, it could be expected that specificity depended on the C-terminus. In the case of excitatory insect-specific Stox (group B), their specificities seem to depend on their atypical disulfide bridge and on their long C-terminus. In the case of LghoIT, increased insect selectivity could be possibly attributed to an extra Pro-residue (residue 56) at the carboxyl-terminus and/or to a mutation that changed Pro residue 41 for Lys, compared to A&HII. Finally, insect selectivity in group 9 Stox seems to depend on their shorter amino terminus and/or atypical J gap and B loop. Although some indicative features of structure-specificity relationship are emerging, much more data will be required in order to acquire a more complete understanding of the structure-function relationship of Stox. Molecular cloning of Stox genes, and selected site-directed mutagenesis should certainly help clarifying some of the structural and functional relationship discussed here.

CONCLUSIONS

Complementary DNA and genomic cloning of Stox genes, specific for Na^+ -channels, has allowed us to deduce the structure of the transcription units encoding toxins from the scorpions of the genus *Tityus* and *Androctonus*. It seems that transcription of these genomic regions generates pre mRNAs of approximately 800nt containing an intron of approximately 470 nt, located close to the end of the region encoding the signal peptide. The processing of these pre mRNAs might give a mature molecule of approximately 330 nt. The translation of the mRNAs give polypeptide precursors of approximately 82-89 aa residues, which in some cases are processed at both ends (amino and C-terminus). At the amino terminus, a 18-21 aa signal peptide is released by a signal peptidase. At the C-terminus, basic extra residues (Lys or Arg) are eliminated by a carboxypeptidase. Finally, when a glycine residue precedes one or two of these extra basic residues, the amino group of this glycine is utilized to amidate the preceding residue. Similarities in structure and aa sequences provided support for the suggestion that Stox genes have evolved from an ancestral gene which seems to have undergone several duplications and each duplicate have evolved independently. Independent evolution of each duplicate by means of insertions, deletions and mutations at different levels and sites, can explain Stox variability in terms of aa sequence and specificity of action.

ACKNOWLEDGEMENTS

This work was partially supported by grants from Howard Hughes Medical Institute No. 75191-527104, DGAPA-UNAM No. 205893, CONACYT No. 0018-N9105 and No. 1105-N9201. The technical assistance of Mr. Fernando Zamudio, Timoteo Olamendi-Portugal and Fredy Coronas-Valderrama is greatly appreciated

REFERENCES

- Miranda, F., Kopeyan, C., Rochat, C. and Lissitzky, S., Purification of animal neurotoxins. Isolation and characterization of eleven neurotoxins from the venom of the scorpions *Androctonus australis* Hector, *Buthus occitanus tunetanus* and *Leiurus quinquestriatus quinquestriatus*, Eur. J. Biochem., 16:514, 1970.
- Zlotkin, E., Miranda, F., and Rochat, C., in Handbook of Experimental Pharmacology: Chemistry and pharmacology of Buthinae scorpion venoms, edited by S. Bettini, p. 317, Springer-Verlag, Berlin, 1978.
- Possani, L. D., in Handbook of Natural Toxins, edited by A. T. Tu, p. 513, Marcel Dekker Inc, New York, 1984. The contribution of Brazilian scientists on this subject can be seen in reviews by: Diniz, C. R., in Arthropod venoms, edited by S. Bettini, p. 379, Springer-Verlag, Berlin, 1978 and Freire-Maia, L. and Campos, J. A., in Natural toxins, edited by C. L. Ownby and G. B. Odell, p. 139, Oxford: Pergamon Press, 1989. Among other additional informations are: Toledo, D. and Neves, A. G. A., Purification and partial characterization of a second toxin from the scorpion *Tityus serrulatus*, Comp. Biochem. Physiol., 55B:249, 1976; Machado, J. C. and Silveira F., J. F., Obtenção experimental da pancreatite hemorrágica aguda no cão por veneno escorpiónico, Mem. Inst. Butantan, 40/41:1, 1976/77; Novães, G., Catanzaro, D. L., Beraldo, W. T. and Freire-Maia, L., Effect of purified scorpion toxin (tityustoxin) on the pancreatic secretion of the rat, Toxicon 20:847, 1982; Sampaio, S. V., Laure, C. J. and Giglio, J. R., Isolation and characterization of toxic proteins from the venom of the Brazilian scorpion *Tityus serrulatus*, Toxicon 21:265, 1983; Marangoni, S., Ghiso, J., Sampaio, S. V., Arantes, E. C., Giglio, J. R., Oliveira, B. and Frangione, B., The complete amino acid sequence of toxin TsTX-VI isolated from the venom of the scorpion *Tityus serrulatus*, J. Prot. Chem., 9:595, 1990.
- Catterall, W. A., Activation of the action potential Na⁺ ionophore by neurotoxins. An allosteric model, J. Biol. Chem., 252:8669, 1977.
- Couraud, F., Jover, E., Dubois, J. M. and Rochat, H., Two types of scorpion receptor sites, one related to activation, the other to inactivation of the action potential sodium channel, Toxicon 20:9, 1982.

6. Carbone, E., Wanke, E., Prestipino, G., Possani, L. D. and Maelicke, A., Selective blockage of voltage-dependent K⁺ channels by a novel scorpion toxin, *Nature* 296:190, 1982.
7. Possani, L. D., Martin, M. B. and Svendsen, I. B., The primary structure of Noxiustoxin: A K⁺ channel blocking peptide, purified from the venom of scorpion *Centruroides noxius* Hoffmann, *Carlsberg Res. Commun.*, 47:285, 1982.
8. Miller, C., Moczydlowski, E., Latorre, R., and Phillips, M., Charybdotoxin, a protein inhibitor of single Ca²⁺-activated K⁺ channels from mammalian skeletal muscle, *Nature* 313:316, 1985.
9. Gimenez-Gallego, G., Navia, M. A., Reuben, J. P., Katz, G. M., Kaczorowski, G. J. and Garcia, M. L., Purification sequence, and model structure of charybdotoxin, a potent selective inhibitor of calcium-activated potassium channels, *Proc. Natl. Acad. Sci. USA* 85:3329, 1988.
10. Strong, P. N., Weir, S. W., Beech, D. J., Hiestand, P. and Kocher, H. P., Effects of potassium channel toxins from *Leiurus quinquestratus hebraeus* venom on responses to cromakalim in rabbit blood vessels, *Br. J. Pharmacol.*, 98:817, 1989.
11. Sands, S. B., Lewis, R. S. and Cahalan, M. D., Charybdotoxin blocks voltage-gated K⁺ channels in human and murine T lymphocytes, *J. Gen. Physiol.*, 93:1061, 1989.
12. DeBin, J. A., Maggio, J. E. and Strichartz, G. R., Purification and characterization of chlorotoxin, a chloride channel ligand from the venom of the scorpion, *Am. J. Physiol.*, 264 (Cell Physiol. 33):C361, 1993.
13. Valdivia, H. H., Fuentes, O., El-Hayek, R., Morrissette, J., and Coronado, R., Activation of the ryanodine receptor Ca²⁺ release channel of sarcoplasmic reticulum by a novel scorpion venom, *J. Biol. Chem.*, 266:19135, 1991.
14. Wheeler, P. K., Watt, D. D. and Lazdunski, M., Classification of Na channel receptors specific for various scorpion toxins. *Pfluegers Arch.*, 397:164, 1983.
15. Jover, E., Couraud, F. and Rochat, H., Two types of scorpion neurotoxins characterized by their binding to two separate receptor sites on rat brain synaptosomes, *Biochem. Biophys. Res. Commun.*, 95:1607, 1980.
16. Nenner, W., in *Advances in Cytopharmacology*, edited by B. Ceccarelli and F. Clementi, v. 3, p. 345, Raven Press, New York, 1979.
17. Rochat, H., Bernard, P. and Couraud, F., in *Advances in Cytopharmacology*, edited by B. Ceccarelli, and F. Clementi, v. 3, p. 325, Raven Press, New York, 1979.
18. Catterall, W. A., The molecular basis of neuronal excitability, *Science* 223:653, 1984.

19. Kharrat, R., Darbon, M., Rochat, H. and Granier, C., Structure/activity relationships of scorpion α -toxins. Multiple residues contribute to the interaction with receptors. *Eur. J. Biochem.*, 181:381, 1989.
20. Eitan, M., Fowler, E., Herrmann, E., Duval, A., Pelhate, M. and Zlotkin, E., A scorpion venom neurotoxin paralytic to insects that affects sodium current inactivation: purification, primary structure, and mode of action. *Biochemistry* 29:5941, 1990.
21. Meves, H., Simard, M. J., and Watt, D. D., in *Annals New York Academy of Sciences*, edited by C. Y. Kao, and S.R. Levinson, v. 479, p. 113, New York Academy of Sciences, New York, 1986.
22. Strichartz, G., Rando, T. and Wang, G. K. An integrated view of the molecular toxinology of sodium channel gating in excitable cells. *Annu. Rev. Neurosci.* 10: 237, 1987.
23. Thomsen, W. J. and Catterall, W. A., Localization of the receptor site for alpha-scorpion toxins by antibody mapping: implication for sodium channel topology. *Proc. Natl. Acad. Sci. USA* 86:10161, 1989.
24. Possani, L. D., Gurrola, G. B., Portugal, T. O., Zamudio, F. Z., Vaca, L. D., Calderon, E. S. A., and Kirsch, G. E., in *Proceedings of the first Brazilian Congress on Proteins*, edited by B. Oliveira and B. Sgarbieri, p. 352, Editora da UNICAMP, Campinas, Brazil, 1991.
25. Gurrola, G. B., Molinar-Rode, R., Sitges, M., Bayon, A., and Possani, L. D. Synthetic peptide corresponding to the sequence of Noxiustoxin indicate that the active site of this K^+ channel blocker is located on its amino terminal position, *J. Neural Transmission* 77:11, 1989.
26. Zamudio, F., Saavedra, R., Martin, B. M., Gurrola-Briones, G., Herion, P. and Possani, L. D., Amino acid sequence and immunological characterization with monoclonal antibodies of two toxins from the venom of the scorpion *Centruroides noxius* Hoffmann, *Eur. J. Biochem.* 204:281, 1992.
27. Becerril, B., Vázquez, A., García, C., Corona, M., Bolívar, F. and Possani, L. D., Cloning and characterization of cDNAs that code for Na^+ -channel-blocking toxins of the scorpion *Centruroides noxius* Hoffmann, *Gene* 128:165, 1993.
28. Vázquez, A., Becerril, B., Martin, B. M., Zamudio, F., Bolívar, F. and Possani, L. D., Primary structure determination and cloning of the cDNA encoding toxin 4 of the scorpion *Centruroides noxius* Hoffmann. *FEBS-Lett.*, 320:43, 1993.
29. Becerril, B., Corona, M., Mejía, M. C., Martin, B. M., Lucas, S., Bolívar, F. and Possani, L. D., The genomic region encoding toxin gamma from the scorpion *Tityus serrulatus* contains an intron, *FEBS-Lett.*, 335:6, 1993.
30. Bougis, P. E., Rochat, H. and Smith, L. A., Precursors of *Androctonus australis* scorpion neurotoxins, *J. Biol. Chem.* 264:19259, 1989.

31. Zilberberg, N., Zlotkin, E. and Gurevitz, M., The cDNA sequence of a depressant insect selective neurotoxin from the scorpion *Buthotus judaicus*. *Toxicon* 29:1155, 1991.
32. Gurevitz, M., Urbach, D., Zlotkin, E. and Zilberberg, N., Nucleotide sequence and structure analysis of a cDNA encoding an alpha insect toxin from the scorpion *Leiurus quinquestriatus hebraeus*. *Toxicon* 29:1270, 1991.
33. Martin-Euclaire, M. F., Céard, B., Ribeiro, A. M., Diniz, C. R., Rochat, H. and Bougis, P. E., Molecular cloning and nucleotide sequence analysis of a cDNA encoding the main β -neurotoxin from the venom of the South American scorpion *Tityus serrulatus*. *FEBS-Lett.*, 302:220, 1992.
34. Zlotkin, E., Gurevitz, M. and Adams, M. E., Depressant insect selective neurotoxins from scorpion venom: chemistry, action and gene cloning. *Arch. Insect Biochem.*, 22:55, 1993.
35. Park, Ch. S., Hausdorff, S. F. and Miller C., Design, synthesis, and functional expression of a gene for charybdotoxin, a peptide blocker of K^+ channels, *Proc. Natl. Acad. Sci. USA*, 88:2046, 1991.
36. Garcia-Calvo, M., Leonard, R. J., Novick, J., Stevens, S. P., Schmalhofer, W., Kaczorowski, G. J. and Garcia, M. L., Purification, characterization, and biosynthesis of Margatoxin, a component of *Centruroides margaritatus* venom that selectively inhibits voltage-dependent potassium channels, *J. Biol. Chem.*, 268:18866, 1993.
37. von Heijne, G., A new method for predicting sequence cleavage sites, *Nucl. Acids Res.*, 14:4683, 1986.
38. Rochat, H., Rochat, C., Sampieri, F., Miranda, F., and Lissitzky, S., The amino acid sequence of neurotoxin II of *Androctonus australis Hector*, *Eur. J. Biochem.* 28:381, 1972.
39. Mains, R. E., Dickerson, I. M., May, V., Stoffers, D. A., Perkins, S. N., Ouafik, L., Husten, E. J. and Eipper, B. A., in *Frontiers in Neuroendocrinology*. edited by L. Martini and M. F. Ganong, Vol. 11, p. 52, Raven Press, Ltd., New York, 1990.
40. Martin-Euclaire, M. F., Delabre, M. L., Ceard, B., Ribeiro, A., Sogaard, M., Svensson, B., Diniz, C., Smith, L. A., Rochat, H. and Bougis, P. E., in *Recent Advances in Toxinology Research*, edited by P. Gopalakrishnakone and C.K. Tan, Vol 1, p. 156, National University of Singapore University, Singapore, 1992.
41. Stamps, P., Kolmakova-Partensky, L., and Miller, C., Intimations of K^+ channel structure from a complete functional map of the molecular surface of charybdotoxin, *Biochemistry* 33:443, 1994.
42. Mevas, H., Simard, M. J., and Watt, D. D., Biochemical and electrophysiological characteristics of toxins isolated from the venom of the scorpion *Centruroides scujpturatus*, *J. Physiol. (Paris)* 79:185, 1984.

43. Darbon, H., Weber, C. and Braun, W., Two-dimensional ^1H nuclear magnetic resonance study of ASHIT, an anti-insect toxin from the scorpion *Androctonus australis* Hector. Sequential resonance assignments and folding of the polypeptide chain, *Biochemistry* 30:1836, 1991.
44. Lima, M. E., Martin, M. F., Hue, B., Loret, E. P., Diniz, C. R. and Rochat, H. On the binding of two scorpion toxins to the central nervous system of the cockroach *Periplaneta americana*, *Insect Biochem.*, 19:413, 1989.
45. Felhate, M. and Zlotkin, E., Voltage-dependent slowing of the turn off of Na^+ current in the cockroach giant axon induced by the scorpion venom "insect toxin", *J. Physiol. (Lond.)* 319:30P, 1981.
46. Zlotkin, E., Kadouri, D., Gordon, D., Felhate, M., Martin, M. F. and Rochat, H. An excitatory and depressant insect toxin from scorpion venom both affect sodium conductance and possess a common binding site, *Arch. Biochem. Biophys.*, 240:877, 1985.
47. Loret, E. P., Martin-Eauclaire, M. F., Mansuelle, P., Sampieri, F., Granier, C. and Rochat, H. An anti-insect toxin purified from the scorpion *Androctonus australis* Hector also acts on the α - and β -sites of the mammalian sodium channel: sequence and circular dichroism study, *Biochemistry* 30:633, 1991.
48. Loret, E. P., Sampieri, F., Roussel, A., Granier, C. and Rochat, H., Conformational flexibility of a scorpion toxin active on mammals and insects: a circular dichroism study, *Proteins* 8:164, 1990.
49. Noda, M., Ikeda, T., Kayano, T., Suzuki, H., Takeshima, H., Kurasaki, M., Takahashi, H. and Shosaku, N., Existence of distinct sodium channel messenger RNAs in rat brain, *Nature* 320:188, 1986.
50. Catterall, W. A., Structure and function of voltage-sensitive ion channels, *Science* 242:50, 1988.
51. Kopeyan, C., Martinez, G., Lissitzky, S., Miranda, F. and Rochat, H. Disulfide bonds of toxin II of the scorpion *Androctonus australis* Hector, *Eur. J. Biochem.* 47:483, 1974.
52. Fontecilla-Camps, J. C., Almassay, R. J., Suddath, F. L., Watt, E. D. and Bugg, C. E. Three dimensional structure of a protein from scorpion venom: a new structural class of neurotoxins, *Proc. Natl. Acad. Sci. USA* 77:6496, 1980.
53. Fontecilla-Camps, J. C., Harbersetzer-Rochat, C. and Rochat, H., X-ray crystallographic structure of the scorpion toxin II of the potent toxin II from the scorpion *Androctonus australis* Hector, *Proc. Natl. Acad. Sci. USA* 85:7443, 1988.
54. Fontecilla-Camps, J. C., Three-dimensional model of the insect-insect scorpion toxin from *Androctonus australis* Hector and its implication for the evolution of scorpion toxins in general, *J. Mol. Evol.*, 29:63, 1989.

55. Bontems, F., Roumestand, Ch., Gilquin, B., Ménez, A. and Toma, F., Refined structure of charybdotoxin: Common motifs in scorpion toxins and insect defensins, *Science* 254:1521, 1991.
56. Kobayashi, Y., Takashima, H., Tamaoki, H., Kyogoku, H., Lambert, P., Kuroda, H., Chino, M., Watanabe, T., Kimura, T., Sakakibara, S. and Moroder, L., The Cysteine-Stabilised α -Helix: A common Structural Motif of Ion-Channel Blocking Neurotoxic Peptides, *Biopolymers* 31:1213, 1991.
57. Bruix, M., Jiménez, M. A., Santoro, J., González, C., Colla, F. J. and Rico, M., Solution structure of 1-H and 1-P thionins from barley and wheat endosperm determined by $^1\text{H-NMR}$: A structural motif common to toxic arthropod proteins, *Biochemistry* 32:715, 1993.
58. Ménez, A., Bontems F., Roumestand, Ch., Gilquin, B. and Toma, F., Structural basis for functional diversity of animal toxins, *Proceedings of the Royal Society of Edinburgh* 99B: 83, 1992.
59. Zlotkin, E., Eitan, M., Adams, M. E., Meyer, M., Burkhart, W. and Fowler, E., Functional duality and structural uniqueness of depressant insect-selective neurotoxins, *Biochemistry* 30: 4814, 1991.

An insect-specific toxin from *Centruroides noxius* Hoffmann cDNA, primary structure, three-dimensional model and electrostatic surface potentials in comparison with other toxin variants

Barbara SELISKO¹, Consuelo GARCIA¹, Baltazar BECERRIL¹, Muriel DELEPIERRE² and Lourival D. POSSANI¹

¹ Department of Molecular Recognition and Structural Biology Institute of Biotechnology, Universidad Nacional Autónoma de México, Cuernavaca, México

² Laboratory of Nuclear Magnetic Resonance (CNRS URA 1129) Pasteur Institute, Paris, France

(Received 5 July 1996) – EJB 96 10003

Scorpion toxins acting on sodium channels differ in their specificity. Toxic peptides specific towards mammals and arthropods (insects and/or crustaceans) have been described. Because of the similar three-dimensional fold of these peptides, the molecular base of their specificity is thought to reside in certain differences at the level of amino acid residues especially within or near the binding site of the toxin to the particular ion channel. The cDNA, amino acid sequence and biological activity of an insect-specific toxin, Cn10, from the scorpion *Centruroides noxius* Hoffmann is reported. The electrostatic potential surface around a three-dimensional model of Cn10 was calculated. It revealed that residues Tyr4, Lys13, Ile18, Leu19, Gly20, Lys43, Leu44, Thr57, Tyr58, Pro59, Thr64 and Cys65, situated at the side of the toxin proposed in the literature to bind to the sodium channel, constitute a positive surface region. Therefore, they may form the site that binds to the channel. Cn10 was included in a comparative analysis of two groups of natural variants, highly similar peptides of the genus *Centruroides* with specificities towards mammals or arthropods. A number of surface-accessible residues, consistently different between the two groups and situated near the putative binding site, may be of importance for the specificity of the analyzed toxins.

Keywords: amino acid sequence; *Centruroides noxius*; scorpion toxin; sodium channel; three-dimensional structure.

Scorpion neurotoxins are basic proteins of low molecular mass which affect ion channels of excitable cells (Possani, 1984). The toxins acting on sodium channels (Lazdunski et al., 1986) characterized to far consist of 60–70 amino acids. They contain four disulfide bridges. The existing three-dimensional structures [Fontecilla-Camps et al., 1980 (first structure solved); Darbon et al., 1991; Zhao et al., 1992; Housset et al., 1994; Jablonsky et al., 1995; Lebreton et al., 1994; Lee et al., 1994a,b] provide strong evidence that they are all folded in a similar way forming a core region of one short α -helix of about 2.5 turns and a three-stranded anti-parallel β -sheet. The core is stabilized by three disulfide bridges shared by all known structures. Two of the core disulfides couple the conserved segment -Cys-Xaa-Xaa-Cys- (where Xaa is variable) of the α -helix to the middle β -strand containing the segment -Cys¹-Naa-Cys² forming a structural motif found also in other cysteine-rich proteins like

insect defensins and plant defense proteins (Bontems et al., 1991; Bruix et al., 1993).

Despite their highly conserved tertiary structure, the toxins acting on sodium channels have shown rather large differences in their functional properties on all three levels in which the function of toxins can be tested: in species specificity and degree of toxicity, in their binding affinities to excitable tissues and in their direct electrophysiological effects on sodium channels. According to their specificity sodium-channel-affecting toxins have been divided into toxins that affect mammals or arthropods (insects and/or crustaceans). Some toxins were reported to affect species of both groups (Eitan et al., 1990; Gurevitz et al., 1991; Gordon and Zlotkin, 1993; De Lima et al., 1986; Loret et al., 1991). Recently, a study of the binding and toxicity of various toxins towards mammals and insects revealed that even toxins known as mammal-specific can show effects towards insects and vice versa (Gordon et al., 1996). Therefore, the concept of specificity seems to be rather a matter of relative toxicities than of exclusive effects.

The molecular basis of toxin specificity has been studied widely for various reasons: (a) in order to understand the toxic effect of scorpion venom in mammals as a prerequisite for the development of more efficient and secure antidotes or vaccines; (b) to study the target molecules of scorpion toxins, i.e. ion channels in excitable tissues, in order to understand their molecular structure and function; (c) to develop insect-specific toxins that can be used as insecticides. However, the existing knowledge is still too limited to give clear answers. Rather small dif-

Correspondence to L. D. Possani, Department of Molecular Recognition and Structural Biology, Institute of Biotechnology/UNAM, Apartado Postal 510–3, Cuernavaca, Morelos 62250, México

Fax: +52 73 172388.

Abbreviations. CIII, *Centruroides infamatus infamatus* toxin 1; CIII1, CIII2 and CIII3, *Centruroides limpidus limpidus* toxins 1, 2 and mammalian toxin 1; CIII1, *Centruroides limpidus texmexanus* toxin 1; Cn2, Cn3, Cn4, Cn5 and Cn10, *Centruroides noxius* toxins 1, 3, 4, 5 and 10, respectively; CnE1, CnE2 and CnE3, *Centruroides sculpturatus* Ewing variants 1, 2 and 3, respectively.

Note. The novel amino acid sequence data published here was deposited with the EMBL sequence data bank and is available under accession number Y08270.

ferences, i.e. single amino acid modifications within or near the critical binding region of the toxin to its receptor, seem to be very important. In order to find these crucial residues one can take two approaches: (a) a theoretical approach by comparing the structure of natural mutants, i.e. highly identical toxins that display differences in their specificity, or (b) an experimental approach, to conduct mutational analysis changing putative critical residues of a specific toxin. The first approach should guide the second, facilitating the design of these point mutation experiments.

In the study presented here we follow the first approach. First, we report the finding of a new toxin of the Mexican scorpion *Centruroides nuxius* Hoffmann that shows specificity towards insects. We have generated a model of its three-dimensional structure and calculated the electrostatic charge distribution on the surface of the molecule (Hong and Siegel, 1995). We then included the new toxin in comparative analysis of mammal-specific versus arthropod-specific toxins acting on sodium channels from the genus *Centruroides* in order to localize residues that might be responsible for the specificity of these toxins.

MATERIALS AND METHODS

Venom and reagents. Scorpion-venom of the species *Centruroides nuxius* Hoffmann were collected in the state of Nayarit, Mexico. The venom was obtained by electrical stimulation of the anaesthetized animals as described in Possani (1984). The dried venom was resuspended in distilled water and centrifuged at 10000×g for 15 min. The supernatant containing the soluble venom was freeze-dried and kept at -20°C until use.

Only analytical grade reagents and distilled water (over quartz) were used throughout.

Isolation and sequencing of cDNA. The construction of the cDNA library from telsons of the scorpion *C. nuxius* Hoffmann, as well as the screening for sequences of putative arthropod-specific toxins acting on sodium channels, has been described in detail by Becerra et al. (1993). As a result, we obtained 13 positive clones. One of these clones was amplified by PCR (GeneAmp PCR System Perkin Elmer, 30 rounds of temperature cycling: 92°C for 1 min, 50°C for 1 min, 72°C for 3 min followed by a final period of 10 min at 72°C) using α gt11 forward primer (5'-GGTGGCCAGCAGCTCTGGAGCCCG) and β gt11 reverse primer (5'-TTGACACAGCACTGGTAAATG) and Vent® DNA polymerase (BioLabs). The PCR product was purified from the gel, cut by EcoRV restriction endonuclease (BioLabs) and subcloned into the EcoRV site of the phagemid BlueScript II KS+ (Stratagene). Subsequently, competent *Escherichia coli* DH5- α cells (Gibco BRL) were transformed. The plasmid DNA of white colonies with an insert of the expected size of around 400 bp was isolated and the inserts sequenced using the Sequenase kit T7 DNA polymerase (US Biochemical) and the above α gt11 primers.

Toxin purification. The first and second step of the toxin purification have been described in detail by Nalvisia et al. (1994). Briefly, the venom was initially purified by size-exclusion chromatography on a Sephadex G-50 column. The second purification step consisted of ion-exchange chromatography on CM-cellulose. The pooled fractions were tested for toxicity. Peak 5 from the second purification step was then further purified by HPLC on a semi-preparative C₈ reverse-phase column (see figure legends). Single peaks were collected and tested for toxicity in mammals, insects and crustaceans. The purification was followed by measuring the absorbance at 230 nm. Recovery was calculated based on the absorbance.

Sequence determination. For sequence determination, the toxin Cn10 was purified, reduced and carboxy-methylated as described by Possani et al. (1985). It was then digested by the endoproteases V8 (Boehringer, Mannheim) in 100 mM ammonium bicarbonate and Asp-N (Boehringer, Mannheim) in 50 mM sodium phosphate pH 8.0 using in both cases a toxin/endoprotease ratio of 200:2 μ g. The resulting peptide mixtures were purified on an analytical C₈ reverse-phase column. N-terminal amino acid sequence analysis of single peaks was obtained by Edman degradation (Edman and Begg, 1967) using a microsequencer model 6400/6600 (MilliGen/Bioscience).

Letality tests. Letality was tested in mammals (mice), insects (crickets, *Acheta* spp.) and crustaceans (crayfishes, *Cambarellus montezumae* spp.), detecting whether a given quantity of toxin was non-toxic, toxic or lethal. Non-toxic meant that the animals showed no symptoms of intoxication within 24 h from the injection, similarly to having been injected with phosphate-buffered saline solution (NaCl/P, = 137.0 mM NaCl, 2.0 mM KCl, 1.0 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4) or water alone. Toxic meant that they showed symptoms of intoxication like excitability, salivation, lacrimation, dyspnoea, temporary paralysis of limbs (mice) or paralysis (crickets and crayfishes). The toxic effect was assigned as lethal if at least one of the test animals died after injection.

Each concentration (mass/average body mass) was tested in parallel with four (in the case of mice) to ten test animals. Female albino mice (average body mass 20 g) of the strain CD 1 (approx. 6 weeks old) were injected intraperitoneally with up to 500 μ g freeze-dried material in 200 μ l NaCl/P. Crickets were injected with 10–50 μ g/animal (average body mass 0.5 g) and freshwater crayfishes with 60–110 μ g/animal (average body mass 1 g) in 20 μ l distilled water.

Modelling and analysis of the three-dimensional structure. The modelling, calculation of energy minimization, molecular dynamics and analysis of the model were performed on a Silicon Graphics IRIS4D35 workstation using the graphic display programme INSIGHT II and the HOMOLOGO, BIOPOLYMER, DISCOVER and ANALOGY packages from Biosym Technologies. The models were built on the basis of known structure of the toxins Cn5-E (Zhuo et al., 1992), Cn11 (Lebreton et al., 1994) and Cn5 (Lebreton, F. Garcia, C. Possani, L. D. and Delpeierre, M., unpublished) which are 78%, 77% and 79% identical to Cn10, respectively. The model structures were extracted from the Protein Data Bank (Bernstein et al., 1977) for Cn5-E (code 25X3), or provided by the authors (Lebreton et al., 1994; Lebreton, F. Garcia, C., Possani, L. D. and Delpeierre, M., unpublished). The sequences were replaced by the sequence of Cn10. The resulting starting structure did not show any unavourable overlapping within a distance of 0.05 nm, was enclosed by a thick shell of a water layer. The structures were optimized with the consistent valence force field using the maximum derivative and the total energy to monitor each step of the modelling. As a first step of minimization the water shell and the hydrogens of the molecule were minimized (all heavy atoms) with steepest descent down to 334.72 kJ/nm and conjugate gradient down to 83.68 kJ/nm. As the following step, an unrestrained minimization was performed with steepest descent down to 334.72 kJ/nm and conjugate gradient with a maximum derivative of 41.84 kJ/nm. Molecular dynamics was performed using a Verlet leapfrog algorithm: the resume phase at 300 K was 20 ps (time step 1 fs) after an equilibration phase of 1 ps to reach constant temperature. The structure given by the last frame of the trajectory constituting an energy minimum was then further minimized to 4.184 kJ/nm in the final step.

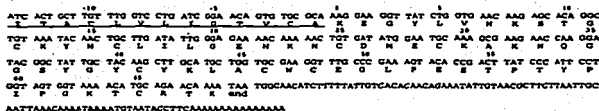


Fig. 1. cDNA sequence of toxin Cn10 and its deduced amino acid sequence. The signal peptide is underlined, the polyadenylation site is printed in bold letters.

Surface accessibilities were determined using the graphic display programme INSIGHT II which calculates the total accessible surface area of single amino acid residues in a protein according to Connolly (1983). Relative surface accessibilities were calculated referring to the total accessible surface area of the corresponding free amino acid.

The electrostatic potential field distribution around the toxins was calculated using additionally to the above-mentioned programmes the DELPHI package from Biosym Technologies (Nicholls and Honig, 1991). The programme uses a macroscopic approach in which the protein and solvent are assigned different dielectric constants (2.0 and 80.0, respectively); the three-dimensional structure of the protein defines the boundaries. We used the default parameters given by DELPHI with the exception of a border space of 1.5 nm. Charge assignments were made before the calculation, setting the pH of the medium to 7.0 using the BUILDER module. The potential surface is calculated by solving numerically the linear form of the Poisson-Boltzmann equation.

RESULTS

Cloning. In order to extend the number of known arthropod-specific toxins from scorpions of the genus *Centruroides* we searched a cDNA library prepared from telsons of the scorpion *Centruroides noxius* Hoffmann using as a probe a cDNA clone which was 98% identical to the C-terminal segment (last 41 residues) of the arthropod-specific toxin CnE3.3 of *Centruroides sculpturatus* Ewing (Becerril et al., 1993). In the second step of the screening we used an oligonucleotide which corresponds to amino acids 10–16 of the same toxin CnE3.3 (Becerril et al., 1993). The search resulted in 13 positive clones of which one was subcloned and sequenced. Fig. 1 shows the sequence of the resulting cDNA clone and its deduced amino acid sequence.

Toxin purification and sequencing. Provided with the knowledge of the new sequence we tried to isolate the toxin from the venom as described in Materials and Methods. The first purification step was performed using size-exclusion chromatography on Sephadex G-50. The elution profile was shown by Valdivia et al. (1994); in total, four fractions were detected of which only the second fraction showed a toxic effect in mice. The second step, ion-exchange chromatography on carboxymethyl-cellulose, resulted in 14 peaks. For the elution profile we refer again to Valdivia et al. (1994). Fraction 5 of this profile was found to be toxic to insects and crustaceans, whereas no toxic effect was detected in mice. This fraction was then further purified by HPLC using reverse-phase chromatography as shown in Fig. 2. Automated Edman degradation indicated that peak 3 contains the new toxin.

Fig. 3 shows the determined amino acid sequence of the toxin which was named Cn10 because it is the tenth distinct

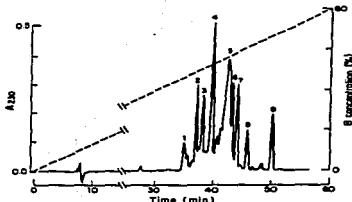


Fig. 2. Reverse-phase HPLC as the third step of the chromatographic purification of Cn10. Freeze-dried material of fraction II-5 (2.2 mg) from the second purification step (Valdivia et al., 1994) was purified by HPLC on a semi-preparative C₁₈ reverse-phase column (25 cm × 1 cm, particle size 10 μm; Vydac) eluted for 60 min with a linear gradient generated by mixing solution A (0.12% trifluoroacetic acid in water) and solution B (acetonitrile with 0.1% trifluoroacetic acid) at a flow rate of 2 ml/min. Peak 3 corresponds to Cn10.

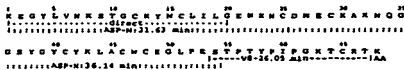


Fig. 3. Determination of the primary structure of Cn10. Amino acids 1–20 were determined by direct sequencing, 1–25 and 26–54 by sequencing the peptide fragments generated by endoprotease Asp-N, 54–67 by sequencing one peptide fragment generated by endoprotease V8. The peptides were purified on an analytical C₁₈ reverse-phase column (25 cm × 0.46 cm, particle size 5 μm; Vydac) with a linear gradient generated by mixing solution A (0.12% trifluoroacetic acid in water) and solution B (acetonitrile with 0.1% trifluoroacetic acid). The flow rate was 1 ml/min. The retention times of the fragments are given. Amino acid 68 was determined by amino acid analysis (AA).

peptide sequence of a putative toxin, acting on sodium channels, of *C. noxius*. The first 20 amino acids were determined by direct sequencing of the reduced and carboxymethylated toxin. The remaining amino acids up to residue 67 were identified by sequencing peptides which resulted from digestion of the denatured toxin with endoproteases Asp-N and V8. Lys68 could not be detected by automated Edman degradation, a common result for the last residue of a peptide due to its desorption from the sequelon membrane. However, the clear detection of the pe-

Table 1. Toxicities of selected toxins from scorpions of the genus *Centruroides*. Data of Cn10, Cn5 and Cn2 from *Centruroides variolus* from this paper; Garcia, C., Becerril, B., Selisko, B., Delapierre, M. and Possani, L. D. (unpublished) and Zambullo et al. (1992), respectively; C111 from *Centruroides limpidus limpidus* from Lobreton et al. (1990); CxEv1, CxEv2 and CxEv3 from *Centruroides sculpturatus* Ewing (Babin et al., 1974). ED₅₀ stands for effective dose (50% of the test animals show symptoms of intoxication), LD₅₀ for lethal dose (50% of the test animals died of intoxication). The average body mass of mouse was 20 g, cricket 0.5 g, crayfish 1 g and chick 350 g.

Toxin	Test animal	Toxicity dose/average body mass			
		non-toxic	toxic	non-lethal	lethal
		μg			
Cn10	mouse	<500	>13		
	cricket		>90	<110	>50
Cn5	mouse	<570	>40	<180	
	cricket				28.5
C111	mouse	<500			
	cricket	< 30	>30		
CxEv1	chick		>25000		5.5
	cricket		>25000		
CxEv2	chick		>25000		
	cricket				10.7
CxEv3	chick		>25000		
	cricket				125
Cn2	mouse				0.4

ultimate residue Thr67 and the detection of an additional Lys by analyzing the amino acid composition of the C-terminal peptide from the V8 digestion established the presence of Lys68 in the mature peptide.

Biological tests. The lethality of the toxin was then tested in mice, crickets and crayfishes (Table 1). We found that Cn10 is not toxic to mice applying up to 500 μg/20 g average body mass. In this respect it is interesting to note that a scorpion of the species *C. notius* Hoffmann releases about 50 μg protein/venom extraction (Possani, 1984). To test the toxic effects in crickets, we applied various doses of 10–50 μg/animal (average body mass 0.5 g). The onset of toxicity was observed at a dose of 15 μg/animal at a dose of 50 μg out of four crickets died within 6 h. To test the toxicity of Cn10 in crayfishes we used doses of 60–110 μg/animal (average body mass 1 g). Toxic symptoms started at 90 μg/animal but doses up to 110 μg/animal did not cause a lethal effect in crayfishes.

Modelling of the three-dimensional structure. A three-dimensional model of Cn10 was generated in order to calculate the electrostatic surface potential map around Cn10 and to localize amino acid variations in Cn10 in comparison to other toxins. We used three structures as bases of the modelling to be able to compare the resulting structures: the crystallographic structure of the toxin CxEv3 of *C. sculpturatus* Ewing (Zhao et al., 1992) and the two NMR structures of C111 of *Centruroides limpidus limpidus* (Lobreton et al., 1994) and Cn5 of *C. notius* Hoffmann (Lobreton, F., Garcia, C., Possani, L. D. and Delapierre, M., unpublished) determined by our group. The models were estab-



Fig. 4. Ribbon diagram of the three-dimensional model of Cn10 presenting the front of the molecule as proposed by Fontecilla-Camps et al. (1980) generated by modelling from Cn5. Only the backbone of the molecule is shown; coloured in white are the α -helix orientated to the back of the molecule and the β -sheet situated on the front side of Cn10 (it includes residues 37, 38 and 49, 30). Additionally, the four disulfide bridges (Cys12-Cys65, Cys16-Cys41, Cys25-Cys64, Cys29-Cys48) are shown in yellow. The C- and the N-terminal are indicated. Residues 26, 32, 35, 37, 38, 49, 50, 57 and 58, which display significant differences between mammal-specific and arthropod-specific toxins (see Discussion and Fig. 6), are represented in red with their side chains.

lished following the protocol outlined in Materials and Methods. Fig. 4 is a stereoview of the Cn10 model derived from Cn5. We will refer to the part of the molecule facing the viewer as the front of the toxin, as proposed by Fontecilla-Camps et al. (1980). The secondary-structure elements, an α -helix including residues 23–30 on the upper part of the back of the toxin and a three-stranded β -sheet consisting of residues 2–4, 37–42 and 46–50 which is situated on the upper part of the front, are retained in this model as in the two models from CxEv3 and C111. The superposition of the three models showed that this core region consisting of the elements of secondary structure superimposes very well (rmsd in these positions when superimposing the trace of amino acids 1–65 are in the range 0.07–0.099 nm) whereas larger differences appear in regions 6–15, 32–37 and especially in the C-terminal 50–68. In order to choose the model which shows the easiest accommodation of the mutations into the underlying structure, we measured the rmsd between the starting and the final structure of each modelling process. Superimposing the entire trace the rmsd of the C α atoms were 0.112 nm, 0.130 nm and 0.133 nm for the models from Cn5, CxEv3 and C111, respectively. In the light of these data, we consider the model structure derived from Cn5 to be the best. The C-terminal sequence after the disulfide bridge 12–65 consists of Cn5 and C111 of only one residue and is totally absent in CxEv3 whereas Cn10 continues up to residue 68. The accommodation of the three C-terminal residues 66–68 is therefore rather arbitrary, which should be considered when comparing structural features of Cn10 with experimentally established structures.

Calculation of the electrostatic surface potential map. Because of the apparent importance of positive residues for the binding of toxins to channels that transfer positively charged ions (Stampe et al., 1993; Goldstein et al., 1994; Hidalgo and MacKinnon, 1995; Lorez et al., 1994; Krezel et al., 1995) we calculated the electrostatic field distribution around Cn10 using

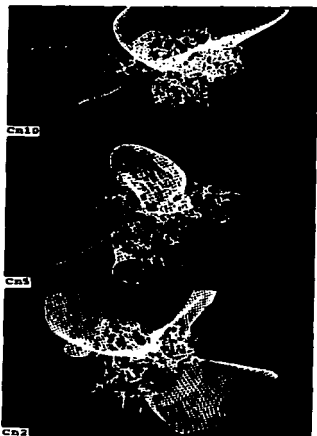


Fig. 5. Electrostatic potential field distribution at the front sides of toxins Cn10 (top), Cn5 (middle) and Cn2 (bottom). Red contours represent the negative equipotential surface at 1 kT/e, blue contours the corresponding positive surface at 1 kT/e and white contours the neutral equipotential surface of 0 kT/e. The proteins trace and side chains without hydrogens are shown in yellow.

the DELPHI program (Nicholls and Honig, 1991). Fig. 5 (top) shows the surface potential map of the front of Cn10. There is a large continuous positive region (shown in blue) on the right side of the molecule comprising the right side of the β -sheet around the loop between the second and the third β -strand and the C-terminal (refer to Fig. 4 for secondary-structure elements). It continues along the left back side of the molecule disturbed by a neutral patch (shown in white) around residues 57–61. The central part of the β -sheet displays an electronegative surface (shown in red) that continues towards the top to the central part of the α -helix.

DISCUSSION

Cloning and sequencing of the purified peptide. In this paper we report the cDNA and amino acid sequence of a novel toxin from the scorpion *Centruroides nexius* Hoffmann which we named Cn10. The cDNA of Cn10 comprises 346 nucleotides including the polyadenylation site. A comparison of the signal peptide with highly similar signal peptides of cDNAs of related

toxins (Becerril et al., 1993) showed that it may be truncated at amino acid -13. The complete signal peptide should consist of 18–21 amino acids. Because of our intention to finally express Cn10 as a fusion protein in *E. coli* cells in order to produce the recombinant toxins and its mutants, we decided not to search for the complete signal peptide sequence.

Biological tests. A sequence comparison with other toxins of the genus *Centruroides* (Fig. 6) shows that Cn10 shares 53% similarity to a consensus sequence of toxins against mammals and 80% similarity to a consensus sequence of toxins against arthropods. The subsequent biological tests proved that Cn10 belongs to the latter group. Table 1 compares the toxicity data of Cn10 with data of other selected toxins of the genus *Centruroides*. In the light of this data Cn10 can be regarded as a toxin specific towards arthropods with a slightly higher activity towards insects. Other toxins of the arthropod-specific group are Cn5 (García, C., Becerril, B., Selisko, B., Delapierre, M., and Pössani, L. D., unpublished), C111 (Lebreton et al., 1994) as well as CsEv1, CsEv2 and CsEv3 (Babin et al., 1974) which were not tested in crustaceans. Cn2 (Zamudio et al., 1992) is given as an example of a toxin with high specificity towards mammals. Further classification of the arthropod-specific toxins into insect-specific and crustacean-specific sub-groups is difficult because the existing data are not complete and sometimes not compatible. However, we could classify Cn10 together with CsEv1 and CsEv2 as specific towards insects whereas Cn5 and C111 seem to be more specific towards crustaceans. For the purpose of this discussion we should refer to specificity more as a matter of relative toxicity than of exclusive effects. It is worth noting that, in the event of prey hunting, the scorpion will not deliver only one toxin since the venom is a complex mixture of toxins, which present different activities, specificities and degrees of toxicity. Therefore, in a safe venom injection several toxins will be acting together against the same organism.

Modelling of the three-dimensional structure and calculation of the electrostatic surface potential map around Cn10. The three-dimensional model of Cn10 was used to calculate the electrostatic surface potential map. Chemical modification experiments (Fontecilla-Camps et al., 1982; Kharat et al., 1989, 1990; El-Ayeb et al., 1986; Darbon and Angelides, 1984) and mutational analysis (Stampe et al., 1994; Goldstein et al., 1994; Hidalgo and MacKinnon, 1993; Lorei et al., 1994) indicate that the biological activity of scorpion toxins require positively charged amino acid residues. The surface potential map of Cn10 in Fig. 5 (top) shows the front of the molecule, the side of scorpion toxin molecules that is of particular interest because there are strong indications that it may bind to the channel. Firstly, several studies showed that the aromatic cluster of α -toxins (toxins affecting the inactivation process of sodium channels) situated on that side of the molecule plays a role in channel binding (Fontecilla-Camps et al., 1982; Kharat et al., 1989). Secondly, for β -toxins (toxins affecting the activation of the channel) that also display a cluster of aromatic residues at the front of the molecule it was shown that hydrophobic residues take part in receptor binding (Jover et al., 1984; Darbon and Angelides, 1984). Additionally, in β -toxins, the residue Lys13 situated in the front near the hydrophobic cluster seems to be involved in binding (Darbon and Angelides, 1984). Thirdly, charybotoxin, a small toxin that has the same three-dimensional fold and is therefore evolutionarily strongly related to the sodium-channel-binding toxins, interacts with the potassium channel via that surface (Stampe et al., 1994; Goldstein et al., 1994). When charybotoxin interacts with the high-conductance Ca^{2+} -activated potassium channel the residues that were found to be crucial for receptor binding

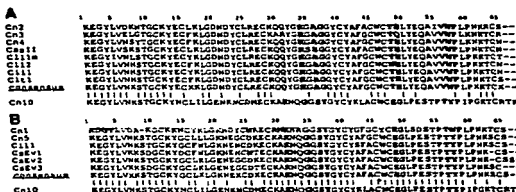


Fig. 6. Sequence comparison of mammal-specific (A) and arthropod-specific (B) toxins of the genus *Centruroides* and Cn10. Gaps (—) were introduced to maximize similarity. Below each toxin family the conserved sequence is shown with the most representative residue with the most representative amino acid in each position. N is a variable residue with no obvious preference. Cn1 (Vazquez et al., 1995), Cn2 and Cn3 (Zamudio et al., 1992), Vazquez et al., 1995), Cn4 (Vazquez et al., 1993) and Cn5 (Garcia, C., Becerra, B., Selisko, B., Delepiere, M., and Possani, L. D., unpublished) are from *C. noxius*, Cn11 from *C. suffusus suffusus* (see review: Possani, 1994), Cn11c, Cn11d and Cn11e are from *C. sculpturus* Ewing (Babin et al., 1974), Cn11 (Lebreron et al., 1994), Cn11f (Ramirez et al., 1994) and Cn12 (Dehesa-Davila et al., 1996) from *C. limpidus limpidus*, Cn11 from *C. infamatus infamatus* (Dehesa-Davila et al., 1996) and Cn11 from *C. limpidus tecomanus* (Martin et al., 1988). Residues that may influence the specificity of a toxin (see Discussion) are given in bold.

are situated in the β -sheet and in the α -helix belonging to the top part of the molecule. The functional epitope of charybotoxin in this case comprises an area of 1.0 nm \times 1.5 nm (see Stampe et al., 1994).

As demonstrated in Fig. 5 (top) the front of Cn10 presents mainly positive or neutral regions. This supports the above-mentioned evidence of its involvement in binding to the sodium channel. However, the central part of the β -sheet and the helix situated towards the top of the molecule are strongly negative due to residues Glu2, Glu21, Asp26, Glu28 and Glu49. Residues Tyr38, Tyr40, Tyr42 and Trp47 are situated within this region and, as outlined above, they are the aromatic residues supposedly involved in the binding of the toxin to the channel. It seems very unlikely that this region could be part of the toxin binding site. We therefore assume that the binding site of Cn10 is situated much more on the right and lower part of the molecule from comprising residues Tyr4, Lys6, Thr18, Glu20, Glu23, Lys43, Leu44, Thr57, Tyr58, Pro59, Thr64 and Cys465. For comparison we calculated the surface potential maps for the NMR structure of Cn5 (Lebreron, F., Garcia, C., Possani, L. D., and Delepiere, M., unpublished) and the three-dimensional model of Cn2 (Garcia et al., 1994) which are shown in Fig. 5 (bottom view) of the toxins Cn5 (middle) and Cn2 (bottom). Both of them show, like Cn10, a negative region at the central part of the β -sheet but with variable size. In Cn2 it is stronger and Tyr4 is also a part of it. The positive equipotential surface is therefore much smaller because of the strong negative region on the right of the molecule due to Glu15 and the influence of the C-terminal. In Cn5 the negative region at the center of the β -sheet is of less intensity and does not include Tyr4. The positive region stretches to the right of the molecule and is shared by Glu15. Cn5 shows neutral regions on the left and bottom part of the molecule.

Comparative analysis of toxins with different specificities. As outlined in the introduction, scorpion toxins share a very similar three-dimensional fold so that their species specificity is probably mediated by rather subtle changes of amino acid residues in selected positions of the primary structure. These crucial residues should belong to the side of the toxin that interacts with

the channel or they could have steric influence on the particular conformation of the receptor binding site.

Following the above-outlined indications for the involvement of the front and top part of the molecule in the binding to the sodium channel, we showed in our comparative analysis of the two toxin families (discussed in Fig. 6) exclusively the variance of residues that are (a) situated in these regions of the molecule, (b) near the aromatic cluster (maximum distance between the C α atoms of a given residue and the next aromatic residue being part of the cluster was chosen to be less than 1.4 nm) and (c) display a surface accessibility above 20%. As an illustration, Fig. 7 shows the front (top model) and top (bottom model) of the Cn10 toxin with the residues surrounding the central cluster of aromatic residues (shown in yellow) coloured according to charge or hydrophobicity (see legend). Including these residues in a comparison of the toxins listed in Fig. 6, clear differences between mammal-specific and arthropod-specific toxins can be stated for residues 26 (Leu/Asp or Asn), 35 (Lys/Gly), 38 (Gly/Tyr), 49 (Thr or Asn/Glu), 50 (His/Gly) and 57 (Val/Thr). Although a marked tendency exists for residues 32 (Gln/Lys), 33 (Ala/Ser) and 35 (Trp/Tyr), their importance seems to be less because of one significant exception in all three cases (see Fig. 6). All these residues are clustered in two regions as demonstrated in Fig. 4. One region is situated on top (see also lower model in Fig. 7) of the molecule comprising a continuous patch of residues 26 (situated in the α -helix), 37 and 38 (second β -strand) as well as 49 and 50 (third (middle) β -strand) and the two residues 32 and 35 situated apart at the turn between α -helix and the second β -strand, and the other at the bottom of the front near the C-terminal region (residues 57 and 58). In conclusion, we assume that these two regions may determine if a toxin from a scorpion of the genus *Centruroides* acts specifically towards mammals or arthropods. Because of the significance of the differences, especially in positions 26, 35, 38, 49, 50 and 57, we propose that these six residues are the most promising candidates for future mutational experiments.

Toxins that act specifically towards arthropods show differences in their fine specificity towards insects or crustaceans. To identify residues that might influence this specificity, we looked for variations within the family of arthropod-specific toxins



Fig. 7. Front (top model) and top view (bottom model) of a space-filling presentation of the model of Cn10. Solvent-accessible residues that are less than 1.4 nm from the nearest residue of the aromatic cluster (shown in yellow, coloured according to charge (red, negatively charged; blue, positively charged) or hydrophobicity (green, hydrophobic; purple, polar uncharged). All other residues are in white.

given in Fig. 6. In general, these toxins are very similar (around 80%). Residues that display variations include the residues 18 and 21, the α -helix residues 23, 26 and 27 and the loop between the second and third β -strand (residues 43 and 44). These regions might determine the fine-tuning of the specificity of toxins active in arthropods.

In conclusion, our analyses indicate that the specificity of a particular toxin of the genus *Centruroides* towards mammals or arthropods is influenced by the character of residues 26, 32, 35, 37, 38, 49, 50, 57 and 58. Of these, residues 26, 32, 35, 37, 38, 49 and 50 are situated within a region of negative electrostatic potential. Considering the fact that scorpion toxins bind to ion channels via regions of positive surface potential, the influence

of these residues might be rather indirect. In contrast, residues 57 and 58, situated in a region of positive surface potential, could directly take part in the binding of these toxins to the sodium channel.

This work was supported partially by a scholarship from the Alexander Humboldt Foundation (Germany) to B. S. and by grants from the Howard Hughes Medical Institute (175191-527104) to L. D. P., from *Consejo Nacional de Ciencia y Tecnología-México* (IN-4734) and *Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México* (IN-205893) to B. B. We greatly appreciate the technical assistance of Timoteo Olamendi, Fernando Zamudio and Cipriano Balderas. We wish to thank Gabriel Alonso-Fajuel for his help with the management of the computer facilities in the laboratory of Dr Xavier Sobrón to whom we are greatly indebted, and Dr Eduardo Horváth for critical reading of the manuscript.

REFERENCES

- Babin, D. R., Watt, D. D., Goos, S. M. & Mlejnek, R. V. (1974) Amino acid sequences of neurotoxic protein variants from the venom of *Centruroides sculpturatus* Ewing. *Arch. Biochem. Biophys.* **164**, 664–706.
- Becerril, B., Vazquez, A., Garcia, C., Corona, M., Bollivar, F. & Possani, L. D. (1993) Cloning and characterization of cDNAs that code for Na⁺-channel-blocking toxins of the scorpion *Centruroides lunatus* Hoffmann. *Gene* **128**, 165–171.
- Bernstein, C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
- Bourenie, F., Roumevrand, C., Gilquin, B., Méner, A. & Toma, F. (1991) Refined structure of charybotoxin, common motif in scorpion toxins and insect detensins. *Science* **254**, 1521–1523.
- Bruix, M., Jimenez, M. A., Santoro, J., Gonzalez, C., Collifa, F. J., Mendez, E. & Risco, M. (1993) Solution structure of 7134 and 9137 diastereoisomers from barley and wheat endoparasites determined by ¹H-NMR: a structural motif common to toxic arthropod proteins. *Biochemistry* **32**, 715–724.
- Connolly, M. L. (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709–713.
- Darbon, H. & Angelides, K. J. (1984) Structural mapping of the voltage-dependent sodium channel. *J. Biol. Chem.* **259**, 6074–6084.
- Darbon, H., Weber, C. & Braun, W. (1991) Two-dimensional ¹H nuclear magnetic resonance study of Aah II, an anti-insect toxin from the scorpion *Androctonus australis* Hector: sequential resonance assignments and folding of the polypeptide chain. *Biochemistry* **30**, 1836–1845.
- Dehes-Dalila, M., Ramirez, A. N., Zamudio, F. Z., Gurrola-Briones, G., Liéano, A., Darazon, A. & Possani, L. D. (1996) Structural and functional comparison of toxins from the venom of the scorpion *Centruroides sublimatus* infonimato, *Centruroides lunatus* Hoffmann and *Centruroides noxius*. *Comp. Biochem. Physiol.* **118B**, 331–339.
- De Lima, M. E., Martín, M. F., Diniz, C. R. & Roehat, H. (1986) Tityus serrulatus toxin VII bears pharmacological properties of both β -toxin and insect toxin from scorpion venoms. *Biochem. Biophys. Res. Commun.* **139**, 296–302.
- Edman, P. & Begg, G. (1967) A protein sequenator. *Eur. J. Biochem.* **1**, 80–91.
- Eran, M., Fowler, E., Herrmann, R., Duval, A., Pelhate, M. & Zlotnik, E. (1990) A scorpion venom neurotoxin paralytic to insects that affects sodium current inactivation: purification, primary structure and mode of action. *Biochemistry* **29**, 5941–5947.
- El-Ayeb, M., Darbon, H., Bahraoui, E., Vargas, O. & Roehat, H. (1986) Differential effects of defined chemical modifications on antigenic and pharmacological activities of scorpion α - and β -toxins. *Eur. J. Biochem.* **155**, 289–294.
- Fonseca-Carrión, J. C., Aljassas, R. J., Sudduth, F. L., Watt, D. D. & Begg, C. E. (1980) Three-dimensional structure of a protein from scorpion venom: A new structural class of neurotoxins. *Proc. Natl. Acad. Sci.* **77**, 6546–6550.

- Foncella-Camps, J. C., Almasy, R. J., Suddath, F. L., & Bugg, C. E. (1982) The three-dimensional structure of scorpion neurotoxins. *Toxicon* 20, 1-7.
- Goldstein, S. A., Pheasant, D. J., & Miller, C. (1994) The eurythmotic receptor of a shaker K⁺ channel: peptide and channel residues mediating molecular recognition. *Neuron* 12, 1377-1388.
- Gordon, D., & Zlotkin, E. (1993) Binding of an α -scorpion toxin to insect sodium channels is not dependent on membrane potential. *FEBS Lett.* 335, 125-128.
- Gordon, D., Martin-Eauclaire, M.-F., Cresté, S., Kopeyan, C., Carlier, E., Khalifa, B., Pelhrai, N.I., & Rochat, H. (1996) Scorpion toxins affecting sodium current inactivation bind to distinct homologous receptor sites on rat brain and insect sodium channels. *J. Biol. Chem.* 271, 8034-8045.
- Gurevitz, D., Urbach, D., Zlotkin, E., & Zilberberg, N. (1991) Nucleotide sequence and structure analysis of a cDNA encoding an insect toxin from the scorpion *Leiurus quinquestriatus hebraeus*. *Toxicon* 29, 1270-1272.
- Gurola, G. B., Moreno-Magallieb, G., Zamudio, F. Z., García, M., Sobrón, X., & Possani, L. D. (1994) The disulfide bridges of toxin 2 from the scorpion *Centruroides notatus* Hoffmann and its three-dimensional structure calculated using the coordinates of variant 3 from *Centruroides sculpturatus*. *FEBS Lett.* 347, 59-62.
- Hidalgo, F., & MacKinnon, R. (1995) REVEALING the architecture of a K⁺ channel pore through mutant cycles with a peptide inhibitor. *Science* 268, 307-310.
- Honig, B., & Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science* 268, 1144-1149.
- Houssert, D., Haberstreter-Rochat, C., Astier, J.-P., & Foncella-Camps, J. C. (1994) Crystal structure of toxin II from the scorpion *Androctonus australis* Hector refined at 1.3 Å resolution. *J. Mol. Biol.* 238, 88-103.
- Jablonsky, M. J., Watt, D. D., & Rama Krishna, N. (1995) Solution structure of an old world-like neurotoxin from the venom of the new world scorpion *Centruroides sculpturatus* Ewing. *J. Mol. Biol.* 245, 445-458.
- Jover, E., Babilio, J., & Couraud, F. (1984) Binding of β -toxin: a physico-chemical study. *Biochemistry* 23, 1144-1152.
- Kharat, R., Darbon, H., Rochat, H., & Granier, C. (1989) Structural relationships of scorpion α -toxins. *Eur. J. Biochem.* 151, 381-390.
- Kharat, R., Darbon, H., Granier, C., & Rochat, H. (1990) Structural relationships of scorpion α -neurotoxins: contributions of arginine residues. *Toxicon* 28, 509-523.
- Krezel, A. M., Kaubhalla, C., Hidalgo, P., MacKinnon, R., & Wagner, G. (1995) Solution structure of the potassium channel inhibitor agitoxin 2: caliper for probing channel geometry. *Protein Sci.* 4, 1478-1489.
- Lazdunski, M., Frelin, C., Barhanin, J., Lombet, A., Meiri, H., Pauron, D., Romey, G., Schmid, A., Schweitz, H., Vigne, P., & Vijverberg, H. P. M. (1986) Poly-peptide toxins as tools to study voltage-sensitive sodium channels. *Ann. NY Acad. Sci.* 479, 204-220.
- Lebreron, F., Delepiere, N., Ramirez, A. N., Balderas, C., & Possani, L. D. (1994) Primary and NMR three-dimensional structure determination of a novel enantioselective toxin from the venom of the scorpion *Centruroides limpidus limpidus* Karsch. *Biochemistry* 33, 11135-11149.
- Lee, W., Moore, C. H., Watt, D. D., & Rama Krishna, N. (1994a) Solution structure of the variant-3 neurotoxin from *Centruroides sculpturatus* Ewing. *Eur. J. Biochem.* 218, 89-95.
- Lee, W., Jablonsky, M. J., Watt, D. D., & Rama Krishna, N. (1994b) Proton nuclear magnetic resonance and distance geometry/simulated annealing studies on the variant-1 neurotoxin from the new world scorpion *Centruroides sculpturatus* Ewing. *Biochemistry* 33, 2468-2475.
- Loret, E. P., Martin-Eauclaire, M.-F., Mansuelle, P., Sampieri, F., Granier, C., & Rochat, H. (1991) An anti-insect toxin purified from the scorpion *Leiurus quinquestriatus* Hector also acts on the α - and β -sites of the mammalian sodium channel: sequence and circular dichroism study. *Biochemistry* 30, 633-640.
- Loret, E. P., Niendev, Soto de Valle, R., Mansuelle, P., Sampieri, F., & Rochat, H. (1994) Positively charged amino acid residues located similarly in sea anemone and scorpion toxins. *J. Biol. Chem.* 269, 16785-16788.
- Martin, B. M., Carbone, E., Ytani, A., Brown, M. A., Ramirez, A. N., Gurola, G. B., & Possani, L. D. (1988) Amino acid sequence and physiological characterization of toxins from the venom of the scorpion *Centruroides limpidus leconteanus* Hoffmann. *Toxicon* 26, 785-794.
- Nicholls, A., & Honig, B. (1991) A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* 12, 435-445.
- Possani, L. D. (1984) Structure of scorpion toxins. In *Handbook of natural toxins* (T. A. D., ed.) vol. 2, pp. 513-550. Marcel Dekker, New York.
- Possani, L. D., Martin, B. M., Svendsen, I., Rode, G. S., & Erickson, B. W. (1985) Scorpion toxin from *Centruroides notatus* and *Tityus serrulatus*. *Biochem. J.* 229, 739-750.
- Ramirez, A. N., Martin, B. M., Gurola, C. B., & Possani, L. D. (1994) Amino acid sequence of a novel toxin from the venom of the scorpion *Centruroides limpidus limpidus* Karsch. *Toxicon* 32, 479-490.
- Stampe, P., Kolmakova-Parensky, L., & Miller, C. (1994) Iniminations of K⁺ channel structure from a complete functional map of the molecular surface of carbonyltoxin. *Biochemistry* 33, 443-450.
- Valdivia, H. M., Martin, B. M., Ramirez, A. N., Fletcher, P. L., & Possani, L. D. (1994) Isolation and pharmacological characterization of four novel Na⁺ channel-blocking toxins from the scorpion *Centruroides notatus* Hoffmann. *J. Biochem. (Tokyo)* 116, 1383-1391.
- Vazquez, A., Becerril, B., Martin, B. M., Zamudio, F., Bolívar, F., & Possani, L. D. (1993) Primary structure determination and cloning of the cDNA encoding toxin 4 from the scorpion *Centruroides notatus* Hoffmann. *FEBS Lett.* 320, 43-46.
- Vazquez, A., Tapia, J. V., Eliason, W. E., Martin, B. M., Lebreron, F., Delepiere, N., Possani, L. D., & Becerril, B. (1995) Cloning and characterization of the cDNAs encoding Na⁺ channel-specific toxins 1 and 2 of the scorpion *Centruroides notatus* Hoffmann. *Toxicon* 33, 1161-1170.
- Zamudio, F., Saavedra, R., Martin, B. M., Gurola-Briones, G., Héron, P., & Possani, L. D. (1992) Amino acid sequence and immunological characterization with monoclonal antibodies of two toxins from the venom of the scorpion *Centruroides notatus* Hoffmann. *Eur. J. Biochem.* 204, 281-292.
- Zhao, B., Carlson, M., Eilick, S. E., & Bugg, C. E. (1992) Structure of scorpion toxin xanti-3 at 1.2 Å resolution. *J. Mol. Biol.* 227, 239-252.

BIBLIOGRAFIA

Adams M. E. and Olivera B. M. (1994). Neurotoxins: overview of an emerging research technology. *TINS*, **17**, No. 4: 151-154.

Almassy R. J., Fontecilla-Camps J. C., Suddath F. L. and Bugg Ch. E. (1983). Structure of variant-3 scorpion neurotoxin from *Centruroides sculpturatus* Ewing, refined at 1.8 Å resolution. *J. Mol Biol.* **170**: 497-527.

Arseniev A. S., Kondakov V. I., Maiorov V. N. and Bystrov V. F. (1984). NMR solution spatial structure of "short" scorpion insect toxin. *FEBS*, **165**: 57-62.

Babin D. R., Watt D. D., Goos S. M. and Mlejnek R. V. (1974). Amino acid sequences of neurotoxic protein variants from the venom of *Centruroides sculpturatus* Ewing. *Arch. Biochem. Biophys.* **164**: 694-706.

Bardwell, J. C. A. (1994). Building bridges: disulphide bond formation in the cell. *Molecular Microbiology*, **14** (2): 199-205.

Becerril B., Vázquez A., García C., Corona M., Bolívar F. and Possani L. D. (1993). Cloning and characterization of cDNAs that code for Na⁺-channel-blocking toxins of the scorpion *Centruroides noxius* Hoffmann. *Gene*, **128**: 165-171.

Becerril B., Corona M., García C., Bolívar F. and Possani L. D. (1995). Cloning of genes encoding scorpion toxins: an interpretative review. *J. Toxicol.-Toxin Reviews*. **14** (3): 339-357.

Blackwell J. R. and Horgan R. (1991). A novel strategy for the production of a highly expressed recombinant protein in an active form. *FEBS Lett.* **295**: 10-12.

Blaustein M. P., Rogowski R. S., Schneider M. J., and Krueger B. K. (1991). Polypeptide toxins from the venom of the old world and new world scorpions preferentially block different potassium channels. *Mol.*

Pharmacol. **40**: 932-942.

Bontems F., Roumestand Ch., Boyot P., Gilquin B., Doljansky Y., Menez A. and Toma F. (1991). Three-dimensional structure of natural charybdotoxin in aqueous solution by $^1\text{H-NMR}$. (1991). Eur. J. Biochem. **196**: 19-28.

Bougis P. E., Rochat H. and Smith L. A. (1989). Precursors of *Androctonus australis* scorpion neurotoxin: Structures of precursors, processing outcomes, and expression of a functional recombinant toxin II. J. Biol. Chem. **264**: 19259-19265.

Bouhaouala-Zahar B., Ducancel F., Zenouaki I., Ben Khalifa R., Borchani L., Pelhate M., Boulain J.-C., El Ayeb M., Ménez A. and Karoui H. (1996). A recombinant insect-specific α -toxin of *Buthus occitanus tunetanus* scorpion confers protection against homologous mammal toxins. Eur. J. Biochem. **238**: 653-660.

Branden C. and Tooze J. (1991). Prediction, engineering and design of protein structures. En: Introduction to protein structure. Chapter 15, pp 247-266. Garland Publishing, Inc. New York.

Brandts J. F., Halvorson H. R. and Brennan M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. Biochemistry, **14** (22): 4953-4963.

Cantor R. Ch. and Schimmel P. R. (1980). Biophysical Chemistry. Part I: the conformation of biological macromolecules. W. H. Freeman and Company, San Francisco.

Carbone E., Wanke E., Prestipino G., Possani L. D. and Maelicke A. (1982). Selective blockage of voltage-dependent K^+ channels by a novel scorpion toxin. Nature, **296**: 90-91.

Caterall W. A. (1991). Structure and function of voltage-gated sodium and calcium channels. Curr. Opin. Neurobiol. **1**: 5-13.

Courad F. and Jover E., Dubois J. M. and Rochat H. (1982). Two types of

scorpion toxin receptor sites, one related to the activation, the other to the inactivation of the action potential sodium channel. *Toxicon*, **20** (1): 9-16.

Creighton, T. E. (1993). Chemical properties of polypeptides. En: *Proteins*. 2nd edition, Chapter 1: 1-48. W. H. Freeman and Company, New York.

De Lima M. E., Martin M. F., Diniz C. R. and Rochata H. (1986). *Tityus serrulatus* toxin VII bears pharmacological properties of both β -toxin and insect toxin from scorpion venoms. *Biochem. Biophys. Res. Commun.* **139** (1): 296-302.

Dee A., Belagaje R. M., Ward K., Chio E. and Lai M-H. T. (1990). Expression and secretion of a functional scorpion insecticidal toxin in cultured mouse cells. *Bio/Technology*, **8**: 339-342.

Delepierre M., Prochnicka-Chalufour A. and Possani L. D. (1997). A novel potassium channel blocking toxin from the scorpion *Pandinus imperator*: a ¹H NMR analysis using a nano-NMR probe. *Biochemistry*, **36**: 2649-2658.

Derman A. I., Prinz W. A., Belin D. and Beckwith J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science*, **262**: 1744-1747.

Dauplais M., Gilquin B., Possani L. D., Gurrola-Briones G., Roumestand C. and Ménez A. (1995). Determination of the three-dimensional solution structure of Noxiustoxin: analysis of structural differences with related short-chain scorpion toxins. *Biochemistry*, **34**: 16563-16573.

Dueñas M., Vázquez J., Ayala M., Söderlind E., Ohlin M., Pérez L., Borrebaeck C. A. K. and Gavilondo J. V. (1994). Intra- and extracellular expression of an scFv antibody fragment in *E. coli*: effect of bacterial strains and pathway engineering using GroES/L chaperonins. *BioTechniques*, **16** (3): 476-482.

Eitan M., Fowler E., Herrmann R., Duval A., Pelhate M. and Zlotkin E. (1990). A scorpion venom neurotoxin paralytic to insects that affects sodium current inactivation: purification, primary structure, and mode of action.

- Biochemistry, **29**: 5941-5947.
- Ellis R. J. (1996). Revisiting the Anfinsen cage. *Folding & Design*, **1**: R9-R15.
- Fontecilla-Camps J. C., Almasy R. J., Suddath F. L., Watt D. D. and Bugg Ch. E. (1980). Three-dimensional structure of a protein from scorpion venom: a new structural class of neurotoxins. *PNAS. USA*, **77** (11): 6496-6500.
- Fischer B., Summer I. and Goodenough P. (1993). Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in *Escherichia coli* as inclusion bodies. *Biotechnol. Bioeng.* **41**: 3-13.
- Fischer G. and Schmid F. (1990). The mechanism of protein folding. Implications of in vitro refolding models for *de novo* folding and translocation in the cell. *Biochemistry*, **29** (9): 2205-2211.
- Freedman R. B. (1994). Folding helpers and unhelpful folders. *Protein Folding. Current Biology*, **4** (10): 933-935.
- García C., Becerril B., Selisko B., Delapierre M. and Possani L. D. (1997). Isolation, characterization and comparison of a novel crustacean toxin with a mammalian toxin from the venom of the scorpion *Centruroides noxius* Hoffmann. *Comp. Biochem. Physiol.* **116B** (3): 315-322.
- Georgiou G. and Valax P. (1996). Expression of correctly folded proteins in *Escherichia coli*. *Curr. Opin. Biotech.* **7**: 190-197.
- Gething M.-J. and Sambrook J. (1992). Protein folding in the cell. *Nature*, **355**: 33-45.
- Granier C., Novotny J., Fontecilla-Camps J.-C., Fourquet P., El Ayeb M. and Bahraqui E. (1989). The antigenic structure of a scorpion toxin. *Mol. Immunol.* **26** (6): 503-513.
- Jablonsky M. J., Watt D. D. and Krishna R. (1995). Solution structure of an old world-like neurotoxin from the venom of the new world scorpion *Centruroides sculpturatus* Ewing. *J. Mol. Biol.* **248**: 449-458.

ESTA TESIS NO DEBE SALIR DE LA BIBLIOTECA

- Jaenicke R. and Rudolph R. (1990). Folding Proteins. En: Protein Structure (Practical approach series). Creighton T. E. ed. IRL Press, Oxford.
- Jover E., Courad F. and Rochat H. (1980). Two types of scorpion neurotoxins characterized by their binding to two separate receptor sites on rat brain synaptosomes. *Biochem. Biophys. Res. Commun.* **95**: 1607-1614.
- Karpp, G. (1987). *Biología Celular. Segunda edición (primera edición en español)*. Mc Graw Hill Inc., USA. pp 950.
- Keegan H. L. (1980). Scorpions of medical importance. University Press of Mississippi. USA. pp. 140.
- Kirsch G. E., Skatteböl A., Possani L. D. and Brown A. M. (1989). Modification of Na channel gating by an α scorpion toxin from *Tityus serrulatus*. *J. Gen Physiol.* **93**: 67-83.
- Kobayashi Y., Takashima H., Tamaoki H., Kyogoku Y., Lambert P., Kuroda H., Chino N., Watanabe T. X., Kimura T., Sakakibara S. and Moroder L. (1991). The cysteine-stabilized α -helix: a common structural motif of ion-channel blocking neurotoxic peptides. *Biopolymers.* **31**: 1213-1220.
- Kyte, J. (1995). Folding and assembly. En: Structure in Protein Chemistry. Chapter 13: 445-500. Garland Publishing, INC. New York & London.
- Lebreton F., Delepierre M., Ramírez A. N., Balderas C. and Possani L. D. (1994). Primary and NMR three-dimensional structure determination of a novel crustacean toxin from the venom of the scorpion *Centruroides limpidus* Karsch. *Biochemistry.* **33**: 11135-11149.
- Li H.-M., Wang D.-Ch., Zeng Z.-H., Jin L. and Hu R.-Q. (1996). Crystal structure of an acidic neurotoxin from scorpion *Buthus martensii* Karsch at 1.85 Å resolution. *J. Mol. Biol.* **261**: 415-431.
- Lille H., McLaughlin S., Fredman R. and Buchner J. (1994). Influence of protein disulfide isomerase (PDI) on antibody folding *in vitro*. *J. Biol. Chem.* **269** (19): 14290-14296.

- Lippens G., Najib J., Wodak S. J. and Tartar A. (1995). NMR sequential assignments and solution structure of chlorotoxin, a small scorpion toxin that blocks chloride channels. *Biochemistry*, **34**: 13-21.
- Loret P. E., Sampieri F., Roussel A., Granier C. and Rochat H. (1990). Conformational flexibility of a scorpion toxin active on mammals and insects: a circular dichroism study. *Proteins*, **8**: 164-172.
- Loret E. P., Martin-Euclaire M-F., Mansuelle P., Sampieri F., Granier C. and Rochat H. (1991). An anti-insect toxin purified from the scorpion *Androctonus australis* Hector also acts on the α - and β -mammalian sodium channel: sequence and circular dichroism study. *Biochemistry*, **30**: 633-640.
- Maeda S., Volrath S. L., Hanzlik T. N., Harper S. A., Majima K., Maddox D. W., Hammock B. D. and Fowler E. (1991). Insecticidal effects of an insect-specific neurotoxin expressed by a recombinant baculovirus. *Virology*, **184**: 777-780.
- Makrides S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **6** (3): 512-538.
- Martin-Eauclaire M-F., Sogaard M., Ramos C., Cestèle S., Bougis P. E. and Svensson B. (1994). Production of active, insect-specific scorpion neurotoxin in yeast. *Eur. J. Biochem.* **223**: 637-645.
- Martínez F., Becerril B., Gurrola G., Martín B. M. and Possani L. D. (1996). Synthesis and expression of the gene coding for noxiustoxin a K⁺ channel-blocking peptide from the venom of the scorpion *Centruroides noxius*. *Toxicon*, **34**: 1413-1419.
- Meves H., Simard J. M. and Watt D. D. (1984). Biochemical and electrophysiological characteristics of toxins isolated from the venom of the scorpion *Centruroides sculpturatus*. *J. Physiol., Paris*, **79**: 185-191.
- Meves H., Simard J. M. and Watt D. D. (1987). Interactions of scorpion toxins with the sodium channel. *Annals New York Academy of Sciences*: 112-132.

Olamendi-Portugal T., Gómez-Lagunas F., Gurrola G. B. and Possani L. D. (1996). A novel structural class of K⁺-channel blocking toxin from the scorpion *Pandinus imperator*. *Biochem. J.* **315**: 977-981.

Park Ch-S and Miller Ch. (1992). Interaction of charybdotoxin with permanent ions inside the pore of a K⁺ channel. *Neuron*, **9**: 307-313.

Pashkov V. S., Maiorov V. N., Bystrov V. F., Hoang A. N., Volkova T. M. and Grishin E. V. (1988). Solution spatial structure of 'long' neurotoxin M₉ from the scorpion *Buthus eupeus* by ¹H-NMR spectroscopy. *Biophys. Chem.* **31**: 121-131.

Possani L. D., Alagón A. C., Fletcher P. L JR. and Rickson B. W. (1977). Purification and properties of mammalian toxins from the venom of the Brazilian scorpion *Tityus serrulatus* Lutz and Mello. *Arch. Biochem. Biophys.* **180**: 394-403.

Possani L. D., Dent M. A. R., Martin B. M., Maelicke A. and Svendsen I. (1981). The amino terminal sequences of several toxins from the venom of the mexican scorpion *Centruroides noxius* Hoffmann. *Carlsberg Res. Commun.* **46**: 207-214.

Possani L. D., Martin B., and Svendsen I. (1982). The primary structure of Noxiustoxin: a K⁺ channel blocking peptide from the venom of the scorpion *Centruroides noxius* Hoffmann. *Carlsberg Res. Comm.* **47**: 285-289.

Possani, L. D. (1983). Las toxinas del veneno de alacranes: estructura y función. *Bol. Estud. Méd. Biol., Méx., Suplemento.* **32**: 285-297.

Possani L. D. (1984). Structure of scorpion toxins. En: *Handbook of Natural Toxins*. Tu A. T. ed. Vol. **2**: 513-550. Marcel Dekker, Inc., New York.

Possani L. D., Martin B. M., Svendsen I., Rode G. S. and Erickson B. M. (1985). Scorpion toxins from *Centruroides noxius* and *Tityus serrulatus*. *Biochem. J.* **229**: 739-750.

Possani L. D., Gurrola G. B., Portugal T. O., Zamudio F. Z., Vaca L. D., Calderón

- E. S. A. and Kirsch G. E. (1991). Scorpion toxins: a model for peptide synthesis of new drugs. En: Proceedings of the First Brazilian Congress on Proteins: 352-3367. Oliveira B. y Sgarbieri V. ed. UNICAMP, Campinas.
- Possani L., Calderón E. S., Olamendi T. P., Dehesa Dávila M. y Gurrola G. B. (1996). Protección contra el alacranismo. En: Vacunas, Ciencia y Salud. Capítulo 44: 553-567. Editores: Escobar A. G., Valdespino J. L. G. y Sepúlveda J. A. Secretaría de Salud. Subsecretaría de coordinación y desarrollo.
- Smith, B. J. (1996). Chemical cleavage of proteins at methionyl residues, p. 369. En: the Protein Protocols handbook, Walker, J. M. ed. Humana Press.
- Rainer J. and Rainer R. (1990). Folding proteins. En: Protein structure (practical approach series). Creighton T. E. ed. Chapter 9: 191- 223. IRL Press at Oxford University Press.
- Rochat H., Bernard P and Courad F. (1979). Scorpion toxins: chemistry and mode of action. En: Advances in Cytopharmacology. Ceccarelli and F. Clementi, eds. Raven Press, New York. 3: 325-334
- Saibil H. R. (1994). How chaperones tell wrong from right. Struct. Biol. 1 (12): 838-842.
- Schindler T., Mayr L. M., Landt O., Hahn U. and Schmid F. X. (1996). The role of a *trans*-proline in the folding mechanism of ribonuclease T1. Eur. J. Biochem. 241: 516-52
- Selisko B., García C., Becerril B., Delepierre M. and Possani L. D. (1996). An insect-specific toxin from *Centruroides noxius* Hoffmann cDNA, primary structure, three dimensional model and electrostatic surface potentials in comparison with other toxin variants. Eur. J. Biochem. 242: 235-242.
- Sossin W. S. and Scheller R. H. (1991). Biosynthesis and sorting of neuropeptides. Curr. Opin. Neurobiol. 1: 79-83.
- Stahnke H. L. (1978). The genus *Centruroides* (Buthidae) and its venom. En: Arthropod Venoms. Vol. 48, chapter 12: 277-307. Bettini S. ed. Springer-

Verlag, Berlin.

Stewart L. M. D., Hirst M., López Ferber M., Merryweather A. T., Cayley P. J. and Possee R. D. (1991). Construction of an improved baculovirus insecticide containing an insect-specific toxin gene. *Nature*, **352**: 85-87.

Texter F. L., Spencer D. B., Rosenstein R. and Matthews C. R. (1992). Intramolecular catalysis of a proline isomerization reaction in the folding of dihydrofolate reductase. *Biochemistry*, **31** (25): 5687-5691.

Vázquez A., Tapia J. V., Eliason W. E., Martin B. M., Lebreton F., Delepierre M., Possani L. D. and Becerril B. (1995). Cloning and characterization of the cDNAs encoding Na⁺ channel-specific toxins 1 and 2 of the scorpion *Centruroides noxius* Hoffmann. *Toxicon*, **33** (9): 1161-1170.

von Heijne G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acid Research*, **14** (11): 4683- 4690.

Weissman, J. S. (1995). All roads lead to Rome? the multiple pathways of protein folding. *Current Biology*, **2**: 255-260.

Wheeler K. P., Watt D. D. and Lazdunski M. (1983). Classification of Na channel receptors specific for various scorpion toxins. *Pflügers Arch.* **397**: 164-165.

Wülfing Ch., Lombardero J. and Plückthun A. (1994). An *Escherichia coli* protein consisting of a domain homologous to FK506-binding proteins (FKBP) and a new mwital binding motif. *J. Biol. Chem.* **269**(4): 2895-2901.

Zilberberg N., Gordon D., Pelhate M., Adams M. E., Norris T. M., Zlotkin E. and Gurevitz M. (1996). Functional expression and genetic alteration of an alpha scorpion neurotoxin. *Biochemistry*, **35**: 10215-10222.

Zlotkin E., Miranda F. and Rochat H. (1978). Chemistry and pharmacology of Buthinae scorpion venoms. En: *Arthropod Venoms*. Vol. 48, chapter 13 C: 317-369. Bettini S. ed. Springer-Verlag, Berlin.

Zlotkin E., Kadouri D., Gordon D., Pelhate M., Martin M. F. and Rochat H.

(1985). An excitatory and a depressant insect toxin from scorpion venom both affect sodium conductance and possess a common binding site. *Arch. Biochem. Biophys.* **240** (2): 877-887.

Zlotkin E., Eitan M., Bindokas V. P., Adams M. E., Moyer M., Burkhart W. and Fowler E. (1991). Functional duality and structural uniqueness of depressant insect-selective neurotoxins. *Biochemistry*, **30**: 4814-4821.