



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS  
QUÍMICAS**

**DISEÑO, EXPRESIÓN, PURIFICACIÓN Y REPLEGAMIENTO PARA LA CARACTERIZACIÓN  
POR RESONANCIA MAGNÉTICA NUCLEAR DE LA DEFENSINA 1**

TESIS  
PARA OPTAR POR EL GRADO DE

**MAESTRO EN CIENCIAS**

PRESENTA

Q. DANIELA FRANCO BODEK



TUTOR: Dr. José Federico del Río Portilla AÑO: 2011

---

Agradezco a:

- La Universidad Nacional Autónoma de México.
- Al Instituto de Química de la U.N.A.M.
- Al proyecto de DGAPA número IN205110.
- Al Laboratorio Nacional de Estructura de Biomacromoléculas.
- Profundamente al Dr. Federico del Río, por sus invaluable consejos, guías y enseñanzas, por ser un ejemplo y por su confianza.
- A mis sinodales: Dr. Miguel Antonio Costas, Dra. Adela Rodríguez, Dr. Abel Moreno, Dr. Carlos Martín Cerda y la Dra. Silvia Pérez.
- Profundamente a la Dra. Patricia Cano, por todas sus enseñanzas aderezadas de buen humor y paciencia.
- A la Dra. Nuria Sánchez, por sus comentarios y consejos.
- A la M. en C. Lucía del Carmen Márquez y a la Q. Eréndira García, por el apoyo técnico y su amabilidad.
- Profundamente a todos los del laboratorio de Química de Biomacromoléculas 1, por hacer del trabajo una experiencia muy agradable, por su buena voluntad, y sobre todo por su compañerismo.
- Profundamente a la fundación Franco-Bodek, por la beca de manutención, el cariño y el apoyo infinito que he recibido siempre.
- A mis amigos, que saben de sobra quienes son y cuanto los quiero.

---

Para mi papá, mi mamá y mi hermano Tomás

---

Este trabajo se realizó en: el Laboratorio de Química de Biomacromoléculas 1 (antes Bioquímica 1) y en el Laboratorio de Biología Molecular

Instituto de Química

Universidad Nacional Autónoma de México

---

## ÍNDICE

Índice.....	v
Índice de tablas .....	vii
Índice de ilustraciones.....	viii
Glosario .....	xi
Resumen.....	1
Introducción.....	2
Antecedentes .....	4
<i>¿Qué son las defensinas?</i> .....	4
<i>Relación entre toxinas y defensinas</i> .....	6
<i>¿Qué es la DEF1?</i> .....	7
<i>Proteínas recombinantes: Expresión en E. coli</i> .....	8
<i>Proteína de fusión</i> .....	10
<i>Plegamiento de proteínas recombinantes</i> .....	12
<i>Resonancia magnética nuclear: Generalidades</i> .....	14
<i>Resonancia Magnética nuclear y proteínas</i> .....	16
Objetivos .....	19
Hipótesis.....	19

---

Metodología Experimental.....	20
<i>Clonación del gen de la Proteína de fusión-DEF1 en el vector pET32-a-mod...</i>	20
<i>Esquema general de expresión y purificación de la DEF1.....</i>	21
<i>Expresión optimizada de proteína soluble.....</i>	22
<i>Purificación optimizada de proteína soluble.....</i>	22
<i>Purificación optimizada de proteína insoluble y plegamiento.....</i>	23
<i>Adquisición de los espectros de RMN de la muestra.....</i>	24
Resultados y discusión.....	25
<i>Análisis de la secuencia de la GS-DEF1 y la proteína de fusión.....</i>	25
<i>Comentarios sobre el sistema de expresión.....</i>	27
<i>Comentarios sobre la purificación por cromatografía de afinidad metal-quelato</i>	30
<i>Caracterización de la GS-DEF1.....</i>	33
<i>Comparación: Soluble y replegada.....</i>	35
<i>Comentarios acerca de la asignación.....</i>	38
Conclusiones.....	52
Bibliografía .....	53
Apéndice .....	I
<i>Abreviatura, Características y estructura de los aminoácidos más comunes .....</i>	I

---

<i>Patrones TOCSY y COSY de los aminoácidos más comunes</i> .....	IV
<i>MALDI-TOF de la DEF1 soluble</i> .....	V
<i>MALDI-TOF de la DEF1 replegada</i> .....	VI
<i>TOCSY de la GS-DEF1 soluble</i> .....	VII
<i>TOCSY de la GS-DEF1 replegada</i> .....	VIII
<i>NOESY de la GS-DEF1 soluble</i> .....	IX
<i>NOESY de la GS-DEF1 replegada</i> .....	X
<i>Datos técnicos</i> .....	XI
<i>Protocolo de transformación de células por choque térmico</i> .....	XIV

## ÍNDICE DE TABLAS

Tabla 1: Sistemas de expresión de proteínas recombinantes, ventajas y desventajas <sup>14</sup> .....	9
Tabla 2: Características y parámetros calculados de la GS-DEF1.....	25
Tabla 3: Características y parámetros calculados de la proteína de fusión .....	26
Tabla 4: Características principales de las cepas de <i>E. coli</i> probadas para la expresión de la DEF1 <sup>30</sup> .....	27
Tabla 5: Comparación de las masas de la GS-DEF1 soluble y replegada. La ligera diferencia entre ellos se debe a que se calibraron de manera distinta.....	37

---

Tabla 6: Datos técnicos de los reactivos y materiales utilizados.....	XI
Tabla 7: Datos técnicos de los equipos utilizados.....	XII
Tabla 8: Composición de las disoluciones utilizadas.....	XIII

## ÍNDICE DE ILUSTRACIONES

Ilustración 1: Mecanismo de acción de reducción de un sustrato por la tiorredoxina <sup>16</sup> . La proteína roja es un sustrato, mientras la azul representa la tiorredoxina.....	11
Ilustración 2: Fracción del sobrenadante purificado por columna de afinidad del extracto de cada una de las cepas utilizadas, crecidas a 28° C durante 6 horas con 0.5 mM de IPTG. ....	28
Ilustración 3: Gel de electroforesis comparativo de la fracción soluble y el pellet .	29
Ilustración4: Cinética de expresión a 20° C. Mezcla de pellet con sobrenadante.	30
Ilustración 5: Gel de electroforesis que muestra la proteína antes de ser pasada por la Hi-Trap (derecha) y después de ser pasada por la Hi-Trap (izquierda).....	31
Ilustración 6: Diferentes etapas del lavado. a) Cuerpos de inclusión recién resuspendidos b) elución de lavados con 10 mM de imidazol c) elución de lavados con 40 mM de imidazol d) elución de 500 mM con lavados previos simples e) elución de 500 mM con lavados previos de 40 mM f)elución de 500 mM con lavados previos de 10 mM .....	32
Ilustración 7: Comparación de los cromatogramas obtenidos antes y después de la segunda purificación por Hi-Trap. Izquierda) Mezcla de producto de corte con	



---

trombina. Centro) Fracción que pasó por la Hi-Trap sin retenerse. Derecha) Fracción que eluyó de la Hi-Trap con 500 mM de imidazol.....	33
Ilustración 8: Espectrometría de masas MALDI-TOF de la GS-DEF1 soluble.....	34
Ilustración 9: Comparación de los perfiles cromatográficos de la GS-DEF1 (izquierda) soluble y (derecha) replegada.....	36
Ilustración 10: Comparación el espectro TOCSY de la GS-DEF1 soluble (izquierda) y la GS-DEF1 replegada (derecha) .....	38
Ilustración 11: Acercamiento a las interacciones HA-HA de algunos aminoácidos involucrados en la estructura secundaria .....	42
Ilustración 12: Acercamiento de la interacción de los hidrógenos alfa de la C33 con algunos HN involucrados en la estructura secundaria.....	43
Ilustración 13: Gráfico de NOES secuenciales en función de la secuencia de aminoácidos. ....	44
Ilustración 14: Gráfico de distribuciones NOE de corto y largo alcance, y número de restricciones para cada residuo.....	44
Ilustración 15: Gráfico de Ramachandran. La zona azul rey muestra las regiones más favorecidas (59.8%), la zona morada muestra las regiones permitidas (40.2%). Las lilas y blancas son las poco permitidas y las no permitidas respectivamente.....	45
Ilustración 16: En la parte superior se muestra la superposición de estructuras calculadas. En la parte inferior, la caricatura de una de las estructuras.....	46
Ilustración 17: Estructua de la GS-DEF1 con los aminoácidos hidrofóbicos explícitos en color naranja.....	47

---

Ilustración 18: Estructuras CS- $\alpha/\beta$ . A la derecha, estructura de una toxina de alacrán. A la izquierda, estructura de una defensina de almeja. .... 48

Ilustración 19: Estructuras terciaria de defensinas de moluscos e insectos. Se listan las estructuras por código PDB. a) 2B68 b) 1FJN c) 1ICA d) 1L4V e) 1MM0 f) 2NY8 g) Cobatoxina h) GS-DEF1 ..... 50

Ilustración 20: Comparación de la GS-DEF1 (verde) y la Cg-Def (azul). ..... 51

---

## GLOSARIO

ABREVIATURA	TÉRMINO COMPLETO
APS	Persulfato de amonio
CS- $\alpha/\beta$	Estructura $\alpha/\beta$ estabilizada por cisteínas
DEF1	Defensina 1
DTT	DL-Ditiotreitol
<i>E. coli</i>	<i>Escherichia coli</i>
HPLC	Cromatografía líquida de alta eficiencia
IPTG	Isopropil- $\beta$ -D-1-tiogalactopiranosida
NOE	Efecto nuclear Overhauser
NOESY	<i>Nuclear Overhauser Effect Spectroscopy</i>
PCR	Reacción en cadena de la polimerasa
PMSF	Fluoruro de fenilmetilsulfonilo
RMN	Resonancia Magnética Nuclear
rpm	Revoluciones por minuto
TEMED	N,N,N',N'-Tetrametiletilendiamina
TOCSY	<i>Total Correlation Spectroscopy</i>
Tris	Tris(hidroximetil)aminometano
Tam	Tamapina

## RESUMEN

Las defensinas son péptidos antimicrobianos que se encuentran en la mayoría de los seres vivos. Se ha propuesto que existe una relación evolutiva entre las toxinas y las defensinas de invertebrados, debido a que tienen el mismo tipo de estructura terciaria. Elucidar una estructura de defensina de alacrán es el primer paso para aportar evidencia experimental a dicha teoría.

Se obtuvo por métodos de biología molecular la defensina de alacrán GS-DEF1 recombinante en *E. coli*, y se logró plegar *in vitro* de manera exitosa. Se purificó por diferentes técnicas cromatográficas, y una vez pura se caracterizó por espectroscopía de masas y por resonancia magnética nuclear.

Los resultados de resonancia de la proteína replegada se utilizaron para elucidar la estructura terciaria y secundaria, utilizando el método de Wütrich para la asignación de los sistemas de espín y luego calcular los ángulos de torsión. La estructura obtenida tiene una hélice alfa y dos hojas beta antiparalelas, y corresponde al motivo alfa/beta-estabilizado por cisteínas como se había esperado. Hasta ahora, ninguna estructura de defensina de alacrán había sido elucidada.

---

## INTRODUCCIÓN

Las defensinas son péptidos antimicrobianos de gran diversidad. Tienen actividad antifúngica y antibacteriana e incluso contra algunos virus a bajas concentraciones. Se encuentran en una amplia variedad de especies, incluyendo mamíferos, pájaros, invertebrados, hongos y plantas. En los últimos años, se ha incrementado el interés en el estudio de las defensinas debido a su potencial uso como antibiótico<sup>1,2</sup>. Sin embargo, a pesar de que se conocen diversos péptidos de esta superfamilia<sup>3</sup>, el conocimiento de las defensinas dista aún de estar completo. Se ha teorizado sobre la relación evolutiva que existe entre las defensinas y las toxinas de los alacranes, con base en que su estructura terciaria que es muy similar<sup>4</sup>.

La obtención de una proteína recombinante en cantidades significativas puede realizarse de manera costeable en el sistema de expresión de *E. coli*<sup>5</sup>. Sin embargo, a pesar de que la producción de proteína de esta forma es alta, en ciertas ocasiones se obtiene agregada y mal plegada<sup>6</sup>. En estos casos el replegamiento *in vitro* de proteínas es una técnica útil para la obtención de una proteína pura en su forma activa, que soluciona dicho problema.

La resonancia magnética nuclear es una espectroscopía que permite elucidar la estructura de las moléculas de manera confiable, y puede utilizarse en proteínas pequeñas. En la determinación de la estructura de proteínas, la resonancia es complementaria a la difracción de rayos X, pues se puede conocer la estructura en disolución acuosa, y de proteínas de las que no se cuente con monocristal<sup>7</sup>.

La defensina 1 es una proteína del alacrán *Centruroides limpidus limpidus* reportada en 2004<sup>8</sup>. Hasta ahora, ninguna estructura terciaria de defensina de alacrán ha sido elucidada. El análisis de la secuencia de aminoácidos de la defensina 1 sugiere que es del tipo CS-  $\alpha/\beta$ , motivo estructural compartido por defensinas y toxinas. Encontrar la estructura terciaria de la defensina 1, de ser la esperada, podría ser evidencia experimental de la relación evolutiva entre estas dos familias de

---

péptidos<sup>8,9</sup>. Es de nuestro interés, por los puntos citados anteriormente, obtener a la defensina 1 mediante técnicas de biología molecular y elucidar su estructura terciaria a través de la técnica de resonancia magnética nuclear.

## ANTECEDENTES

### ¿QUÉ SON LAS DEFENSINAS?

Los péptidos antimicrobianos son catiónicos y anfipáticos, es decir, que los aminoácidos hidrofóbicos y catiónicos están espacialmente organizados en secciones de tal manera que pueden interactuar con las membranas biológicas y ser solubles en agua<sup>1</sup>. Presentan una gran diversidad de tamaños (desde doce hasta cientos de aminoácidos) y diferentes grupos estructurales. Los péptidos antimicrobianos mejor conocidos en los mamíferos son las defensinas y las catelicidinas<sup>1</sup>.

Las defensinas son pequeños péptidos de estructura compacta, cuya estructura tridimensional general consta de un centro de tres hojas beta y una hélice de tamaño variable en el amino terminal, estabilizadas por tres puentes disulfuro<sup>3</sup>. Esta estructura está altamente conservada; sin embargo, existe mucha variabilidad con respecto a la secuencia primaria de aminoácidos<sup>9</sup>, dando como resultado mucha diversidad de propiedades fisicoquímicas. Las defensinas de mamíferos se dividen en dos grupos: las  $\alpha$ -defensinas y las  $\beta$ -defensinas. Éstas se diferencian por el tamaño de la cadena, así como por la conectividad entre los residuos de cisteínas para la formación de los puentes disulfuro. Las primeras tienen una unión del tipo C1-C6, C2-C4, C3-C5 mientras que las segundas tienen una unión C1-C5, C2-C4 y C3-C6. Un tercer grupo de defensinas estructuralmente diferentes, las  $\theta$ -defensinas, ha sido identificado, las cuales se caracterizan por tener una estructura cíclica<sup>3</sup>.

Las defensinas se encuentran en todos los mamíferos en los gránulos de los leucocitos en altas concentraciones y en el tejido epitelial, dando protección contra infecciones. Además, la presencia de defensinas puede ser de tipo constitutivo (es decir, que se produce continuamente) o en respuesta a una infección<sup>1</sup>.

Las defensinas tienen actividad antifúngica y antibacteriana e incluso contra algunos virus a bajas concentraciones; sin embargo, su óptimo desempeño se da en

condiciones de fuerza iónica baja, así como bajas interferencias de cationes divalentes y otros péptidos<sup>3</sup>.

A pesar de que se conoce que las defensinas, de carga positiva, interaccionan con la membrana del agente infeccioso, de carga negativa, no se encuentra por completo elucidado el mecanismo de acción de las mismas. Se ha propuesto que el péptido se intercala en la membrana desestabilizándola. Esto podría llevar a diferentes mecanismos, como por ejemplo: 1) que la membrana se fragmente; 2) que el contenido de la célula se drene por agujeros físicos; 3) que se reacomoden los componentes de la membrana alterando la función de la misma; 4) que se despolarice la membrana; 5) que se internalice el compuesto y éste ataque funciones específicas<sup>1</sup>. Una comparación de las interacciones entre las defensinas y sitios específicos de la bicapa lipídica, ha llevado a identificar receptores específicos que permiten a las defensinas matar a ciertos organismos agresores sin causar citotoxicidad<sup>10</sup>.

El origen de las defensinas ha sido estudiado en los últimos años, y se han propuesto varios mecanismos para su diversificación, como evolución convergente y barajeo de exones, que a su vez han sido discutidos por otros autores como poco probables<sup>11</sup>. Otra propuesta es que las defensinas, a pesar de su amplia variedad, provienen de un ancestro común. Prueba de esto, es que se han encontrado péptidos tipo defensina en mixobacterias<sup>12</sup>.

Existe un creciente interés por el estudio de las defensinas, debido a que se considera que el profundo conocimiento de éstas será clave para el desarrollo de nuevos antibióticos, por lo que se han construido bases de datos actualizadas en donde se recopila información acerca de los últimos avances<sup>2</sup>.



## RELACIÓN ENTRE TOXINAS Y DEFENSINAS

Se ha observado que la gran mayoría de las defensinas de insectos y toxinas de alacrán comparten el motivo de plegamiento CS- $\alpha/\beta$  independientemente del tamaño del péptido, secuencia o función del mismo. Esta observación ha llevado a la hipótesis de que tanto las defensinas como las toxinas comparten un ancestro en común. Como soporte de esta hipótesis, se han encontrado péptidos tipo defensina en el veneno de alacrán <sup>4</sup>.

Se considera que los péptidos antimicrobianos con motivos CS- $\alpha/\beta$  son moléculas de rápida evolución debido a la competencia con los patógenos, por lo que las relaciones filogenéticas no suelen ser muy claras<sup>9</sup>; en este sentido, comparar las secuencias no suele ser muy útil para resolver las relaciones filogenéticas, y por tanto al tratarse de una secuencia nueva, aunque el motivo estructural sea el mismo, suele encontrarse poca identidad con los péptidos ya registrados en bases de datos. Se ha encontrado que en muchos casos, los péptidos que comparten este motivo tienen únicamente las posiciones de las cisteínas conservadas <sup>9</sup>.

En particular, las defensinas de los invertebrados son de tipo ancestral. La actividad de las defensinas se propone que se debe a su capacidad de interaccionar con la membrana citoplasmática. Por otro lado, existen principalmente dos tipos de neurotoxinas en el veneno de los alacranes: las de cadena larga, que modulan las funciones de los canales de sodio, y las de cadena corta que bloquean los canales de potasio. Se han encontrado similitudes altas entre la estructura y secuencia de algunas defensinas de alacrán y toxinas de cadena corta. Incluso se han encontrado defensinas con cierta capacidad moduladora de canales iónicos <sup>13</sup>.

Debido a que la secuencia primaria de las defensinas es muy variada y la relación entre toxinas y defensinas está evidenciada únicamente en su estructura terciaria, es clave resolver la estructura terciaria de las defensinas de alacrán para corroborar que

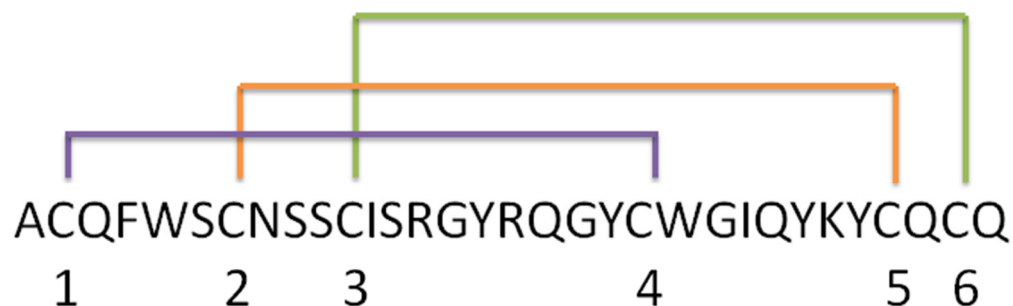
la estructura es del tipo CS- $\alpha/\beta$ , puesto que la secuencia misma no es suficiente para poder analizar a profundidad sus orígenes filogenéticos ni su relación con las toxinas.

### ¿QUÉ ES LA DEF1?

El nombre Defensina 1, acortado como DEF1, identifica al péptido aislado del alacrán *Centruroides limpidus limpidus*. También identificada como CII-dlp (*Centruroides limpidus limpidus-defensine like peptide*), está registrada con el número de secuencia Q6GU94 en la base de datos de UniProt<sup>1</sup>, pero también se puede encontrar con su código secundario P83738. Fue reportada por primera vez en 2004.

El *Centruroides limpidus limpidus* (Buthidae, Scorpiones, Arachnida, Artropoda, Animalia) es un alacrán nativo de México que habita principalmente en el estado de Guerrero. La DEF1 se extrajo de la hemolinfa de dicho alacrán, y se encontró que su secuencia madura es ACQFWSCNSSCISRGYRQGYCWGIQYKYCQCQ con tres puentes disulfuro del tipo C1-C4, C2-C5, C3-C6, que se encuentran comúnmente en las defensinas de invertebrados<sup>8</sup>.

Dicho péptido de 32 residuos de aminoácidos, puede esquematizarse de la siguiente manera:



<sup>1</sup><http://www.uniprot.org/uniprot/Q6GU94>

Rodríguez de la Vega *et al* (2004)<sup>8</sup> estudiaron los mecanismos de síntesis de la defensina, y encontraron que es de tipo inducida, es decir, no se encuentra de manera cotidiana en la hemolinfa del alacrán, sino que se acumula una vez inoculado el mismo; pero presenta una actividad antimicrobiana marginal. Sin embargo, al asociarse con dos péptidos aún no identificados (llamados por los autores del trabajo como HCII-II.12a y HCII-II.12b) que se encuentran de manera constitutiva en la linfa (y que por si mismos no presentan actividad) en cantidades equitativas, presenta actividad biológica contra *E. coli*, *B. subtilis*, *K. pneumoniae* y *S. aureus* en concentración de péptido total de 20 µg/mL.

Los autores del trabajo compararon la secuencia primaria de la defensina 1 con las secuencias de otras defensinas conocidas de invertebrados, así como de toxinas de veneno. Se encontró una similitud baja con ambas; sin embargo, el parecido es mayor a las defensinas que a las toxinas. Únicamente los residuos de cisteína se encuentran completamente conservados. A pesar de no contar con datos sobre la estructura secundaria o terciaria, a partir de la información filogenética de la alta conservación del motivo CS- $\alpha/\beta$  (*Cysteine stabilized  $\alpha/\beta$* ), los autores sugieren que la DEF1 sea de éste tipo. Dicha estructura es la más común entre los péptidos antimicrobianos y se define por dos puentes disulfuro que unen la secuencia Cys<sub>i</sub>-Xaa-Xaa-Xaa-Cys<sub>1+4</sub> con Cys<sub>j</sub>-Xaa-Cys<sub>j+2</sub> a través de los puentes formados por Cys<sub>i</sub>-Cys<sub>j</sub> y Cys<sub>1+4</sub>-Cys<sub>j+2</sub>. Esto se cumple para la DEF1, siendo Cys<sub>i</sub>= Cys<sub>2</sub> y Cys<sub>j</sub>= Cys<sub>5</sub><sup>8</sup>.

#### **PROTEÍNAS RECOMBINANTES: EXPRESIÓN EN *E. COLI***

Se conocen<sup>5</sup> como proteínas recombinantes a las proteínas expresadas por la traducción de material genético recombinante. Durante las últimas tres décadas, se han estudiado y perfeccionado los modos y la tecnología para expresar una proteína dada, en un organismo anfitrión ajeno a la producción de la misma en condiciones naturales. La técnica general consiste en introducir el material genético por medio de un vector a un sistema de expresión, que comúnmente puede ser una bacteria, levadura, célula de insecto o de mamífero. Estos sistemas, aunque no son los únicos,

son fáciles de conseguir comercialmente, y se conocen protocolos sencillos, baratos y probados para llevar a cabo la expresión de manera exitosa para producción a escala de laboratorio de investigación la proteína de interés

El sistema de expresión<sup>5</sup> de la proteína deseada se elige en función de las características de la proteína a producir. Como regla general, las modificaciones post traduccionales (glicosilaciones, amidaciones, hidroxilaciones, etc.) se dan en sistemas eucariontes, y varían el tipo de modificaciones según el sistema de expresión. En la siguiente tabla, se resumen las características (ventajas y desventajas) de los diferentes sistemas de expresión que existen de manera comercial.

Tabla 1: Sistemas de expresión de proteínas recombinantes, ventajas y desventajas <sup>14</sup>

SISTEMA	VENTAJAS	DESVENTAJAS
<i>E. coli</i>	Bajo costo: medios baratos y alto rendimiento	Modificaciones post traduccionales limitadas
	Expresión rápida (días)	Dificultad para producir algunas proteínas de manera soluble y bien plegada
	Fácil de escalar	
	Genética bien conocida	
<i>P. pastoris</i>	Bajo costo: medios baratos y alto rendimiento	Unión N-glicano diferente a la de mamíferos
	Expresión moderada (semanas)	Requiere medidas de seguridad al producir en gran escala por la formación de metanol
	Muchas de las modificaciones post traduccionales y buen plegamiento	
<i>Células de insecto</i>	Muchas de las modificaciones post traduccionales y buen plegamiento	Unión N-glicano diferente a la de mamíferos
	Expresión moderada (semanas)	Alto costo, bajo rendimiento

<b>Células de mamífero</b>	El tiempo de expresión puede ser moderado (semanas) a lento (meses) dependiendo del tipo de expresión	
	Todas las modificaciones post traduccionales y buen plegamiento	Alto costo, bajo rendimiento
		Dificultad para escalar

Aunque existen lineamientos para elegir el mejor sistema de expresión para una proteína problema, en algunos casos es necesario probar diferentes sistemas para poder encontrar el sistema óptimo de expresión.

Para proteínas que no tienen modificaciones post traduccionales, el sistema de expresión en *E. coli* es el más deseable, debido a su manejo sencillo y de bajo costo. Los tiempos de obtención de la proteína son rápidos y se obtiene en muy altos rendimientos de hasta 30% de la biomasa. Sin embargo, en ciertas proteínas como las ricas en puentes disulfuro, o tóxicas para la bacteria, la sobre-acumulación de proteína genera agregados insolubles llamados cuerpos de inclusión. En los cuerpos de inclusión, las proteínas se pueden encontrar mal plegadas y por lo tanto no en su forma activa, por lo que es necesario un tratamiento posterior. Para evitar la formación de los cuerpos de inclusión y obtener la mayor cantidad de proteína soluble en agua, suele bajarse la temperatura de expresión. Se considera que, de esta manera se disminuye la velocidad de transcripción, traducción y plegamiento, además de bajar la actividad de las proteasas. Otra estrategia es utilizar proteínas chaperonas que aumenten la solubilidad de la proteína de fusión. Por último, la formación apropiada de puentes disulfuro (y con esto, el mejor plegamiento de una proteína) está limitada debido a la naturaleza reductora del citoplasma y la carencia de retículo endoplasmático<sup>14</sup>.

### **PROTEÍNA DE FUSIÓN**

La expresión de proteínas con formación de puentes disulfuro con la conectividad correcta en *E. coli* ha sido un obstáculo para la producción de proteínas bien plegadas

en éste sistema, y se han desarrollado diferentes estrategias para conseguir cantidades significantes de proteína bien plegada. Sin embargo, a pesar de la gran cantidad de información que hay acerca de estrategias que han funcionado para producir algunas proteínas de manera plegada, no hay manera certera de predecir si cierto sistema va a funcionar para producir una proteína dada, pues los mecanismos involucrados en el plegamiento de las proteínas siguen sin haber sido descritos por completo de manera precisa<sup>6,15</sup>.

Los compartimentos intracelulares tienen potenciales redox específicos, lo cual genera un ambiente reductor u oxidante. La formación de puentes disulfuro está limitada debido al ambiente reductor citoplasmático, pues se evita la oxidación de los grupos tiol. El intercambio tiol-disulfuro en sistemas biológicos es altamente específico con respecto a los grupos tiol involucrados, y está controlado por catálisis enzimática de oxidoreductasas conocidas como tiorredoxinas<sup>16,17</sup>.

La tiorredoxina (Trx) es una enzima que reduce puentes disulfuro por medio de un mecanismo “ping-pong”, como se muestra a continuación:

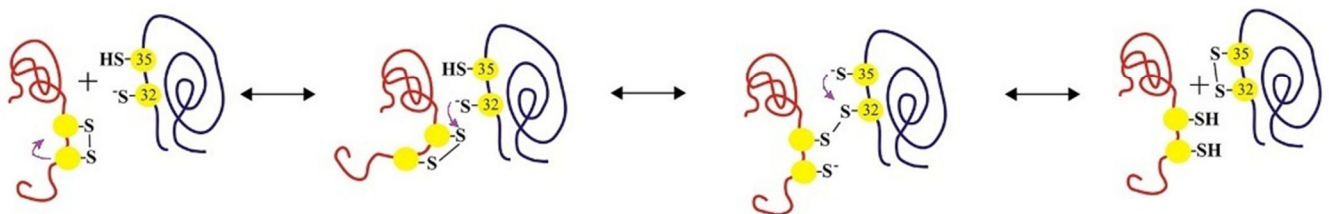


Ilustración 1: Mecanismo de acción de reducción de un sustrato por la tiorredoxina<sup>16</sup>. La proteína roja es un sustrato, mientras la azul representa la tiorredoxina.

Sin embargo, a pesar de que la tiorredoxina principalmente reduce los puentes disulfuro, se ha visto que es un agente auxiliar en el plegamiento de las proteínas, ya sea por la formación de los puentes disulfuro en el citoplasma de *E. coli*, o por ayudar a que exista un rearrreglo de los puentes disulfuro. *In vivo*, la proteína funciona como chaperona aumentando la solubilidad de la proteína de fusión, lo que hace que las

proteínas se plieguen mejor. Esto es independiente de su actividad redox, pues se ha visto que el aumento en solubilidad se da también en los análogos con sitios activos mutados. En el caso de que la proteína no aumente su solubilidad *in vivo*, la tiorredoxina es auxiliar para realizar una solubilización y replegamiento *in vitro*<sup>16,18</sup>

Para simplificar la purificación de las proteínas, se emplea un sistema de cromatografía de afinidad metal quelato, en el cual se usa una columna empacada cargada con iones de níquel (II), el cual se coordina al nitrógeno imidazólico de las histidinas. Para conseguir una mayor afinidad, se diseña que en la proteína de fusión haya una cola de histidinas de seis residuos, de manera que el ángulo de anclaje es óptimo y la afinidad por el níquel (II) es significativamente mayor que la de cualquier proteína nativa.

Por último, es necesario una estrategia en la que pueda desprenderse la proteína chaperona de la proteína deseada. Para esto, se diseña que en la proteína de fusión exista un sitio donde pueda actuar una proteasa (sitio de corte). En particular, se va a utilizar un sitio de corte de trombina para separar la proteína DEF1 del resto de la proteína de fusión, cuyo sitio de reconocimiento es LVPR:GS, por lo cual la proteína final quedará como GS-DEF1 en lugar de simplemente DEF1.

### **PLEGAMIENTO DE PROTEÍNAS RECOMBINANTES**

La obtención de una proteína sobreexpresada en forma de cuerpos de inclusión puede ser ventajosa en algunos casos debido a que son: resistentes a la proteólisis, fáciles de concentrar por centrifugación, poco contaminados con otras proteínas, y pueden ser replegados a su forma activa<sup>14</sup>.

Sin embargo, los porcentajes de reacción del replegamiento pueden ser muy variados. El principal problema para replegar, son los subproductos cinéticos que forman agregados insolubles. Para evitar este problema, es necesario trabajar en

disoluciones muy diluidas, estrategia que a su vez tiene la desventaja de que el producto final estará a muy baja concentración <sup>19</sup>.

De manera general, la metodología del replegamiento de una proteína consiste en la solubilización de los cuerpos de inclusión con agentes caotrópicos y su desnaturalización con algún reductor de puentes disulfuro. Luego se realiza un intercambio de disolución amortiguadora, o *buffer*, a un medio naturalizante, para obtener la proteína nuevamente oxidada y plegada de manera correcta. Se han probado muchas estrategias para replegar proteínas, cambiando toda clase de variables. Algunos de los parámetros de la matriz de reacción que pueden afectar decisivamente el replegamiento de una proteína son: fuerza iónica, pH, temperatura, potencial redox, concentración de la proteína y presencia de agentes caotrópicos. Además, el cambio de *buffer* puede darse por diferentes métodos que a su vez tendrán un efecto determinante en el éxito del replegamiento, como son la dilución, diálisis, filtración e inmovilización en una fase sólida. Debido a que no hay un método universal de plegamiento y el éxito de un método dado es dependiente de la proteína objetivo, en muchos casos es necesario probar diferentes métodos para encontrar qué variables son claves en la proteína particular con la que se está trabajando. Existen de manera comercial *kits* especializados en optimizar el plegamiento de las proteínas, como el QuickFold® de AthenaES. También hay proyectos a gran escala (REFOLD database) que buscan compilar la mayor cantidad de información posible sobre intentos exitosos y fallidos de replegamiento de proteínas para elucidar los mecanismos. Debido a que se desconoce las razones por las que algunos métodos funcionan con ciertas proteínas y con otras no, se considera que al recopilar y homogeneizar información subida por los propios usuarios, se generará una base de datos lo suficientemente completa para arrojar luz acerca de este tema<sup>20,21</sup>. Por lo tanto, el replegamiento es una estrategia deseable para la purificación de proteínas; sin embargo, es difícil de diseñar pues es particular para cada caso.



## RESONANCIA MAGNÉTICA NUCLEAR: GENERALIDADES

La resonancia magnética nuclear (RMN) es una espectroscopía de amplio uso actualmente, que revela información crucial sobre la estructura, dinámica y reactividad de un compuesto químico. La espectroscopía es por definición, el estudio de la interacción de la materia con la radiación electromagnética. En particular, el principio de la RMN se debe a que los isótopos observables ( $^1\text{H}$ ,  $^{13}\text{C}$ , etc.) con un espín nuclear de  $\frac{1}{2}$ , al ser sometidos a un campo magnético externo, los niveles energéticos del número cuántico  $m$  se separan, dando  $m=-\frac{1}{2}$  y  $m=\frac{1}{2}$  cuya separación es dependiente de la fuerza del campo, el primero representa la orientación antiparalela al campo, y la segunda la paralela al campo, que es la preferencial. La transición entre los dos niveles de energía se inducen por irradiación, y la transición cumple con la regla de selección  $\Delta m = \pm 1$ , y puede observarse gracias a la diferencia de poblaciones en cada nivel<sup>22</sup>.

El experimento de RMN puede esquematizarse del siguiente modo: se introduce una muestra en disolución en un campo magnético estático  $B_0$  y se irradia con una o más radiofrecuencias  $B_1$  ( $B_2$ ,  $B_3$ , etc.). Los espines nucleares de la muestra se polarizan en presencia de  $B_0$  lo cual resulta en una magnetización macroscópica  $M$  que está alineada con el campo magnético. Al aplicar la radiofrecuencia  $B_1$  en el eje perpendicular al campo, se ejerce una torca en la magnetización, de manera que ésta rota  $90^\circ$  y queda en el plano perpendicular a  $B_0$ . La magnetización transversal precesa bajo la influencia de  $B_0$  a la frecuencia de resonancia  $\nu_0$ , lo cual genera una corriente eléctrica que es detectada. Debido a que el sistema tiende a llegar al equilibrio termodinámico, la magnetización transversal decae con el tiempo. Se le llama FID (por sus siglas en inglés) al decaimiento de inducción libre, que se mide en un tiempo determinado, y cuyos datos se procesan con transformada de Fourier para obtener el espectro final<sup>7</sup>.

Las propiedades que determinan los espectros de RMN se enlistan a continuación<sup>23</sup>:

- Desplazamiento químico ( $\delta$ ): Puesto que el campo magnético externo no es idéntico al interno de la muestra, debido al campo inducido por los electrones orbitando, existe un efecto de protección magnética en los núcleos. Esto se traduce en que no todos los núcleos aparecen en la misma posición en el espectro, y las diferencias entre las posiciones se conocen como desplazamientos químicos.
- Constante de acoplamiento espín-espín (J): El acoplamiento espín espín, también conocido como escalar o indirecto, tiene como resultado la división de la banda de una señal, y la distancia entre las bandas resultantes es el valor de la constante de acoplamiento. Se considera que es un acoplamiento indirecto debido a que está mediado por los electrones del enlace químico. Se consideran acoplamientos de corto alcance los que involucran a tres enlaces o menos.
- Efecto Nuclear Overhauser (NOE): Se refiere a los cambios de intensidad de una señal debido al acoplamiento dipolar. Se debe a los efectos de relajación a través del espacio, y es dependiente de la distancia entre los núcleos.

La observación del  $^1\text{H}$  es la más alta de todos los isótopos que pueden obtenerse, además de que al encontrarse en la parte periférica de la estructura, las señales en este tipo de resonancia tienen una buena sensibilidad para ver interacciones de no enlace<sup>7</sup>.

El experimento TOCSY es un experimento bidimensional en el cual el sistema evoluciona bajo la influencia únicamente de las constantes de acoplamiento. El experimento NOESY, también bidimensional, establece conectividades entre espines a través del relajamiento cruzado (que es la transferencia de polarización entre dos núcleos que no están acoplados escalarmente, pero que se encuentran cercanos en el espacio<sup>22</sup>). La velocidad del relajamiento cruzado es dependiente de las distancias nucleares, por lo que la señal que se observa en este experimento depende de la distancia espacial (no vía enlace) de los núcleos. Idealmente, para la asignación de

una estructura debe utilizarse algún experimento (como el TOCSY) de determinación indirecta para asignar las señales de los núcleos correspondientes a las señales observadas, y así encontrar la conectividad, y utilizar las señales NOE restantes del espectro (que son únicamente espaciales) para determinar la orientación, ángulo y distancia de separación que existe entre los núcleos involucrados<sup>23</sup>.

## **RESONANCIA MAGNÉTICA NUCLEAR Y PROTEÍNAS**

El inicio<sup>24</sup> de la aplicación de la RMN en la biología, se dio en 1957 con la publicación del primer espectro de una proteína. Se considera a Oleg Jardetzky, Mildred Cohn y Robert Shulman como los fundadores de esta rama. En la década de los sesenta, se reconoció que la técnica de la RMN podía proveer información acerca de la estructura secundaria y terciaria de una proteína, así como de la dinámica molecular interna de ésta. También se concibió la idea de marcar isotópicamente una proteína para estudiar detalladamente estos aspectos. Una de las primeras observaciones importantes fue que el espectro de una proteína desnaturalizada es la suma de los espectros de los aminoácidos que la conforman, no así una proteína en su forma nativa. A partir de la década de los setenta, empezaron a proliferar los avances en la tecnología de la RMN, haciendo posible que diversos grupos estudiaran proteínas con dicha técnica, pero no fue hasta que en los ochenta, cuando Wüthrich y Wagner elucidaron la estructura de la misma proteína de manera simultánea y ciega, por RMN y difracción de rayos X, respectivamente, que se aceptó el potencial de la RMN en este campo de manera generalizada

La determinación de la estructura terciaria de las proteínas ha sido llevada a cabo por técnicas de difracción de rayos X, décadas antes de que la tecnología de la RMN pudiera ser utilizada para este fin. Ambas técnicas son complementarias; mientras que en difracción de rayos X es necesario tener la muestra en forma de monocristal, en RMN se puede estudiar la estructura en disolución acuosa, lo cual permite estudiar muestras que no se han logrado cristalizar, además de que pueden hacerse estudios

de la muestra a diferentes condiciones de pH, temperatura, fuerza iónica, e incluso hacer estudios de la dinámica biomacromolecular interna <sup>7</sup>.

La RMN de sólidos se ha utilizado también de manera exitosa para determinar la estructura de muestras sólidas de proteínas que no se pueden cristalizar ni mantenerse en disolución<sup>25</sup>.

En el anexo se muestran las estructuras de los 20 aminoácidos más comunes, así como el patrón COSY y TOCSY asociado a cada uno de ellos.

Debido a que existe un plegamiento en las cadenas de biopolímeros, hay una dispersión de los desplazamientos químicos, es decir, no se conserva el valor del desplazamiento químico de un monómero libre al monómero concatenado. Además, protones que normalmente son lábiles en la forma monomérica y por lo tanto no se ven en la RMN, una vez que se encuentran en la cadena pueden ser observados<sup>7</sup>.

En una proteína no plegada, los residuos del mismo tipo se encuentran rodeados principalmente de disolvente y por lo tanto tienen un ambiente químico muy parecido, lo cual se traduce en que aparecen en el mismo desplazamiento químico. Sin embargo, en una proteína plegada, estos mismos residuos se encontrarán rodeados de otros residuos y tendrán por lo tanto un microambiente diferente entre si, lo cual tiene como resultado que cada resto de aminoácido aparecerá en un desplazamiento químico diferente<sup>7</sup>.

La RMN también se utiliza para conocer las dinámicas internas de una proteína, así como el estudio de la interacción de una proteína con moléculas pequeñas. Se ha hecho el seguimiento del plegamiento de una proteína por esta técnica *in vitro*, para encontrar los patrones de las estructuras secundarias y patrones de los residuos hidrofóbico e hidrofílicos<sup>26,27</sup>.

Se han desarrollado métodos de etiquetamiento isotópico para experimentos de RMN en proteína que se llevan a cabo en medio mínimo, y que facilitan la producción de

proteínas enriquecidas en  $^{15}\text{N}$  y  $^{13}\text{C}$  para su mejor asignación<sup>28</sup>, puesto que permiten resolver sistemas de espín superpuestos. Cabe mencionar que se ha reportado la estructura de una defensina de invertebrado por RMN, y se obtuvo una estructura del tipo CS- $\alpha/\beta$ <sup>29</sup>.

## OBJETIVOS

- El presente trabajo tiene como objetivo producir la defensina 1 de alacrán de manera recombinante (GS-DEF1) y purificarla.
  - Para cumplir este objetivo, es necesario adicionalmente plegar la proteína *in vitro*.
- Una vez obtenida, es de nuestro interés caracterizarla por resonancia magnética nuclear y calcular los ángulos de torsión, de manera que se obtenga la estructura terciaria de la GS-DEF1; siendo así la primera estructura terciaria de defensina de alacrán en ser elucidada.

## HIPÓTESIS

- Se espera encontrar que la estructura terciaria de la defensina GS-DEF1 presente el plegamiento del tipo CS-  $\alpha/\beta$ , que consta de una hélice alfa y dos hojas beta estabilizados por cisteínas.

## METODOLOGÍA EXPERIMENTAL

Los datos técnicos, composición de disoluciones, pureza y marca de reactivos y especificaciones de los equipamientos, se encuentran detallados en el apéndice.

Un resumen del protocolo experimental se encuentra en el esquema de la siguiente página.

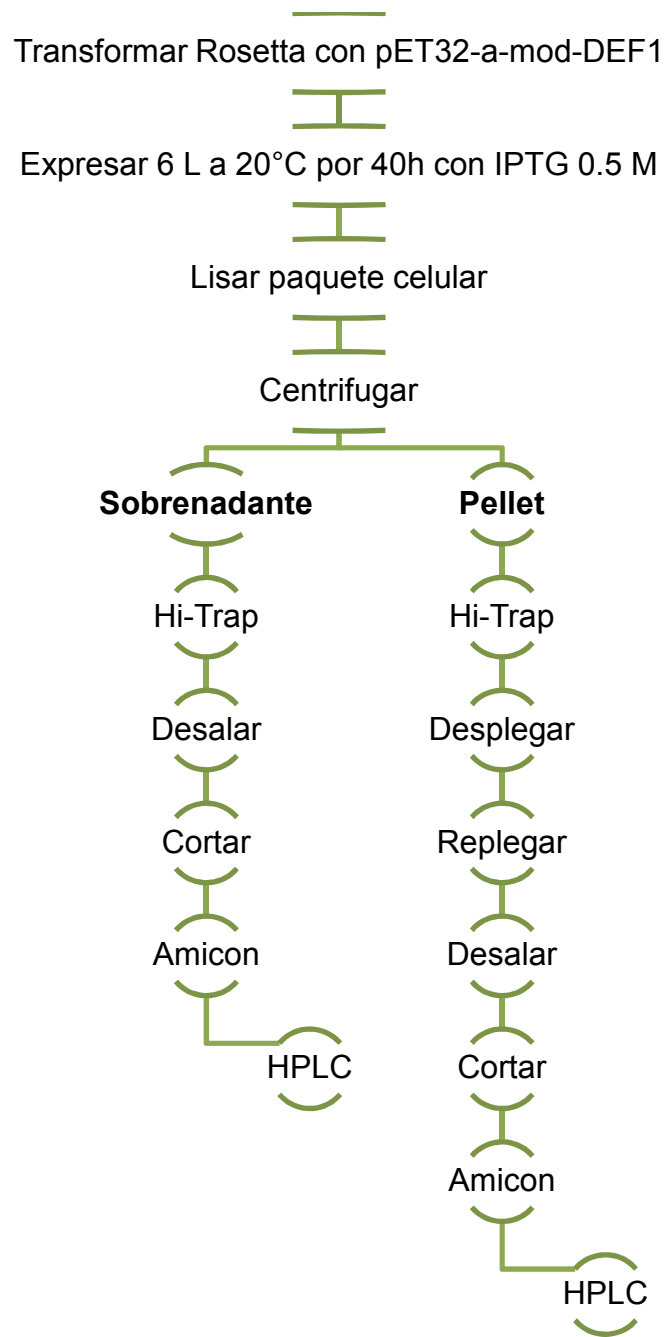
### **CLONACIÓN DEL GEN DE LA PROTEÍNA DE FUSIÓN-DEF1 EN EL VECTOR PET32-A-MOD**

Con base en la secuencia de aminoácidos reportada para Rodríguez de la Vega *et al*<sup>8</sup>, se diseñó el gen de la DEF1 optimizando los codones para su expresión en *E. coli*, agregando los sitios de restricción Nco1 y BamH1 señalados en negritas y un codón de paro señalado en cursivas. El gen se sintetizó en los laboratorios Epochlab (CA, USA),

**CCATGGGCGCGTGCCAATTTTGGTCTTGCAACTCTTCTTGCATCTCTCGTGGTTA**  
**CCGTCAGGGTTACTGCTGGGGTATCCAGTACAAATACTGCCAGTGCCAGTAGGG**  
**ATCC**

El gen sintético clonado en el vector pBSK--- se subclonó en el vector pET32a-mod previamente digerido con Nco1/BamH1 usando T4 DNA ligasa (Biolab) generando el nuevo vector (pET32aMod-DEF1) el cuál fue transformado en células de *E. coli* DH5 $\alpha$  (Novagen). De las clonas obtenidas se realizó una purificación del DNA plasmídico, el cual fue secuenciado para corroborar marco de lectura.

ESQUEMA GENERAL DE EXPRESIÓN Y PURIFICACIÓN DE LA DEF1





### **EXPRESIÓN OPTIMIZADA DE PROTEÍNA SOLUBLE**

Con la construcción del pET32-1-mod-DEF1 se transforman células competentes de Rosetta por el método del choque térmico. Se deja un preinóculo toda la noche (*overnight*) a 37° C en agitación y con éste se inocula el volumen total a tratar en agitación a 37° C hasta que alcance una densidad óptica (O.D.) de 0.5-0.7 a 600 nm; punto en el cual se reduce la temperatura a 20° C, se induce la expresión con una concentración de IPTG 0.5 M final y se deja el cultivo en agitación de 30-40 horas.

Una vez que se haya terminado la expresión de la proteína, se centrifuga el medio de cultivo a 8,000 x g durante al menos 20 minutos para separar el paquete celular del medio líquido. Se congela el paquete celular (lo que facilita su posterior lisis) para después resuspender con buffer de lisis (500 mM cloruro de sodio, 50 mM Tris), con un aproximado de 20 mL de buffer por cada gramo de paquete celular. Se dan golpes sónicos a potencia 6 Watts durante 20 segundos, con 1.5 minutos de descanso durante 10 minutos netos, o hasta que se aprecie que el paquete celular se encuentra del todo lisado (Sonicator 3000 Misonix). Finalmente, se centrifuga a 16,500 rpm por una hora (BeckmancoulterAvanti J-25, rotor JA-17). La fase líquida se denominará a partir de ahora sobrenadante, o fracción soluble, y la fase sólida se denominará pellet.

### **PURIFICACIÓN OPTIMIZADA DE PROTEÍNA SOLUBLE**

Se filtra con membranas *Milipore* el sobrenadante y se hace pasar por una columna de afinidad metal-quelato (Hi-Trap cargada con níquel (II)) previamente equilibrada con el buffer de lisis. Se realiza un lavado simple con buffer de lisis y se eluye con buffer de elución (buffer de lisis con 500 mM de imidazol). En este punto, la proteína se encuentra considerablemente más concentrada, aunque presenta aún contaminantes. Se remueve el imidazol de la proteína usando membranas de diálisis de 10,000 kDa para quedar nuevamente en buffer de lisis. Es importante en este punto no bajar la concentración de cloruro de sodio, pues éste ayuda a mantener a la proteína soluble. La proteína de fusión se corta con trombina (Sigma) a temperatura

ambiente en buffer de corte (buffer de lisis con 10 mM de cloruro de calcio), de 16 a 24 horas. Se centrifuga para separar de la trombina, se ultrafiltra con membrana de Amicon de 10,000 para retirar los contaminantes y residuos de corte, de manera que se colecta únicamente la fracción que pasa el filtro. La fracción colectada se inyecta al cromatógrafo para separar en columna analítica de fase reversa (Jupiter 4u Proteo 90 A250 x 4.60 nm Phenomenex) con un gradiente 20-50% de acetonitrilo TFA 0.05% y un flujo de 1 mL/min, durante 30 minutos. La fracción colectada se liofiliza, y se obtiene GS-DEF1 pura.

#### **PURIFICACIÓN OPTIMIZADA DE PROTEÍNA INSOLUBLE Y PLEGAMIENTO**

Los cuerpos de inclusión que están en el pellet se solubilizan con una disolución de buffer de lisis con urea 8 M (agente caotrópico). El extracto se filtra (puede antes centrifugarse para facilitar éste paso) y se hace pasar por una columna de afinidad Hi-Trap cargada con níquel (II) equilibrada con el mismo buffer. Se realiza un lavado simple y se eluye con buffer de lisis con urea 8 M y 500 mM de imidazol. La proteína se desnaturaliza con DTT 1 mM durante una hora, y se intercambia el buffer por medio de diálisis con membranas de 10,000 al buffer de replegamiento (Tris 100 mM, arginina 600 mM, cistina 5 mM, cisteína 0.5 mM). Se dializa contra la disolución en tres pasos con agitación constante, y en el último recambio de buffer se deja toda la noche a temperatura ambiente. Cada cambio de buffer debe ser contra 50 veces el volumen de la muestra. En este punto se pierde el agente reductor lentamente y se retira también la urea y el imidazol, imponiendo un potencial redox con el par cistina/cisteína con arginina como agente caotrópico. Una vez terminado este paso, se mantiene la disolución en la membrana de diálisis y se cambia el buffer por el de corte con trombina, por lo menos 16 horas. La muestra se ultrafiltra por una membrana de 10,000 para quitar el residuo de corte y agentes contaminantes, y la fracción eluída se purifica por cromatografía líquida de alta eficiencia de fase reversa en las mismas condiciones que se especificaron para la proteína soluble. La fracción colectada se liofiliza y se obtiene GS-DEF1 pura y replegada.

## ADQUISICIÓN DE LOS ESPECTROS DE RMN DE LA MUESTRA

Los experimentos de RMN de protón se realizaron en un equipo Varian INOVA de 500 MHz, en un tubo Shigemi con una disolución de agua con agua deuterada al 5%. La adquisición fue de tres días. El experimento NOESY tuvo un tiempo de mezclado de 300 y 100 ms y el experimento TOCSY tuvo un mezclado de 0.074 s. Tanto el TOCSY como el NOESY tuvieron 512 incrementos complejos y 1024 puntos complejos en la dirección de detección. Se utilizó esta proporción de H<sub>2</sub>O/D<sub>2</sub>O para que los protones lábiles no se intercambiaran por deuterio, y pudieran por tanto ser observados. Estos espectros se utilizaron para realizar la asignación de los sistemas de espín. Se realizó un experimento homólogo en el equipo de 700 MHz Varian VNMRS, el cual se utilizó para calcular los ángulos de torsión y definir la estructura terciaria. Se procesaron los datos adquiridos en NMRDraw y NMRPipe; se asignaron los picos utilizando *software* de CARA 1.5. Posteriormente se analizaron los ángulos de torsión con el programa CYANA.

## RESULTADOS Y DISCUSIÓN

### ANÁLISIS DE LA SECUENCIA DE LA GS-DEF1 Y LA PROTEÍNA DE FUSIÓN

La expresión de la DEF1 recombinante da dos residuos extra (GS) en el amino terminal, y los parámetros calculados utilizando ExPASyCA > Tools > Primary structure analysis > ProtParam para el péptido GS-DEF1 son<sup>2</sup>:

Tabla 2: Características y parámetros calculados de la GS-DEF1

CARACTERÍSTICA	VALOR
Número de aminoácidos	34
De los cuales:	
Neutros	11
Positivos	3
Especiales	10
Hidrofóbicos	10
pI teórico	8.65
Fórmula	C <sub>170</sub> H <sub>243</sub> N <sub>49</sub> O <sub>50</sub> S <sub>6</sub>
Total de átomos	518
Peso molecular	3,965.4
Peso molecular oxidado (-6 H)	3,959.4
Coeficiente de extinción molar a 280 nm (M <sup>-1</sup> cm <sup>-1</sup> ):	
Todas las cisteínas reducidas	17,335
Todas las cisteínas apareadas	16,960

---

<sup>2</sup> <http://ca.expasy.org/tools/protparam.html>

La construcción final de la proteína de fusión tiene la siguiente secuencia de aminoácidos: MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY QCGKLTVAKL NIDQNPGTAP KYGIRGIPTL LLFKNGEVAA TKVGALSKGQ LKEFLDANLA GSGSGHMHSHHSSGLVPH SSGMKETAAA KFERQHMDSP DLGTGLVPRGSACQFWSCNSSCISRGYRQG YCWGIQYKYC QCQ

Y se puede esquematizar de la siguiente manera, tomando en cuenta la funcionalidad de cada sección, con el mismo código de color.



Se muestra a continuación una tabla con las características de la proteína de fusión, calculadas de igual manera que se calcularon los mismos para la GS-DEF1.

Tabla 3: Características y parámetros calculados de la proteína de fusión

CARACTERÍSTICA	VALOR
Número de aminoácidos	193
pI teórico	6.53
Fórmula	$C_{926}H_{1429}N_{259}O_{277}S_{14}$
Total de átomos	2905
Peso molecular	21070.9
Peso molecular oxidado (-8 H)	21062.9
Coeficiente de extinción molar a 280 nm ( $M^{-1} cm^{-1}$ ):	
Todas las cisteínas reducidas	30940
Todas las cisteínas apareadas	31440

**COMENTARIOS SOBRE EL SISTEMA DE EXPRESIÓN**

Para producir a la DEF1 se probaron diferentes cepas de *Escherichia coli* con la finalidad de encontrar la cepa que produjera mayor cantidad de defensina soluble. Las características de las cepas de *E. coli* se encuentran resumidas en la tabla 4.

Tabla 4: Características principales de las cepas de *E. coli* probadas para la expresión de la DEF1<sup>30</sup>

CEPA	RESISTENCIA	CARACTERÍSTICAS <sup>3</sup>	CLAVE
BL21(DE3)	CAM	Es la cepa más ampliamente utilizada, de ésta se derivan la ρ, τ, C. Deficiente en las proteasas <i>lon</i> y <i>opm-t</i> . Expresa proteínas de vector pET por inducción con IPTG de manera controlada.	π
C41	-	Sobre expresión de proteínas tóxicas y proteínas de membrana.	C
Tuner	-	La expresión es dependiente de la concentración de IPTG.	τ
Rosetta (DE3)	CAM	Ideal para proteínas eucariotas con codones raros.	ρ
Rosetta gami (DE3)	KAN, CAM, TET	Favorece la formación de puentes disulfuro en proteínas no nativas. Rosetta + Origami	γ

Para esta prueba, se llevó a cabo la expresión a 28° C con 0.5 mM de IPTG durante 6 horas, tiempo en el cual se obtiene la máxima cantidad de proteína en Rosetta gami según experimentos preliminares. Todas las muestras se sometieron al mismo proceso de purificación.

<sup>3</sup> Un catálogo resumido en la página: <http://wolfson.huji.ac.il/expression/bac-strains-prot-exp.html> de The Hebrew University of Jerusalem, por la Dra. Nurit Doron.

En la siguiente ilustración (2), se muestra el SDS-PAGE de las muestras purificadas por cromatografía de afinidad. Se observa, que las intensidades relativas de la banda que corresponde a la proteína de fusión son muy parecida para todos los sistemas de expresión probados, por lo que no puede utilizarse este criterio para seleccionar la cepa más apropiada para la expresión de la defensina.

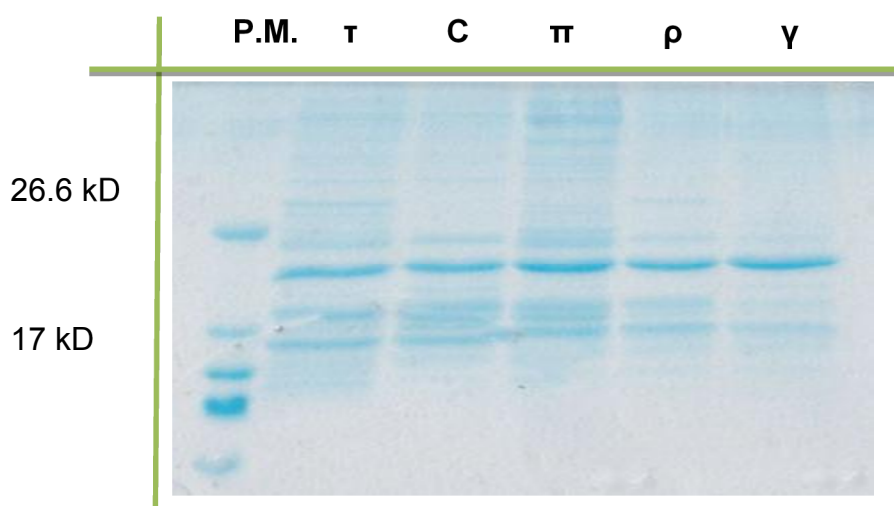


Ilustración 2: Fracción del sobrenadante purificado por columna de afinidad del extracto de cada una de las cepas utilizadas, crecidas a 28° C durante 6 horas con 0.5 mM de IPTG.

Debido a que el objetivo del trabajo es la obtención de la estructura terciaria, podría ser necesario el enriquecimiento isotópico, para el cual es necesario utilizar medio mínimo de cultivo. Debido a que Rosetta gami ha mostrado dificultad para crecer en dichos medios, y se cuenta con conocimiento previo acerca del buen funcionamiento de Rosetta como medio de expresión, se eligió utilizar dicha cepa como sistema. A partir de este punto, todos los resultados mostrados son para la cepa Rosetta.

Originalmente se buscaba que la proteína de interés se encontrara mayoritariamente en el sobrenadante, debido a que de esta manera es más factible que el plegamiento se llevara *in vivo* y únicamente sería necesario, a partir de este punto, obtener la defensina pura y en las cantidades más abundantes posibles para su posterior caracterización. Se observó, sin embargo, que la fracción mayoritaria de la proteína

se iba a cuerpos de inclusión, a pesar de la proteína chaperona. Esto se ve claramente en la ilustración 3, pues mientras que en la fracción soluble no se aprecia significativamente la proteína sobreexpresada, en el pellet (donde están los cuerpos de inclusión) se observa una proteína del peso esperado con una concentración relativa muy alta. Por lo tanto, la fracción mayoritaria de la defensina se encuentra en los cuerpos de inclusión.

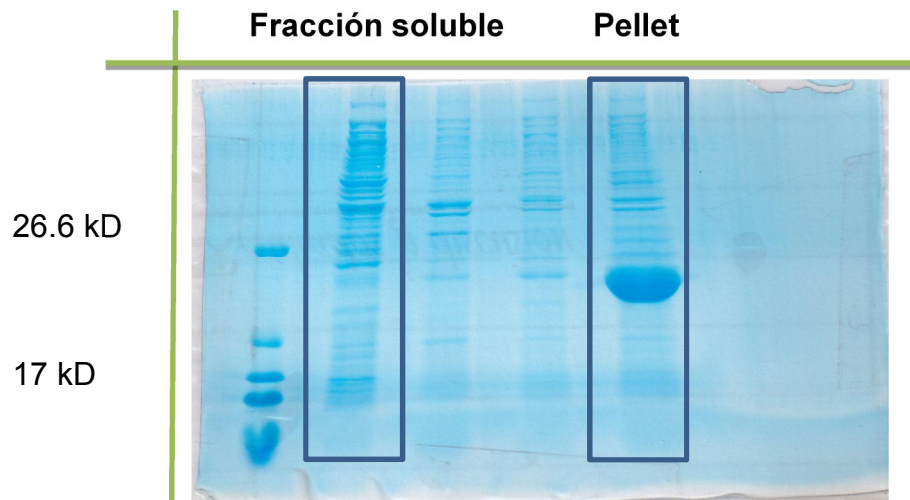


Ilustración 3: Gel de electroforesis comparativo de la fracción soluble y el pellet

Para contrarrestar este efecto, se probó la producción a bajas temperaturas, siendo la más baja alcanzable 20° C, y se llevó a cabo un seguimiento cinético.



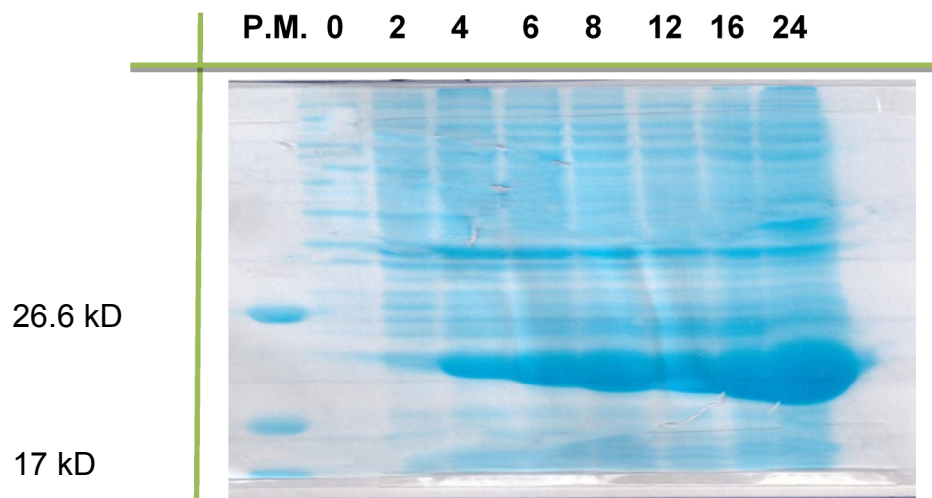


Ilustración4: Cinética de expresión a 20° C. Mezcla de pellet con sobrenadante.

A las 24 horas de expresión, no se llegó a la saturación. Se repitió la cinética, con mediciones espaciadas durante tres días. Se concluyó que para la obtención a gran escala de proteína de fusión se debían utilizar 6 L de cultivo a 20° C durante 40 horas con 0.5 mM de IPTG. A pesar de que hay un ligero incremento en la fracción soluble al trabajar con baja temperatura, el rendimiento es muy bajo, con 300 µg de proteína por los 6 litros de cultivo.

#### **COMENTARIOS SOBRE LA PURIFICACIÓN POR CROMATOGRAFÍA DE AFINIDAD METAL-QUELATO**

Una vez obtenida y filtrada la fracción soluble de la proteína de fusión, se hace pasar por una columna de 5 mL equilibrada con el buffer de lisis. Al retenerse la proteína en la columna, se observa que el color cambia de un verde pálido a café oscuro. Se espera que únicamente las proteínas con alta afinidad química por el níquel (II) sean retenidas en la columna (específicamente, con His-tag), y que las demás proteínas del lisado celular eluyan junto con la matriz. Se lava la columna con 5 volúmenes de buffer de lisis, y se procede a eluir la muestra con 1 a 5 volúmenes de buffer de lisis y 500 mM de imidazol.

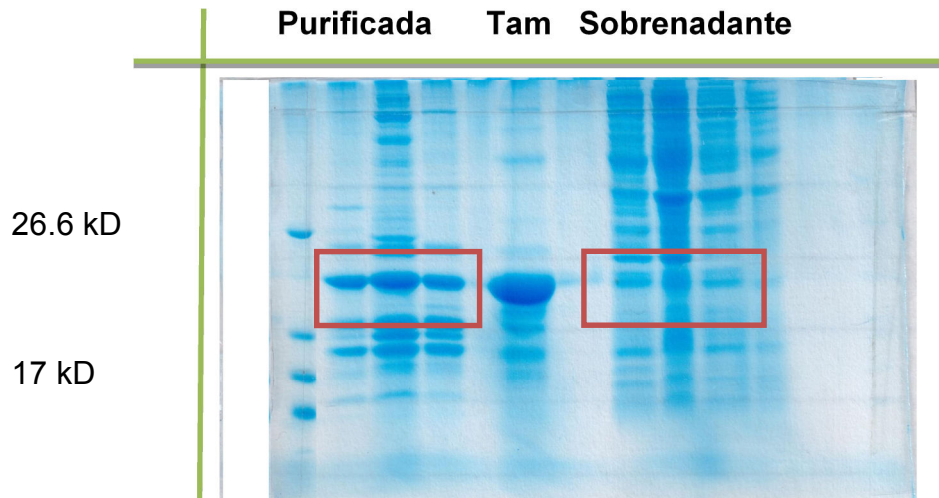


Ilustración 5: Gel de electroforesis que muestra la proteína antes de ser pasada por la Hi-Trap (derecha) y después de ser pasada por la Hi-Trap (izquierda).

La purificación de la proteína soluble por cromatografía de afinidad no es tan eficiente como podría desearse. En la ilustración 5 se observa lo siguiente: La proteína del lisado celular (sobrenadante) se encuentra muy diluida y a simple vista no puede apreciarse de manera significativa. Una vez concentrada en la columna y eluída con 500 mM de imidazol, la fracción de la proteína se vuelve observable; sin embargo, existen diversas proteínas contaminantes en concentraciones apreciables.

Para mejorar la pureza de la proteína, se probó realizar diferentes lavados con un buffer de lisis e imidazol a diferentes concentraciones en la fracción obtenida de los cuerpos de inclusión, de manera que pudiera apreciarse en qué concentraciones de imidazol puede lavarse la proteína sin que ésta se desprenda de la columna de afinidad. Se muestra en la ilustración 6 el resultado de la misma. Se muestra primero la proteína sin pasar por la Hi-Trap. Luego se hizo un lavado con 10 mM de imidazol, y como se observa en el gel, la proteína de interés no eluye en esta concentración. Los contaminantes de la fracción insoluble tampoco eluyen, pero esto no es necesariamente igual para la fracción soluble. Se probó nuevamente utilizando un buffer de lisis con 40 mM de imidazol, pero los contaminantes eluyen junto con la

proteína deseada. Finalmente, se compararon las fracciones de proteína de fusión eluída con imidazol 500 mM con un lavado simple de buffer de lisis, un lavado de buffer de lisis con 10 mM de imidazol y un buffer de lisis con 40 mM de imidazol, y se observa que la pureza relativa de la proteína es semejante.

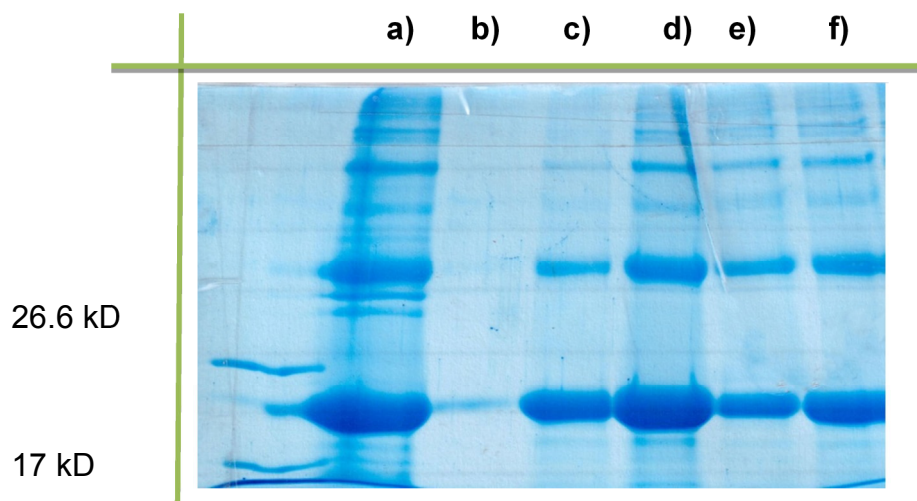


Ilustración 6: Diferentes etapas del lavado. a) Cuerpos de inclusión recién resuspendidos b) elución de lavados con 10 mM de imidazol c) elución de lavados con 40 mM de imidazol d) elución de 500 mM con lavados previos simples e) elución de 500 mM con lavados previos de 40 mM f) elución de 500 mM con lavados previos de 10 mM

Por lo tanto, para purificar la proteína obtenida de manera soluble, debe lavarse con buffer de lisis de manera sencilla, y puede hacerse un lavado adicional con buffer de lisis e imidazol 10 mM. Para purificar la proteína obtenida de los cuerpos de inclusión, basta con hacer el lavado simple.

Una vez que la trombina corta la proteína de fusión se obtienen dos fragmentos principalmente: la GS-DEF1 y el residuo de corte, el cual contiene tanto a la trombina como a la His-Tag, por lo que una manera sencilla de purificar la defensina sería utilizar nuevamente la Hi-Trap para retener el residuo de corte (y de paso las proteínas afines a la columna que quedaron desde la primera purificación por Hi-Trap) puesto que ésta simplemente eluiría. Sin embargo, se observó que inesperadamente

la GS-DEF1 tiene afinidad por la columna de níquel (II). Esto se muestra de manera simple en los siguientes cromatogramas:

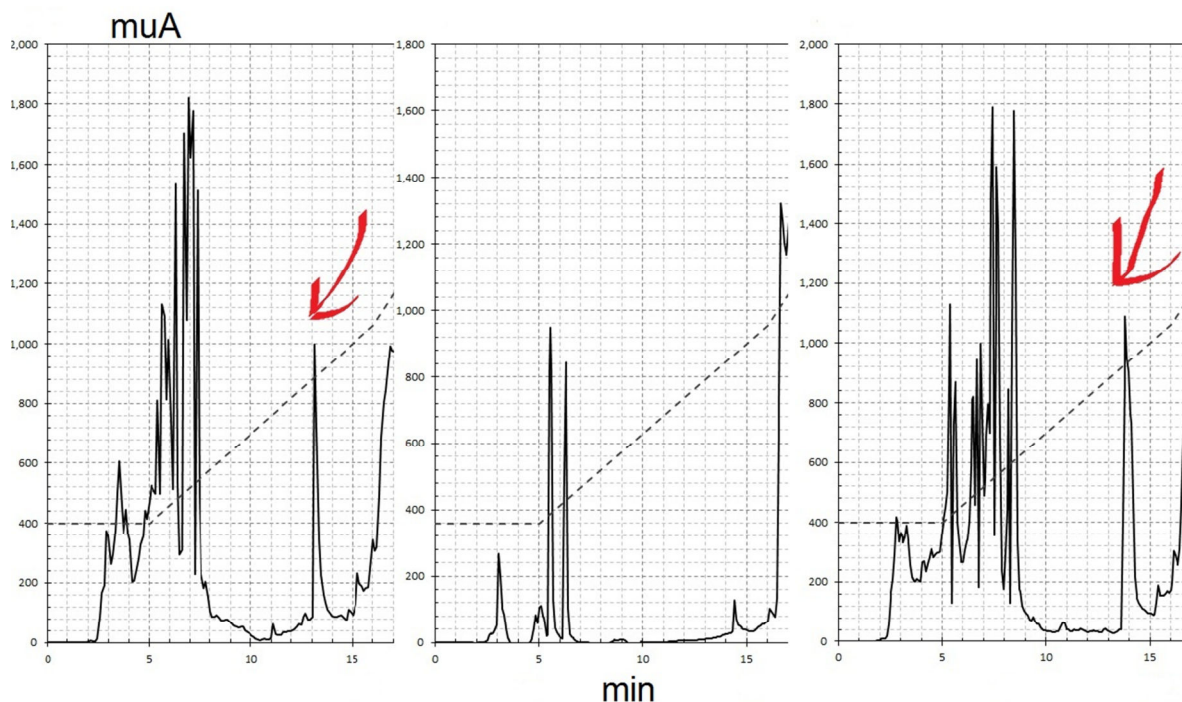


Ilustración 7: Comparación de los cromatogramas obtenidos antes y después de la segunda purificación por Hi-Trap. Izquierda) Mezcla de producto de corte con trombina. Centro) Fracción que pasó por la Hi-Trap sin retenerse. Derecha) Fracción que eluyó de la Hi-Trap con 500 mM de imidazol

Las impurezas residuales de la proteína se remueven de la siguiente manera: los péptidos menores a 10 kDa se pierden durante la desalación por diálisis antes del corte con trombina, y los mayores a 10 kDa (residuo de corte, entre ellos) durante la ultrafiltración posterior al corte. De esta manera, la proteína queda con una pureza relativa alta antes de ser sometida a cromatografía líquida de fase reversa.

### CARACTERIZACIÓN DE LA GS-DEF1

Durante los experimentos de purificación de la defensina 1, se utilizaron geles de acrilamida para monitorearla. Sin embargo, el valor del peso de la proteína de fusión es lejano al de los dos compuestos de referencia, y debido a que la escala de peso no

es lineal, no puede asegurarse que el peso sea cercano al correcto. La proteína GS-Tamapina, que es producida en el laboratorio y está bien caracterizada, se usó a modo de referencia para corroborar que la masa en la que aparece la GS-DEF1 está en el intervalo esperado.

Se obtuvo un perfil cromatográfico reproducible que fue auxiliar en el monitoreo de los diferentes procesos de purificación, así como en el proceso de plegamiento. El pico aislado por HPLC de fase reversa fue analizado por espectrometría de masas MALDI-TOF (la masa esperada es de 3,959.4). En el apéndice se encuentran los espectros completos.

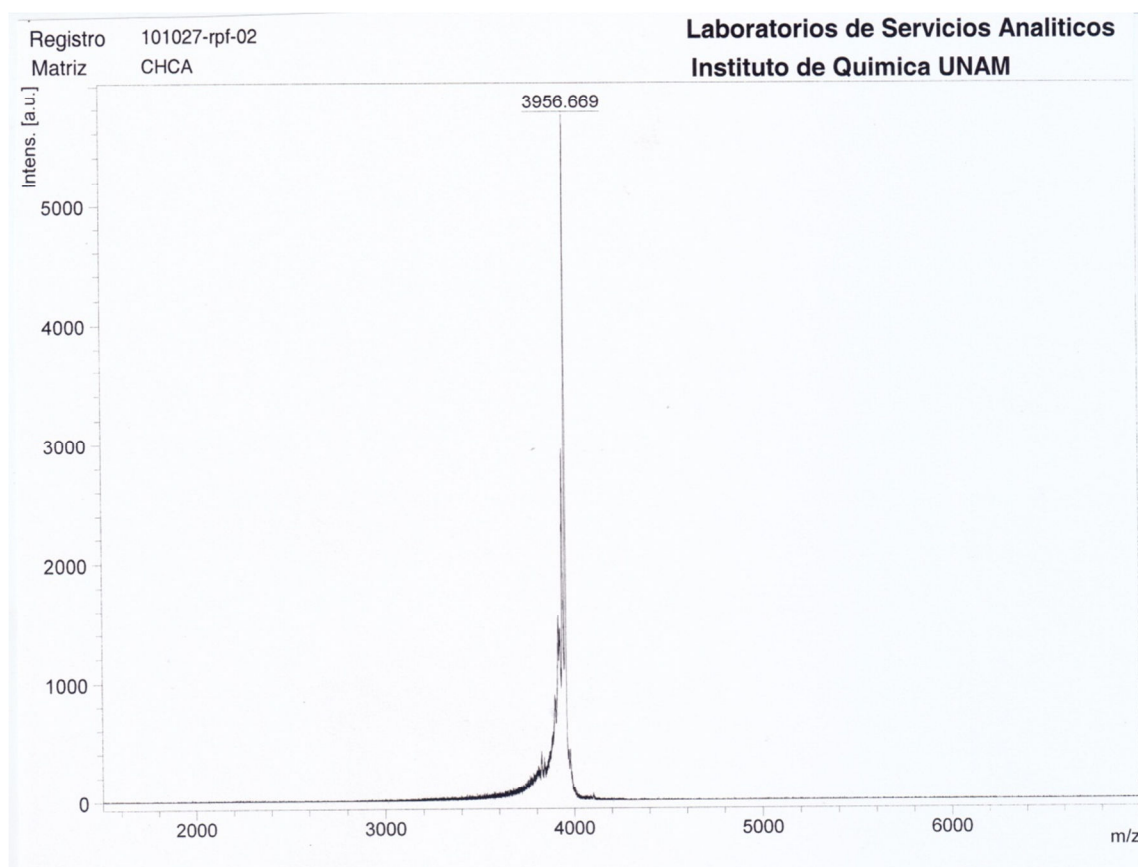


Ilustración 8: Espectrometría de masas MALDI-TOF de la GS-DEF1 soluble

El proceso de obtención y purificación de la defensina requirió de la producción de más de 24 L de medio de cultivo, debido al muy bajo porcentaje de recuperación. Se obtuvo en total 500 µg de muestra de 12 L de cultivo por el método de replegamiento. La defensina 1 es muy insoluble y precipita con mucha facilidad y todo ensayo en el que se mantenga en disolución acuosa más de un día (como en la obtención de los espectros de RMN) se precipita. Una vez precipitada, no es útil para su posterior caracterización.

Con la proteína pura se obtuvo el espectro por resonancia magnética nuclear por los experimentos TOCSY y NOESY para hacer la caracterización de la estructura, los cuales se discuten más adelante.

#### **COMPARACIÓN: SOLUBLE Y REPLEGADA**

La masa esperada para la GS-DEF1 es de 3,959.4 tomando en cuenta que todas las cisteínas se encuentren oxidadas, es decir, formando los tres puentes disulfuro que se sabe la defensina debe tener. Durante el proceso de replegamiento, se busca oxidar las cisteínas lentamente para obtener el plegamiento nativo. Por lo mismo, se espera que tanto la defensina obtenida de manera soluble (oxidada y con el plegamiento deseado) como la replegada tengan el mismo peso molecular, que corresponde a la existencia de los tres puentes disulfuro.

Al trabajar en la purificación de la GS-DEF1 soluble, se obtuvo un perfil cromatográfico en condiciones conocidas que es reproducible si dichas condiciones se mantienen constantes. El valor del tiempo de retención de un analito en una columna es dependiente de la interacción química del adsorbente con éste. En el caso de la cromatografía de fase reversa, el analito tendrá una interacción más fuerte cuando su superficie de contacto sea de menor polaridad, por lo que los compuestos polares eluirán rápidamente mientras que los no polares se retendrán. A su vez, la

naturaleza de la polaridad de una proteína está dada por su plegamiento; las proteínas plegadas orientan los residuos al exterior o al interior de ésta dependiendo de la naturaleza polar de los mismos, mientras que las proteínas desplegadas tienen todos los residuos en contacto con la matriz. Por lo tanto, se espera que una proteína plegada y la misma desplegada, o incluso plegada con un plegamiento distinto, tengan un comportamiento cromatográfico diferente. Una vez que se logró obtener la proteína GS-DEF1 replegada, se purificó por cromatografía líquida de fase reversa en las mismas condiciones de trabajo (columna, flujo, gradiente de disolventes, etc.) y se obtuvo un perfil cromatográfico similar al obtenido con la GS-DEF1 soluble. Ambos cromatogramas muestran un pico en el mismo tiempo de retención (~16.5 min) que presenta una masa molecular muy similar. Esto es evidencia fuerte, aunque no concluyente, de que el replegamiento de la proteína se hizo de manera correcta.

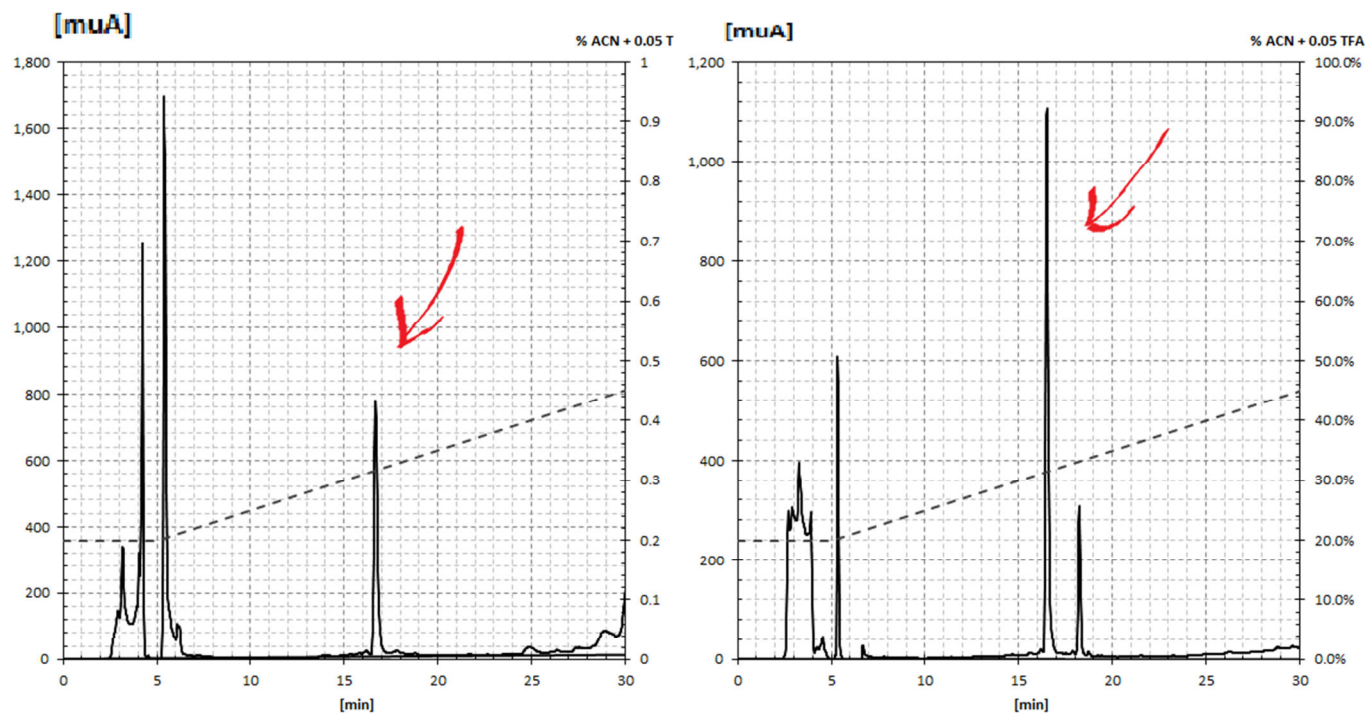


Ilustración 9: Comparación de los perfiles cromatográficos de la GS-DEF1 (izquierda) soluble y (derecha) replegada.

Se resume lo discutido anteriormente en la tabla 5.

Tabla 5: Comparación de las masas de la GS-DEF1 soluble y replegada. La ligera diferencia entre ellos se debe a que se calibraron de manera distinta.

DEFENSINA	MASA POR MALDI-TOF Y REFERENCIA	TIEMPO DE RETENCIÓN
Soluble	3,956.669 ref. GS-Tam	16.6 min
Replegada	3,959.204 ref. insulina	16.5 min

Finalmente, se obtuvo el espectro de resonancia magnética nuclear de la GS-DEF1, tanto en la forma soluble como en la forma replegada. Al comparar ambos espectros y encontrar que son superponibles, se puede aseverar que se trata del mismo compuesto, con la misma orientación espacial. Por lo tanto, el replegamiento de la GS-DEF1 fue el esperado. Los espectros TOCSY completos se encuentran en el apéndice.

La ilustración 10 muestra el espectro de TOCSY con una buena diferenciación de los sistemas de espín de los aminoácidos, es decir, las correlaciones de los hidrógenos del amino del enlace peptídico se encuentran esparcidos en casi tres partes por millón, lo que es indicativo de que la proteína se encuentra plegada. Por tanto, incluso si no se contara con el espectro de la proteína soluble, el espectro TOCSY de la replegada sería suficiente para afirmar que existe algún plegamiento. El hecho de que estos sistemas se encuentren en el mismo desplazamiento químico, en ambos casos muestra que el plegamiento no sólo existe, sino además es el deseado. La diferencia principal que se observa en los espectros al centro de los mismos se debe a la supresión de agua, que no es igual en ambos casos. En el espectro de la proteína soluble se utilizó la secuencia de pulsos Watergate para suprimir el agua, mientras que en el de la derecha se utilizó la secuencia de *excitation sculpting*, con el cual se obtuvo mejores resultados.



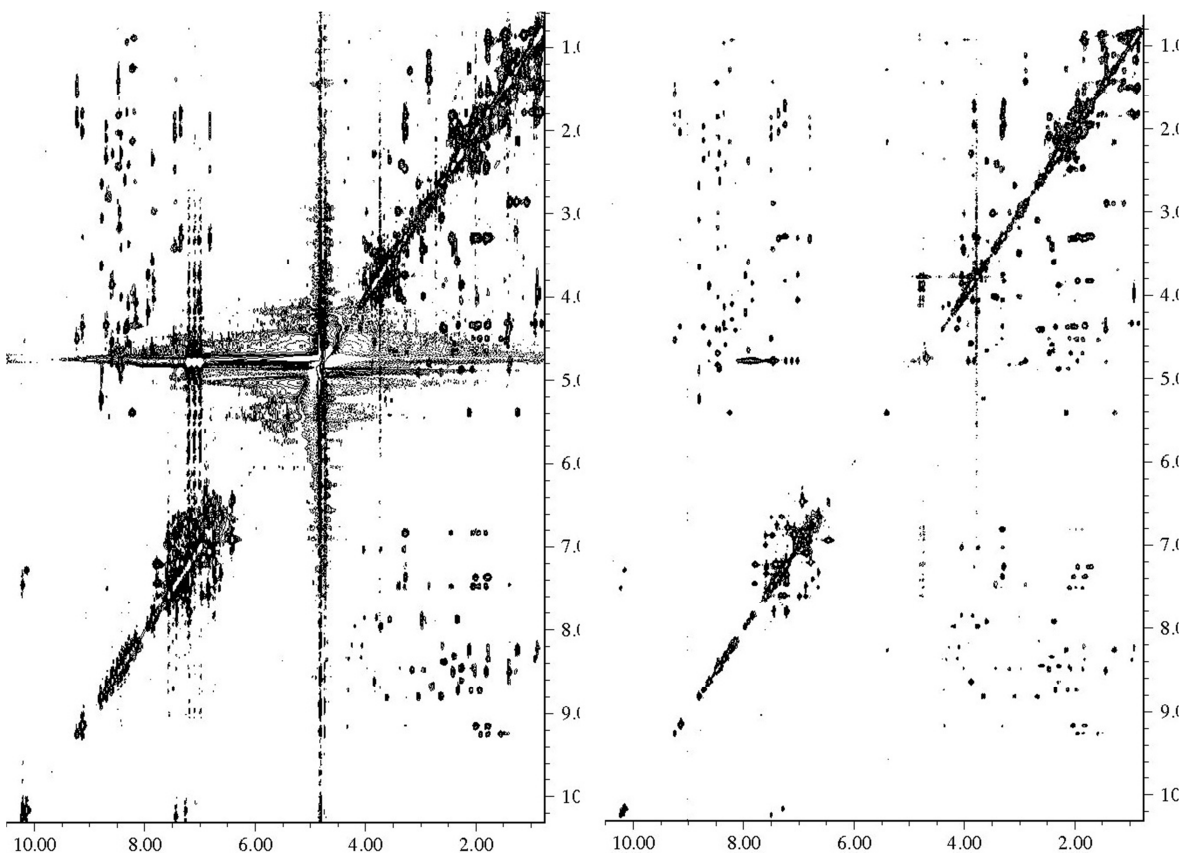


Ilustración 10: Comparación el espectro TOCSY de la GS-DEF1 soluble (izquierda) y la GS-DEF1 replegada (derecha)

Una vez que se comprobó que el replegamiento es un método adecuado para obtener a la defensina en la forma deseada, se produjo una mayor cantidad de péptido para adquirir los espectros de RMN necesarios para realizar la asignación de los aminoácidos y sus sistemas de espín, debido a que se obtienen cantidades más generosas por este método.

#### COMENTARIOS ACERCA DE LA ASIGNACIÓN

Mientras el experimento TOCSY muestra la interacción de los protones de un mismo sistema de espín, y únicamente por acoplamiento escalar, el experimento NOESY permite la observación de las consecuencias del acoplamiento dipolar. Por esto, mientras en el TOCSY se aprecian principalmente los sistemas de espín intraresiduo,

en el NOESY pueden verse tanto éstos como los interresiduo. Se utilizó el método de Wüthrich para realizar la asignación <sup>7</sup>.

En la región de 9.5-5.5 ppm se observan las interacciones que existen entre los hidrógenos de los nitrógenos, tanto los del esqueleto de la proteína como los propios del residuo, y los sistemas de espín correspondientes, sean intra o interresiduo. Aproximadamente de 3.5-5.3 ppm se observan los hidrógenos alfa de cada residuo, región que coincide con la señal del agua en 4.75 ppm (que fue suprimida tanto en la adquisición como en el procesamiento), por lo que las intensidades de estas señales pueden verse afectadas. Los intervalos normales de observaciones de hidrógenos alifáticos y aromáticos son iguales que para las moléculas pequeñas. La interacción entre los hidrógenos alfa de diferentes residuos es difícil de observar debido a que pueden estar en la zona del agua. La interacción de los hidrógenos del esqueleto con los de los residuos muestra conexiones secuenciales, es de especial importancia si aparece en el sistema de espín del hidrógeno del nitrógeno peptídico la señal del carbono alfa del aminoácido anterior.

Para obtener la estructura terciaria de la proteína GS-DEF1 se utilizó el programa CYANA<sup>31</sup>, que calcula los ángulos de torsión entre los átomos a partir de las señales NOE del espectro NOESY, debido a que éstas guardan información relativa a la distancia entre los átomos involucrados.

Aunque las señales se encuentran extendidas a lo largo de la ventana espectral, es posible que dos señales se encuentren en el mismo desplazamiento químico, lo cual vuelve un reto la interpretación de ciertos fragmentos. Un ejemplo de esto es la señal que corresponde simultáneamente a la interacción entre el hidrógeno alfa (HA) de la Y30 con el hidrógeno de la amina del enlace peptídico (NH) y también a la interacción de los hidrógenos beta (QB) de la S8 con su NH respectivo. Esto se comprueba de la siguiente manera: la señal citada (7.7, 3.7 ppm) interacciona tanto con la señal (7.7, 3.9) como con (7.7, 2.1), pero éstas no tienen interacción entre sí, de tal suerte que las tres señales no pueden pertenecer al mismo sistema de espín. Para discernir

entre los sistemas, primero se analiza qué señales sí tienen correlación entre sí, y se agrupan como del mismo sistema. Se buscan señales complementarias que indiquen la concatenación de los aminoácidos, y así se etiquetan a las señales con el aminoácido correcto. Puesto a que todas las señales debidas a estructura secundaria y a secuenciación se encuentran en el mismo desplazamiento químico, es necesario discernir también entre qué señales interaccionan con el NH de un aminoácido y cuales con el otro. En este caso, se identificó que las señales interresiduo que se encuentran en el desplazamiento químico 7.7 ppm pertenecen a los QB del W7 (lo que sugiere uno de los sistemas pertenece a la S8) y a su vez, que las señales de (7.77, 3.7 ppm) y (7.7, 3.9 ppm) se repiten en el desplazamiento químico del NH de la C31, lo que sugiere que el otro sistema pertenece a la Y30. Una vez hechas estas suposiciones, se comprueban con las interacciones NH-NH de los aminoácidos secuenciales, y se busca la relación con los aminoácidos aromáticos de la Y30. Al constatar estos puntos, puede afirmarse que los sistemas de espín involucrados son efectivamente Y30 y S8. Para conseguir una estructura terciaria, es necesario analizar qué interacciones a largo plazo puede tener cada uno de los aminoácidos y esto se facilita con ayuda de un modelo o simulación, por lo que se realizan cálculos preliminares que sirven de guía para analizar qué interacciones son factibles.

Otra dificultad que se encuentra al momento de realizar las asignaciones es que no hay tantos sistemas de espín como los que se esperan. Existen sistemas que no se observan en el espectro NOESY pero sí en el TOCSY, como la I26, y que por tanto no pueden asignarse restricciones de distancia y ángulo. Hay sistemas de espín que no se observan en ninguno de los dos espectros, y presentan los mismos problemas. Esto dificulta el cálculo de una estructura, puesto que no se cuenta con datos para trabajar. Como resultado de esto, se entiende que existe una sección flexible de la proteína, es decir, no estructurada. El fragmento GS en el extremo N-terminal se espera no afecte la estructura terciaria y no tenga estructura. No se encontraron señales que correspondan a estos residuos, lo cual sugiere que no tienen una estructura rígida. No pudieron asignarse los sistemas correspondientes a: Q5, C23,

G25. Parte de los sistemas F6, W24 tampoco pudieron asignarse. Se trabajó con un 86% de asignación completa inicial (anterior al cálculo).

La información de la estructura secundaria y terciaria está implícita en las interacciones que existen entre dos átomos no acoplados por enlace químico que se encuentran espacialmente cercanos. Estas distancias, aunque cortas en relación a las distancias que existen entre los átomos de un punto de la proteína a otro, son largas en relación de las magnitudes observables. Esto es, porque la intensidad de la señal es inversamente proporcional al radio elevado a la sexta potencia, por lo que la intensidad decrece dramáticamente al alejar dos núcleos. Por lo tanto, si un núcleo tiene interacción con otro núcleo lejano en la cadena, quiere decir que es cercano en el espacio. La estructura secundaria en particular, se elucida a partir de la interacción de los hidrógenos NH y HA de dos aminoácidos diferentes. Ejemplo de esto, es el fragmento del espectro que se muestra a continuación en la ilustración 11. La estructura secundaria está determinada por dichas señales, y se han descrito de la siguiente manera: En las hojas beta, los residuos se encuentran muy cerca, por lo que el HA y NH de los dos residuos diferentes interaccionan, y puede verse una señal en el espectro NOESY. Otra señal indicativa de hoja beta es cuando existe una interacción entre los HA de éstos residuos. Por otro lado, un indicativo de hélice alfa es que exista una interacción entre los NH de dos aminoácidos secuenciales intensa. Nuevamente, esta intensidad es debida a la corta distancia entre los mismos.

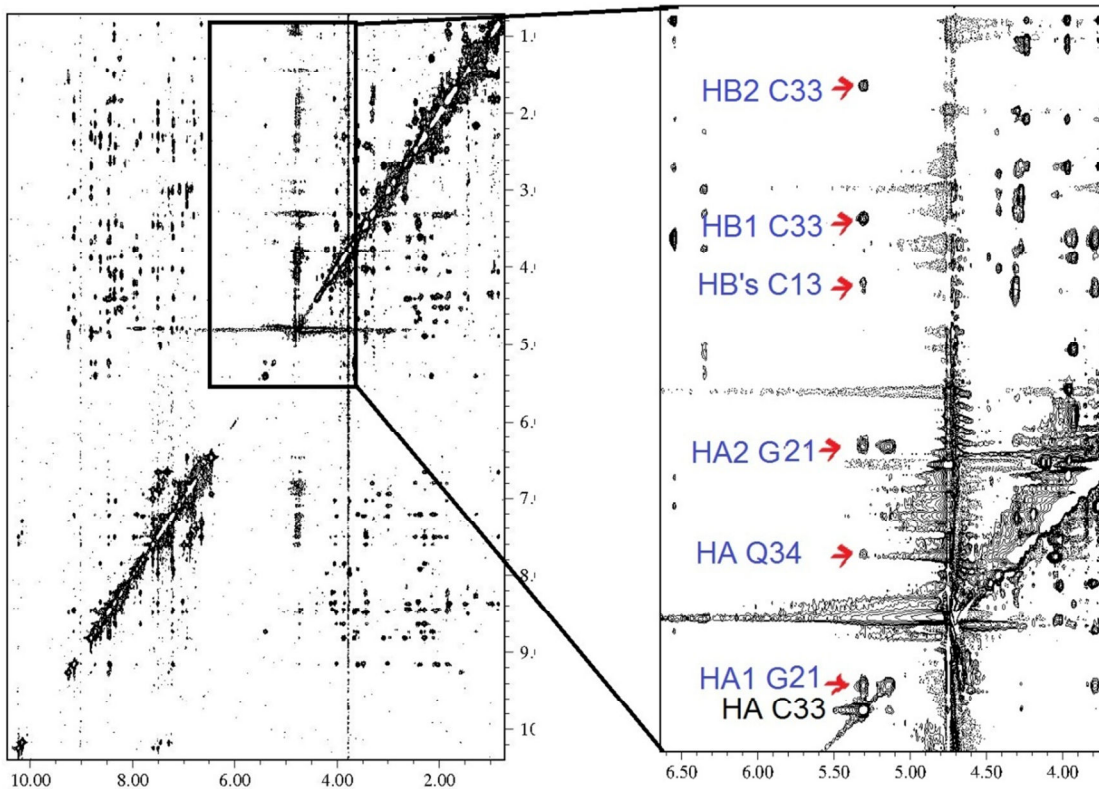


Ilustración 11: Acercamiento a las interacciones HA-HA de algunos aminoácidos involucrados en la estructura secundaria

Se observa que el hidrógeno alfa de la C33 interacciona con los hidrógenos beta de la C13, lo cual confirma la existencia del puente disulfuro esperado. Se observa además la interacción entre los hidrógenos alfa de la C33 con los de la G21 lo que indica estructura secundaria entre estos, específicamente, una hoja beta. Estas correlaciones se corroboran buscando interacciones de los aminoácidos adyacentes.

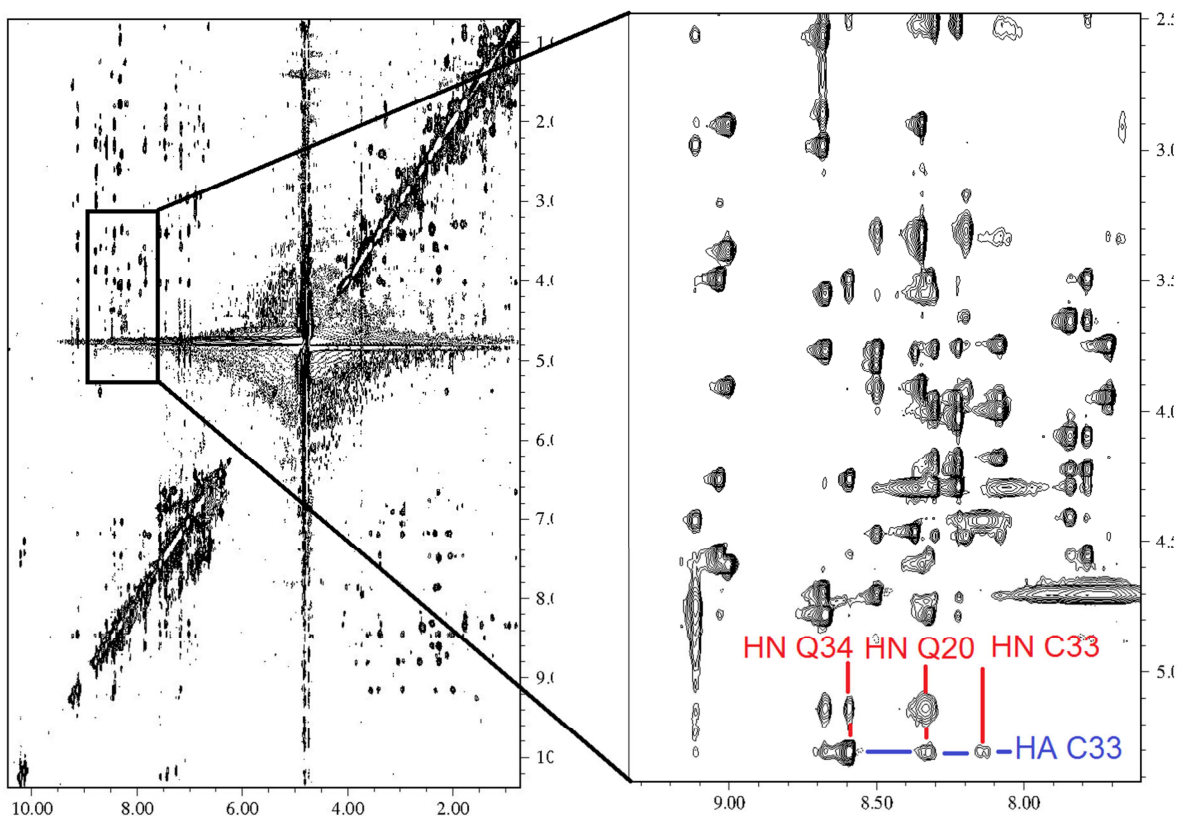


Ilustración 12: Acercamiento de la interacción de los hidrógenos alfa de la C33 con algunos HN involucrados en la estructura secundaria.

Se observan las interacciones del hidrógeno alfa de la C33 con los núcleos vecinales en el espacio, y esto confirma la existencia inequívoca de la hoja beta, además de especificar qué núcleos se encuentran involucrados. Éste mismo análisis se realiza para todas las interacciones de esta naturaleza.

Una vez asignados todos los sistemas de espín encontrados, así como la mayor posibilidad de interacciones entre sistemas, se calcularon con el programa CYANA 200 estructuras, de las cuales se seleccionaron automáticamente las 20 mejores. Se tomaron 316 restricciones de distancia máxima, de las cuales 193 ( $|i-j| \leq 1$ ) son de corto alcance, 64 ( $1 < |i-j| < 5$ ) de mediano y 59 de largo ( $|i-j| \geq 5$ ).

Se muestran a continuación los gráficos referentes a los cálculos realizados.

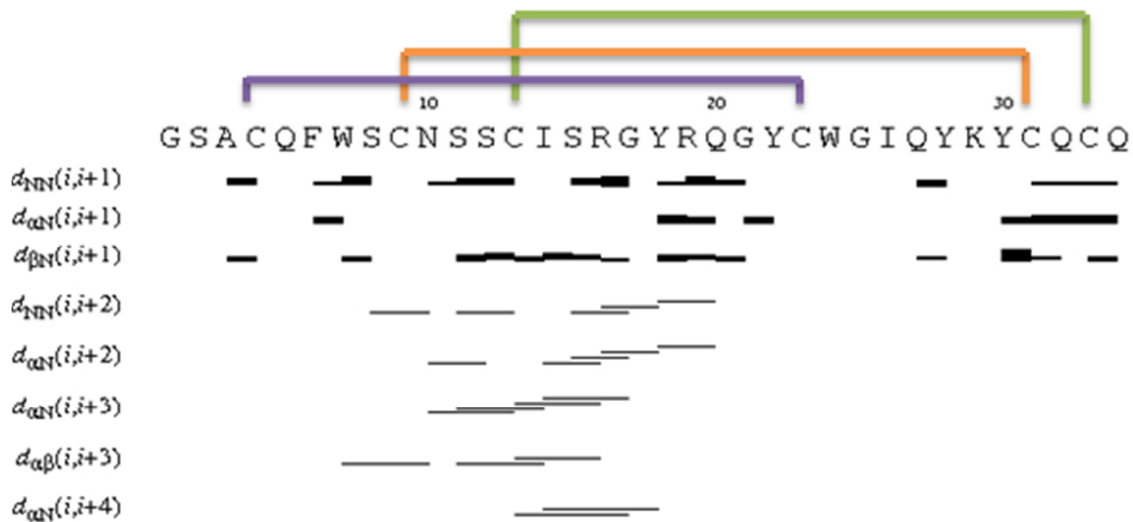


Ilustración 13: Gráfico de NOEs secuenciales en función de la secuencia de aminoácidos.

La ilustración 13 muestra la intensidad de los NOEs secuenciales en el grosor de las líneas, los cuales sirven para predecir la estructura secundaria, junto con la información de la ilustración 14.

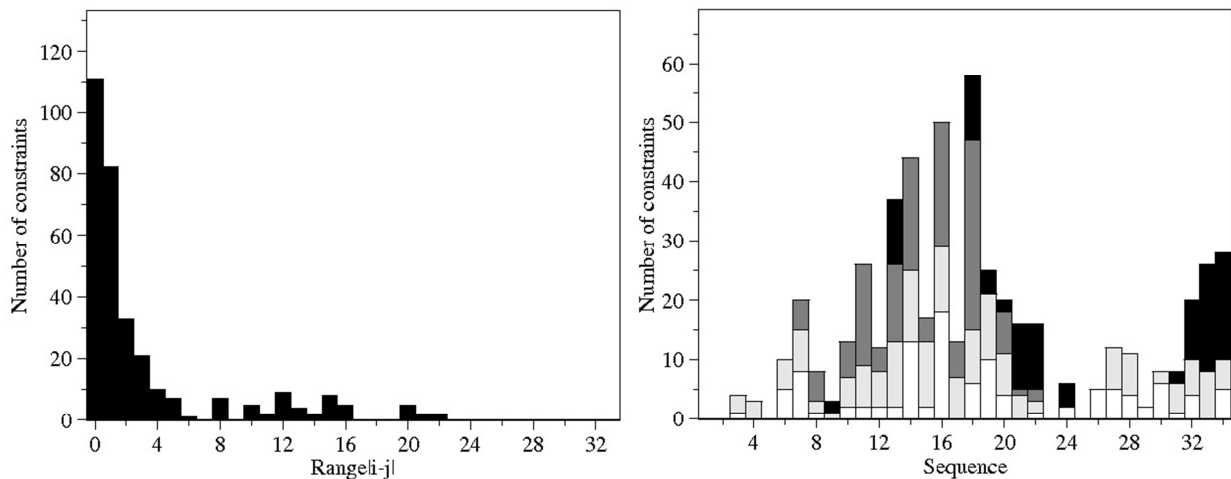


Ilustración 14: Gráfico de distribuciones NOE de corto y largo alcance, y número de restricciones para cada residuo

La ilustración 14 muestra las distribuciones NOE para la secuencia de aminoácidos, donde en blanco se definen los intrarresiduo, en gris claro los secuenciales. En gris oscuro los de mediano alcance y en negro los de largo alcance.

La ilustración 15 muestra el gráfico de Ramachandran para el ensamble de las veinte estructuras calculadas, donde se observan las regiones permitidas en una estructura, y que relaciona los ángulos de la cadena principal con la estructura secundaria.

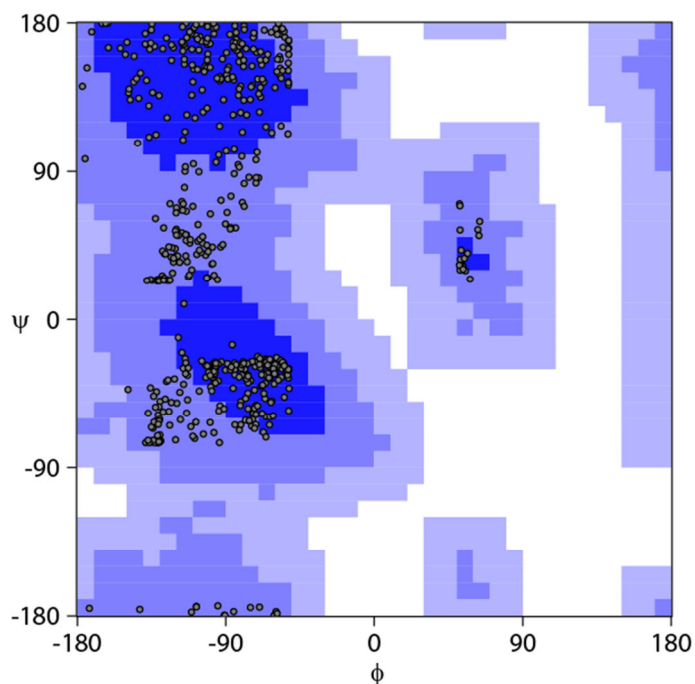


Ilustración 15: Gráfico de Ramachandran. La zona azul rey muestra las regiones más favorecidas (59.8%), la zona morada muestra las regiones permitidas (40.2%). Las lilas y blancas son las poco permitidas y las no permitidas respectivamente.

Estos cálculos llevan a una estructura terciaria. Se muestran a continuación las estructuras calculadas superpuestas y la caricatura de una de las estructuras como ejemplo.



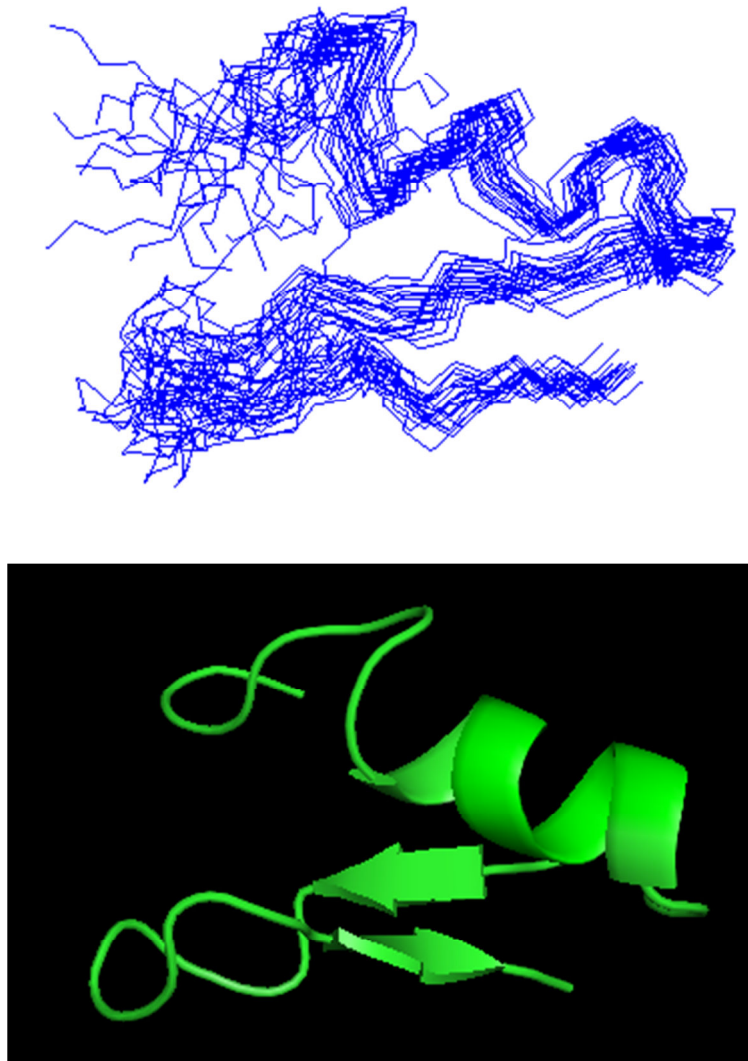


Ilustración 16: En la parte superior se muestra la superposición de estructuras calculadas. En la parte inferior, la caricatura de una de las estructuras.

Se aprecian tres secciones rígidas en la defensina: una alfa hélice, dos hebras beta torcidas y dos zonas flexibles: el extremo amino terminal y un asa no estructurada.

Se escriben de manera explícita únicamente los residuos hidrofóbicos en la ilustración 17, a continuación.

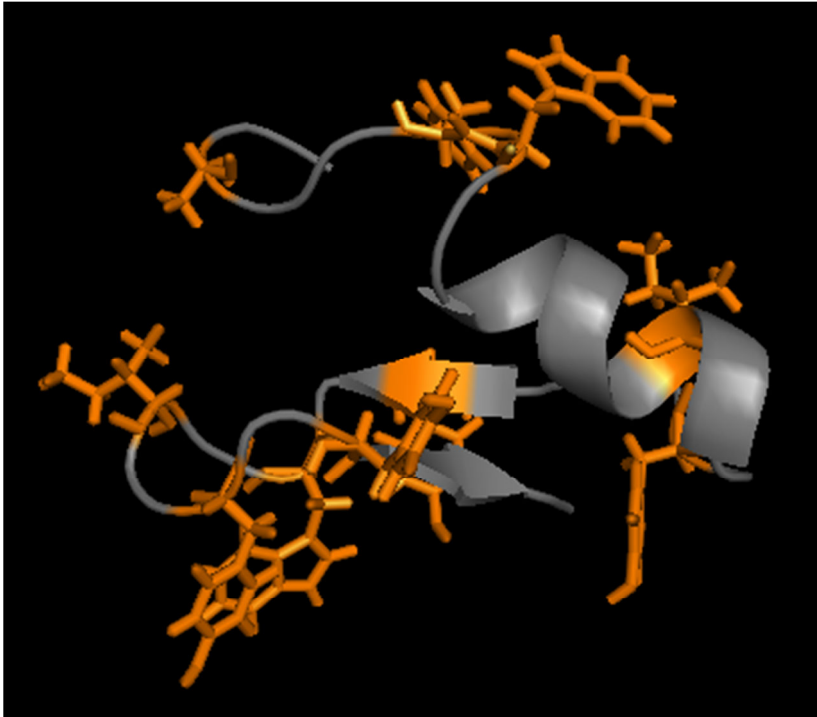


Ilustración 17: Estructura de la GS-DEF1 con los aminoácidos hidrofóbicos explícitos en color naranja.

La estructura terciaria explica una de las propiedades observadas de la defensina: su insolubilidad. Los residuos hidrofóbicos (que se encuentran en número mayoritario con respecto a los cargados) se encuentran expuestos a la superficie, lo que la hace insoluble.

La estructura terciaria de la DEF1 muestra similitud con otras estructuras conocidas, pues comparte el motivo CS- $\alpha/\beta$  que es recurrente en las defensinas de invertebrados y toxinas. La ilustración 18, muestra la estructura de dos proteínas previamente reportadas que tienen este motivo, y que fueron resueltas por resonancia magnética nuclear. A la izquierda se muestra la estructura de una toxina de un alacrán *Centruroides*, cobatoxina 1<sup>32</sup>(PDB= 1PJV) y a la derecha una defensina de almeja, Cg-Def<sup>29</sup> (PDB= 2B68). En esta imagen queda clara la razón estructural por la cual se ha propuesto que existen similitudes importantes entre las toxinas y las defensinas.

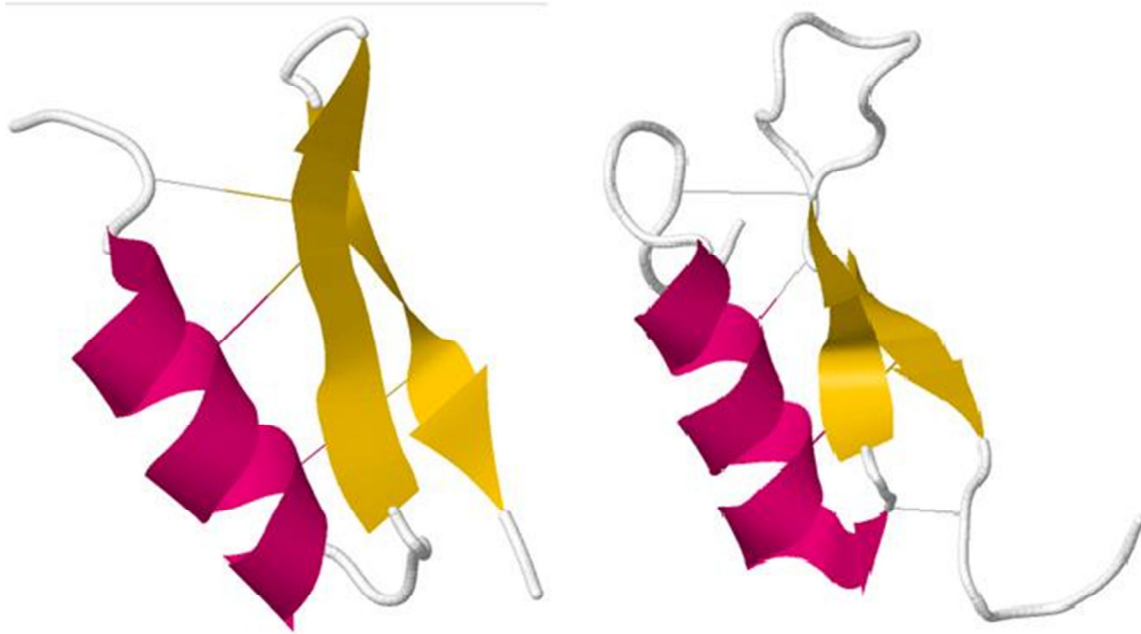


Ilustración 18: Estructuras CS- $\alpha/\beta$ . A la derecha, estructura de una toxina de alacrán. A la izquierda, estructura de una defensina de almeja.

La estructura de la cobatoxina 1 es relevante porque se encuentra en la misma familia de alacranes que la DEF1, y tiene el mismo motivo estructural que la misma. Esto hace que la defensina en cuestión sea un blanco interesante para el estudio de la relación defensinas-toxinas, como han propuesto ya algunos autores<sup>8,11</sup>.

En la búsqueda de estructuras reportadas de defensinas de insecto en el PDB (al 7 de septiembre de 2011) se encuentra que únicamente hay diez, de las cuales seis son mutantes de una de éstas. Relevantes por su estructura para el análisis son: la sapecina (PDB= 1LV4) que es una defensina de la *Sarcophaga peregrina* que es una mosca; la termicina (PDB= 1MM0) que proviene de la termita *Pseudacanthotermes spiniger*; la DEF-AAA (PDB= 2NY8) del mosquito *Anopheles gambiae* del cual provienen las mutaciones mencionadas; y la defensina A (PDB= 1ICA) de otra mosca *Protophormia terraenovae*. La búsqueda de estructuras para defensinas en la misma base en general da un resultado de 98 estructuras, de las cuales se descartan todas

las estructuras de mamíferos y plantas por falta de relevancia, y también los resultados propios de las defensinas de insecto ya mencionadas. De los resultados restantes, resaltan la de dos moluscos: Cg-Def (PDB=2B68), de la almeja *Crassostrea gigas*; la defensina MGD-1 (PDB= 1FJN), del mejillón mediterráneo *Mytilus galloprovincialis*.

En la ilustración 19 se comparan todas las estructuras de las defensinas antes mencionadas como relevantes. Éstas están alineadas para poder mostrar una vista comparable, teniendo en la parte superior la hélice alfa con el amino terminal del lado izquierdo. En general, las defensinas de los moluscos son estructuralmente más parecidas a la defensina 1, debido a la presencia de un asa que conecta a las hojas beta. Únicamente la defensina de la termita presenta esta región tan marcada dentro de las estructuras de insectos. Se aprecia que en general que existe una región no estructurada en el amino terminal, luego una hélice alfa que en todos los casos tiene un número de vueltas similar, luego las hojas beta que cambian en longitud y definición y están conectadas por una región no estructurada de diferentes longitudes (como se mencionó anteriormente) y finalmente las proteínas pueden terminar en la misma hoja beta, o tener un fragmento no estructurado. En la ilustración se omiten los puentes disulfuro.

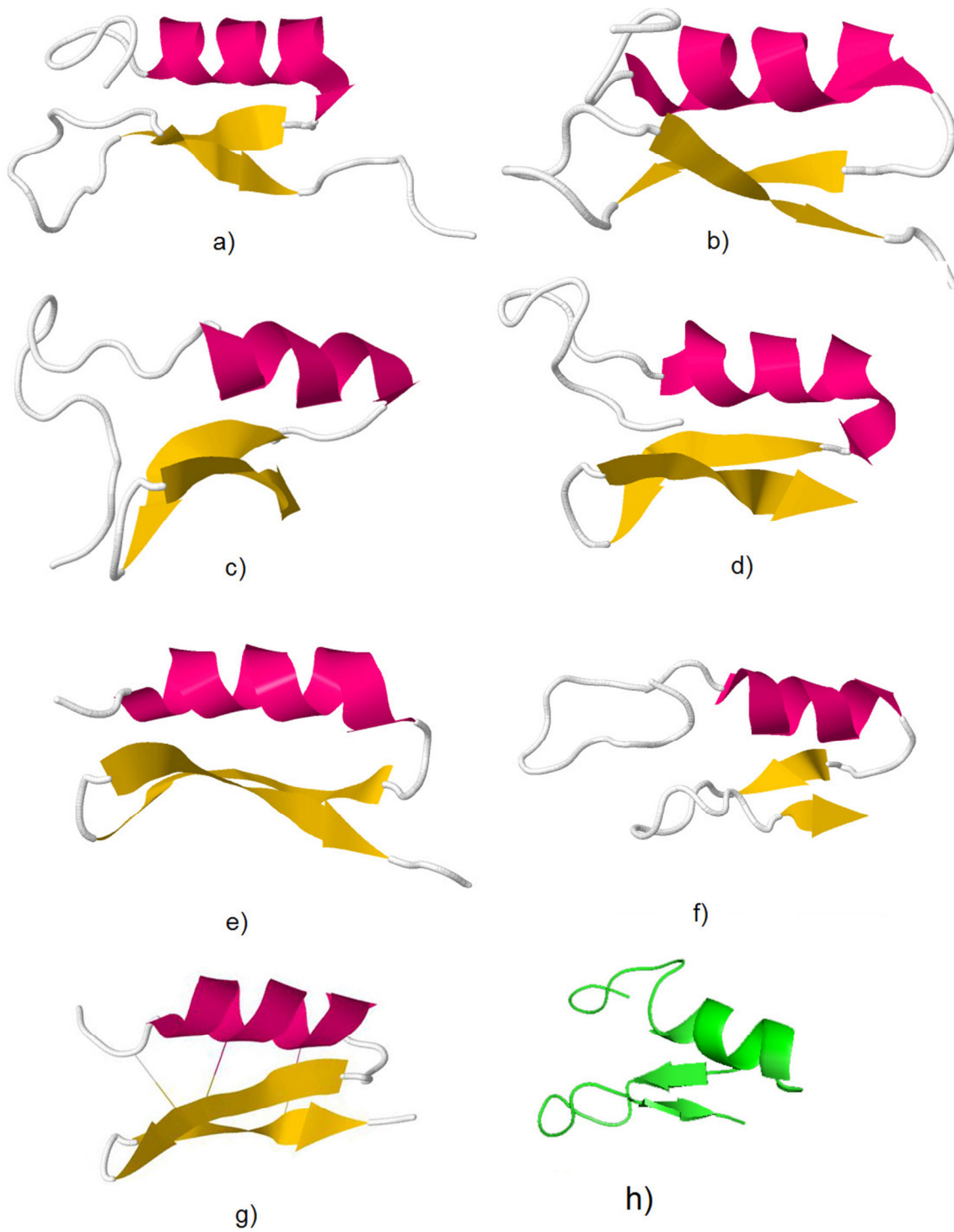


Ilustración 19: Estructuras terciaria de defensinas de moluscos e insectos. Se listan las estructuras por código PDB. a) 2B68 b) 1FJN c) 1ICA d) 1L4V e) 1MM0 f) 2NY8 g) Cobatoxina h) GS-DEF1

Una comparación particular se hace entre la defensina estudiada en este trabajo con la de mayor parecido estructural de las antes descritas.

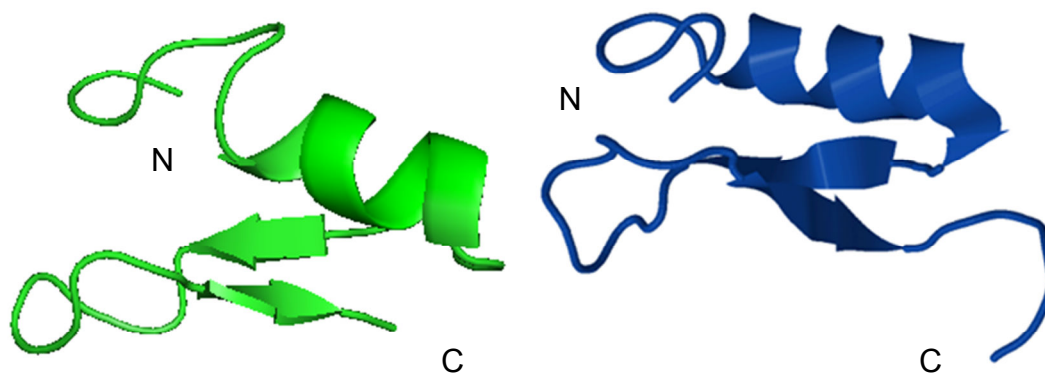


Ilustración 20: Comparación de la GS-DEF1 (verde) y la Cg-Def (azul).

En la ilustración 20, se observan diversas similitudes entre las estructuras terciarias de ambas defensinas. En el N-terminal hay una secuencia de aminoácidos no estructurada, y luego una alfa hélice. Seguida de ésta, está la primera hebra de la hoja beta, luego hay un asa que es una zona flexible y poco estructurada, y termina con la segunda hebra beta. La Cg-Def presenta una tercera zona importante no estructurada. Cabe destacar que aunque a primera vista las hojas beta de las defensinas son diferentes, en realidad tienen una estructura parecida, puesto que ambas muestran una secuencia de aminoácidos rígidos antiparalelos entre sí, que no necesariamente son planos. Un análisis detallado de la estructura de la Cg-Def, revela que las hojas betas están torcidas.

## CONCLUSIONES

El objetivo del presente trabajo fue obtener la GS-DEF1 para su caracterización por resonancia magnética nuclear, y así corroborar el motivo de plegamiento CS-  $\alpha/\beta$  que se esperaba tuviese.

Se logró obtener la proteína deseada GS-DEF1 por expresión en *E. coli*. Se caracterizó dicha proteína por secuenciación del gen expresado, el peso de la molécula por espectrometría de masas (que muestra el estado de oxidación esperado), resonancia magnética nuclear y cálculos de ángulos de torsión con puentes disulfuro.

Se logró obtener la GS-DEF1 pura, tanto por un procedimiento de plegamiento *in vivo* como *in vitro*, y ambas muestras mostraron tener el mismo plegamiento. Por lo tanto, el método de replegamiento fue exitoso.

Se asignaron los sistemas de espín de la defensina en el espectro NOESY de la GS-DEF1 replegada, y se asignaron también las señales cruzadas entre los sistemas de espín. Procesando las señales NOE en CYANA con dichas asignaciones, se obtuvo la estructura terciaria de la proteína. Se corrobora de esta manera que existe el motivo de plegamiento esperado.

## BIBLIOGRAFÍA

Algunas ligas de interés de bases de datos de defensinas son:

*Defensins Knowledgebase*: <http://defensins.bii.a-star.edu.sg/>

*Antimicrobial Peptides Laboratory*: <http://www.bbcm.univ.trieste.it/~antimic/home.html>

1. Zasloff M. Antimicrobial peptides of multicellular organisms. *Nature*. 2002;415(6870):389–395. Available at: <http://www.nature.com/nature/journal/v415/n6870/abs/415389a.html> [Accessed April 19, 2011].

2. Seebah S, Suresh A, Zhuo S, et al. Defensins knowledgebase: a manually curated database and information source focused on the defensins family of antimicrobial peptides. *Nucleic acids research*. 2007;35(Database issue):D265-8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1669742&tool=pmcentrez&rendertype=abstract> [Accessed April 18, 2011].

3. Ganz T. Defensins: antimicrobial peptides of innate immunity. *Nature reviews Immunology*. 2003;3(9):710-20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12949495> [Accessed July 16, 2010].

4. Zhu S, Li W, Jiang D, Zeng X. Evidence for the existence of insect defensin-like peptide in scorpion venom. *IUBMB life*. 2000;50(1):57-61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11087122>.

5. Brondyk WH. *Selecting an appropriate method for expressing a recombinant protein*. 1st ed. Elsevier Inc. 2009:131-47. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19892171> [Accessed June 16, 2011].



6. Marco A de. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microbial cell factories*. 2009;8:26. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2689190&tool=pmcentrez&rendertype=abstract>.
7. Wüthrich K. *NMR of proteins and nucleic acids*. Zürich: John Wiley & sons; 1986:292.
8. Rodríguez de la Vega RC, García BI, D'Ambrosio C, et al. Antimicrobial peptide induction in the haemolymph of the Mexican scorpion *Centruroides limpidus limpidus* in response to septic injury. *Cellular and molecular life sciences : CMLS*. 2004;61(12):1507-19. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15197474> [Accessed February 3, 2011].
9. Zhang H, Kato Y. Common structural properties specifically found in the CSalphabeta-type antimicrobial peptides in nematodes and mollusks: evidence for the same evolutionary origin? *Developmental and comparative immunology*. 2003;27(6-7):499-503. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12697306>.
10. Wilmes M, Cammue BPA, Sahl H-G, Thevissen K. Antibiotic activities of host defense peptides: more to it than lipid bilayer perturbation. *Nat. Prod. Rep*. 2011;28:1350-1358.
11. Rodríguez de la Vega RC, Possani LD. On the evolution of invertebrate defensins. *Trends in genetics: TIG*. 2005;21(6):330-2. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15922831>.
12. Zhu S. Evidence for myxobacterial origin of eukaryotic defensins. *Immunogenetics*. 2007;59(12):949-54. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18058146> [Accessed January 5, 2011].

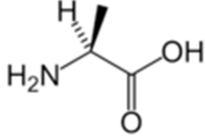
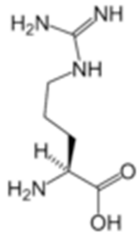
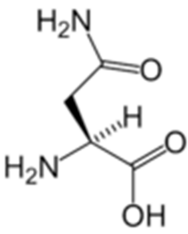
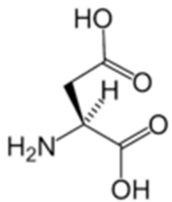
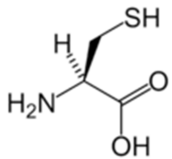
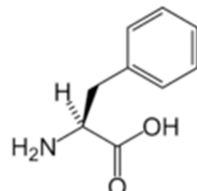
13. Froy O, Gurevitz M. Arthropod defensins illuminate the divergence of scorpion neurotoxins. *Journal of peptide science: an official publication of the European Peptide Society*. 2004;10(12):714-8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15635623> [Accessed April 19, 2011].
14. Figueredo S, Mastroianni JR, Tai KP, Ouellette J. Antimicrobial Peptides. In: Giuliani A, Rinaldi AC, eds. *Methods in Molecular Biology*. Vol 618. Totowa, NJ: Humana Press; 2010. Available at: <http://www.springerlink.com/index/10.1007/978-1-60761-594-1> [Accessed June 17, 2011].
15. Ren G, Bardwell JC a. Engineered Pathways for Correct Disulfide Bond Oxidation. *Antioxidants & redox signaling*. 2011;14(12). Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21250836> [Accessed April 19, 2011].
16. Berndt C, Lillig CH, Holmgren A. Thioredoxins and glutaredoxins as facilitators of protein folding. *Biochimica et biophysica acta*. 2008;1783(4):641-50. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18331844> [Accessed March 30, 2011].
17. Inaba K. Disulfide bond formation system in Escherichia coli. *Journal of biochemistry*. 2009;146(5):591-7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19567379> [Accessed June 21, 2011].
18. Bobovnikova Y, Graves PN, Vlase H, Davies TF. Characterization of soluble, disulfide bond-stabilized, prokaryotically expressed human thyrotropin receptor ectodomain. *Endocrinology*. 1997;138(2):588-93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9002990>.
19. Thomas JG, Ayling A, Baneyx F. Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli. To fold or to refold. *Applied Biochemistry And Biotechnology*. 1997;66(3):197-238. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9276922>.

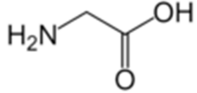
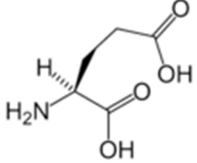
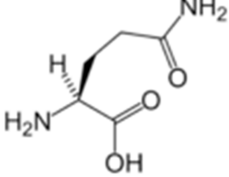
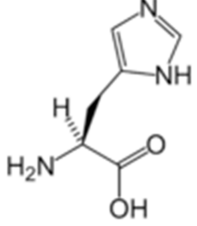
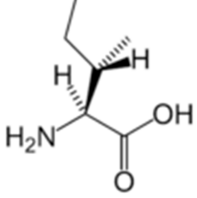
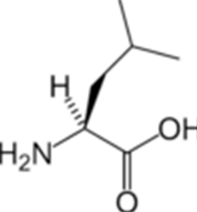
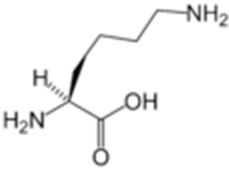
20. Chow MKM, Amin AA, Fulton KF, et al. The REFOLD database: a tool for the optimization of protein expression and refolding. *Nucleic Acids Research*. 2006;34(Database issue):D207-D212. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16381847>.
21. Bulaj G. Formation of disulfide bonds in proteins and peptides. *Biotechnology advances*. 2005;23(1):87-92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15610970> [Accessed June 21, 2011].
22. Friebolin H. *Basic one- and two- dimensional NMR spectroscopy*. Third rev. Wiley-VCH; 1998:373.
23. Ven FJM van de. *Multidimensional NMR in liquids. Basic Principles and Experimental Methods*. New York: Wiley-VCH; 1995.
24. Jardetzky O. Simple Insights from the Beginnings of Magnetic Resonance in Molecular Biology. In: Markley J, Opella S, eds. *Biological NMR spectroscopy*. New York: Oxford University Press; 1997:3-15.
25. Castellani F, Rossum BV, Diehl A, et al. NMR spectroscopy. *Nature*. 2002;420(November):23-26. Available at: <http://www.nature.com/nature/journal/v420/n6911/abs/nature01070.html>.
26. Dobson CM. The Role of NMR Spectroscopy in Understanding How Proteins Fold. In: Markley J, Opella S, eds. *Biological NMR spectroscopy*. New York: Oxford University Press; 1997:82-93.
27. Juneja J, Udgaonkar JB. NMR studies of protein folding. *Curr Sci*. 2003;84(2):157-172. Available at: <http://www.ias.ac.in/currsci/jan252003/157.pdf>.

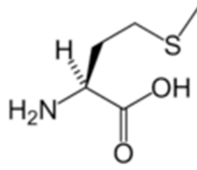
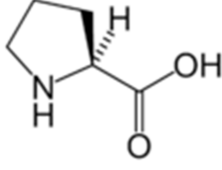
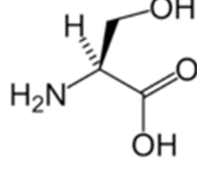
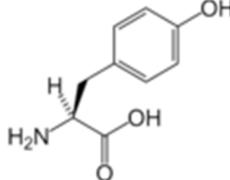
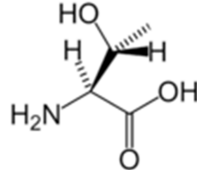
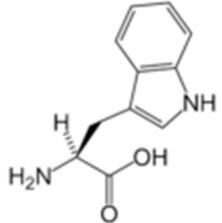
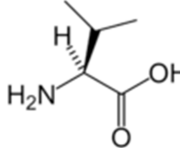
28. Marley J, Lu M, Bracken C. A method for efficient isotopic labeling of recombinant proteins. *Journal of biomolecular NMR*. 2001;20(1):71-5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11430757>.
29. Gueguen Y, Herpin A, Aumelas A, et al. Characterization of a defensin from the oyster *Crassostrea gigas*. Recombinant production, folding, solution structure, antimicrobial activities, and gene expression. *The Journal of biological chemistry*. 2006;281(1):313-23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16246846> [Accessed January 23, 2011].
30. Novagen. Novagen. *Components*. 2004:1-23. Available at: [www.novagen.com](http://www.novagen.com).
31. Güntert P. Automated NMR structure calculation with CYANA. *Methods in molecular biology (Clifton, N.J.)*. 2004;278:353-78. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19137264>.
32. Jouirou B, Mosbah A, Visan V, et al. Cobatoxin 1 from *Centruroides noxius* scorpion venom: chemical synthesis, three-dimensional structure in solution, pharmacology and docking on K<sup>+</sup> channels. *The Biochemical journal*. 2004;377(Pt 1):37-49. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1223841&tool=pmcentrez&rendertype=abstract>.

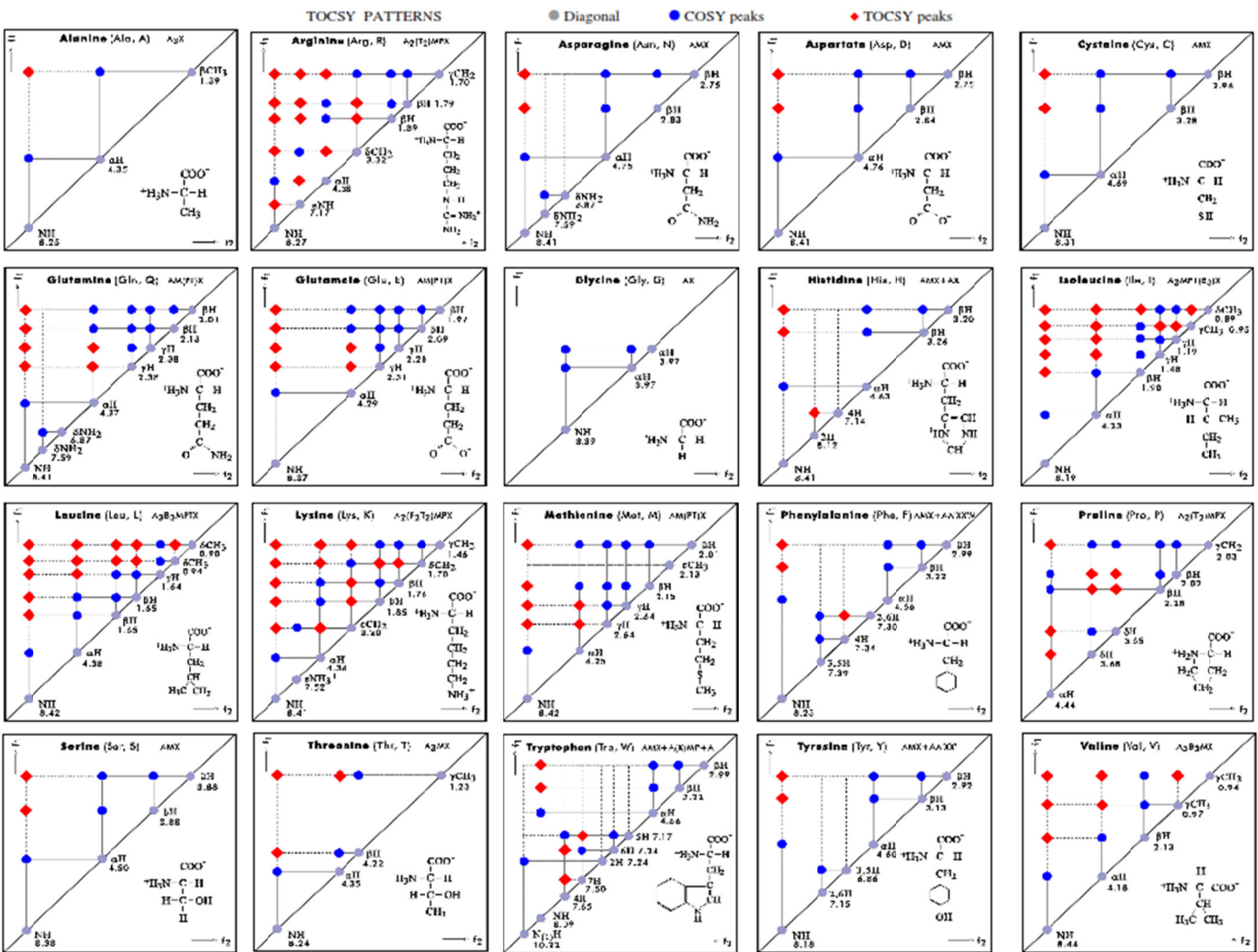
## APÉNDICE

### ABREVIATURA, CARACTERÍSTICAS Y ESTRUCTURA DE LOS AMINOÁCIDOS MÁS COMUNES

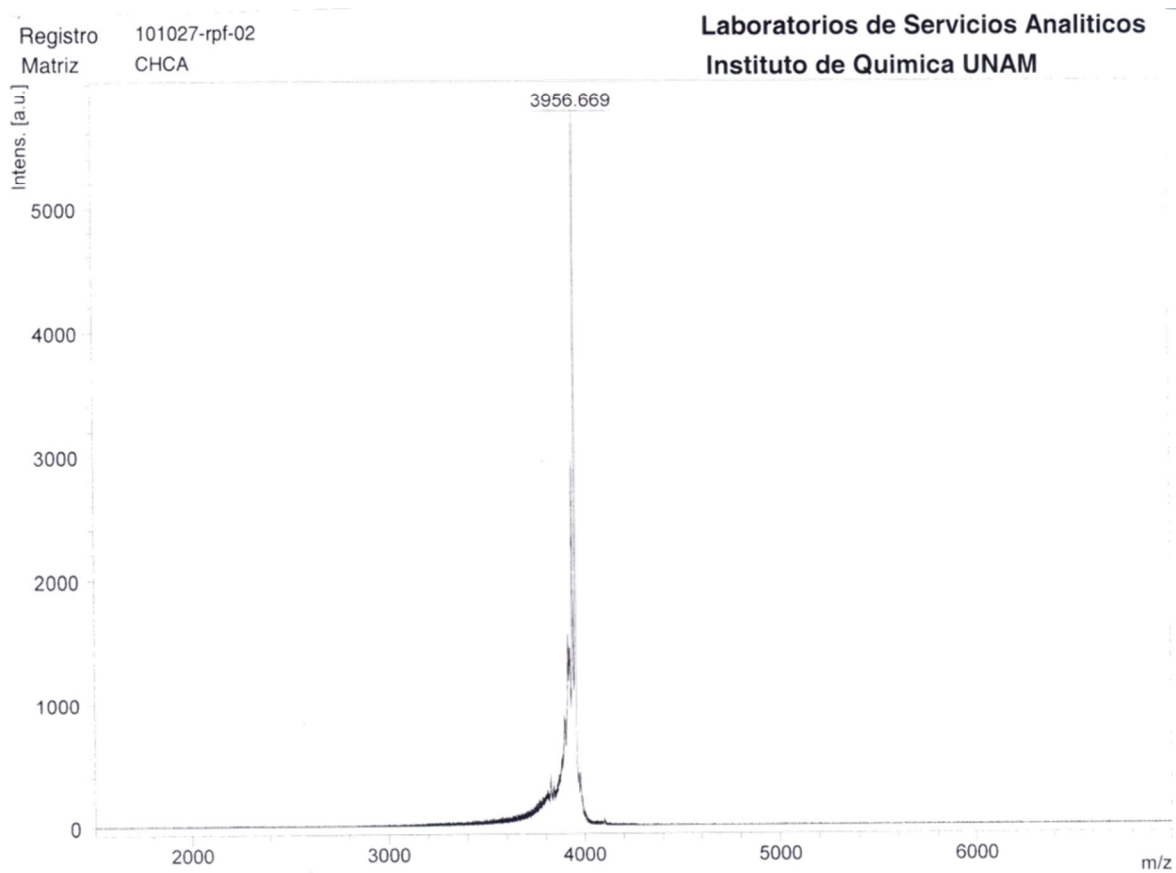
NOMBRE	ABREVIATURA		CARACTERÍSTICA	ESTRUCTURA
Alanina	<b>A</b>	ALA	Alifático no polar	
Arginina	<b>R</b>	ARG	Cargado positivamente	
Asparagina	<b>N</b>	ASN	Polar sin carga	
Aspartato	<b>D</b>	ASP	Cargado negativamente	
Cisteína	<b>C</b>	CYS	Polar sin carga	
Fenilalanina	<b>F</b>	PHE	Aromático	

NOMBRE	ABREVIATURA		CARACTERÍSTICA	ESTRUCTURA
Glicina	<b>G</b>	GLY	Alifático no polar	
Glutamato	<b>E</b>	GLU	Cargado negativamente	
Glutamina	<b>Q</b>	GLN	Polar sin carga	
Histidina	<b>H</b>	HIS	Cargado positivamente	
Isoleucina	<b>I</b>	ILE	Alifático no polar	
Leucina	<b>L</b>	LEU	Alifático no polar	
Lisina	<b>K</b>	LYS	Cargado positivamente	

NOMBRE	ABREVIATURA		CARACTERÍSTICA	ESTRUCTURA
Metionina	<b>M</b>	MET	Alifático no polar	
Prolina	<b>P</b>	PRO	Alifático no polar	
Serina	<b>S</b>	SER	Polar sin carga	
Tirosina	<b>Y</b>	TYR	Aromático	
Treonina	<b>T</b>	THR	Polar sin carga	
Triptófano	<b>W</b>	TRP	Aromático	
Valina	<b>V</b>	VAL	Alifático no polar	





**MALDI-TOF DE LA DEF1 SOLUBLE****Acquisition Parameter**

Date of acquisition 2010-10-27T13:30:17.812-05:00  
Acquisition method name D:\Methods\flexControlMethods\LNM\_UNAM\RP\_PepMix\_eren.par

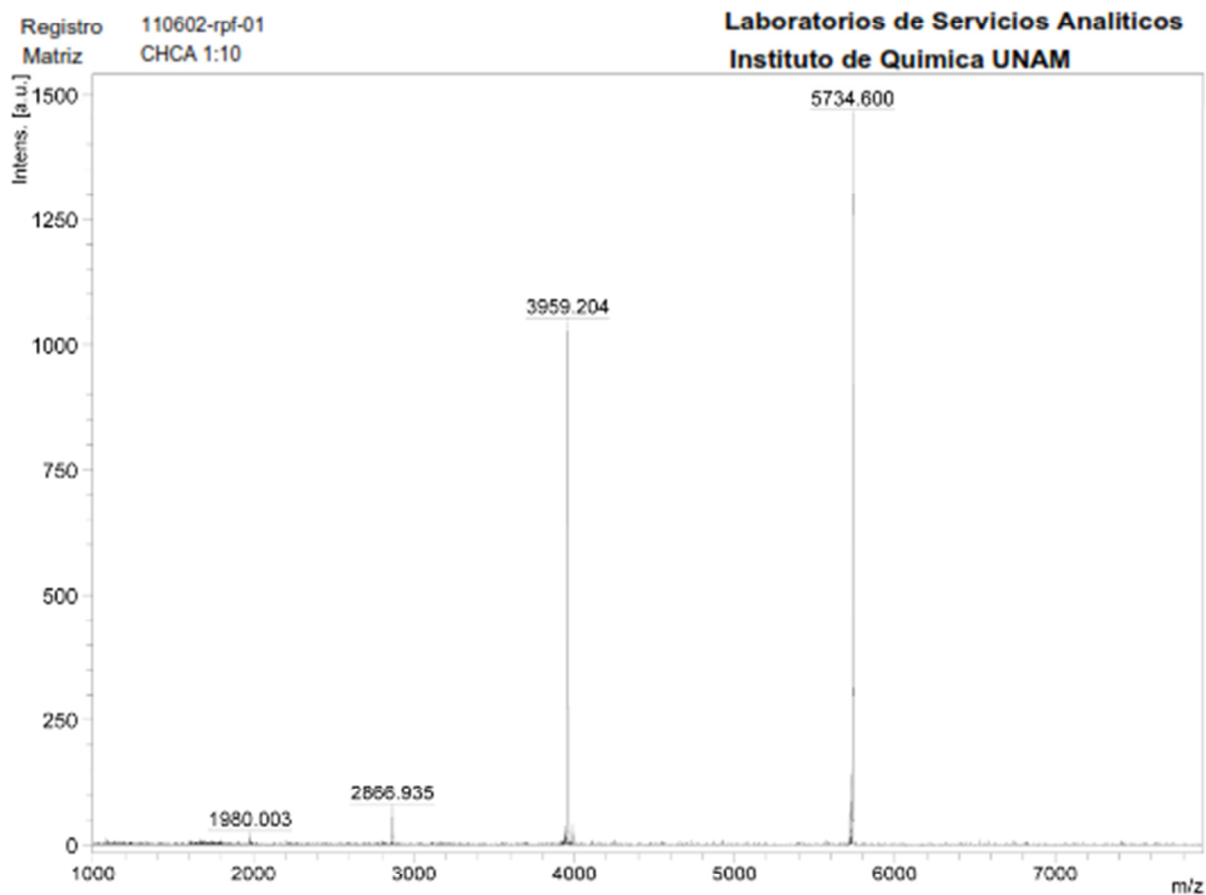
Acquisition operation mode Reflector  
Voltage polarity POS  
Number of shots 100  
Name of spectrum used for calibration  
Calibration reference list used

**Instrument Info**

User Fremont Demo  
Instrument FLEX-PC  
Instrument type microflex

D:\data\LSA\25\10\_E6\1

## MALDI-TOF DE LA DEF1 REPLEGADA

**Acquisition Parameter**

Date of acquisition 2011-06-06T12:22:52.937-05:00  
Acquisition method name D:\Methods\flexControlMethods\LNM\_UNAM\RP\_PepMtx\_eren.par

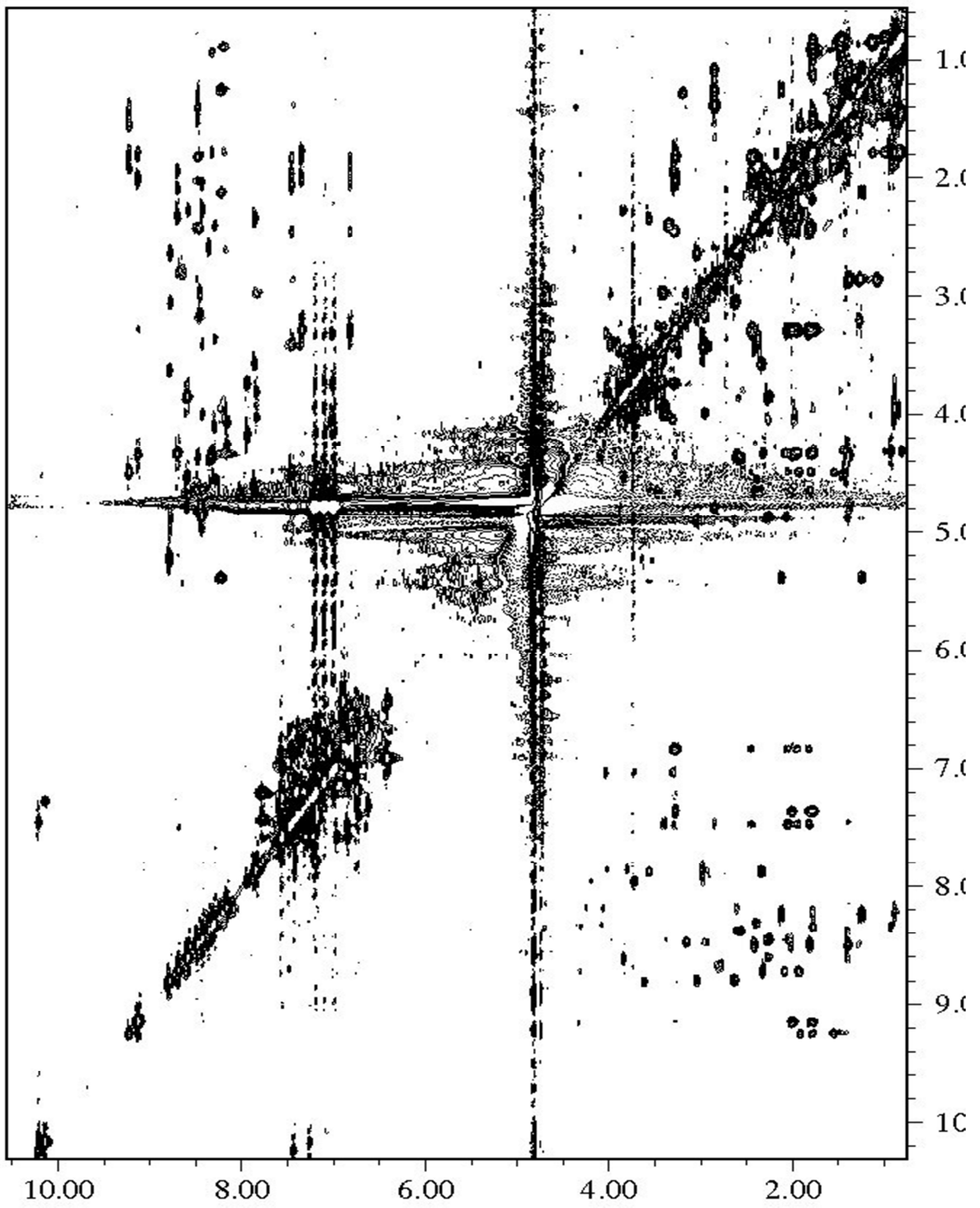
Acquisition operation mode Reflector  
Voltage polarity POS  
Number of shots 300  
Name of spectrum used for calibration  
Calibration reference list used 1-9 KDa CalbStandard\_eren

**Instrument Info**

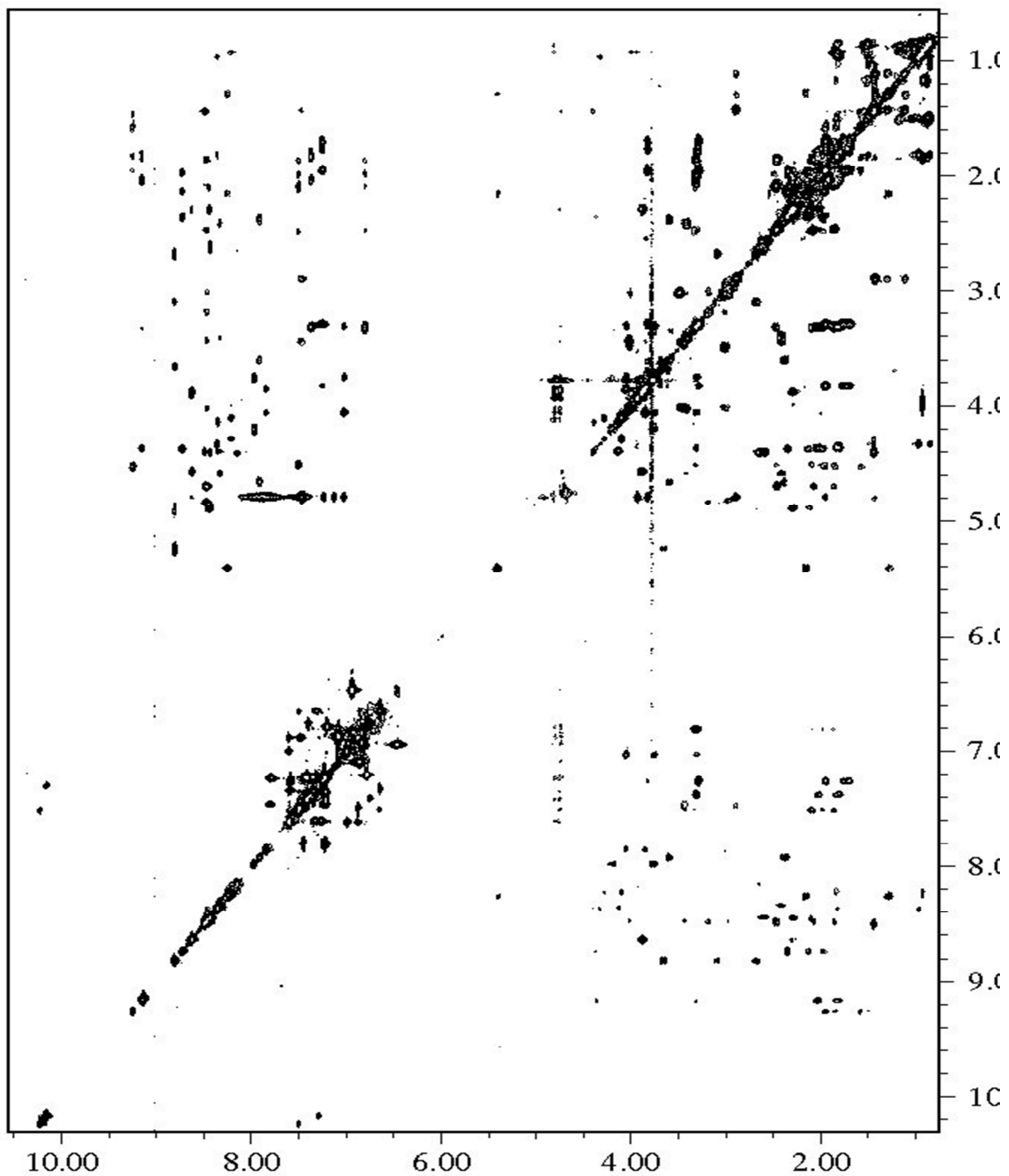
User UNAM  
Instrument FLEX-PC  
Instrument type microflex

D:\data\LSA\DEF1+Insulina0\_F8\2

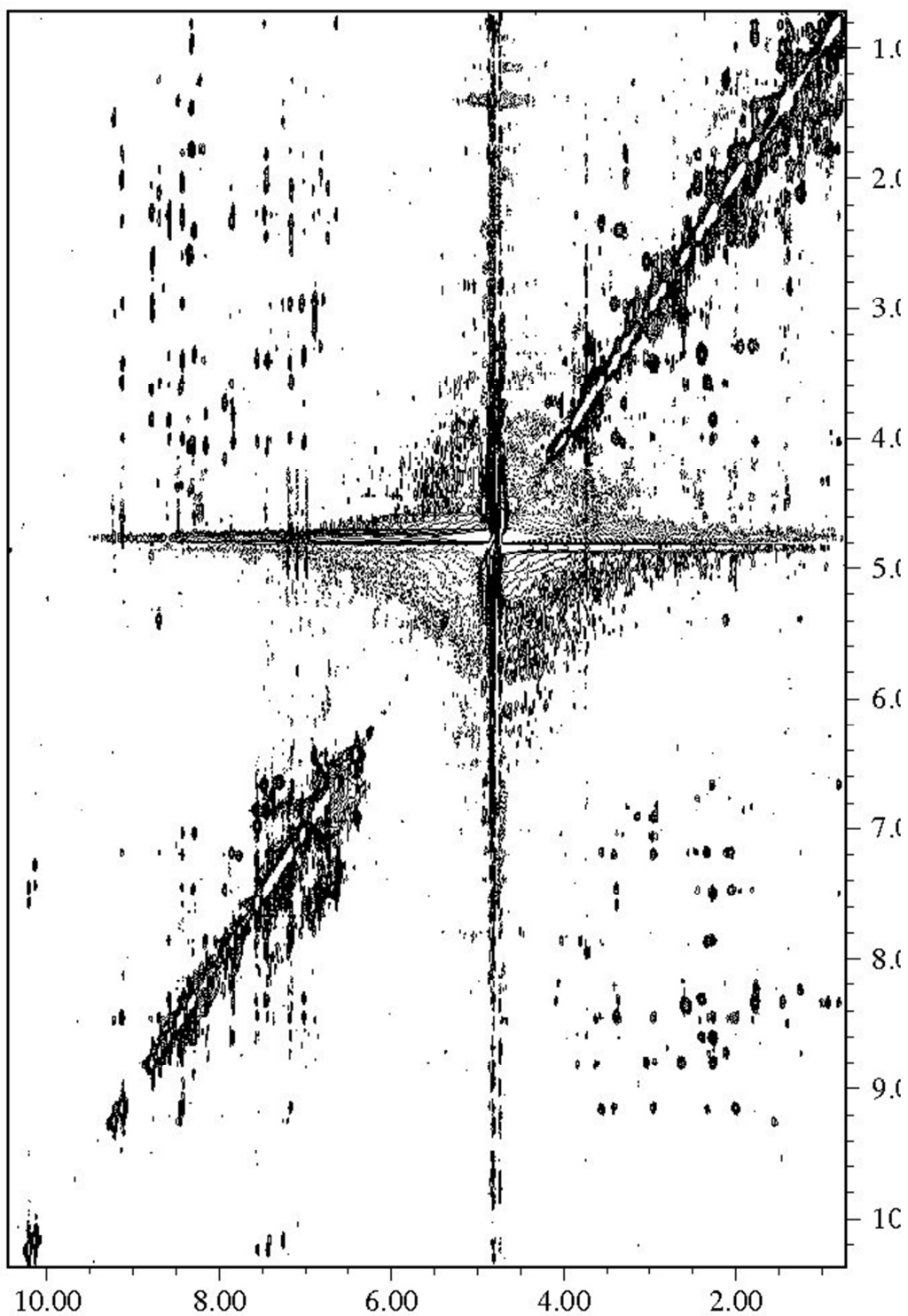
TOCSY DE LA GS-DEF1 SOLUBLE



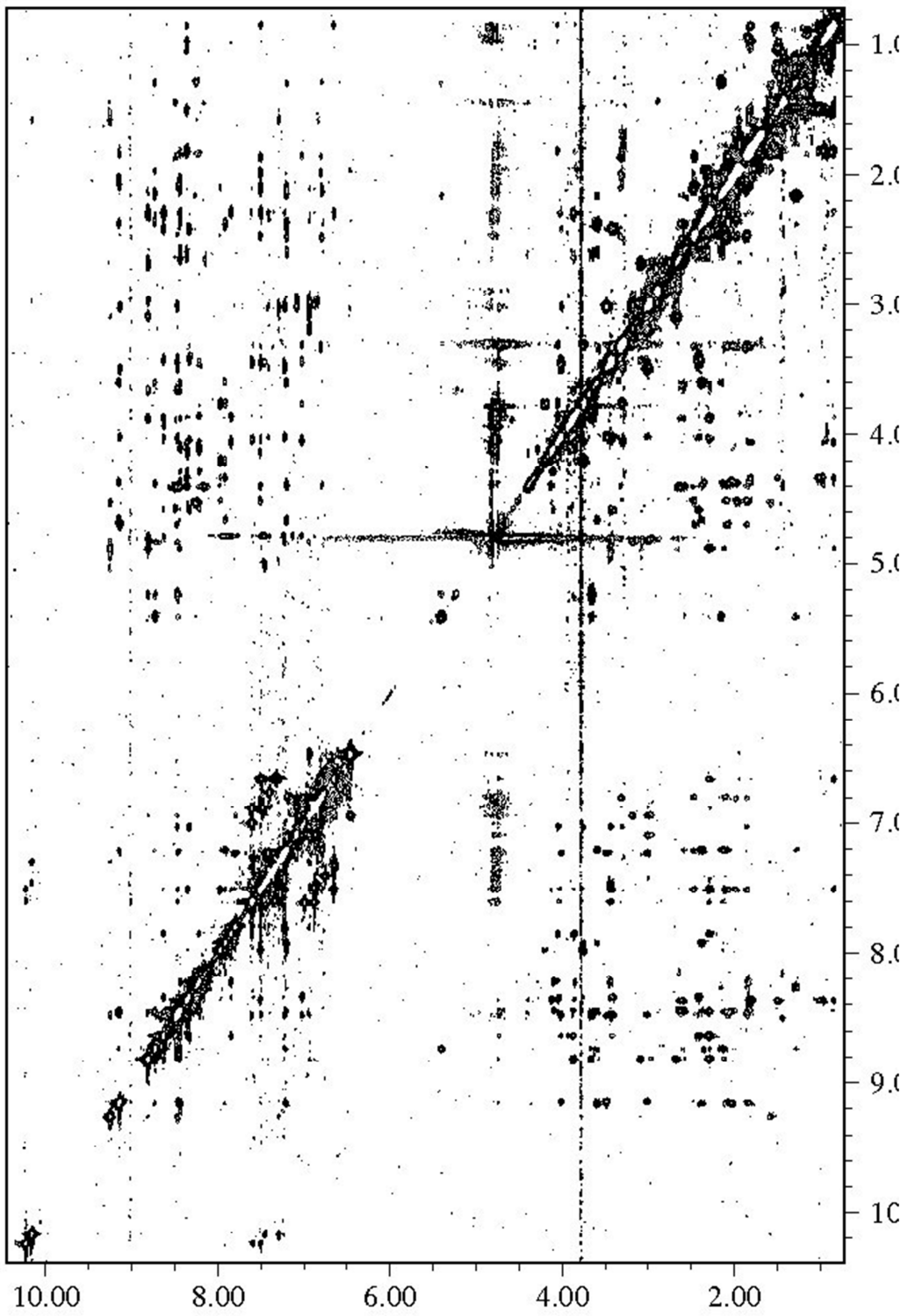
TOCSY DE LA GS-DEF1 REPLEGADA



NOESY DE LA GS-DEF1 SOLUBLE



NOESY DE LA GS-DEF1 REPLEGADA



## DATOS TÉCNICOS

Tabla 6: Datos técnicos de los reactivos y materiales utilizados

PROCESO	NOMBRE DEL REACTIVO	MARCA	PUREZA
Obtención del gen	Gen de la DEF-1	Epochlab	-
	Vector pET32-a-mod	Novagen	-
	BamHI	Biolab	-
	NcoI	Biolab	-
	Ligasa	Biolab	-
	E. coli DH5 $\alpha$	Novagen	-
Expresión de proteína	E. coli Rosetta	Novagen	-
	Medio L.B.	Sigma	Biología molecular
	Medio L.B. Agar	Sigma	Biología molecular
	IPTG	Sigma	99%
	Cloranfenicol	Sigma	-
	Ampicilina	Goldbio	-
Purificación por afinidad	Trizmahidrocloruro	Sigma	99%
	Cloruro de sodio	Reasol	R.A.
	Urea	Sigma	98%+
	Imidazol	Sigma-Aldrich	99%
	Cloruro de níquel	Reasol	R.A.
	Hi-Trap	Sephadex	-
Corte de la proteína	Trombina	Sigma	-

PROCESO	NOMBRE DEL REACTIVO	MARCA	PUREZA
Plegamiento	DL-Ditiotreitól	Sigma-Aldrich	99%
	L-Cistina	Sigma	Reactivo
	L-CisteínaHClmonohidrato	Sigma	>98%
	Arginina	Sigma	Reactivo
Purificación por HPLC	Acetonitrilo	Fermont	HPLC
	Ácido Trifluoroacético	Sigma-Aldrich	99%

Tabla 7: Datos técnicos de los equipos utilizados

Uso	EQUIPO	OPERACIÓN
Cromatografía	Detector: ProStar 325 UV-Vis detector (Varian)	Usuario
	Bombas: ProStar 210 Solvent Delivery Module (Varian)	
	Degasificador: DegasexMod.DG-4400 (Phenomenex)	
	Columna: Jupiter 4u Proteo 90 A250 4.60 nm (Phenomenex)	
Resonancia Magnética Nuclear	Varian Unity Inova 500 MHz	Dr. Federico del Río
	Varian 700 MHz	Dr. Federico del Río
Espectrometría de masas	Espectrómetro de masas MALDI-TOF Bruker microflex	Q. Eréndira García Ríos
Sonicador	Sonicator 3000 Misonix	Usuario
Incubadora con agitación	Innova 4000 Incubatorshaker	Usuario



Uso	EQUIPO	OPERACIÓN
	New Brunswick scientific	
Ultrafiltración	Amicon	Usuario
Centrifuga	BeckmancoulterAvanti J-25 I Rotores: JA-14 y JA-17	Usuario

Tabla 8: Composición de las disoluciones utilizadas

NOMBRE DE LA DISOLUCIÓN	COMPONENTE	CONCENTRACIÓN
Buffer de lisis	Tris	50 mM
	NaCl	500 mM
Buffer de lisis con urea	Tris	50 mM
	NaCl	500 mM
	Urea	8 M
Buffer de elución	Tris	50 mM
	NaCl	500 mM
	Imidazol	500 mM
Buffer de elución con urea	Tris	50 mM
	NaCl	500 mM
	Imidazol	500 mM
Buffer de plegamiento	Tris	100 mM
	Cistina	5 mM
	Cisteína	0.5 mM
	Arginina	600 mM
Buffer de corte	Tris	50 mM

NOMBRE DE LA DISOLUCIÓN	COMPONENTE	CONCENTRACIÓN
	CaCl <sub>2</sub>	10 mM

#### PROTOCOLO DE TRANSFORMACIÓN DE CÉLULAS POR CHOQUE TÉRMICO

1. Se toman 50  $\mu$ L de la disolución de bacterias competentes, y se agregan 2  $\mu$ L del plásmido pET32-a-mod-DEF1 y se mezclan por inversión; se deja incubar en hielo media hora. Se recomienda tener 50  $\mu$ L de bacterias competentes a las que se les someta al mismo procedimiento, sin plásmido, a modo de control.
2. Se realiza un choque térmico. Se transfieren los tubos a un baño a 42° C durante 45 segundos, y se regresa al hielo para incubar 2 minutos.
3. Se añaden 200  $\mu$ L de disolución L.B. y se incuban a 37° C en agitación una hora.
4. Se siembran en una caja Petri con medio L.B.-agar con cloranfenicol y ampicilina. Se deja incubando toda la noche.