

**Centro de Investigación Científica y de Educación
Superior de Ensenada, Baja California.**



**Doctorado en Ciencias
Ciencias de la Vida
con Orientación en Biotecnología Marina**

**Transcriptómica del ducto venenoso de cuatro especies de la
familia Conidae: su importancia para elucidar sus
componentes y los estudios de Proteómica.**

Tesis
para cubrir parcialmente los requisitos necesarios para obtener el grado de
Doctor en Ciencias

Presenta:

Andrea Figueroa Montiel

Ensenada, Baja California, México
2017.

Tesis defendida por

Andrea Figueroa Montiel

y aprobada por el siguiente Comité

Dr. Alexei Fedoróvish Licea Navarro

Director de tesis

Dr. Carlos Alberto Brizuela Rodríguez

Dr. Luis Javier González López

Dra. María Asunción Lago Lestón

Dr. Marco Antonio Ramos Ibarra



Dra. Clara Elizabeth Galindo Sánchez

Coordinador del Posgrado en Ciencias de la vida

Dra. Rufina Hernández Martínez

Directora de Estudios de Posgrado

Andrea Figueroa Montiel © 2017

Queda prohibida la reproducción parcial o total de esta obra sin el permiso formal y explícito del autor y director de la tesis.

Resumen de la tesis que presenta **Andrea Figueroa Montiel** como requisito parcial para la obtención del grado de Doctor en Ciencias de la Vida con orientación en Biotecnología Marina.

Transcriptómica del ducto venenoso de cuatro especies de la familia Conidae: su importancia para elucidar sus componentes y los estudios de Proteómica.

Resumen aprobado por:

Dr. Alexei Fedoróvish Licea Navarro
Director de tesis

Las conotoxinas de los caracoles marinos venenosos de la Familia Conidae, que tienen como blanco varios receptores y canales iónicos de importancia terapéutica, se distinguen por ser muy disímiles entre sí: conservan únicamente el péptido señal y determinados marcos de cisteína. A pesar de que estos venenos han sido estudiados de manera sistemática durante los últimos 60 años, sólo ~2% de los conopéptidos han sido identificados. La principal limitante es que el método tradicional de descubrimiento requiere de tiempo y de una gran cantidad de veneno difícil de obtener. Además, esta metodología ha sesgado el estudio de los venenos, favoreciendo a las conotoxinas y olvidando al resto de los componentes. La mejora de las técnicas de secuenciación masiva ha permitido obtener transcriptomas a un costo y tiempo razonable. En este trabajo se secuenció el transcriptoma de la glándula venenosa de cuatro cónidos pobremente estudiados. Se construyeron bibliotecas de cDNA para cada especie que se secuenciaron utilizando Illumina, posteriormente las lecturas se ensamblaron *de novo* con Trinity y se identificaron las conotoxinas utilizando pHMM. El número de secuencias obtenidas (de 20 a 200) fue menor al número de conotoxinas esperadas (de 100 a 200), no obstante, representa una primera aproximación del contenido venómico de estas especies. Se realizó un análisis exhaustivo del transcriptoma de *Conus regularis* y se identificaron 57 conopéptidos, clasificados en 18 superfamilias. También se identificaron secuencias que no pudieron ser clasificadas pero que presentaban características típicas de conotoxinas; así como se confirmó la existencia de nuevos marcos de cisteínas. Por otro lado, se sabe que las PDIs son indispensables para el plegamiento adecuado de péptidos con residuos de cisteínas. Se utilizó un motivo sumamente conservado de estas enzimas (APWCGHCK) para recuperar las familias de PDIs contenidas en los venenos estudiados. La caracterización de las familias de PDIs encontradas se realizó mediante un análisis *in silico* basado en la presencia de dominios funcionales similares a la Tiorredoxina. En total se recuperaron 41 secuencias que, en promedio, tenían un 65 % de identidad con PDIs de otras especies. Por lo menos una secuencia de cada familia es capaz de adquirir la estructura tridimensional esperada para la Tiorredoxina, sugiriendo que los cónidos requieren de estas enzimas para realizar un plegamiento correcto de sus toxinas. Finalmente, se comprobó la utilización de *sequence tags* como un método eficiente para obtener secuencias completas de cónidos combinando el análisis de LC-MS/MS con los transcriptomas obtenidos. Se reporta la primera conotoxina aislada del veneno de *Conasprella ximenes*, I1_xm11a, un péptido de 37 aminoácidos (PM 4109.69 Da) y ocho residuos de cisteína. Este péptido no sólo es la primera conotoxina descrita capaz de inhibir el crecimiento de Mtb, sino que lo hace a concentraciones similares a las de dos fármacos utilizados en el tratamiento clínico de esta enfermedad. Por último, se identificó O2_rs15a una conotoxina con un número inusual de MPT: ocho residuos de cisteínas, dos hidroxiprolinas y tres residuos de ácido gamma carboxiglutámico.

Palabras clave: Conotoxinas, pHMM, proteómica, transcriptómica, PDIs.

Abstract of the thesis presented by **Andrea Figueroa Montiel** as a partial requirement to obtain the Doctor of Science degree in Life Science with orientation in Marine Biotechnology

Venom duct transcriptomic of four species from the Conidae family: it's importance in elucidating its components and the relevance in Proteomic studies.

Abstract approved by:

Dr. Alexei Fedoróvish Licea Navarro
Thesis Director

Conotoxins of the venomous marine snails from the Conidae Family target several receptors and ion channels of therapeutic relevance. Conopeptides sequences present a lot of variation from one another, only preserving a highly conserved signal peptide and certain cysteine frameworks. Although these venoms have been systematically studied over the past 60 years, only ~2% of conopeptides have been identified. The main limitation is that the traditional method of discovery requires time and large amounts of venom difficult to obtain. Moreover, this methodology has biased the study of cone snails venoms towards the discovery of conotoxins and limiting the elucidation of other components. With the decreasing costs and increasing efficiency of NGS techniques it is easier to obtain complete transcriptomes. In this work, the venom gland transcriptome of four poorly studied cone snails was sequenced. cDNA libraries were constructed for each species and sequenced using Illumina, reads were assembled into contigs with Trinity in a *de novo* fashion and pHMM were used in order to identify conotoxins. The number of sequences obtained (20-200) was smaller than the one expected (100-200), however it represents the first approximation of the venom content of these species. A comprehensive analysis of *Conus regularis* transcriptome was performed, yielding 57 conopeptides classified in 18 different superfamilies. Sequences that couldn't be classified but exhibited characteristics typical of conotoxins were retrieved. Also, new cysteine frameworks were confirmed. On the other hand, PDIs are required for proper folding of "cysteine-rich" peptides, so in order to capture potential PDIs, a highly conserved motif (APWCGHCK) was used. Characterization of PDIs families was performed *in silico* based in the presence of functional Thioredoxin-like domains. 41 sequences were retrieved which had on average 65% identity with other PDIs. At least one sequence of each family showed the expected thioredoxin fold, suggesting that cone snails require these enzymes to perform a correct folding of their toxins. Last, the use of "sequence tags" was established as an efficient method to obtain complete conotoxins sequences combining LC-MS/MS analysis with the obtained transcriptomes. The first conotoxins isolated from *Conasprella ximenes* venom is reported, I1_xm11a, a 37-mer peptide (MW 4109.69 Da) with eight cysteine residues. This peptide is not only the first conotoxin described that is capable of inhibit the growth of Mtb, but it also does it at concentrations similar to two drugs used in clinical treatments. Lastly, O2_rs15a a conotoxin with unusual PTMs was described: eight cysteine residues, two hydroxiprolines and three gamma-carboxyglutamic acids.

Keywords: Conotoxins, transcriptomic, proteomics, pHMM, PDIs.

Dedicatoria

A Eduardo,

Este logro es tan mío como tuyo, porque sin tu apoyo no hubiera sido posible. Eres la persona que me inspira a ser mejor cada día, compartir mi vida contigo ha sido mi mejor decisión. Tengo fe en el camino que sigue por recorrer porque sé que será a tu lado. Te amo.

When no-one else can understand me
 When everything I do is wrong
 You give me hope and consolation
 You give me strength to carry on
 And you're always there to lend a hand
 In everything I do
 That's the wonder
 The wonder of you
 And when you smile the world is brighter
 You touch my hand and I'm a ~~king~~ queen
 Your kiss to me is worth a fortune
 Your love for me is everything
 I'll guess I'll never know the reason why
 You love me like you do
 That's the wonder
 The wonder of you

Baker Night

A Manuel Figueroa, porque para que yo llegara aquí fuiste campesino, profesor y científico, pero sobre todo has sido un padre amoroso que fue capaz de regalarme la Luna.

A Astrid Montiel, mi mami, porque estás hecha
 de olas y espuma
 de palabras sabias y abrazos que curan
 de carcajadas al viento que no dejan duda.

A Ana Sofía, la chaparrita que vino a hacer mis días más brillantes.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada para la realización de este proyecto de doctorado.

Al Centro de Investigación Científica y Educación Superior de Ensenada (CICESE) por permitirme el uso de sus instalaciones y recursos para poder realizar mis estudios de posgrado.

Al Dr. Alexei Licea Navarro por permitirme formar parte de su equipo de trabajo, abrirme las puertas a varias oportunidades que se presentaron a lo largo de estos años y por encontrar la manera de cumplir las metas.

A los miembros de mi comité de tesis: Dr. Carlos Brizuela Rodríguez, Dr. Luis Javier González López, Dra. Asunción Lago Lestón y Dr. Marco Ramos Ibarra porque tuvieron la mejor disposición de aclarar mis dudas. Muchas gracias por sus observaciones que enriquecieron este trabajo y por recibirme siempre con calidez.

Un especial agradecimiento al Dr. Luis Javier González López, no sólo porque fue imprescindible en el desarrollo de esta tesis, sino porque es un excelente maestro que transmite su pasión por la ciencia.

Al Dr. Marco A. Ramos y la Dra Rosa E. Mares, que fueron piezas clave en entender el mundo de las PDIs.

Al M.C. Armando Beltrán Verdugo cuya colaboración fue esencial para la identificación de las conotoxinas descritas en esta tesis.

Del IBT, al Dr. Possani, y en particular al Dr. Ernesto Ortiz, por ayudarme con las extracciones de ARN, pero sobre todo por contestar mis correos llenos de dudas.

A las personas que me ayudaron a resolver los múltiples problemas que se presentaron durante la realización de este trabajo: M.C. Hugo Guillén, M.C. Sylvia Camacho, M.C. Isabel Echevarría y Dr. Óscar Juárez. Así como a todos mis compañeros de la DIB, en particular a Salvador, Jahaziel, Félix, Mimí y Ricardo.

A Adriana Mejía y Gaby Altamirano por ayudarme en todos los trámites administrativos siempre con la mejor disposición. A las doctoras Fabiola Lafarga y Clara Galindo por ser mujeres líderes. Así como a Charles Darwin, uno de mis científicos favoritos, cuya biografía me ayudó en el último año de esta tesis.

A mi compañera, pero sobre todo amiga, Johanna por estar al pendiente de mí, echándome porras en los momentos claves y resolviendo mis dudas. A Samanta, que es una gran amiga y que me acompañó a varias aventuras de esta tesis, sin ti no hubiera sido ni la mitad de divertido.

Agradecimientos personales

A Eduardo, por toda tu luz. A mis padres por su amor incondicional. A mi hermana por las risas.

A mi suegra Griselda por estar al pendiente de mí e invariablemente recibirme con una sonrisa. Y a mis cuñis Raquel y Reyna, a las que tengo la fortuna de llamar familia.

A los amigos de mi esposo, que ahora son mis amigos, y que hicieron los fines de semana más divertidos. Particularmente a Mario y a Felipe por acompañarnos en tantas pláticas y comidas. Por último, si bien no menos importante a mis amigas, parte indispensable de mí: Anette, Bonnie, Sole, Angie, Tali, Cris, Sofi, Edna y las Cundi. Especialmente a Ana Elena e Irasema que me acompañan en esta nueva etapa de mi vida, por ser una amigas *cool* que hacen los momentos difíciles menos amargos, por escucharme, por el chisme y las “pláticas *fit*”, que fueron la mejor distracción. Las quiero muchísimo.

Tabla de contenido

	Página
Resumen en español	ii
Resumen en inglés	iii
Dedicatoria.....	iv
Agradecimientos	v
Lista de figuras	xi
Lista de tablas	xiii
 Parte I. Identificación de las conotoxinas que conforman el veneno de cuatro especies de cónidos de la costa de Baja California.	 1
Capítulo 1. Introducción	1
1.1 Antecedentes.....	2
1.1.1 Generalidades de la familia Conidae	2
1.1.2 La diversidad del veneno de los cónidos	2
1.1.3 El aparato venenoso	4
1.1.4 Clasificación de los componentes del veneno de cónidos	5
1.1.5 Importancia del estudio de las conotoxinas.....	9
1.1.6 RNA-seq	10
1.2 Justificación	11
1.3 Hipótesis.....	11
1.4 Objetivo general	11
1.4.1 Objetivos particulares	12
Capítulo 2. Metodología	13
2.1 Obtención de los especímenes	13
2.2 Extracción de ARN total de la glándula venenosa y secuenciación de su transcriptoma	13
2.3 Entrega de secuencias.....	14
2.4 Ensamble del transcriptoma empleando el programa de cómputo Trinity v.2.0.3.....	15

2.5 Búsqueda de conotoxinas con el programa de cómputo ConoSorter	16
2.6 Construcción de perfiles de Modelos Ocultos de Markov	16
Capítulo 3. Resultados	17
3.1 Extracción de ARN total	17
3.2 Análisis de la calidad de las secuencias.....	18
3.2.1 Calidad de las secuencias por base	19
3.2.2 Contenido de bases por secuencia	19
3.2.3 Contenido GC: distribución GC a lo largo de las secuencias.....	20
3.3 Análisis del BLAST del ensamble del IBT	21
3.4 Resultados obtenidos del ensamble con el programa de cómputo Trinity v.2.0.3	22
3.5 Resultados obtenidos con el programa de cómputo ConoSorter	23
3.6 Resultados obtenidos con los pHMM	25
3.7 Análisis del transcriptoma de <i>C. regularis</i>	26
3.7.1 Superfamilia O1	27
3.7.2 Superfamilia O2	28
3.7.3 Superfamilia O3	29
3.7.4 Superfamilia C/Contulacina y superfamilia D	30
3.7.5 Superfamilia H	31
3.7.6 Superfamilia I1.....	31
3.7.7 Con-ikot-ikot.....	32
3.7.8 Superfamilia L.....	33
3.7.9 Superfamilia M	33
3.7.10 Superfamilia B2	35
3.7.11 Superfamilia N	35
3.7.12 Superfamilia P	36
3.7.13 Superfamilia S.....	37
3.7.14 Superfamilia T.....	37
3.7.15 Superfamilias Divergentes.....	38
3.7.15.1 Superfamilia MSTLGMTLL-	38
3.7.15.2 Superfamilia M---L-LTVA	40
3.7.16 Secuencias que no pueden ser clasificadas.....	42
3.7.16.1 Superfamilia D putativa.....	42
3.7.16.2 Superfamilias I1, B1 y B2 putativas	43

3.7.16.3 Superfamilia A putativa	44
3.7.17 Contenido total del veneno de <i>C. regularis</i>	45
Capítulo 4. Discusión	49
Capítulo 5. Conclusiones.....	53
 Parte II. Las proteínas disulfuro isomerasas y su papel esencial en el plegamiento adecuado de las conotoxinas.	54
Capítulo 1. Introducción	54
1.1 Antecedentes.....	55
1.1.1 PDI típica	55
1.1.2 PDI de cónidos.....	56
1.2 Justificación	57
1.3 Hipótesis.....	57
1.4 Objetivos	58
1.4.1 Objetivo general.....	58
1.4.2 Objetivos específicos.....	58
Capítulo 2. Metodología	59
2.1 Ensamble del transcriptoma.....	59
2.2 Análisis Bioinformático	59
2.3 Predicción de la estructura 3D de las PDI de cónidos	60
2.4 Construcción de árbol filogenético de las PDI de cónidos reportadas a la fecha.....	61
Capítulo 3. Resultados	62
3.1 Ensamble del transcriptoma con el programa de cómputo Trinity	62
3.2 Análisis estructural de las familias de PDI de los cónidos.....	62
3.2.1 PDIs de <i>C. ximenes</i>	65
3.2.2 PDIs de <i>C. regularis</i>	66
3.2.3 PDIs de <i>C. mahogani</i>	67
3.2.4 PDIs de <i>C. californicus</i>	68
3.3 Alineamiento de las secuencias de PDI.....	68
3.4 Predicción de la estructura tridimensional de las PDI de los cónidos.	70

3.5 Árbol filogenético	70
Capítulo 4. Discusión	74
Capítulo 5. Conclusiones.....	77
 Parte III. Combinación de estrategias de identificación en la secuenciación de conotoxinas ..	78
Capítulo 1. Introducción	78
1.1 Antecedentes.....	79
1.1.1 <i>Mycobacterium tuberculosis</i>	79
1.1.2 Los venenos como fuentes de fármacos	80
1.1.3 Elucidación de la estructura primaria de una conotoxina.	80
1.2 Justificación	82
1.3 Hipótesis.....	82
1.4 Objetivos	82
1.4.1 Objetivo general	82
1.4.2 Objetivos específicos.....	82
Capítulo 2. Metodología	83
2.1 Colecta de especímenes y extracción del veneno	83
2.2 Purificación de los péptidos de interés.	84
2.3 Reducción y S-alquilación de las conotoxinas.	84
2.4 Análisis mediante LC-MS/MS	85
2.5 Transcriptómica de la glándula venenosa	86
2.6 Búsqueda en la base de datos y análisis de la secuencia.....	86
2.7 Verificación de la secuencia de la toxina I1_xm11a mediante la digestión trípica y análisis LC-MS/MS	87
Capítulo 3. Resultados	88
3.1 Purificación de los péptidos de interés.	88
3.2 Identificación de la conotoxina de <i>C. ximenes</i>	89
3.2.1 Cálculo del número de cisteínas.....	89
3.2.2 Obtención de la <i>sequence tag</i> e identificación de la secuencia.	91

3.2.3 Verificación de I1_xm11a mediante la digestión trípica y análisis LC-MS/MS de los fragmentos derivados del péptido reducido y S-alquilado.	92
3.2.4 Análisis de la conotoxina I1_xm11a.	94
3.3 Identificación de la conotoxina de <i>C. regularis</i>.	95
3.3.1 Cálculo del número de cisteínas.....	95
3.3.2 Obtención de la <i>sequence tag</i> e identificación de la secuencia	96
3.3.3 Análisis de la secuencia identificada	99
Capítulo 4. Discusión	101
Capítulo 5. Conclusiones.....	104
Literatura citada.....	105
Anexo 1. Construcción de los pHMM	114
Anexo 2. Artículo publicado.....	116
Anexo 3. Gráficas de la calidad de las secuencias.....	117

Lista de figuras

Figura 1. Aparato venenoso de los cónidos	4
Figura 2. Maduración del conopéptido y diferentes maneras en que puede ser clasificado	5
Figura 3. Electroferograma y representación en gel de agarosa de las cuatro extracciones de ARN	18
Figura 4. Gráfica de calidad por base para la secuenciación de <i>C. ximenes</i>	19
Figura 5. Contenido de bases a lo largo de la secuencia correspondiente al cebador R1 de <i>C. regularis</i> ...	20
Figura 6. Distribución GC de la secuenciación con el cebador R1 para la especie <i>C. californicus</i>	21
Figura 7. Secuencias para la superfamilia O1 de <i>C. regularis</i>	28
Figura 8. Secuencias identificadas para la superfamilia O2 de <i>C. regularis</i>	29
Figura 9. Conotoxinas identificadas para la superfamilia O3 de <i>C. regularis</i>	29
Figura 10. Cotulacina y miembros de la superfamilia D identificadas para <i>C. regularis</i>	30
Figura 11. Conotoxinas de la superfamilia H.....	31
Figura 12. Secuencias de la superfamilia I1	32
Figura 13. Secuencias con-ikot-ikot de <i>C. regularis</i>	32
Figura 14. Secuencias de la superfamilia L de <i>C. regularis</i>	33
Figura 15. Superfamilia M	34
Figura 16. Superfamilia B2	35
Figura 17. Superfamilia N.....	36
Figura 18. Superfamilia P	36
Figura 19. Superfamilia S.....	37
Figura 20. Superfamilia T.....	38
Figura 21. Superfamilia MSTLGMTLL-	40
Figura 22. Superfamilia M---L-LTVA.	41
Figura 23. Conotoxinas que no pudieron ser clasificadas	42
Figura 24. Conopéptidos sin poder ser clasificados	44
Figura 25. Superfamilia A putativa	45
Figura 26. Distribución del número de conotoxinas encontradas para cada una de las superfamilias de genes recuperadas.	46
Figura 27. Organización de los dominios (a, b, b', a' y c) en una PDI típica	55
Figura 28. Comparación de los dominios catalíticos de las secuencias de PDIs.....	69
Figura 29. Patrón de plegamiento de las PDIs de los cónidos.....	72
Figura 30. Árbol filogenético obtenido con las secuencias de Safavi-Hemmai y colaboradores (2016) y las obtenidas en este trabajo.	73

Figura 31. Fraccionamiento mediante RP-HPLC del veneno de <i>C. ximenes</i>	88
Figura 32. Fraccionamiento mediante RP-HPLC del veneno de <i>C. regularis</i>	89
Figura 33. Espectro ESI-MS del péptido nativo y reducido S-carboamidometilado de la conotoxina I1_xm11a.....	90
Figura 34. Reducción y alquilación de I1_xm11a	92
Figura 35. Espectro ESI-MS del péptido triptico derivado de la reducción y S-alquilación de la conotoxina I1_xm11a.....	93
Figura 36. Alineamiento de I1_xm11a con otros miembros de la superfamilia I1 con marco de cisteínas XI ...	94
Figura 37. Espectro ESI-MS de la conotoxina O2_rs15a nativa y reducida S-carboamidometilada.....	95
Figura 38. Estructura del ácido glutámico carboxilado y su transformación en ácido glutámico en las condiciones de fragmentación por CID	97
Figura 39. Espectro ESI-MS/M de O2_rs15a reducida y S-alquilada	98
Figura 40. Espectro ES-MS/MS de los iones 5 ⁺ y 7 ⁺ del péptido rs15a	99
Figura 41. Alineamiento del péptido O2_rs15a con otros miembros de la superfamilia O2 con marco de cisteínas XV.....	100
Figura 42. Flujo de trabajo para la anotación de las conotoxinas en las diversas superfamilias de genes utilizando pHMMs.	114
Figura 43. Gráfica de calidad por base para la secuenciación de <i>C. californicus</i>	117
Figura 44. Gráfica de calidad por base para la secuenciación de <i>C. regularis</i>	117
Figura 45. Gráfica de calidad por base para la secuenciación de <i>C. mahogani</i>	117
Figura 46. Contenido de bases a lo largo de la secuencia de <i>C. mahogani</i>	118
Figura 47. Contenido de bases a lo largo de la secuencia de <i>C. ximenes</i>	118
Figura 48. Contenido de bases a lo largo de la secuencia de <i>C. californicus</i>	118
Figura 49. Contenido de bases a lo largo de la secuencia del cebador R2 de <i>C. regularis</i>	119
Figura 50. Distribución GC de la secuenciación del cebador R2 para <i>C. californicus</i>	119
Figura 51. Distribución GC de la secuenciación para <i>C. ximenes</i>	120
Figura 52. Distribución GC de la secuenciación para <i>C. mahogani</i>	120
Figura 53. Distribución GC de la secuenciación para <i>C. regularis</i>	120

Lista de tablas

Tabla 1. Marcos de cisteína que han sido descritos.....	8
Tabla 2. Archivos generados en el IBT después de realizar el ensamble del RNA-seq del transcriptoma de los cuatro ductos venenosos de las especies utilizadas en este trabajo.....	15
Tabla 3. Concentración y calidad de cada extracción de ARN total.....	17
Tabla 4. Total de secuencias obtenidas para cada uno de los archivos entregados por el IBT.....	21
Tabla 5. Total de secuencias obtenidas para los archivos de aminoácidos que tuvieron anotación.....	22
Tabla 6. Resultados de las estadísticas obtenidas de los 4 ensambles realizados con Trinity.	23
Tabla 7. Número de secuencias encontradas para los archivos Regex y los archivos pHMM de cada uno de los ensambles.....	24
Tabla 8. Secuencias obtenidas en el ensamble y secuencias obtenidas después de eliminar las redundantes para cada especie.	25
Tabla 9. Comparación de secuencias encontradas con diferentes metodologías.	25
Tabla 10. Diversidad funcional encontrada en el transcriptoma de la glándula venenosa de <i>C. regularis</i>	46
Tabla 11. Resultados de las estadísticas obtenidas de los 4 ensambles realizados con Trinity.....	62
Tabla 12. Análisis de las secuencias obtenidas para cada una de las especies estudiadas.....	63
Tabla 13. Análisis estructural de los diferentes miembros de la familia de PDI de <i>C. ximenes</i>	65
Tabla 14. Análisis estructural de los diferentes miembros de la familia de PDI de <i>C. regularis</i>	66
Tabla 15. Análisis estructural de los diferentes miembros de la familia de PDI de <i>C. mahogani</i>	67
Tabla 16. Análisis estructural de los diferentes miembros de la familia de PDI de <i>C. californicus</i>	68
Tabla 17. Resumen de las secuencias de conopéptidos recuperadas de ConoServer.....	115

Parte I. Identificación de las conotoxinas que conforman el veneno de cuatro especies de cónidos de la costa de Baja California.

Capítulo 1. Introducción

El veneno producido por animales resulta de interés científico porque es una fuente vasta de componentes bioactivos, potentes y selectivos, con gran potencial terapéutico y son excelentes herramientas que permiten estudiar interacciones moleculares (Favreau y Stöcklin, 2009).

Dentro de los venenos de animales, se encuentran las toxinas conocidas como conopéptidos o conotoxinas que conforman el veneno de los caracoles marinos de la familia Conidae. Se ha demostrado que éstas tienen como blanco una amplia variedad de receptores y canales iónicos de relevancia biológica y que son atractivos desde el punto de vista terapéutico (Favreau y Stöcklin, 2009). Tradicionalmente, el descubrimiento de estas toxinas se ha basado en el fraccionamiento cromatográfico del veneno seguido de una evaluación de la actividad biológica de cada una de las fracciones. Posteriormente se realiza una purificación, empleando incluso distintos principios de separación, hasta llegar a una preparación homogénea que solo contiene la biomolécula de interés la cual se caracteriza estructural y funcionalmente para conocer su identidad y blancos moleculares. Sin embargo, este proceso es muy lento y requiere grandes cantidades de veneno crudo, que no sólo es difícil de obtener sino que la colecta constante de organismos puede ocasionar un impacto ambiental sino se realiza de manera controlada (Prashanth et al., 2012). Además esto ha ocasionado que el estudio de estas especies se centre únicamente en la identificación de sus toxinas y deje de lado la identificación del resto de los componentes que se encuentran en el veneno. A diferencia del genoma, que representa el contenido completo de genes de un organismo, el transcriptoma representa los genes que están siendo activamente transcritos en un tiempo determinado en un órgano, tejido, o célula. Esto en particular es de gran importancia en las especies venenosas donde el veneno producido se restringe a la actividad fisiológica de la glándula venenosa (Rendón-Anaya et al., 2013). La secuenciación del transcriptoma de esta glándula se está convirtiendo cada vez más en una herramienta rutinaria para el estudio de las posibles especies de naturaleza peptídicas presentes en el veneno.

En este trabajo se propone realizar un análisis exhaustivo del transcriptoma de la glándula venenosa de cuatro especies de cónidos (*Conasprella ximenes*, *Californiconus californicus*, *Conus regularis* y *Conus*

mahogani) de mares circundantes a la península de Baja California. Esto facilitaría la búsqueda de componentes biológicamente activos que podrán ser utilizados en estudios posteriores.

1.1 Antecedentes

1.1.1 Generalidades de la familia Conidae

Los miembros de la familia Conidae son gasterópodos marinos carnívoros que se caracterizan por tener un aparato venenoso complejo. La familia se conforma por 4 géneros diferentes: *Conus*, *Conasprella*, *Profundiconus* y *Californiconus* (Puillandre et al., 2015). El hábitat de los cónidos es diverso ya que se encuentran asociados a rocas, corales o arenas; y se distribuyen desde el intermareal hasta el mar profundo. Sin embargo, la mayoría habita las comunidades arrecifales de las aguas tropicales, a profundidades no mayores a los 100 m. (Olivera, 2002). Los cónidos suelen ser predadores nocturnos y, dependiendo de su tipo de dieta, pueden clasificarse en tres grupos principales. El grupo mayoritario son los vermívoros, cuya dieta principal son los gusanos, seguido de los moluscívoros, que se alimentan de otros gasterópodos, y los piscívoros, que depredan peces (Olivera, 1997). El veneno de algunos de estos organismos puede llegar a ser letal para los humanos, siendo el piscívoro *Conus geographus* el más peligroso ya que en ausencia de asistencia médica, el 70 % de las picaduras de este cónido resultan letales (Hu et al., 2012).

1.1.2 La diversidad del veneno de los cónidos

Durante los últimos 60 años se ha intentado caracterizar el veneno de las más de 700 especies que conforman a la familia Conidae, no obstante se calcula que sólo una pequeña fracción (~2 %) de los conopéptidos han sido identificados (Kaas et al., 2010). Se estima que cada especie de cónido produce de 100 a 200 conopéptidos diferentes muy estructurados y de tamaño pequeño (tienen en promedio de 10 a 80 aminoácidos) (Olivera, 2002). Sin embargo, la elevada sensibilidad de la espectrometría de masas ha revelado que probablemente el número de conotoxinas sea mayor al inicialmente estimado (alrededor de los miles por especie) (Dutertre et al., 2013; Lavergne et al., 2015). Esta gran diversidad de los péptidos que componen su veneno, optimizado durante 33 millones de años de evolución (Duda y Kohn, 2005), se

debe a diferentes razones. En primer lugar, el veneno de los cónidos tiene diferentes funciones: no sólo les permite capturar y digerir a su presa, sino que también es utilizado para combatir a sus depredadores y delimitar su territorio. En segundo lugar, la forma en que los cónidos utilizan su veneno en el momento de atrapar a su presa, consiste en inyectar una determinada combinación de toxinas que tienen como blanco la musculatura y el sistema nervioso central, lo que permite una rápida y contundente inmovilización. Esto quiere decir que el veneno de los caracoles está compuesto por péptidos muy disímiles que actúan sobre distintos blancos moleculares. Por esta razón, el veneno de los caracoles funciona utilizando grupos de péptidos que actúan de manera conjunta y sinérgica para lograr el mismo efecto fisiológico final: paralizar a la presa (Olivera y Cruz, 2001). Se dice que los cónidos son expertos en farmacología porque utilizan una combinación de distintos péptidos, cada uno destinado a interactuar específicamente con un blanco molecular, siendo capaz de diferenciar entre distintos subtipos de un mismo objetivo.

A raíz de esto surge la siguiente pregunta: ¿Cómo son capaces de lograr esta diversidad de toxinas tan amplia? En primer lugar, las toxinas que conforman el veneno de los cónidos se caracterizan por la presencia de una alta tasa de modificaciones postraduccionales (MPT). Hasta la fecha se han identificado 14 diferentes tipos de MPT en las toxinas maduras de los cónidos (Dutertre et al., 2013), que permiten generar distintas conotoxinas a partir de un mismo transcrito. Se estima que alrededor del 10 % de los conopéptidos contienen residuos de γ -carboxiglutamato que se cree facilitan la formación de α -hélices (Bandyopadhyay et al., 1998; 2002). La hidroxilación de los residuos de prolina también suele ser común, así como la amidación del grupo carboxilo terminal (Gayler et al., 2005). Sin embargo, la MPT más frecuente y de relevancia es la formación de enlaces disulfuros: de los 234 conopéptidos aislados a nivel de proteína y registrados en la base de datos de conotoxinas de ConoServer (Kaas et al., 2010), 220 presentan esta modificación. Las funciones de la mayoría de las MPT aún no están claras; sin embargo, es muy probable que facilite el plegamiento e incremente la afinidad hacia su blanco molecular. El procesamiento de diferentes MPT explica parcialmente la complejidad del veneno, ya que la mayoría de los conopéptidos presentan en promedio dos posiciones modificadas (sin tomar en cuenta los enlaces disulfuro), esto generaría alrededor de 400 péptidos a partir de 100 genes (Dutertre et al., 2013). Por esta razón, Dutertre y colaboradores (2013) proponen el mecanismo del Procesamiento Variable del Péptido (VPP, por sus siglas en inglés) donde los conopéptidos pueden ser procesados de manera diferente tanto en sus extremos amino y carboxilo. Estas posibles variantes corresponden a procesamientos alternativos en los sitios de corte arginina (R) o lisina (K), que se presentan en la secuencia.

1.1.3 El aparato venenoso

En los animales venenosos, la biosíntesis del veneno y su liberación es lograda gracias a un vasto rango de estructuras y mecanismos, pero comúnmente incluyen una glándula venenosa para la síntesis, almacenamiento y procesamiento del veneno, así como de un aparato venenoso especializado que variará dependiendo del tipo de dieta que tenga la especie. La forma de la glándula venenosa es muy heterogénea dentro de los animales venenosos; en el caso de los cónidos, tiene una estructura de ducto, por lo que muchas veces a la glándula se le llama ducto venenoso (Rendón-Anaya et al., 2013).

Las conotoxinas son sintetizadas en las células secretoras epiteliales del ducto venenoso, un órgano que puede llegar a medir cuatro veces el largo del cuerpo del caracol (Safavi-Hemami et al., 2014). El ducto venenoso tubular desemboca en la faringe del organismo que está situada entre la probóscide y el esófago (Olivera, 2002). En el extremo opuesto de la faringe se encuentra el bulbo venenoso, el cual es una estructura de tejido muscular desarrollado (Norton y Olivera, 2006). Los dientes de los cónidos distan mucho de parecerse a las rádulas del resto de los gasterópodos: tienen forma de arpón hueco, están constituidas por quitina, tienen la función de inyectar el veneno y son desechadas en cada ataque. Los dientes se producen en el saco radular, que se divide en dos compartimientos, el brazo largo que sintetiza las rádulas y el brazo corto que contiene las rádulas que están listas para ser cargadas con el veneno; además el tipo de arpón es característico de cada especie (Figura 1).

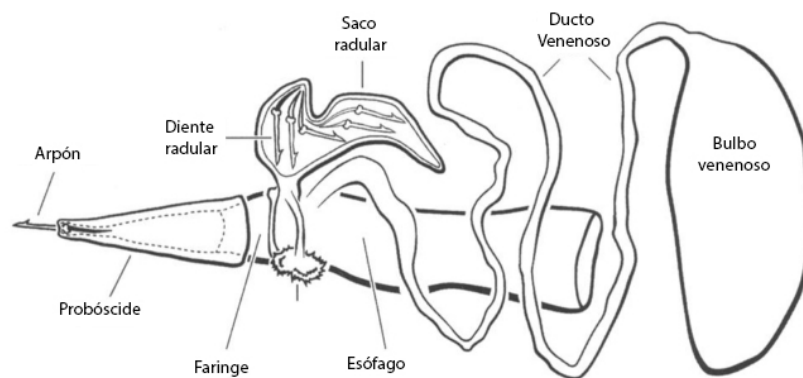


Figura 1. Aparato venenoso de los cónidos. Modificado de Olivera, 2002

La composición de las células epiteliales, la ultraestructura y el contenido granular varían entre las porciones proximales y distales de la glándula, con los cambios morfológicos más prominentes en la parte cercana a la faringe (Marshall et al., 2002). Estas diferencias morfológicas se cree que se ven reflejadas en la síntesis, procesamiento, empaquetamiento y secreción del veneno. Las células epiteliales sintetizan

cientos o miles de conotoxinas y las liberan hacia el lumen del ducto. Estudios morfológicos han indicado que el veneno es liberado de las células epiteliales mediante la ruptura de la membrana celular (Marshall et al., 2002). El movimiento peristáltico de las células musculares de la glándula y la contracción del bulbo muscular son las fuerzas que impulsan el veneno hacia la faringe donde es cargado en el diente radular para después ser inyectado en la presa.

1.1.4 Clasificación de los componentes del veneno de cónidos

Existen tres tipos de clasificaciones para describir diferentes aspectos de los conopéptidos (Figura 2): (1) el esquema de clasificación de la “superfamilia de genes” se centra en las relaciones evolutivas entre los conopéptidos, (2) el esquema de “marcos de cisteína” los clasifica de acuerdo a sus arreglos de cisteínas y (3) el esquema de “familias farmacológicas”, refleja la especificidad del blanco sobre el que cada conopéptido actúa (Kaas et al., 2010). A continuación, se explican más detalladamente:

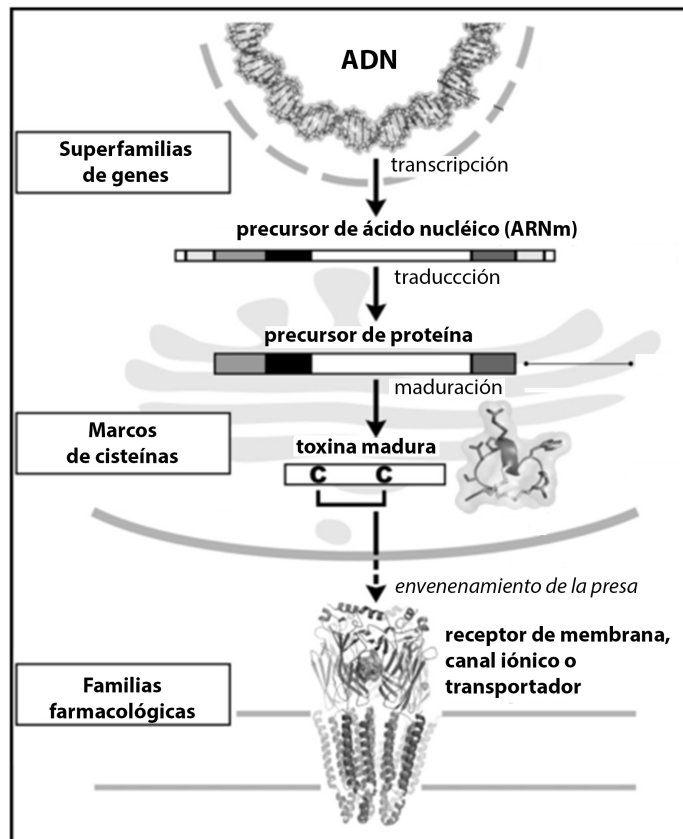


Figura 2. Maduración del conopéptido y diferentes maneras en que puede ser clasificado. Modificado de Kaas et al., 2010.

- **Superfamilias de genes**

Cada uno de los genes de conopéptidos es transcrito como ARN mensajero (ARNm) que a su vez es traducido como un precursor prepropeptídico con una longitud de 70 a 120 aminoácidos, y se caracteriza por un gradiente de divergencia del amino al carboxilo terminal. Es decir, los conopéptidos están integrados por una región precursora que consiste en una secuencia señal muy conservada, seguida por una región propeptídica menos conservada. La región pre le permite al péptido dirigirse al retículo endoplasmático (RE) y la región pro se cree que juega un papel importante en su plegamiento correcto; estas dos regiones son escindidas en el aparato de Golgi para dar lugar a una toxina madura hipervariable muy divergente (con excepción de unos cuantos residuos claves de cisteína que son conservados entre ciertos grupos) (Violette et al., 2012).

- **Familias farmacológicas**

Cuando la toxina madura adquiere su estructura tridimensional, entonces es capaz de actuar sobre los receptores de su presa. Los conopéptidos pueden clasificarse en familias dependiendo del blanco molecular sobre el que estén actuando. Estas familias se designan con letras griegas, por ejemplo aquellas conotoxinas que bloquean receptores nicotínicos de acetilcolina (nAChR) pertenecen a la familia alfa (α); aquellas que bloquean canales de sodio dependientes de voltaje pertenecen a la familia mu (μ) y mientras que aquellos que son agonistas de estos canales (realizan una inactivación tardía) pertenecen a la familia delta (δ). De acuerdo a Kaas y colaboradores (2010) se tienen reportadas 11 familias. Conocer la familia de un conopéptido es una labor difícil, por lo que muy pocas secuencias llegan a tener este tipo de clasificación.

Se cree que los conopéptidos emergieron de la diversificación de unas cuantas superfamilias de genes, cuyos miembros son identificados gracias a la conservación de su secuencia señal (Olivera, 2002). Esta secuencia señal permite clasificar a las conotoxinas en superfamilias empíricas de genes que son designadas con una letra mayúscula y en ciertos casos acompañadas con un número arábigo. Las superfamilias de genes de las que más conopeptidos existen reportados son: A, D, I1, I2, J, M, O1, O2, O3 y T. Las superfamilias han ido creciendo en la misma medida en que se profundiza el estudio de las conotoxinas, y en los últimos dos años se ha llegado a proponer 10 nuevas superfamilias (Lavergne et al., 2013) (Tabla 17, Anexo 1).

- **Marcos de cisteína**

Las conotoxinas maduras tienen residuos de cisteínas en una frecuencia sumamente elevada (se han reportado toxinas con 10 cisteínas) y pueden estar separadas por 0 a 6 aminoácidos. El marco de cisteínas de una conotoxina hace referencia al arreglo que presenten dichos residuos a lo largo de su estructura primaria, independientemente de la conectividad que exista entre los enlaces disulfuro. Hasta la fecha existen descritos 26 marcos de cisteína distintos los cuáles son asignados con números romanos (Tabla 1), sin embargo existen más marcos de cisteína a los que no se les ha designado un nombre (Lavergne et al., 2015; Phuong et al., 2016). Aunque no existe una correlación como tal entre la superfamilia de genes y su función biológica (Kaas et al., 2010); si se ha observado que las superfamilia de genes y los marcos de cisteína puede ser útiles predictores de la función biológica.

Tabla 1. Marcos de cisteína que han sido descritos. Los guiones entre cisteínas (C) son indicadores de la existencia de uno o varios aminoácidos intermedios. [PO] indica la posición que puede estar ocupada por una prolina (P) o una hidroxiprolina (PO). Modificado de Akondi et al, 2013.

Marco de cisteína	Patrón de cisteínas	No. de cisteínas	Superfamilia de genes
I	CC-C-C	4	A, M, O1, T
II	CCC-C-C-C	6	A, M
III	CC-C-C-CC	6	M
IV	CC-C-C-C-C	6	A, M
V	CC-CC	4	T
VI/VII	C-C-CC-C-C	6	H, I1, I3, M, O1, O2, O3
VIII	C-C-C-C-C-C-C-C-C-C	10	B2, S
IX	C-C-C-C-C-C	6	M, P
X	CC-C.[PO]C	4	T
XI	C-C-CC-CC-C-C	8	I1, I2, I3
XII	C-C-C-C-CC-C-C	8	I2, O1
XIII	C-C-C-CC-C-C-C	8	G
XIV	C-C-C-C	4	A, I2, J, L, M, O1, O2
XV	C-C-CC-C-C-C-C	8	D, N, O2, V
XVI	C-C-CC	4	M, T
XVII	C-C-CC-C-CC-C	8	Y
XVIII	C-C-CC-CC	6	
XIX	C-C-C-CCC-C-C-C-C	10	
XX	C-CC-C-CC-C-C-C-C	10	D
XXI	CC-C-C-C-CC-C-C-C	10	
XXII	C-C-C-C-C-C-C-C	8	E
XXIII	C-C-C-CC-C	6	K
XXIV	C-CC-C	4	B3
XXV	C-C-C-C-CC	6	
XXVI	C-C-C-C-CC-CC	8	

1.1.5 Importancia del estudio de las conotoxinas

Determinar cuáles son los conopeptidos que componen el veneno de cada uno de los cónidos es interesante desde el punto de vista tanto biológico como biomédico-biotecnológico (Lluisma et al., 2012). Por un lado, los conopeptidos reflejan el nicho ecológico del cónido; la identificación y enumeración de los componentes del veneno resultan una “lectura molecular” de las interacciones bióticas de cada especie (Olivera, 2002). Esta información a nivel molecular provee la pauta para entender la evolución de las toxinas, así como de los mecanismos moleculares que generan su diversidad. Por otro lado, los componentes peptídicos de los caracoles venenosos son de sumo interés en el campo del descubrimiento y desarrollo de nuevos fármacos. Las conotoxinas han sido estudiadas como posibles fármacos debido a que tienen un tamaño pequeño, son sumamente estables y son selectivas con sus blancos moleculares. Se ha demostrado que los conopeptidos tienen como blanco un amplio rango de receptores y canales iónicos de importancia terapéutica (Lluisma et al., 2012); además cabe destacar que existe una alta similitud entre los blancos de las conotoxinas y los receptores de mamíferos (Lewis y Garcia, 2003). Una alta proporción de los primeros conopeptidos que fueron caracterizados han alcanzado desarrollo clínico: de los primeros 30 péptidos purificados a partir de veneno de cónidos, el 10 % ha alcanzado por lo menos la fase I en ensayos clínicos (Olivera, 2006). Un ejemplo exitoso es el de Prialt® (nombre genérico Ziconitide), desarrollado en el 2004 por la farmacéutica Elan, y aprobado como agente terapéutico por la FDA (Food and Drug Administration) de los Estados Unidos. Éste fármaco es idéntico a la toxina MVIIA, una ω -conotoxina del caracol *Conus magus* que sirve como un analgésico intratecal para combatir el dolor crónico severo que ya no puede ser mitigado con la morfina (Terlau y Olivera, 2004). Esta conotoxina pertenece a un cónido piscívoro y presenta una alta especificidad por los canales de calcio tipo N (Cav2.2). Otro ejemplo de una conotoxina con resultados prometedores es el de Xen2174, análogo del conopeptido MrIA de *Conus marmoreus* que se encuentra en la fase II de pruebas clínicas. Este péptido tiene como blanco el transportador neuronal de norepinefrina (Lewis et al., 2012). Finalmente, la conotoxina GVIA del caracol *C. geographus* es el conopeptido más utilizado en neurociencia -con más de 2, 000 publicaciones en la literatura- como una herramienta farmacológica (Olivera, 2006). Esto nos demuestra que incluso cuando las conotoxinas no puedan ser desarrolladas directamente como terapéuticos sí pueden ser extremadamente valiosas para identificar moléculas blanco con potencialidades para el descubrimiento de nuevos fármacos.

1.1.6 RNA-seq

La secuenciación de ARN (RNA-seq, por su abreviación en inglés) produce una gran cantidad de información con respecto a los ARNm (y por inferencia de las proteínas) producidas en el tejido de elección. El análisis del transcriptoma es utilizado para descifrar los componentes peptídicos expresados en el veneno de una manera más rápida y eficiente (Fry et al., 2012). En el caso de las especies que no cuentan con un genoma de referencia, como lo son los cónidos, se requieren de estrategias bioinformáticas especializadas con las que se pueda realizar un ensamblaje de los transcritos secuenciados *de novo*.

La primera secuenciación de un ducto venenoso de cónido permitió obtener un panorama global de las conotoxinas que conforman el veneno de *Conus bullatus*, un caracol perteneciente al grupo de los pobremente caracterizados Textiles (Hu et al., 2011). En ese trabajo se pudo conocer que existía una alta cantidad de las conotoxinas de la superfamilia A, las cuales normalmente se encontraban en tasas de expresión muy bajas. Debido a esta capacidad de identificar una plétora de conotoxinas anteriormente desconocidas, el número de estudios enfocados en el descubrimiento de conopeptidos mediante un análisis del transcriptoma de su veneno ha crecido rápidamente, por lo que es posible encontrar más de un artículo por año en revistas académicas. La transcriptómica es una disciplina que nos permite entender cómo un organismo combina sus toxinas en base al ambiente en el que se encuentre (Jin et al., 2014; 2015), dependiendo de la dieta que presente (Phuong et al., 2016) o simplemente elucidar la mayor cantidad de componentes que puede tener un veneno (Robinson et al., 2014, 2017).

1.1.7 Especies de interés para este trabajo

Las especies de interés de este trabajo son *C. ximenes* y *C. mahogani*, presuntamente especies moluscívoras, *C. regularis* posiblemente una especie vermívora y *C. californicus*, especie única de su género y atípica de la Familia porque es, entre otras cosas, capaz de tener una dieta sumamente diversa. Estos cónidos presentan amplia distribución en la península de Baja California y, por ende, el acceso a ellos es relativamente fácil en comparación a otras especies de cónidos. Estas especies han sido utilizadas por el laboratorio de Inmunología Molecular y Biotoxinas de CICESE, por lo que conocer las secuencias que conforman su veneno facilitará los proyectos de dicho laboratorio. Con excepción de *C. californicus*, el veneno de las especies utilizadas en este trabajo ha sido pobremente explorado, y debe de tenerse en cuenta que la mayoría de los estudios de cónidos ha sido sesgado ya que se ha centrado en las especies

que habitan las áreas del Indo-Pacífico (Elliger et al., 2011; Kaas et al., 2010; Neves et al., 2013). Por tal motivo, la información que sea producida en este trabajo es valiosa tanto para análisis comparativos del veneno como para la toxicología de la familia Conidae en general.

1.2 Justificación

A pesar del potencial que los conopéptidos tienen como agentes terapéuticos y como herramientas de investigación, sólo un número muy pequeño ha sido caracterizado en detalle. El flujo de trabajo tradicional para identificar (basado en ensayos de bioactividad) ha limitado el número de conopéptidos que han sido descritos. Las mejoras en el campo de la secuenciación de nueva generación, que han permitido por primera vez secuenciar transcriptomas enteros a un costo razonable, nos permiten descubrir de una manera más rápida y completa cuáles son los componentes que conforman el veneno de los cónidos. Cabe mencionar que esta estrategia introduce nuevos retos: el número de herramientas bioinformáticas disponibles para el estudio de las toxinas es limitado.

1.3 Hipótesis

Los genes que codifican conotoxinas pueden ser elucidados a partir del estudio de los transcriptomas del ducto venenoso.

1.4 Objetivo general

Ensamblar el transcriptoma del ducto venenoso de cuatro especies de cónidos de la península de Baja California para conocer los genes que codifican conotoxinas.

1.4.1 Objetivos particulares

- Secuenciar el transcriptoma del veneno de cuatro especies diferentes de la familia Conidae a partir de extracciones de ARN del ducto venenoso.
- Ensamblar el transcriptoma *de novo*.
- Predecir la mayor cantidad de conotoxinas y agruparlas de acuerdo a superfamilias.

Capítulo 2. Metodología

2.1 Obtención de los especímenes

Los especímenes colectados se distribuyen en dos zonas diferentes de la Península de Baja California. El caracol marino de la especie *C. californicus* se distribuye desde Los Farallones (Sur de California, EUA) hasta Cabo San Lucas (Baja California, México); y el punto de colecta se ubicó en el Km 58 de la carretera Tijuana-Ensenada. Los organismos de las especies *C. ximenes*, *C. regularis* y *C. mahogani* se distribuyen en el lado este de la península y fueron colectados a 544 km al sur de Ensenada, en Bahía de Los Ángeles. Ambas colectas fueron planeadas consultando un calendario de predicción de mareas para que coincidieran con mareas bajas, los organismos fueron colectados en base a las características morfológicas de su especie. La colecta se realizó durante el mes de Enero del 2014, lo que corresponde al período de invierno. Todos los organismos fueron transportados vivos al laboratorio en trapos húmedos con agua de mar, ya que al ser organismos del intermareal pueden sobrevivir sin estar en contacto directo con el mar. Una vez en laboratorio, se colocaron en peceras con agua de mar, a una temperatura promedio de 24 °C y aireación constante.

Debido a que no existe ningún transcriptoma de las especies en cuestión y a que el objetivo de esta tesis era obtener el mayor número de genes codificantes de conotoxinas, podemos establecer la secuenciación del ARN del ducto venenoso de estas especies como condición cero o específica del ambiente; por lo que podrá ser utilizada en trabajos posteriores como punto de comparación.

2.2 Extracción de ARN total de la glándula venenosa y secuenciación de su transcriptoma

La disección de los organismos se llevó a cabo bajo condiciones libres de ARNasas, procesando la totalidad del ducto (desde donde termina el bulbo venenoso hasta el inicio de la faringe, Figura 1). Se realizaron extracciones de la glándula venenosa por duplicado para las especies *C. ximenes*, *C. californicus* y *C. regularis*, mientras que para *C. mahogani* se realizó una sola extracción. Se utilizaron 6 organismos por réplica para las especies *C. regularis*, *C. mahogani* y *C. ximenes*; mientras que para *C. californicus* se procesaron 15 organismos por réplica debido a que presentan un menor tamaño. La extracción del ARN

de los ductos se realizó con ayuda del kit *SV Total RNA Isolation System* de Promega, siguiendo el protocolo que ofrece el fabricante y únicamente aumentando los tiempos de centrifugación al doble de lo que se indica. La primera verificación de la calidad y cantidad de ARN se determinó mediante la razón de la absorbancia a 260/280 nm (1.8 - 2.0), no obstante la integridad se reconfirmó utilizando el Bioanalizador 2100 (Agilent Technologies).

Se construyeron bibliotecas de ADN complementario (ADNc) para cada una de las especies utilizando el kit de preparación de *Illumina TruSeq Stranded mRNA Sample*, siguiendo las instrucciones del manual del fabricante. La secuenciación de ADN se llevó a cabo utilizando la plataforma de secuenciación Genome Analyzer IIX (GAIIIX) de la compañía Illumina. El procedimiento lo realizó la unidad de secuenciación del *Instituto de Biotecnología (IBT)* de la Universidad Autónoma de México (UNAM) en el Campus Morelos. El tipo de secuenciación realizada fue de 72 pb desde los dos extremos de los transcritos (pair-end sequencing, PE), con un tamaño de inserto de 200 a 400 pb.

2.3 Entrega de secuencias

Cada biblioteca consistió de dos archivos en formato FASTQ (R1 y R2, lecturas sentido y antisentido), libres de adaptadores. La unidad de secuenciación del IBT realizó un ensamble de los transcriptomas utilizando una versión anterior (v. 2.0.2) del ensamblador Trinity (Grabherr et al., 2011) a la que se usó posteriormente en este trabajo (v. 2.0.3). En el ensamble del IBT se utilizaron los parámetros predeterminados con la variante de un tamaño mínimo de *contig* de 100 pb (el predeterminado es de 200 pb). Después de ensamblar, hicieron una traducción de las secuencias de nucleótidos y una anotación de los péptidos obtenidos mediante un script propio utilizando la base de datos del NCBI y del SwissProt (2013) (Tabla 2). Dentro de las secuencias que conforman los archivos (archivo.annot.fasta, archivo.annot.pep y archivo.someannot.pep) se buscaron aquellas secuencias que hubieran tenido alguna anotación relacionada con cónidos o con conotoxinas.

Tabla 2. Archivos generados en el IBT después de realizar el ensamble del RNA-seq del transcriptoma de los cuatro ductos venenosos de las especies utilizadas en este trabajo.

Tipo de archivo	Contenido del archivo
archivo.fasta	secuencias de nucleótidos totales
archivo.annot.fasta	secuencias de nucleótidos que tuvieron anotación en el NCBI
archivo.noannot.fasta	secuencias de nucleótidos que no tuvieron anotación en el NCBI
archivo.pep	secuencias de aminoácidos que presentan un marco de lectura abierto
archivo.annot.pep	secuencias de aminoácidos con anotación en la base de datos de SwissProt
archivo.someannot.pep	secuencias de aminoácidos sin anotación en la base de datos de SwissProt pero que presentaron un dominio en Pfam. Tienen un péptido señal predicho o una región transmembranal
archivo.noannot.pep	Secuencias de aminoácidos con un marco de lectura abierto que no presentaron anotación

2.4 Ensamble del transcriptoma empleando el programa de cómputo Trinity

v.2.0.3 (Grabherr et al., 2011)

En primer lugar se obtuvo la calidad de las lecturas crudas utilizando el programa FastQC (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>). Como se mencionó anteriormente, no existe un genoma de referencia disponible para la familia Conidae por lo que las lecturas cortas generadas por el secuenciador (archivos fastq R1 y R2) se ensamblaron *de novo* utilizando el software Trinity (v. 2.0.3). Después de varias pruebas se decidió utilizar el protocolo estándar (Grabherr et al., 2011), ejecutando el parámetro específico de cadena, realizando una normalización *in silico* de las lecturas y utilizando un *contig* mínimo de 100. El ensamble del transcriptoma se llevó a cabo utilizando los siguientes parámetros: *Trinity --seqType fq --max_memory 60G --min_contig_lenght 100 --normalize_max_read_cov 30 --left R1.fastq --right R2.fastq --SS_lib_type RF --CPU 27 --no_bowtie*. Con el propósito de evaluar la calidad del ensamble, se ejecutó el script TrinityStats.pl sobre el archivo fasta generado. Este paso permitió determinar la estadística del número de genes e isoformas así como de la contigüidad del ensamble. Finalmente, utilizando el script Transdecoder (incluido en Trinity) se identificaron los marcos de lectura abiertos (ORF por sus siglas en inglés) mayores a 50 aminoácidos. Después de eliminar las secuencias redundantes se utilizaron las secuencias para crear las bases de datos de cada especie.

2.5 Búsqueda de conotoxinas con el programa de cómputo ConoSorter (Lavergne et al., 2013)

Se ejecutó el programa ConoSorter (Lavergne et al., 2013) en la terminal utilizando el comando que permite trabajar con nucleótidos. Se utilizaron los *contigs* ensamblados por Trinity. Los dos archivos que arroja el programa (RegEx y pHMM) se visualizaron en Excel. La cantidad de secuencias en estos archivos rondaba hasta el medio millón de secuencias por especie, la cuál es mucho mayor a la esperada (100 a 200 conotoxinas). Se siguieron las siguientes pautas para eliminar aquellas secuencias que no correspondían a conopéptidos:

1. Se eliminaron secuencias menores a 50 aminoácidos.
2. Se eliminaron aquellas secuencias que en el archivo Excel estaban clasificadas con una señal de hidrofobicidad menor a 50 %.
3. Se acomodaron los datos por superfamilia (señal, región pro y toxina madura) y se dejaron sólo aquellas secuencias que tuvieran un puntaje de 2.
4. Se eliminaron todas aquellas secuencias mayores a 150 aminoácidos.
5. Se agruparon las secuencias de acuerdo a la similitud que conservaba el péptido señal.

2.6 Construcción de perfiles de Modelos Ocultos de Markov

Se contruyeron perfiles de Modelos Ocultos de Markov (pHMM, por sus siglās en inglés) para cada superfamilia de genes de conotoxinas reportadas en ConoServer. Se siguió una metodología similar a la propuesta por Peng et al, 2016 y Robinson et al, 2014b, la cuál está explicada en el Anexo 1.

Capítulo 3. Resultados

3.1 Extracción de ARN total

En la 3 se pueden observar los resultados obtenidos de las extracciones de ARN: los organismos procesados, la verificación de la cantidad y calidad mediante la razón de la absorbancia 260/280; así como las concentraciones obtenidas con el Bioanalizador 2100 (Agilent Technologies). Como era la primera vez que se realizaba un trabajo de este tipo en el laboratorio era necesario realizar el procedimiento de extracción por duplicado para asegurarnos que la cantidad y la calidad obtenida fueran las adecuadas para la secuenciación. Se escogió la mejor muestra de cada especie para verificar su calidad por el Bioanalizador 2100 y todas las muestras fueron guardadas a -70 °C hasta su uso.

Tabla 3. Concentración y calidad de cada extracción de ARN total.

Especie	Organismos procesados	[] ng/μL	A ₂₆₀ / A ₂₈₀	[] en Bioanalizador
<i>C. ximenes</i>	6	48.09	2.05	44.0 ng/μL
	6	14.63	1.85	No analizada
<i>C. regularis</i>	6	126.83	1.93	94.0 ng/μL
	6	324.88	2.02	No analizada
<i>C. mahogani</i>	6	177.09	2.14	168.0 ng/μL
<i>C. californicus</i>	15	564.20	2.05	378 ng/μL
	15	1,108.45	2.16	No analizada

El ARN total (ARN_T) se compone de ARN ribosomal (ARN_r), ARN de transferencia (ARN_t) y ARN mensajero (ARN_m). El ARN_r representa un 90 % del total de la muestra, por lo que es este tipo de ARN el que nos permite inferir la calidad de la extracción de ARN_m. Los electroferogramas obtenidos con el Bioanalizador 2100 de las muestras con mejor calidad de ARN (Figura 3) mostraron una ausencia del pico que corresponde al 28S ribosomal, así como la presencia de un pico abundante y bien definido a la altura del 18S. Este electroferograma puede ser traducido a una imagen tipo gel de agarosa en el que se puede observar una banda definida a la altura que corresponde al 28S. La ausencia del 18S impidió el cálculo del número RIN (RNA Integrity Number), el cuál es el cociente entre la señal obtenida del 28S y del 18S y es

utilizado como un indicador de la calidad del ARN. No obstante, la ausencia de ruido y la presencia de picos limpios y claros permitieron determinar que la calidad del ARN era la adecuada para ser secuenciado.

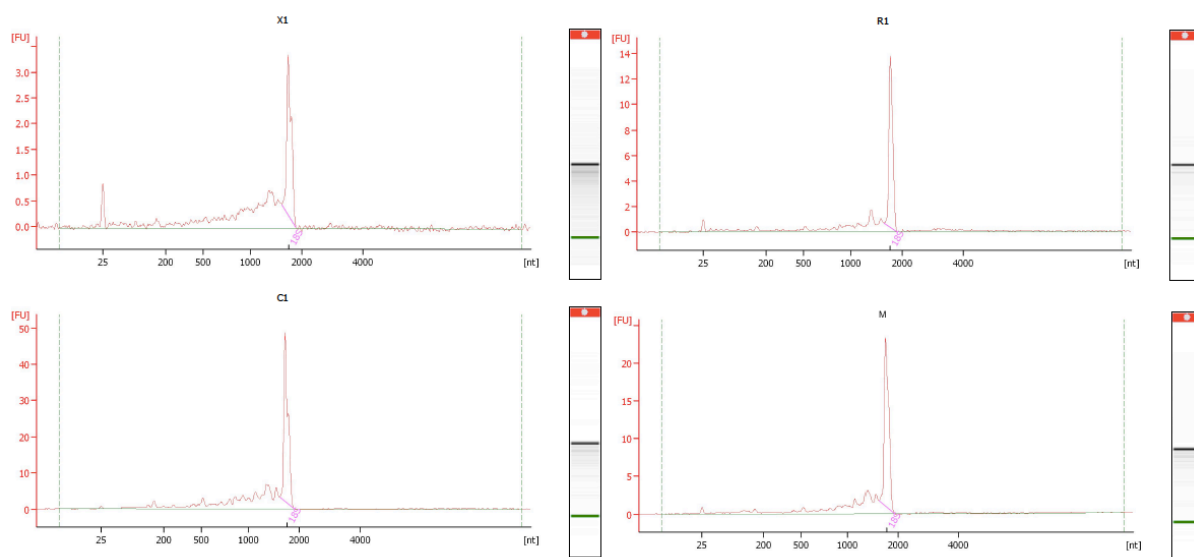


Figura 3. Electroferograma y representación en gel de agarosa de las cuatro extracciones de ARN utilizadas para la construcción de las bibliotecas. *C. ximenes* (X1), *C. regularis* (R1), *C. californicus* (C1) y *C. mahogani* (M).

3.2 Análisis de la calidad de las secuencias

La secuenciación de ARN produjo un vasto número de lecturas para cada una de las bibliotecas, este número rondaba alrededor de los 50 millones (42'340,967 para *C. californicus*; 44'590,225 para *C. mahogani*; 50'360,289 para *C. regularis*; y 56'363,079 para *C. ximenes*), el cual es el esperado para este tipo de experimentos (Robinson et al., 2017). La calidad de estas lecturas fue analizada utilizando el programa FastQC. Este programa califica a cada una de las pruebas de análisis de calidad de las secuencias: en buenas; precautorias, e incorrectas. Las tres gráficas que se utilizaron para definir la calidad de las secuencias se describen a continuación, tomando sólo una secuenciación como ejemplo. El resto de las gráficas pueden ser consultadas en el Anexo 3 del formato digital de esta tesis.

3.2.1 Calidad de las secuencias por base

En este tipo de gráficas se representa la calidad promedio por cada base secuenciada. En el eje de abscisas se representa la posición de cada base, en este caso es de 72 pb secuenciadas. Por cada posición se representa un diagrama de caja y bigotes. En dicho diagrama la línea central roja es el valor de la mediana. La caja amarilla representa el valor del rango del intercuartil (la dispersión estadística del 25-75 %). Los bigotes superiores e inferiores representan los puntos 10 y 90 %. La línea azul que atraviesa toda la gráfica representa la media. El eje de ordenadas representa la calidad de la secuencia en formato Phred, entre más alto sea, mejor será la calidad de la base. Este eje se divide en tres colores: verde si la calidad es muy buena, naranja si la calidad es lo suficientemente buena y rojo si la calidad es sumamente baja. La calidad de la mayoría de las plataformas decae conforme la corrida avanza, por lo que es común ver bases en la zona naranja hacia el final de la lectura. Para las cuatro secuenciaciones la calidad de las secuencias por base fue la adecuada, cayendo siempre dentro del color verde. Únicamente se muestran los resultados obtenidos para las secuencias de *C. ximenes* (Figura 4). Cabe mencionar que en todas las bibliotecas se vio una pérdida de calidad al final de la secuenciación para el cebador R1, esto es probablemente debido a un problema con el cebador y no con la secuenciación como tal.

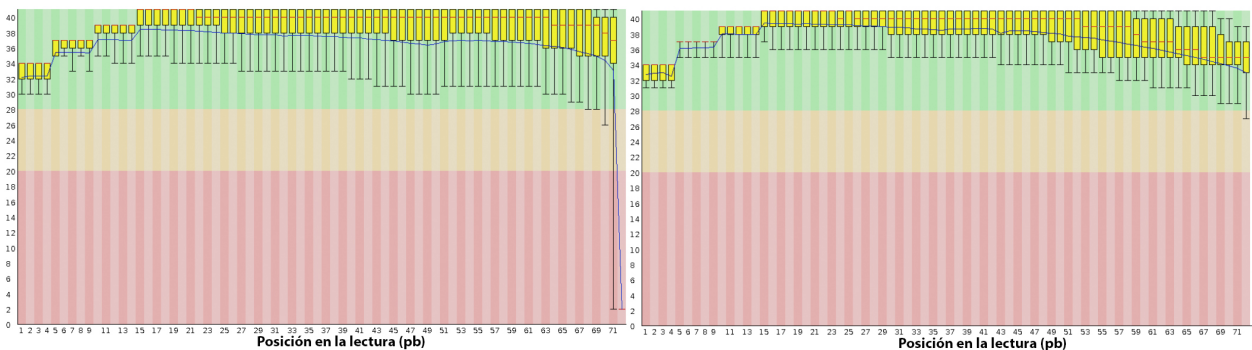


Figura 4. Gráfica de calidad por base para la secuenciación con el cebador R1 (izquierda) y con el cebador R2 (derecha) de *C. ximenes*.

3.2.2 Contenido de bases por secuencia

Este tipo de gráfica muestra la proporción de cada base dentro de la secuencia. En una biblioteca realizada al azar se espera que no exista diferencia o que la diferencia sea mínima entre la proporción de las bases de una corrida, por lo que las líneas de las bases deben de correr de manera paralela entre ellas. Las líneas

no deben de estar desbalanceadas entre ellas, si la diferencia entre ellas es más del 20 % entonces existe una falla en la secuenciación. No obstante, debido al uso de hexámeros al azar para la cebación en la secuenciación, en las bibliotecas de ARN-seq es común encontrar problemas dentro de las primeras 12 pb (Figura 5). No es un problema que pueda ser solucionado con procesamiento, pero parece ser que no afecta de manera negativa los análisis posteriores. En este caso sólo se muestra la gráfica para el primer cebador de *C. regularis* (Figura 5), pero el resto de las gráficas fueron muy similares.

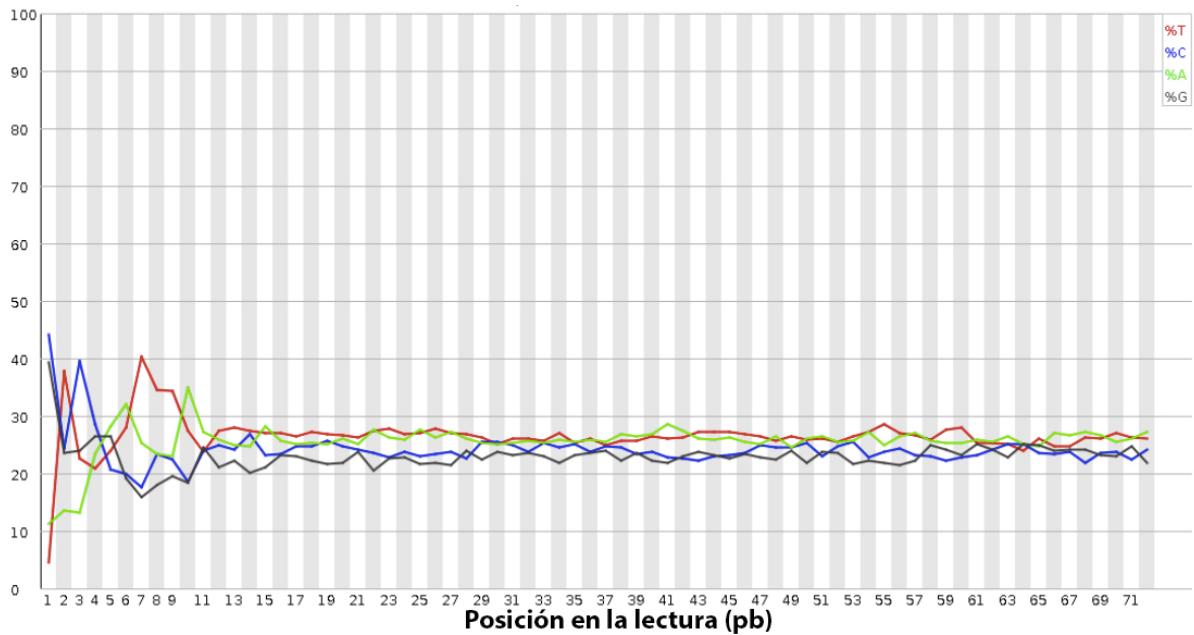


Figura 5. Contenido de bases a lo largo de la secuencia correspondiente al cebador R1 de *C. regularis*.

3.2.3 Contenido GC: distribución GC a lo largo de las secuencias

Este tipo de gráficas miden el contenido GC a lo largo de la secuencia (Figura 6, línea roja) y lo compara con un modelo de distribución normal de GC (Figura 6, línea azul), buscando que sean lo más similares posibles. Para el caso del cebador R1 las cuatro bibliotecas presentaron una distribución del contenido GC muy similar a la distribución normal; no obstante, en el caso del cebador R2 se pudieron observar comportamientos no del todo ideales (valles dentro de la gráfica) que lo más probable es que se deban a un error que ocasionó el cebador y no a la secuenciación en sí (como es el caso de las gráficas de la calidad de las secuencias). Se muestra la gráfica del cebador R1 para la especie *C. californicus* (Figura 6). Las gráficas de las muestras se acercaban a una distribución normal y por lo tanto fueron consideradas como óptimas en la secuenciación.

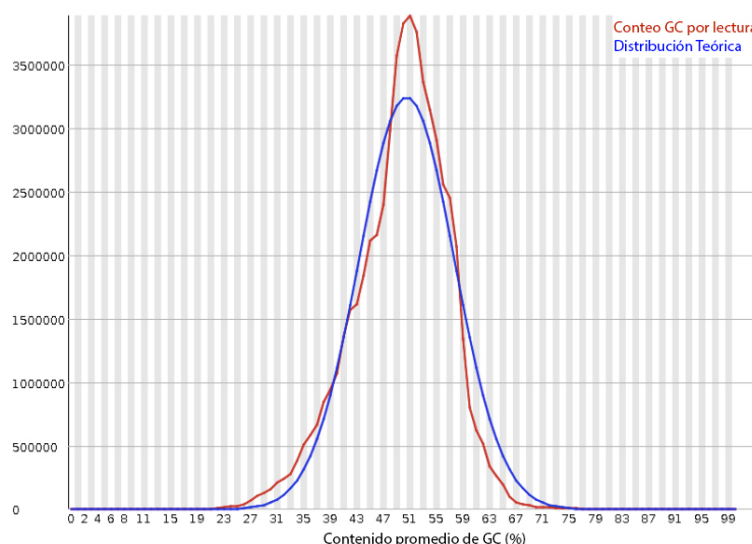


Figura 6. Distribución GC de la secuenciación con el cebador R1 para la especie *C. californicus*.

3.3 Análisis del BLAST del ensamble del IBT

La unidad de secuenciación del IBT nos entregó una serie de archivos (Sección 2.3) cuyo número de secuencias se ve reflejado en la Tabla 4.

Tabla 4. Total de secuencias obtenidas para cada uno de los archivos entregados por el IBT. La extensión *.fasta hace referencia a secuencias de nucleótidos, mientras que la extensión *.pep corresponden a secuencias de aminoácidos.

Tipo de archivo	<i>C. californicus</i>	<i>C. regularis</i>	<i>C. ximenes</i>	<i>C. mahogani</i>
archivo.fasta	85, 381	144, 030	190, 240	183, 168
archivo.annot.fasta	7, 544	15,487	19, 615	19, 254
archivo.noannot.fasta	72, 214	115, 847	155, 179	148, 412
archivo.pep	11, 723	26, 204	31, 472	33, 474
archivo.annot.pep	5, 192	10, 321	11, 853	14,149
archivo.someannot.pep	1, 225	2, 144	2, 499	2,522
archivo.noannot.pep	4, 968	12, 569	15, 665	15, 434

Dentro de las secuencias que tuvieron anotación (archivo.annot.fasta, archivo.annot.pep y archivo.someannot.pep) se buscaron tanto las conotoxinas como aquellas secuencias que hubieran estado

relacionadas con la familia Conidae (enzimas, conopéptidos sin cisteínas) (Tabla 5). Se encontraron mucho menos secuencias de las esperadas, sobre todo para las especies de las que no existe ningún reporte en la literatura (*C. ximenes*, *C. regularis* y *C. mahogani*).

Tabla 5. Total de secuencias obtenidas para los archivos de aminoácidos que tuvieron anotación. Para *C. californicus* existen secuencias de conotoxinas repetidas.

Tipo de archivo	Búsqueda	<i>C. californicus</i>	<i>C. regularis</i>	<i>C. ximenes</i>	<i>C. mahogani</i>
Archivo.annot.pep	Conotoxina	92	15	7	4
	Conidae	15	13	10	9
Archivo.someannot.pep	Conotoxina	2	0	0	1
	Conidae	0	0	0	0

Como se puede observar en la Tabla 5 la cantidad de conotoxinas disminuye drásticamente entre *C. californicus* y el resto de los organismos. Esto posiblemente se debe a que *C. californicus* es una especie que ha sido sumamente estudiada y de la que existen varias secuencias reportadas, ya que existen bibliotecas de ADNc del veneno (Biggs et al., 2010; Elliger et al., 2011). La utilización de un BLAST (Basic Local Alignment Search Tool, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) no es suficiente para reconocer todas las conotoxinas presentes en el ensamble.

3.4 Resultados obtenidos del ensamble con el programa de cómputo Trinity

v.2.0.3

Debido al número reducido de conotoxinas encontradas en los ensambles entregados por el IBT (Tabla 5), se decidió realizar nuevamente el ensamble, utilizando el servidor Ixachi de CICESE, cuyas estadísticas se muestran en la Tabla 6. Se realizaron varias pruebas de ensambles, y aunque las estadísticas resultaban mejores cuando se utilizaba un *contig* de 200 (Sección 3.1, Parte II), se decidió trabajar utilizando un *contig* de 100 porque al comparar los ensambles se identificaban un mayor número de conotoxinas en el que utilizaba un *contig* de 100. En la Tabla 6 podemos ver que el N50 dista mucho del obtenido al realizar un ensamble con un *contig* de 200 (Tabla 11), pero sí podemos observar que se mantiene constante entre las

diferentes especies. En el caso de los transcriptomas, a diferencia del ensamble de un genoma, maximizar el N50 no es garantía de un buen ensamble. En estos casos el N50 debe de ser utilizado como un índice entre especies que están relacionadas. Por lo tanto, para el ensamble de transcriptomas de la misma familia (Conidae), puede ser utilizado como un indicador de que el ensamble fue exitoso.

Tabla 6. Resultados de las estadísticas obtenidas de los 4 ensambles realizados con Trinity.

	<i>C. californicus</i>	<i>C. mahogani</i>	<i>C. regularis</i>	<i>C. ximenes</i>
Transcritos totales	92,675	95,212	74,035	15,005
Contig promedio	71.15	76.30	76.46	165.85
N10	151	212	210	470
N50	66	71	71	156

3.5 Resultados obtenidos con el programa de cómputo ConoSorter (Lavergne et al., 2013)

ConoSorter (Lavergne et al., 2013) es un software que traduce las secuencias de ADNc en los 6 marcos de lecturas (aunque también puede trabajar con secuencias de proteínas) y, con el objeto de obtener precursores de proteínas completos, recorta aquellas secuencias de aminoácidos que están delimitadas por un codón de metionina y un codón de *stop*, desechando el resto. Estas secuencias son sometidas a una búsqueda booleana para encontrar regiones pre, pro y maduras que tengan una alta identidad con superfamilias de conopéptidos establecidas. De esta manera las secuencias son clasificadas en dos archivos: 1) el archivo RegEx contiene las secuencias de conotoxinas que fueron identificadas mediante una búsqueda de expresiones regulares 2) el archivo pHMM, como su nombre lo indica, contiene las secuencias que no fueron alineadas y que por lo tanto son sometidas a un análisis utilizando pHMM. A las secuencias del archivo RegEx se les asigna un puntaje en base a cada una de las tres regiones de la conotoxinas. El puntaje es de 1 (si se encontró la región) o de 0 (si no se encontró), por lo que el máximo puntaje por secuencia es de 3 (en dado caso de que se encontrara una región pre, pro y madura similar a alguna ya establecida). Por otro lado, el grupo de secuencias que no tuvieron acierto es sometido a una búsqueda estocástica más flexible que utiliza pHMM, con un e-value de 10, para identificar secuencias que pudieran estar lejanamente relacionadas con grupos de conopéptidos que ya estén bien establecidos. El total de puntaje de las secuencias es calculado como la sumatoria del e-value de las tres regiones de

manera independiente. Además, ConoSorter cataloga a los péptidos en base a características claves de la secuencia (frecuencia relativa, longitud, número de cisteínas, hidrofobicidad del N-terminal, puntaje de similitud) y automáticamente busca en la base de datos de cónidos llamada ConoServer (Kaas et al., 2010) de acuerdo a precursores de secuencias ya conocidos, facilitando así la identificación de conopéptidos.

Al ejecutar el programa ConoSorter se obtuvieron dos archivos para cada especie: uno que correspondía a la búsqueda booleana de secuencias (Regex) y otro que correspondía a la búsqueda utilizando pHMM. Después de revisar estos archivos (ver en Tabla 7 “primer filtro”) empleando los criterios enunciados en la sección 2.5 seleccionamos una menor cantidad de secuencias. Como era de esperarse *C. californicus* fue la única especie en la que el programa pudo identificar conotoxinas ya reportadas. No obstante las secuencias obtenidas para cada transcriptoma eran muy difíciles de clasificar ya que muy pocas secuencias obtuvieron un puntaje total de 3. Además, las características de las secuencias obtenidas con el ConoSorter resultaban muy disímiles a las esperadas para las conotoxinas, existían secuencias sumamente largas o los péptidos señales no eran posibles de agruparse. Por lo tanto, se decidió desechar estos resultados y construir pHMM propios para analizar nuestros transcriptomas.

Tabla 7. Número de secuencias encontradas para los archivos Regex y los archivos pHMM de cada uno de los ensamblajes.

Especie	Análisis	Regex	pHMM	Total
<i>C. ximenes</i>	Total de secuencias	65, 536	3, 309	68, 845
	Primer filtro	105	333	438
	Segundo filtro (150 a.a.)	42	329	371
<i>C. regularis</i>	Total de secuencias	362, 380	2, 441	364, 821
	Primer filtro	398	262	660
	Segundo filtro (150 a.a.)	209	253	462
<i>C. mahogani</i>	Total de secuencias	446, 765	2, 905	449, 670
	Primera limpieza	509	285	794
	Segunda limpieza (150 a.a.)	269	279	548
<i>C. californicus</i>	Total de secuencias	169, 214	1, 205	170, 419
	Primera limpieza	215	195	410
	Segunda limpieza (150 a.a.)	148	191	339

3.6 Resultados obtenidos con los pHMM

Una vez obtenidas las secuencias del ensamble se eliminaron todas aquellas secuencias que estuvieran repetidas utilizando el programa fastaRD (creado por M.C. Jesús Armando Beltrán Verdugo) (Tabla 8).

Tabla 8. Secuencias obtenidas en el ensamble y secuencias obtenidas después de eliminar las redundantes para cada especie.

Especie	N.º de secuencias del ensamble	N.º de secuencias sin repetir
<i>C. californicus</i>	92, 700	86, 197
<i>C. mahogani</i>	239, 595	215, 349
<i>C. regularis</i>	178, 562	163, 872
<i>C. ximenes</i>	27, 747	24, 284

Además se conservaron únicamente aquellas secuencias que comenzaran con una metionina y a las que se les pudiera identificar un péptido señal. Posteriormente los pHMM nos permitieron identificar las secuencias de conotoxinas encontradas para cada especie, y se decidió compararlas con las obtenidas anteriormente mediante los otros métodos mencionados (Tabla 9). Se muestra un número aproximado porque cada transcriptoma debe de ser analizado de manera individual para poder obtener el número real. Sólo en el caso de *C. regularis* se muestra el número definitivo ya que fue la única especie de la que se realizó un análisis completo.

Tabla 9. Comparación de secuencias encontradas con diferentes metodologías.

Especie	BLAST	ConoSorter	ConoSorter filtradas	pHMM	Análisis final de los pHMM
<i>C. ximenes</i>	7	441, 773	419	20 (aprox)	-
<i>C. regularis</i>	15	305, 594	317	70 (aprox)	57
<i>C. mahogani</i>	5	377, 825	395	100 (aprox)	-
<i>C. californicus</i>	92*	66, 516	286	280 (aprox)	-

*Este número contiene secuencias repetidas.

Un problema con el que nos enfrentamos fue cómo nombrar las conotoxinas de manera adecuada. Es complicado tratar de ser riguroso con un sistema de nomenclatura impuesto a algo que se encuentra en la naturaleza. De acuerdo a lo establecido (Walker et al., 1999) la secuencia encontrada recibirá nombres distintos si se identifica a nivel de proteína que si se identifica a nivel de cDNA. En ambos casos, si son piscívoros recibirán una letra para designar la especie de cónido y dos letras para designar si la secuencia pertenece a otro tipo de cónido (moluscívoro o vermívoro). Aquellas secuencias encontradas a nivel de cDNA utilizarán letras mayúsculas y aquellas que sean identificadas a nivel de proteína utilizarán letras minúsculas. En las dos situaciones se utilizará un número arábigo para designar el marco de cisteínas; no obstante si pertenece a una secuencia de ADN estará seguida de un número separado por un punto indicando el orden de identificación, y si fuese identificado a nivel de proteína estará seguido por una letra empezando por la letra “a” (sin utilizar el punto). Un ejemplo de secuencia encontrada a nivel de ADN sería P5.1, mientras que es encontrada a nivel de proteína sería p5a. Por último Robinson y colaboradores (2014) sugieren que se incorpore al principio del nombre la superfamilia a la que pertenece la conotoxina.

Ahora bien, esto es en el caso de que la conotoxina presente enlaces disulfuro. La nomenclatura no está clara para aquellos conopéptidos que no presenten marcos de cisteína. Por lo tanto, surgen las siguientes dudas: ¿cómo se nombra aquellos conopéptidos que carecen de cisteínas? ¿cómo se nombran los conopéptidos con sólo dos cisteínas? ¿cómo se nombran aquellos de los que no existe reportado el marco de cisteínas? ¿cómo nombrar secuencias con cisteínas impares de las que cada vez vemos más casos presentes? ¿qué sucede con aquellas secuencias a las que no se les puede asignar una superfamilia? Finalmente, cabe destacar que para el caso de *C. californicus* resulta todavía más complicado ya que las dos bibliotecas de ADNc construidas para esta especie (Biggs et al., 2010; Elliger et al., 2011) fueron nombradas sin considerarse la una con la otra. Por lo tanto, existen reportadas las mismas secuencias con nombres diferentes.

3.7 Análisis del transcriptoma de *C. regularis*

El programa de pHMM entregó un archivo con 106 posibles secuencias de conotoxinas para *C. regularis*, utilizando un valor de e-value de 0.011 para el umbral de corte. El mismo programa identificó la secuencia señal de cada conotoxina utilizando el algoritmo de signalP (<http://www.cbs.dtu.dk/services/SignalP/>).

Cada una de las secuencias fue nombrada con una letra identificando al cónido (Rs para *C. regularis*) seguida de un número ascendente para indicar el orden de descubrimiento (ej 001, 002, 003, etc). En los casos que pudiera ser asignado, se le otorgaba el nombre de acuerdo a la nomenclatura descrita, y se tomó como referencia todos los puntos comentados en la sección anterior.

Después de analizar cada una de las secuencias individualmente, se eliminaron aquellas secuencias que no correspondían a una conotoxina y que habían sido erróneamente clasificadas. Además, se analizó la asignación a los grupos de superfamilia y en caso de ser necesario se realizó una reclasificación. Cada grupo de conotoxinas fue alineado contra las secuencias reportadas en el ConoServer utilizando Multalin (<http://multalin.toulouse.inra.fr/multalin/>). Se realizó una investigación de aquellas secuencias reportadas recientemente en la literatura que están en el NCBI (<https://www.ncbi.nlm.nih.gov/>) y en el UNIPROT (<http://www.uniprot.org/>). Los alineamientos finales se realizaron con ayuda del software CLC Sequence Viewer 7.6 (QIAGEN) con los parámetros establecidos por omisión.

No existe un porcentaje de identidad establecido como tal para poder incluir una secuencia dentro de un grupo de genes; no obstante, Kaas y colaboradores (2010) mencionan que las secuencias que conforman la superfamilia O1 conservan el 76 % de sus aminoácidos (con una desviación estándar del 12 %). Esto es un punto de referencia para los análisis realizados en esta tesis. De acuerdo con el NCBI, la identidad es la medida en que dos secuencias (de nucleótidos o de aminoácidos) tienen los mismos residuos en las mismas posiciones en un alineamiento, y es expresada como un porcentaje.

$$\text{Porcentaje de identidad} = (\text{aminoácidos idénticos} \times 100) / \text{longitud de la región alineada (con gaps)}$$

3.7.1 Superfamilia O1

La superfamilia O1 fue una de las primeras superfamilias de genes en ser caracterizada. Es de la que más se han reportado secuencias (599 conotoxinas en ConoServer) no obstante sólo una pequeña porción ha sido caracterizada a nivel de proteína. Los marcos de cisteína asociados a esta superfamilia son varios: I, VI/VII, IX, XII, XIV y XVI, aunque la mayoría presenta el marco VI/VII.

Se encontraron 5 secuencias de la superfamilia O1 para *C. regularis* (Figura 7), convirtiéndola en la segunda familia de genes con mayor diversidad de conotoxinas para esta especie. La mayoría de las secuencias

presentaron un arreglo de cisteínas dentro del marco VI/VII (C-C-CC-C-C), no obstante, la secuencia O1_Rs1 no presentó cisteínas. Aunque en Conoserver no se reportan conotoxinas sin cisteínas para esta superfamilia de genes, Phuong y colaboradores (2016) mencionan haber encontrado secuencias sin cisteínas para O1; además Robison y colaboradores reportan una secuencia sin cisteínas para *Conus victoriae*. No obstante, la secuencia O1_Rs1 (Figura 7b) no presenta identidad con la de este cónido.

Por otro lado, se encontró una secuencia (Rs6.16) que no está comprendida dentro del umbral establecido por los pHMM para ser clasificado dentro de la superfamilia pero que fue reconocida como O1. El péptido señal de esta secuencia tiene una 79.18 % de identidad con una conotoxina de *Conus betulinus*, sus regiones pro también conservan varios aminoácidos en común, el arreglo de las cisteínas es idéntico y presentan ciertos aminoácidos en común en la región de la toxina madura. El péptido señal de ambas conotoxinas presentan cierta identidad (37.5 %) con el péptido señal de una conotoxina de la superfamilia O1, además presentan el marco de cisteínas típico. Estas dos toxinas podrían ser un subgrupo dentro de esta superfamilia (Figura 7c).

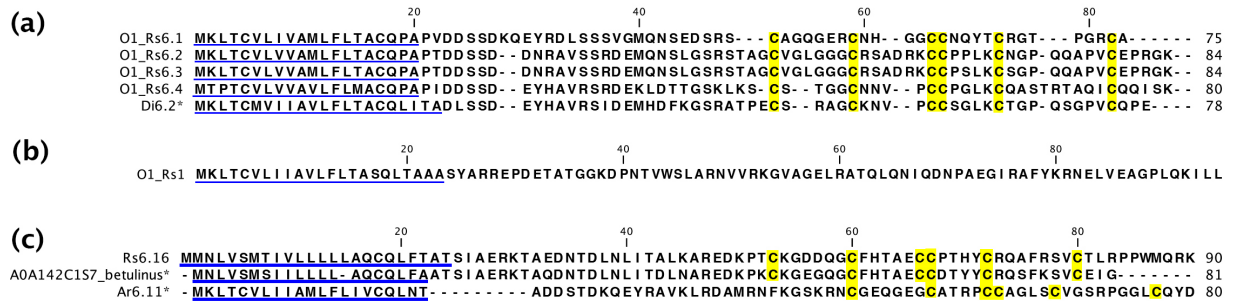


Figura 7. Secuencias para la superfamilia O1 de *C. regularis*. (a) con marco de cisteínas VI/VII (b) sin cisteínas y (c) que no entró en el umbral establecido para los pHMM. Las secuencias Di6.2* (*Conus distans*), A0A142C1S7_betulinus* (*C. betulinus*) y Ar6.11* (*Conus arenatus*) sirven de comparación. Las cisteínas se muestran en color amarillo, el péptido señal está indicado con una línea azul. Este esquema de colores es utilizado en las figuras siguientes.

3.7.2 Superfamilia O2

La mayoría de las conotoxinas identificadas para la superfamilia O2 presentan los marcos de cisteínas VI/VII o XV (Robinson y Norton, 2014). Dos de las conotoxinas de esta superfamilia con marco VI/VII (TxVIIA y PnVIIA) han sido caracterizadas a nivel de proteína y ambas presentan actividad específica en moluscos, ya que producen un efecto paralítico en ellos pero no en artrópodos o vertebrados (Fainzilber et al., 1998; 1991).

Las secuencias de *C. regularis* presentaron los marcos de cisteína esperados para esta superfamilia (Figura 8). Se identificaron 3 secuencias pertenecientes a la superfamilia O2, una con marco VI/VII (Figura8b) y dos con marco XV (Figura 8a), una de las cuales pudo ser validada a nivel de proteína (Parte III de esta tesis). Para las secuencias que tienen un marco de cisteínas XV se conservan los aminoácidos ECCPNLXCKC en la región central de la toxina madura.

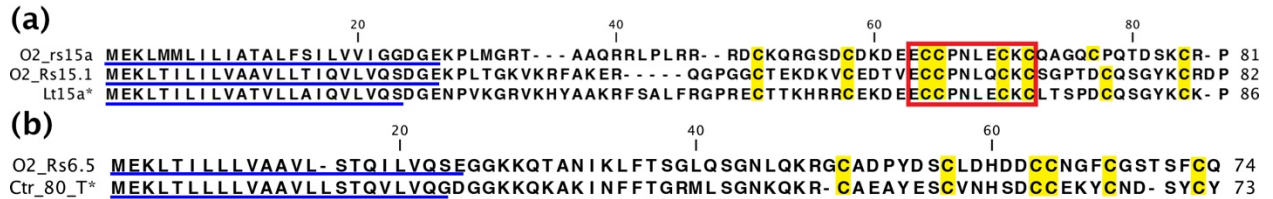


Figura 8. Secuencias identificadas para la superfamilia O2 de *C. regularis* (a) con marco de cisteínas XV (b) con marco de cisteínas VI/VII. Las secuencias Lt15a* (*Conus litteratus*) y Ctr_80_T* (*Conus tribblei*) se muestran como referencia. Los aminoácidos conservados son señalados con un rectángulo color rojo, este esquema será utilizado para las figuras subsecuentes.

3.7.3 Superfamilia O3

Casi la totalidad de conotoxinas reportadas para la superfamilia O3 presentan un marco de cisteínas VI/VII. Recientemente se reportó que existen secuencias sin cisteínas clasificadas para la superfamilia O3 en dos especies diferentes: *C. victoriae* y *Conus gloriamaris* (Robinson et al., 2014, 2017). Para *C. regularis* se recuperaron dos secuencias, cada una representante de lo que se ha descrito para esta superfamilia: O3_Rs6.6 (Figura9a), con el marco VI/VII, y O3_Rs1 (Figura 9b), sin cisteínas. O3_Rs1 presentó una mayor identidad con la secuencia de *C. victoriae* (85 %) que con la de *C. gloriamaris* (60.86 %).

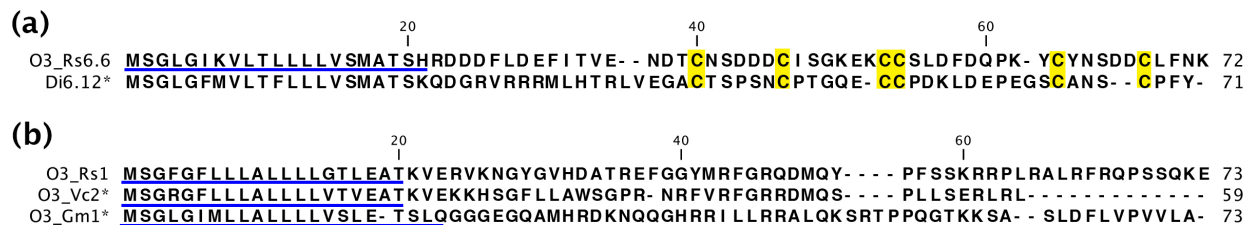


Figura 9. Conotoxinas identificadas para la superfamilia O3 de *C. regularis* (a) secuencia identificada con un marco de cisteínas VI/VII (b) secuencia sin cisteínas. Las conotoxinas Di6.12* (*C. distans*), O_Vc2* (*C. victoriae*) y O3_Gm1* (*C. gloriamaris*) se muestran como punto de comparación.

3.7.4 Superfamilia C/Contulacina y superfamilia D

Sólo cuatro secuencias están reportadas en ConoServer para la superfamilia C, conocida como las contulacinas, en únicamente dos especies: *C. geographus* y *C. litteratus*. El pHMM de esta superfamilia fue construido con estas secuencias, no obstante, en la literatura también existen descritas secuencias de esta superfamilia para *Conus parius* (Jimenez et al., 2007) y *C. pulicarius* (Lluisma et al., 2012).

Las contulacinas pertenecen a la familia de las neurotensinas las cuales son analgésicos que activan los receptores de neurotensinas acoplados a proteínas G (GPCRs), jugando un papel importante en neurotransmisión y neuromodulación (Robinson y Norton, 2014). Los conopeptidos de esta superfamilia pueden carecer de cisteínas, presentar un solo enlace disulfuro o tener los marcos de cisteínas V y XIV.

En este trabajo se reporta la primera contulacina para *C. regularis* la cual presenta un solo enlace disulfuro (Figura 10a). Se observó una conservación de 13 aminoácidos en la región C-terminal (CIWKVCPPXPWRR) de la toxina madura al ser comparada con la Contulacina Lt1.

Por otro lado, la superfamilia D se ha encontrado mayoritariamente en especies vermívoras, como es el caso de *C. regularis*. Los marcos de cisteína relacionados con esta superfamilia son IV, XIV, XV, XX, XXIV. Se identificaron dos secuencias pertenecientes a esta superfamilia ambas con un marco de cisteínas XX (Figura 10b), una toxina está posiblemente incompleta del amino terminal y le falta una pequeña parte del precursor (D_Rs20.1).

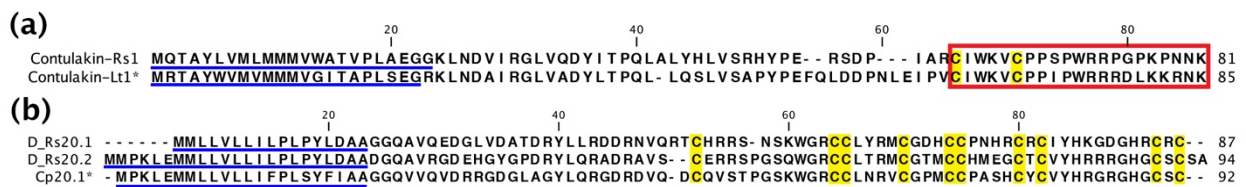


Figura 10. Contulacina y miembros de la superfamilia D identificadas para *C. regularis* (a) contulacina (b) miembros de la superfamilia D con un marco de cisteínas XX. Contulacina-Lt1* (*C. litteratus*) y Cp20.1* (*Conus capitaneus*) fueron utilizadas como referencia.

3.7.5 Superfamilia H

Es una superfamilia recientemente descrita, se han encontrado secuencias para *C. marmoreus*, *C. gloriamaris* y *C. victoriae* (Dutertre et al., 2013; Robinson et al., 2014; 2017). La mayoría de las secuencias descritas presentan el marco de cisteínas VI/VII; no obstante, se han reportado dos secuencias sin cisteínas. En el transcriptoma de *C. regularis* se encontraron dos secuencias representativas de cada una: sin cisteínas (Figura 11a) y con el marco VI/VII (Figura 11b). Para las secuencias sin cisteínas se observó una elevada conservación de aminoácidos en lo que puede ser la región pro y otra conservación en lo que puede ser la toxina madura.

Por otro lado, se identificaron dos secuencias incompletas que posiblemente pertenezcan a la superfamilia H (Figura 11c). El péptido señal de ambas presentó un gran parecido (70.83 % de identidad) con una conotoxina H de *C. tribblei* que no está reportada en ConoServer, pero que fue recuperada de UNIPROT.

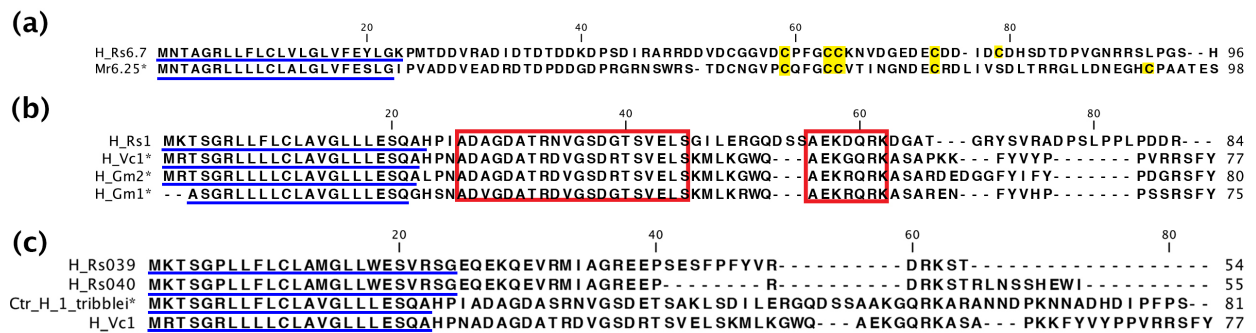


Figura 11. Conotoxinas de la superfamilia H (a) secuencia con marco VI/VII (b) secuencia identificada sin cisteínas (c) posibles secuencias de la superfamilia que se encontraron incompletas. Las conotoxinas utilizadas como referencia son Mr6.25* (*C. marmoreus*), H_Vc1* (*C. victoriae*), H_Gm2* y H_Gm1* (*C. gloriamaris*).

3.7.6 Superfamilia I1

La mayoría de las conotoxinas reportadas para la superfamilia I1 presentan un marco de cisteínas XI (C-C-CC-CC-C-C). No obstante, existe una sola secuencia de *C. californicus* dentro de esta superfamilia reportada con un marco de cisteínas VI/VII; este es el caso de las cuatro secuencias recuperadas para *C. regularis* (Figura 12). Las secuencias de *C. regularis* confirman la presencia del marco de cisteínas VI/VII dentro de la superfamilia I1 no obstante, la diferencia del péptido señal con los otros miembros de la superfamilia hacen pensar que podría tratarse de un subgrupo de genes dentro de esta superfamilia.

```

                20                40                60                80
I1_Rs6.9 - MKLSVTFVTSLLIVMLLPVINGRELSG-----SSALRKVIQR-VKRDCSYSSCTS-LTCCGEYLCNAN-KDCVDPGPGG-PFGK 75
I1_Rs6.8 - MKLSMTV--LLILMLLPVIAGEVEPG-----QHALRDLLKRSVKHRCIFGCLDG-SCCAGFCY-GGSLCR-----61
I1_Rs6.10 MMKVLMTF--LLVLMILPLCQPIK-----FGEQLTMKPAART-----GDTCIRAPCESKDNCCKNYFCDTCELRQCCISNARKTFKNGK 75
I1_Rs6.11 ----MTF--LLVLMILPLCQPTMMVGPLPVFVQNLKVTAPTT-----GDSCCIRAPCCDRKDNCCKNYFCDQELGLCISNAWRPFRNGK 77
Cl6.15* - MKLSVKF--LLFLMILPLIAGEDMS--DNDAPKSVDVQRNVKRGQSQFGEQCTGHLCDCFGDCLCFDGYCIMTSWIWPCEN-W-----77

```

Figura 12. Secuencias de la superfamilia I1, todas presentaron un marco de cisteínas VI/VII. Se utilizó como referencia Cl6.15* (*C. californicus*) la única conotoxina de esta superfamilia con marco VI/VII.

3.7.7 Con-ikot-ikot

El pHMM construido para la superfamilia K identificó 3 secuencias en el transcriptoma, dos de las cuales se clasificaban en esta superfamilia con un umbral por debajo del establecido para ser clasificadas dentro del grupo. El análisis realizado a estas secuencias reveló que en realidad pertenecían al grupo de las con-iko-ikot. Las toxinas de este tipo se han identificado porque son más largas que el resto de las conotoxinas y presentan arreglos de cisteínas inusuales. Aquellas que han sido caracterizadas, inhiben la capacidad de regresar a su estado original a los receptores AMPA (ácido α -amino-3-hidroxy-5-metil-4-isoxazolepropiónico) (Walker et al., 2010). Existen dos tipos de con-ikot-ikot descritas: las que presentan 13 residuos de cisteínas y las que tienen 10 cisteínas con un marco XXI (Moller y Marí, 2011; Walker et al., 2010). Sólo una de las secuencias encontradas para *C. regularis* presentó este tipo de arreglos (con-ikot-ikot-Rs21.1); mientras que con-ikot-ikot-Rs1 presentó 14 residuos de cisteínas y con-ikot-ikot-Rs2 presentó 11 cisteínas las cuales no entraban dentro de ninguno de los marcos de cisteínas establecidos (Figura 13). Estas secuencias confirman lo que se había visto previamente para *C. victoriae*: que las con-ikot-ikot no son exclusivas de los cónidos piscívoros sino que pueden presentarse en aquellos que tienen dieta vermívora (Robinson et al., 2014).

```

                20                40                60                80
con-ikot-ikot-Rs1  MTVNMSMTLCVFVMVVI AAT I VGFTPSEDKDLYCEGRE--YHDCCKHE--MQRCCMEE--CTSVFTKEY-----CW 64
con-ikot-ikot-Rs2  MAMNWMWTLCINVLVVVMAATVDGSTPLQERDNR-----ECCIND--MYRC-----LQGY-----PG 50
con-ikot-ikot-Rs21.1 MAMNWMWTLCSVLVVVLTITDCCSLLEDKREALKRRGDTTSLCCFSVFGCIRECKAKGGSNCWPLCCELEYSDVCCKGDFSVYCCWSFGNCCY 91
con-ikot-ikot-G1*  MAMNMSVMLSCAFVMVVVSACTVTGFTHLQEPDLCSRMERSPPPHNDCCK--MKECCAQ--TTELCCKEF-----PN 65

                100                120                140                160
con-ikot-ikot-Rs1  DR-----CYNCSAATNCCG-CHPTDCGCCPDFLEQYSSCCLFP--QDRGVF--ECWERSKCFVPCW 115
con-ikot-ikot-Rs2  QEQVYYTK-----CHHDAANACG-VSAAHGCCNG--YMICIGMTV-DQGLMAHNYCCKKLCRCDPVGC 108
con-ikot-ikot-Rs21.1 EECIKTKSEEHCFKGCRYEAKLDCIE-EGSIGCCCPQFMTCYCYECF--EKTGRYSCSTHCCKDVHC-----152
con-ikot-ikot-G1*  EEHCIYCYST-----CYQRASHACGGQFNEICVGCCYG--YRQCMLQNVQNLGLNCWANQCCKEWNCCLNPCCE 125

```

Figura 13. Secuencias con-ikot-ikot de *C. regularis* comparadas con una secuencia del mismo grupo de *C. geographus* (Con-iko-ikot-G1*).

3.7.8 Superfamilia L

M_Rs16.1, no tienen un nombre asignado de acuerdo a la nomenclatura establecida y no han sido curadas por Conoserver. Es probable que este grupo de conotoxinas pertenezca a un subgrupo de la familia M.

La conotoxina M_Rs027 (Figura 15d), presenta 8 cisteínas dentro de un acomodo no reportado hasta la fecha (CC-CC-C-C-CC), por lo que no pudo ser nombrada de acuerdo a la nomenclatura convencional de conotoxinas, no obstante, es evidente que sí pertenece a la superfamilia M.

Finalmente, para el péptido M_Rs8.1 no se encontró una conotoxina que conservara de la misma manera el péptido señal y el marco de cisteínas como con el grupo de M_Rs27.1, pero sí presentó 42.81 % de identidad con una conotoxina para *C. betulinus* reportada en UNIPROT.

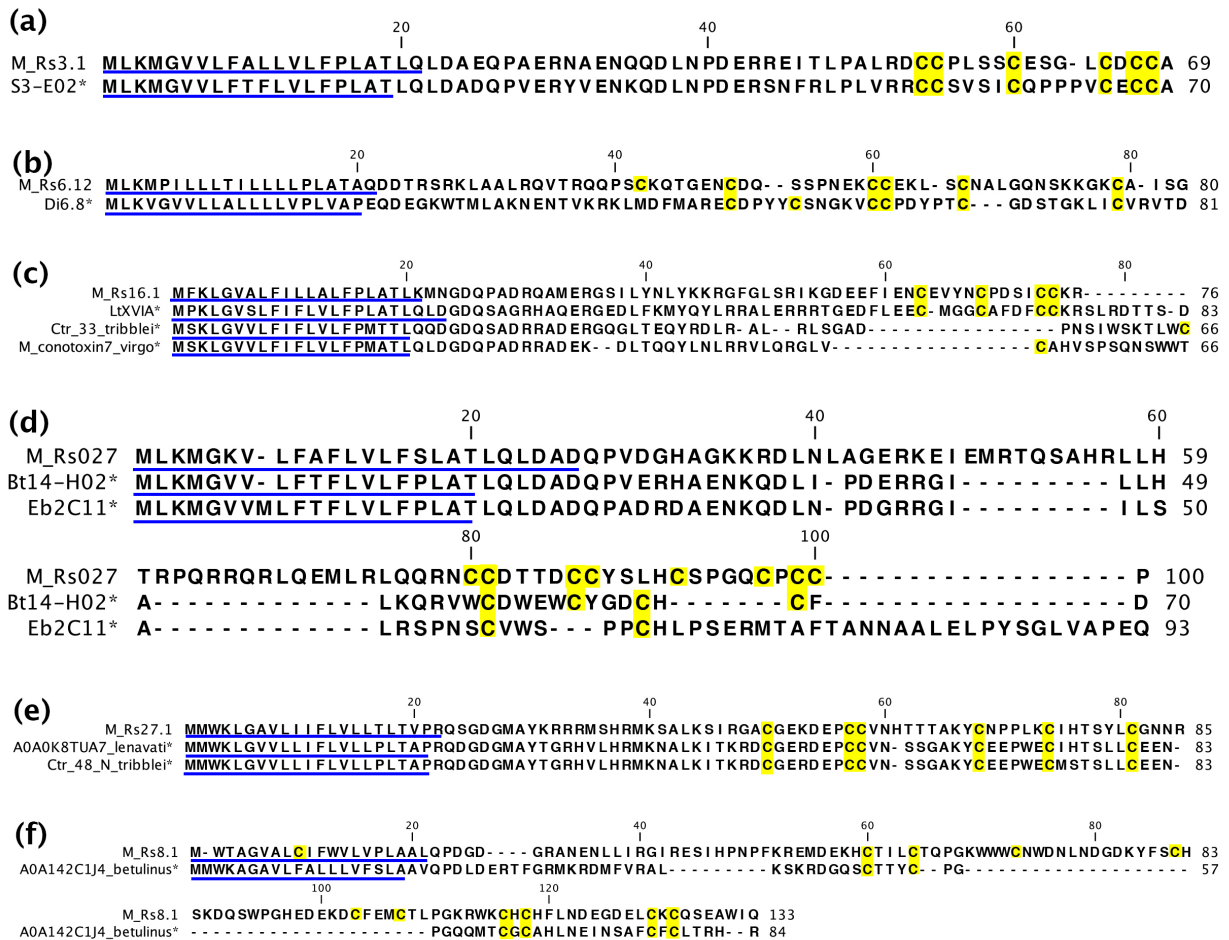


Figura 15. Superfamilia M (a) y (b) son secuencias que presentan marcos de cisteínas reportados para esta superfamilia (III, VI/VII); (c) y (d) es un grupo de secuencias que confirma la presencia de arreglo de cisteínas no convencionales para esta superfamilia; (e) y (f) son nuevos marcos de cisteína reportados para esta superfamilia. Las secuencias S3-E02* (*Conus striatus*); Di6.8* (*C. distans*); LtxVIA* (*C. litteratus*); Ctr_33_tribblei* y Ctr_48_N_tribblei* (*C. tribblei*); M_conotoxin7_virgo* (*Conus virgo*); Bt14-H02* y A0A142C1J4_betulinus* (*C. betulinus*); Eb2C11 (*Conus ebraeus*) y A0A0K8TUA7_lenavati* (*C. lenavati*) son mostradas como comparación.

3.7.10 Superfamilia B2

La superfamilia B2 fue identificada recientemente en *Conus consors* (Violette et al., 2012) como un conopéptido sumamente abundante que era muy similar a una conotoxina de *C. litteratus*, previamente identificada en una biblioteca de cDNA (Pi et al., 2006). Aunque existen varias secuencias reportadas para la superfamilia B2 en la literatura para diversos cónidos (*C. marmoreus*, *C. victoriae*, *C. gloriamaris*, entre otros), en Conoserver solo están listadas dos: una de *C. marmoreus* y una de *C. victoriae*. Por lo que al igual que en el caso de las contulacinas, los pHMM de esta superfamilia fueron construidos con menos secuencias de las que realmente existen. La función de esta superfamilia de conotoxinas permanece desconocida, pero se sabe que pueden presentarse sin cisteínas o con un marco VIII.

Al hacer el alineamiento de B2_Rs1 con los otros miembros de la superfamilia (Figura 16) se pudo observar una alta conservación de aminoácidos a lo largo de toda la secuencia: desde el péptido señal (82.35 %) hasta la región de la toxina madura.

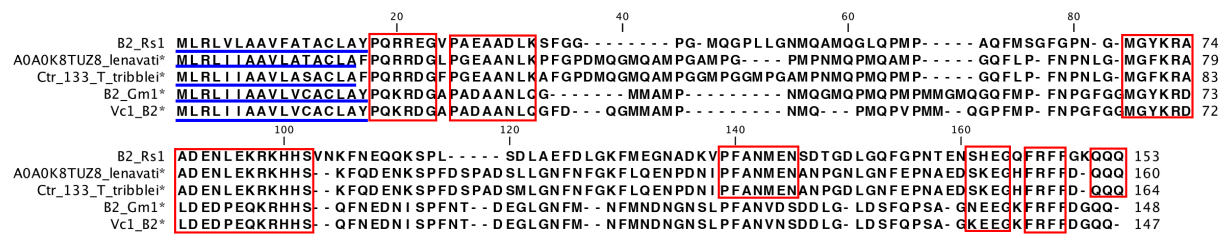


Figura 16. Superfamilia B2. Varios de los aminoácidos que componen a las secuencias se conservan, en este caso sólo se marcaron algunos. Las secuencias A0A0K8TUZ8_lenavati* (*C. lenavati*), Ctr_133_T_tribblei* (*C. tribblei*), B2_Gm1* (*C. gloriamaris*) y Vc1_B2* (*C. victoriae*) son conotoxinas reportadas para la superfamilia B2.

3.7.11 Superfamilia N

La superfamilia N está conformada por unos cuantos conopéptidos que fueron identificados en el transcriptoma de la glándula venenosa de *C. marmoreus* (Dutertre et al., 2013). El patrón de cisteína que presentan estos péptidos pertenece al marco XV. Recientemente se encontró una secuencia para esta superfamilia con el mismo patrón de cisteínas para *C. gloriamaris* (Robinson et al., 2017). En el caso de *C. regularis* (Figura 17) se recuperó una sola secuencia que conservaba el péptido señal (75 % de identidad) con los miembros de la superfamilia, no obstante presenta un arreglo de cisteínas que no ha sido nombrado (C-C-CC-C-C-C-C). Después de una búsqueda bibliográfica se encontró una conotoxina (Lt15.2)

que presenta el mismo patrón de cisteínas que el encontrado para la conotoxina N_Rs032, lo que confirma la existencia de dicho patrón de cisteínas.

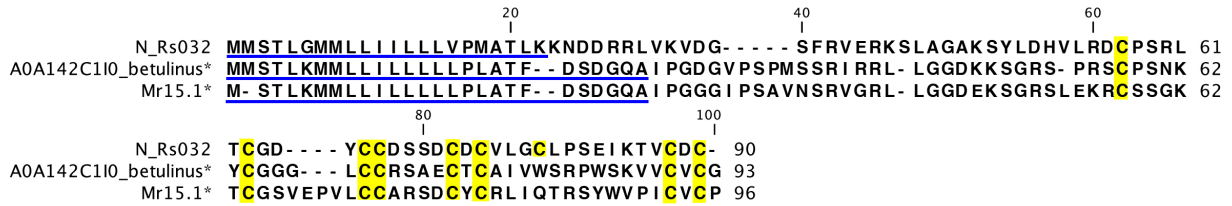


Figura 17. Superfamilia N. La secuencia N_Rs032 presenta un marco de cisteínas no convencional. Las secuencias A0A142C1I0_betulinus* (*C. betulinus**) y Mr15.1* (*C. marmoreus*) presentan similitud con la conotoxina reportada.

3.7.12 Superfamilia P

La superfamilia P presenta un marco de cisteína muy característico denominado IX (C-C-C-C-C) (Lirazan et al., 2000), aunque recientemente se identificaron dos conotoxinas con marco XIV para esta superfamilia (Robinson et al., 2014; 2017). Son pocos los miembros de la superfamilia P que han sido identificados, por lo que resulta interesante que *C. regularis* presente 3 conotoxinas de este grupo, todas con el nuevo marco de cisteínas identificado XIV. La conotoxina P_Rs9.2 (Figura 18a) se alineó con GmIXA, uno de los miembros más estudiados de esta superfamilia, la cual presenta efecto espasmódico al ser inyectado en ratones y que posiblemente tenga como blanco un receptor de glicina.

Por otro lado, las conotoxinas P_Rs9.1 y P_Rs9.2 (Figura 18b) se alinearon considerablemente bien (hasta 81 % de identidad en el péptido señal) con dos conotoxinas que no están reportadas en ConoServer y que posiblemente también sean miembros de la superfamilia P. Estas conotoxinas pertenecen a *C. tribblei* y *C. betulinus*, dos especies con las que las conotoxinas de *C. regularis* han tenido similitud, pero cuyos conopéptidos no han sido nombrados de acuerdo a la nomenclatura establecida. Aunque el péptido señal presenta 59 % de identidad con otros miembros de la superfamilia P, como puede ser Lt9a, es posible que estas conotoxinas sean un subgrupo de esta superfamilia.

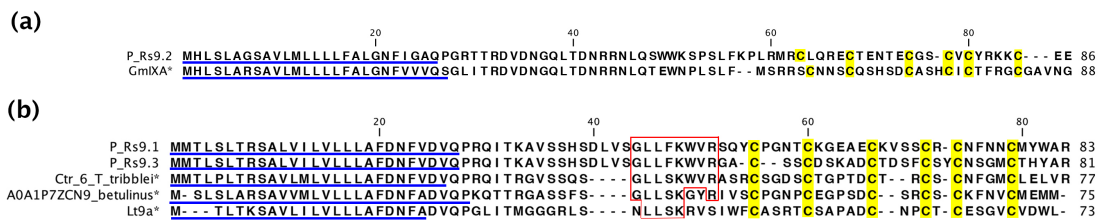


Figura 18. Superfamilia P (a) nuevo arreglo de cisteínas para esta superfamilia (b) posible nuevo grupo. La conotoxina GmIXA* (*C. gloriamaris*), Ctr_6_T_tribblei* (*C. tribblei*), A0A1P7ZCN9_betulinus* (*C. betulinus*) y Lt9a* (*C. litteratus*) se muestran como comparación.

3.7.13 Superfamilia S

Al igual que en los transcriptomas de *C. gloriamaris* y de *C. victoriae* (Robinson et al., 2014; 2017), solamente se identificó una sola secuencia (S_Rs8.2) de esta superfamilia en el transcriptoma de *C. regularis* (Figura 19). No son muchas las conotoxinas identificadas para este grupo, pero todas han presentado hasta la fecha un marco de cisteínas VIII (C-C-C-C-C-C-C-C-C). Las dos secuencias de esta superfamilia que han sido caracterizadas presentan actividades diferentes: GVIIA es un inhibidor de los receptores de serotonina (Oers et al., 1998), mientras que RVIIA inhibe los receptores nicotínicos de acetilcolina (Teichert et al., 2005).

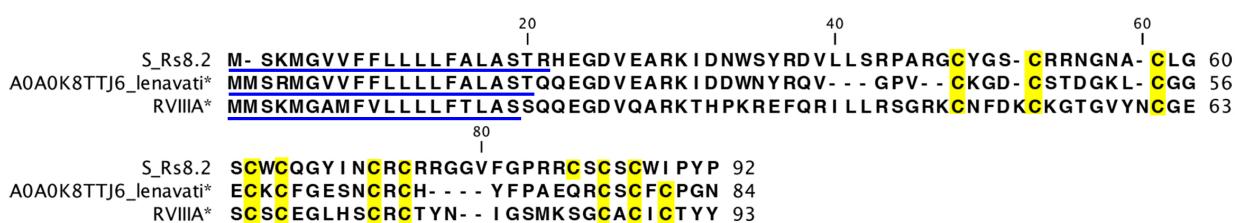


Figura 19. Superfamilia S. S_Rs8.2 Presenta el marco de cisteínas esperado para esta superfamilia. Las conotoxinas A0A0K8TTJ6_lenavati* (*C. lenavati*) y RVIIA* (*Conus radiatus*) se muestran como comparación.

3.7.14 Superfamilia T

Es una de las superfamilias que más miembros tiene reportados, existen 239 secuencias en el ConoServer. No obstante, en el transcriptoma de *C. regularis* únicamente se encontraron dos secuencias. La secuencia T_Rs15.1 presenta un marco de cisteínas (CC-CC) característico para esta superfamilia y la región de la toxina madura presentó varios aminoácidos en común con la conotoxina Rt5.1 (Figura 20a).

La otra conotoxina identificada para esta superfamilia (T_Rs1) no presenta cisteínas. Existen ya conotoxinas de *Conus ventricosus* reportadas para esta superfamilia que no presentan cisteínas (Figura 20b). Además, Phuong y colaboradores (2016) reportan haber identificado secuencias sin cisteínas para esta superfamilia. La secuencia de *C. regularis* confirma la existencia de conotoxinas sin cisteínas para esta superfamilia.

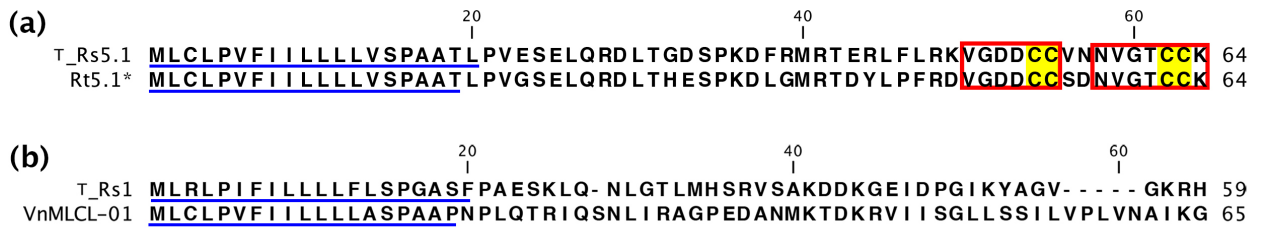


Figura 20. Superfamilia T (a) secuencia identificada con un marco típico para la superfamilia (b) secuencia que confirma la existencia de conotoxinas sin cisteínas para esta superfamilia. Las conotoxinas Rt5.1* (*Conus rattus*) y VnMLCL-01* (*C. ventricosus*) se utilizan como punto de comparación.

3.7.15 Superfamilias Divergentes

Existen ciertas conotoxinas que no han podido ser clasificadas en las superfamilias establecidas hasta la fecha, por lo que se han clasificado en superfamilias de conotoxinas denominadas como divergentes. Estos grupos de conotoxinas no tienen un nombre como tal, sino que han sido nombradas de acuerdo a como inicia el péptido señal (Tabla 17, Anexo 1). La mayoría de las conotoxinas clasificadas dentro de estas superfamilias divergentes pertenecen a la especie atípica de los cónidos *C. californicus*; no obstante, una de las superfamilias divergentes (MSTLGMTLL-) presenta secuencias de otras especies.

Se construyó un pHMM de cada una de las superfamilias de genes divergentes, aunque cabe mencionar que la mayoría tiene registrada sólo una secuencia lo que hace que el modelo no sea óptimo. Se obtuvieron conotoxinas de *C. regularis* para tres superfamilias divergentes, las cuáles se analizan a continuación.

3.7.15.1 Superfamilia MSTLGMTLL-

Esta superfamilia divergente, que se compone de 6 secuencias, es la única que contiene conotoxinas que no son de *C. californicus* ya que también están registradas una conotoxina de *Conus pulicarius* y una conotoxina de *C. distans*. Los marcos de cisteína reportados para esta superfamilia son IX, XIX y XXII.

Se recuperaron 4 conotoxinas de *C. regularis* para esta superfamilia (Figura 21). La conotoxina Rs19.1 se alineó perfectamente con las secuencias Di19A y Pu19.1 (Figura 21a), las cuales son las únicas conotoxinas reportadas para esta superfamilia que no pertenecen a *C. californiconus*. Existe una conservación tanto del péptido señal (85.71 % de identidad), como del arreglo de cisteínas, así como de ciertos aminoácidos

en la región pro y en la toxina madura. Esto es un indicador de que este grupo de secuencias probablemente pertenezcan a una nueva superfamilia de genes que no había sido descrita.

Por otro lado, al realizar un BLAST con la base de datos de UNIPROT de la secuencia Rs20.3 (Figura 21b) se obtuvieron varias secuencias con un alto porcentaje de identidad. Estas secuencias no han sido nombradas bajo la nomenclatura convencional de los cónidos, y por lo tanto no han sido curadas en ConoSever. Al igual que en el caso anterior se conserva el péptido señal (77.27 % de identidad), el arreglo de cisteínas es idéntico, así como existe una conservación de varios aminoácidos de la región pro y de la toxina madura. Es muy probable que este grupo de secuencias sea una nueva superfamilia de cónidos no reportada hasta la fecha.

Se realizó un BLAST de la secuencia Rs6.12 contra la base de datos de UNIPROT y se identificaron dos secuencias que tenían un alto porcentaje de identidad con ellas. Una de las secuencias (Fla6.10) está reportada como un miembro de la superfamilia V (con sólo dos miembros) no obstante en ConoServer no se encuentra listada. Por esta razón se hizo un alineamiento de Rs6.12 con las dos secuencias de la superfamilia V y con los dos péptidos con los que obtuvo una mayor identidad (Figura 21c). Es probable que la conotoxina Rs6.12 pertenezca a un subgrupo de la superfamilia V que aún no ha sido reportado.

Finalmente, la secuencia Rs6.13 se alineó perfectamente, tanto en el péptido señal (91.66 % de identidad) como en el arreglo de cisteínas, con 4 secuencias reportadas en UNIPROT que no han sido clasificadas (Figura 21d). Este grupo de secuencias podría ser una nueva superfamilia de genes.

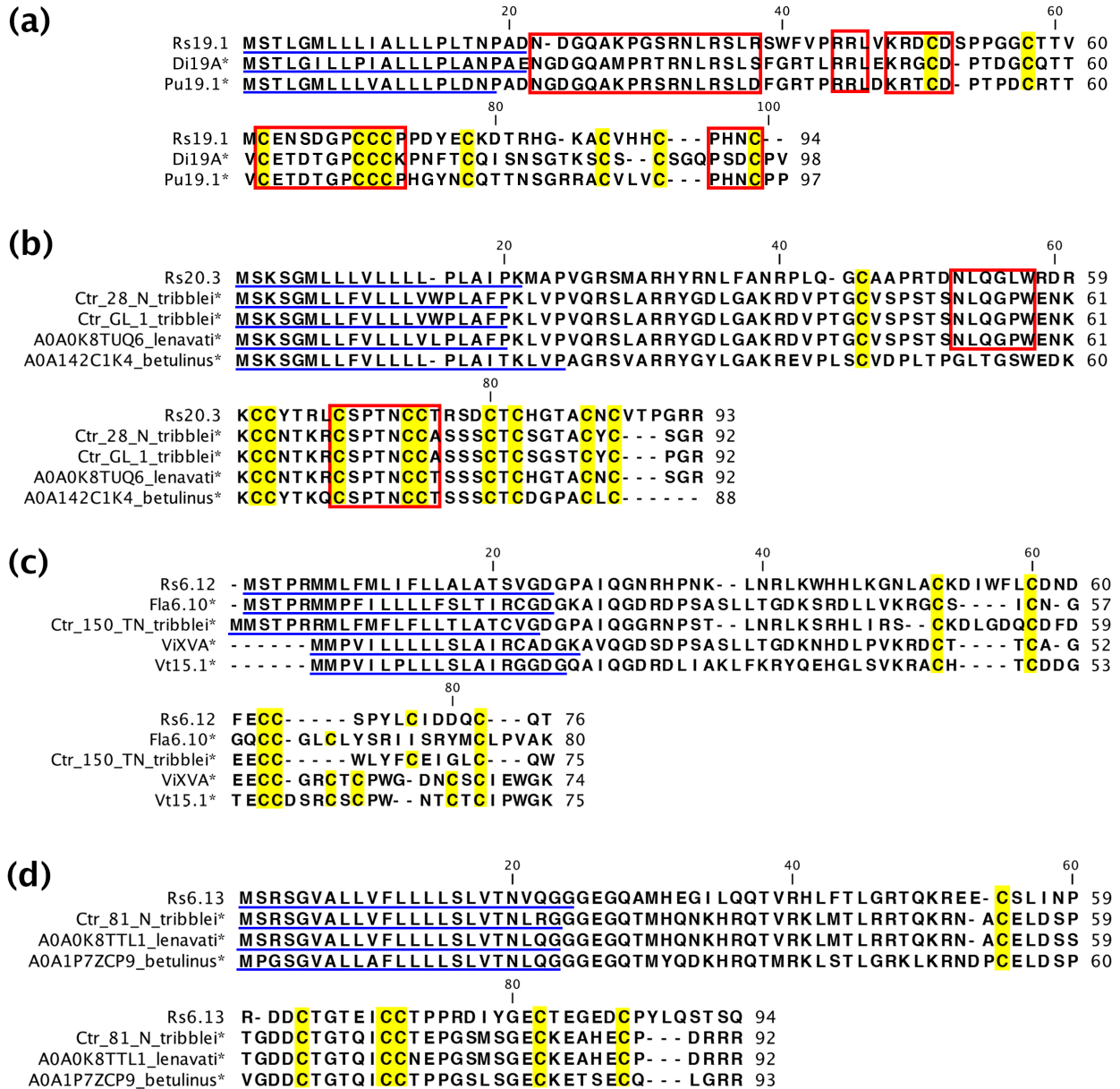


Figura 21. Superfamilia MSTLGMTLL- (a) (b) y (d) son grupos de secuencias que podrían ser miembros de nuevas superfamilias de genes. (c) es un grupo de secuencias que podría pertenecer a un subgrupo de la superfamilia V. Las secuencias Di19A* (*C. distans*); Pu19.1* (*C. pulicarius*); Ctr_28_N_tribblei*, Ctr_GL_1_tribblei*, Ctr_81_N_tribblei* y Ctr_150_TN* (*C. tribblei*); A0A0K8TTL1_lenavati* y A0A0K8TUQ6_lenavati* (*C. lenavati*); A0A1P7ZCP9_betulinus* y A0A142C1K4_betulinus* (*C. betulinus*); Fla6.10 (*Conus flavidus*); ViXVA* (*C. virgo*) y Vt15.1 (*Conus vitulinus*) son mostradas como comparación.

3.7.15.2 Superfamilia M---L-LTVA

Los 9 miembros de esta superfamilia son sumamente diversos. Aunque el pHMM reconoció 4 secuencias diferentes, el análisis individual de cada conotoxina nos permitió identificar que posiblemente

pertenezcan a diferentes grupos. Las conotoxinas Rs6.14 y Rs.15 alinearon perfectamente con el péptido señal de una conotoxina de *C. betulinus* recuperada del UNIPROT (Figura 22a), el marco de cisteínas de las tres secuencias fue VI/VII.

La secuencia Rs046 presentó un marco de cisteína atípico (C-C-C-C-C-C-C-C-C-C) que fue reportado previamente por Phuong y colaboradores (2016). Este marco de cisteínas se encontró en dos secuencias de *C. betulinus* y en una secuencia nueva de *C. geographus*, todas ellas no catalogadas en ninguna de las superfamilias ya establecidas (Figura 22b). Existe una conservación del péptido señal (65 % de identidad) de Rs046 con esas secuencias, lo que hace suponer que sean miembros de una nueva superfamilia de genes.

Finalmente, la secuencia Rs047 (Figura 22c) presenta cierta similitud con las secuencias Cal14a y Cl14.4 que son miembros de la superfamilia divergente M---L-LTVA. No obstante, presenta un arreglo de cisteínas que no está dentro de los marcos de cisteína establecidos (C-C-C-CCC-C-C) pero que si ha sido previamente reportado por Phuong et al (2016).

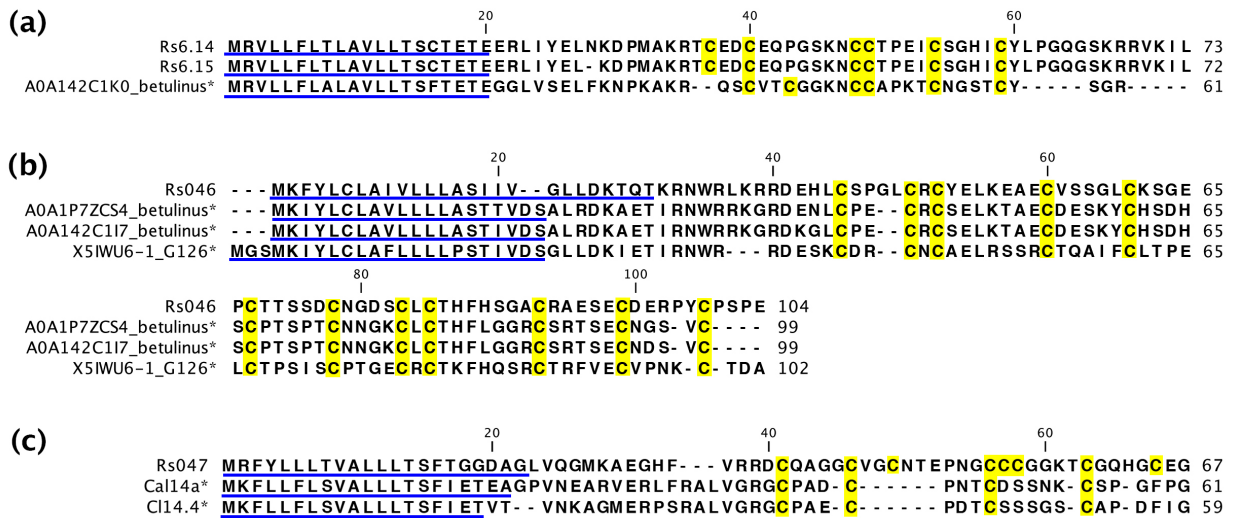


Figura 22. Superfamilia M---L-LTVA (a) posible nueva superfamilia de genes (b) y (c) confirmación de dos nuevos marcos de cisteínas y posible nuevas superfamilias de genes. Las secuencias A0A142C1K0_betulinus* y A0A1P7ZCS4_betulinus* (*C. betulinus*); X5IWU6-1_G126* (*C. geographus*); Cal14a* y Cl14.4* (*C. californiconus*) se muestran como comparación.

3.7.16 Secuencias que no pueden ser clasificadas.

Varias secuencias fueron consideradas conotoxinas por las características que presentaban; no obstante, el umbral establecido para los pHMM no permite que sean clasificadas adecuadamente dentro de una superfamilia de genes. Las conotoxinas putativas de las superfamilias O1 y H fueron mostradas dentro de las figuras de las conotoxinas que sí pudieron ser clasificadas para estos grupos (Figuras 7 y 11). El resto se muestra a continuación.

3.7.16.1 Superfamilia D putativa

Se identificaron dos posibles secuencias para la superfamilia D que presentaban marcos de cisteínas ya reportados hasta la fecha. La secuencia Rs9.4 (Figura 23a) fue clasificada como una posible conotoxina de la superfamilia D, no obstante, tuvo un mayor parecido con dos secuencias que no han sido clasificadas. El péptido señal conserva el 94.7 % de sus aminoácidos con una conotoxina de *C. tribblei*, que no ha sido nombrada de acuerdo a la nomenclatura establecida y que sólo está reportado en el UNIPROT. Además, el número de cisteína también se conserva. El péptido señal presentó cierto grado de identidad (47.36 %) con una conotoxina de *C. californicus* reportada para la superfamilia divergente MSKLVLAVL-, con la que también comparte el arreglo de cisteínas.

La conotoxina Rs14.3 (Figura 23b) también fue identificada como una posible conotoxina de la superfamilia D sin embargo, al ser comparada con una conotoxina de esta superfamilia (Cp20.1) la identidad con el péptido señal es muy baja.

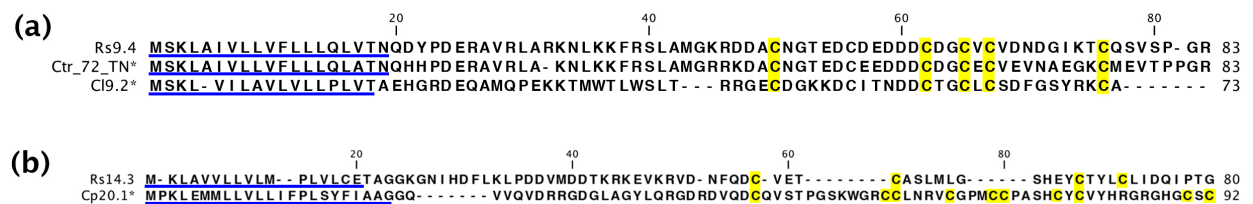


Figura 23. Conotoxinas que no pudieron ser clasificadas (a) Rs9.4 presenta un alto grado de similitud con Ctr_72_TN*, una conotoxina de *C. tribblei*. La secuencia Cl9.2 (*C. californicus*) presenta una conservación en el péptido señal y comparte el arreglo de cisteínas. (b) La conotoxina Rs14.3 no puede ser clasificada dentro de la superfamilia D. La secuencia Cp20.1 (*Conus capitaneus*), miembro de dicha superfamilia, se muestra como comparación.

3.7.16.2 Superfamilias I1, B1 y B2 putativas

Se identificó una secuencia que posiblemente pertenecía a la superfamilia I1, una para la superfamilia B1 y una para la superfamilia B2. Las conotoxinas de I1 y B1 presentaban marcos de cisteína ya establecidos, no obstante, la conotoxina B2 presenta un arreglo de cisteínas no reportado (C-C-CC).

La conotoxina Rs22.1 (Figura 24a) presentó identidad con miembros de la superfamilia I1 (CI11.1 y RXIA): el 40% de los aminoácidos del péptido señal se conservaban entre las secuencias. Además, el marco de cisteínas que presenta (XXII) la secuencia Rs22.1 no ha sido reportado para la superfamilia I1, lo que la excluye aún más de este grupo.

La conotoxina Rs14.4 (Figura 24b) no presentó identidad con las secuencias de la superfamilia B1, también conocidas como conantocinas. Este grupo se ha caracterizado por no presentar cisteínas, el cuál no era el caso de la secuencia Rs14.4.

Finalmente, la conotoxina Rs055 (Figura 24c) no presentó un arreglo de cisteínas que hubiera sido reportado y el péptido señal no es parecido al que está reportado para la superfamilia B2 por lo que no puede ser catalogado para esta superfamilia.

Posiblemente las 3 secuencias identificadas pertenezcan a otras superfamilias de genes que aún no han sido descritas. Como es el caso de las secuencias de las conotoxinas de *C. regularis* que presentaron similitud con las conotoxinas de *C. tribblei* que hasta la fecha no habían podido agruparse con otros conopéptidos.

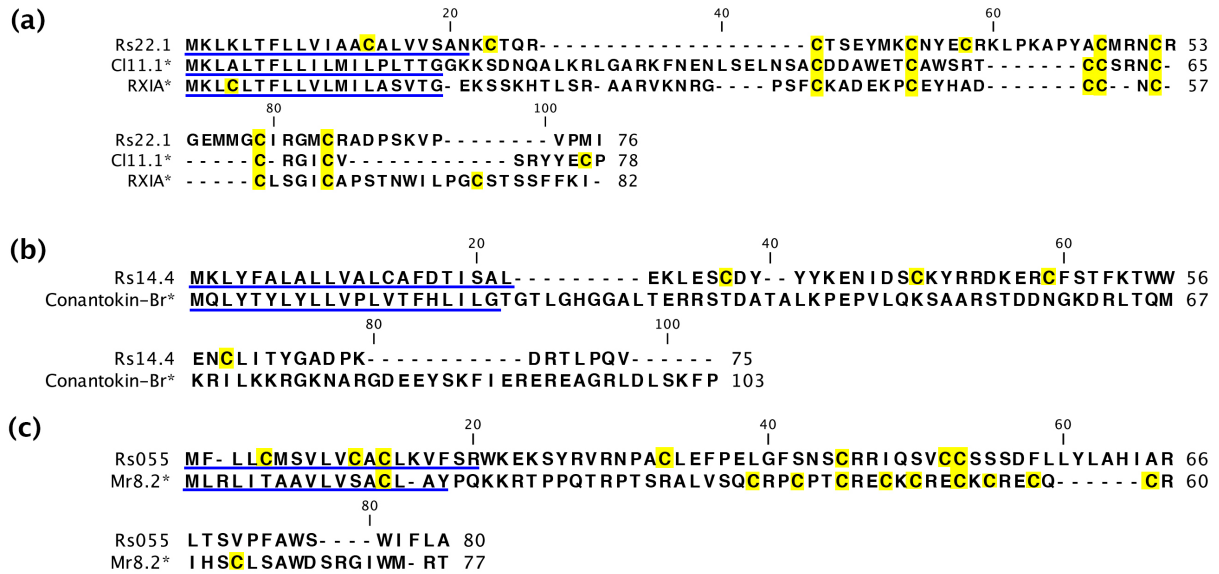


Figura 24. Conopéptidos sin poder ser clasificados (a) Rs22.1 es un posible miembro de la superfamilia D (b) Rs14.4 es un posible miembro de las conantocinas (c) Rs055 es un posible miembro de la superfamilia B2. Las conotoxinas Cl11.1* (*C. californicus*), RXIA* (*C. radiatus*), Conantokin-Br (*Conus sulcatus*) y Mr8.2* (*C. marmoreus*) se muestran como comparación.

3.7.16.3 Superfamilia A putativa

Se recuperaron dos secuencias que posiblemente pertenezcan a la superfamilia A. La secuencia Rs056 (Figura 25a) es una conotoxina incompleta cuyo péptido señal presenta una identidad del 71.42 % con el péptido señal de otro miembro de la superfamilia A (Fla1.8) pero que sobre todo se conserva con dos secuencias reportadas en UNIPROT que no han sido clasificadas correctamente. Estas secuencias son prácticamente idénticas a Fla1.8 con sólo 1 a 3 cambios de aminoácidos.

Por otro lado, la secuencia Rs22.2 presentó una alta identidad con una secuencia de *C. betulinus*, tanto en el péptido señal (75 %) como en el arreglo de sus cisteínas. Además, el péptido señal presentó similitud con otras secuencias que sólo han sido reportadas en el UNIPROT.

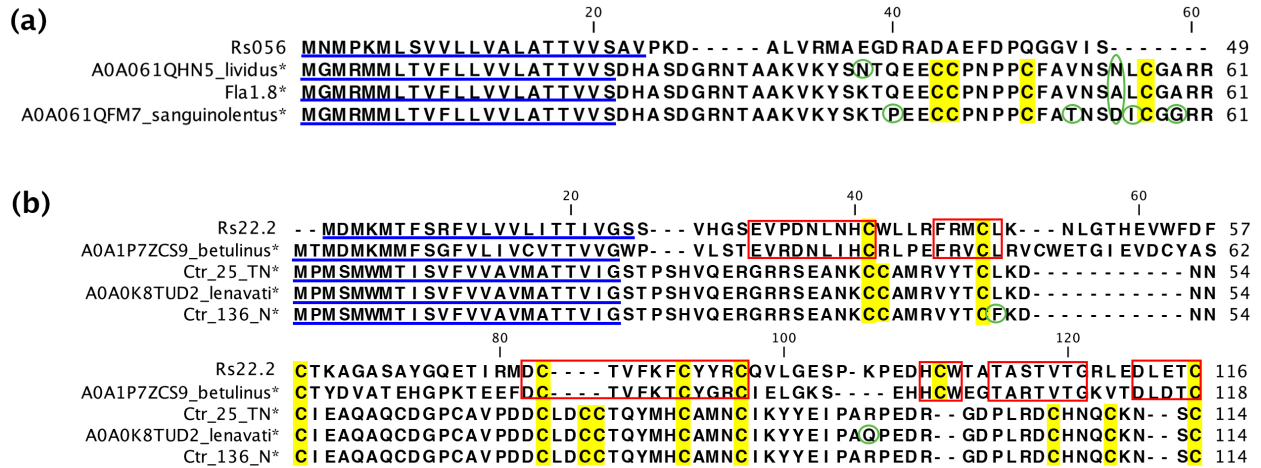


Figura 25. Superfamilia A putativa. (a) conotoxina incompleta que posiblemente sea miembro de la superfamilia A (b) La secuencia Rs22.2 presentó una similitud muy grande con la conotoxina A0A1P7ZCS9_betulinus* (*C. betulinus*). Las secuencias A0A061QHN5_lividus* (*Conus lividus*), Fla1.8* (*C. flavidus*), A0A061QFM7_sanguinolentus* (*Conus sanguinolentus*), A0A0K8TUD2_lenavati* (*C. lenavati*) Ctr_25_TN* y Ctr_136_N* (*C. tribblei*) se muestran como comparación. En color verde se muestran los aminoácidos que no se conservan dentro de ciertas secuencias.

3.7.17 Contenido total del veneno de *C. regularis*

Se identificaron un total de 18 superfamilias de genes de conotoxinas en base a la conservación del péptido señal para *C. regularis*. Además, se identificaron 7 grupos de secuencias que no pudieron ser clasificados dentro de los grupos ya establecidos pero que presentaban características típicas de conotoxinas. El número de secuencias recuperadas varió para cada superfamilia (Figura 26), siendo las que tuvieron un mayor número de secuencias las superfamilias M, O1 y las superfamilias divergentes. Cabe aclarar que esto no es representativo de los niveles de expresión que existen de cada superfamilia dentro del veneno. Por otro lado, vale la pena hacer mención que se encontraron secuencias de familias que tienen pocos miembros o que sólo están representadas en unas cuantas especies, como lo sería la superfamilia N.

En la Tabla 10 se muestra un resumen de las características de las conotoxinas que se encontraron para cada superfamilia, así como las actividades asociadas a cada una de ellas. Varias secuencias permitieron confirmar marcos de cisteínas que no han sido nombrados y además se observaron subgrupos o nuevos grupos con secuencias que sólo han sido reportadas en el UNIPROT pero que no han sido nombradas correctamente.

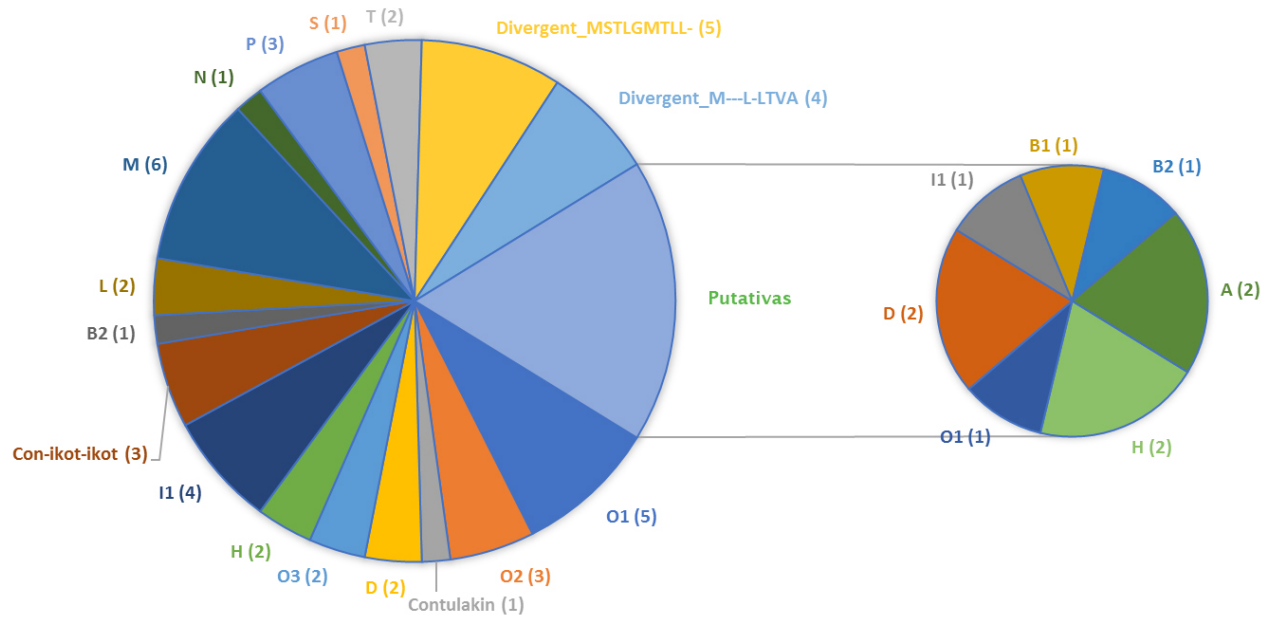


Figura 26. Distribución del número de conotoxinas encontradas para cada una de las superfamilias de genes recuperadas.

Tabla 10. Diversidad funcional encontrada en el transcriptoma de la glándula venenosa de *C. regularis*. Se muestra cada una de las superfamilias identificadas con el marco de cisteínas que presentaron, así como su total de secuencias. Se presenta un resumen de las actividades asociadas para cada grupo.

Superfamilia	Marcos de cisteínas	Secuencias identificadas	Actividades asociadas	Referencia	Anotaciones
O1	VI/VII Sin cisteínas	5	Agonistas de Na _v , bloqueadores de K _v , Na _v y Ca _v .	(Azam y McIntosh, 2009; Callaghan et al., 2008; Sharpe et al., 2001)	Presenta un posible subgrupo
O2	VI/VII XV	3	Moduladores de marcapasos neuronales	(Fainzilber et al., 1991)	Conotoxina identificada a nivel de proteína
C/Contulacina	C-C	1	Receptores de neurotensinas; inhibidores musculares de nAChRs	(Craig et al., 1999; Jimenez et al., 2007)	-
D	XX	2	Inhibidores de nAChRs	(Loughnan et al., 2006)	-

O3	VI/VII Sin cisteínas	2	ND	-	Confirma conotoxinas sin cisteínas para O3
H	VI/VII Sin cisteínas	2	ND	-	Presentó dos posibles secuencias más
I1	VI/VII	4	Agonistas de Na _v	(Fiedler et al., 2008)	Confirma marco VI/VII para I1 Posible subgrupo
Con-ikot-ikot	XXI C-CC-C-C-C-C-C-C- CC-C-C-C CC-C-C-C-CC-C-C-C- C	3	Moduladores de los receptores AMPA	(Walker et al., 2010)	Nuevos marcos de cisteína para la superfamilia
L	XIV	2	Inhibe receptores de nAChRs	(Peng et al., 2006)	-
M	III VI/VII XXVII VIII	6	Síntomas excitatorios en ratones (IC). Agonistas de Na _v	(Mcdougal et al., 2008)	Posible subgrupo Nuevos marcos de cisteína para M.
B2	Sin cisteínas	1	ND	-	Alta conservación de aminoácidos
N	C-C-CC-C-C-C-C-C	1	ND	-	Nuevo marco de cisteínas
P	IX	3	Hiperactividad y espasticidad en ratones (IC)	(Miles et al., 2002)	Posible nuevo grupo
S	VIII	1	Inhibidor de receptores 5-HT ₃ y nAChR	(England et al., 1998; Teichert et al., 2005)	-
T	XV Sin cisteínas	2	Inhibidor de Na _v y de canales presinápticos de Ca ⁺ (o modulador de GPCR), antagonista de sst3 GPCR	(Liu et al., 2007; Petrel et al., 2013; Rigny et al., 1999)	Confirma conotoxinas sin cisteínas para T.

Divergente MSTLGMTLL-	VI/VII XIX XX C-C	5	ND	-	3 posibles nuevas superfamilias de genes. Nuevo miembro de V.
Divergente M---L-LTVA	VI/VII C-C-C-CCC-C-C C-C-C-C-C-C-C-C- C-C-C	4			Posible nueva superfamilia. Confirma nuevo marco de cisteínas.
Posible H	NA	2	NA	NA	Incompletas
Posible O1	VI/VII	1	NA	NA	Posible subgrupo
Posible D	IX	2	NA	NA	Posible MSKL divergente
Posible I1	XXII	1	NA	-	Marco no reportado para I1
Posible B1	XIV	1	NA	-	Marco no reportado para B1
Posible B2	C-C-CC	1	NA	-	Arreglo de cisteínas no reportado
Posibles A	Sin cisteínas XXII	2	Inhibidor de nAChRs, agonista de receptores GABA _B , inhibidores de α 1- adrenorreceptores	(Azam y Mcintosh, 2009; Callaghan et al., 2008; Sharpe et al., 2001)	1 incompleta Posible nuevo grupo

NA, no aplica; ND, no determinado; IC, inyección intracaneal; Ca_v, canales de calcio dependientes de voltaje; K_v, canales de potasio dependientes de voltaje; Na_v, canales de sodio dependientes de voltaje; AMPA, ácido α -amino-3-hidroxi-5-metil-4-isoxazolepropionico; GABA, ácido γ -aminobutírico; GPCR, receptor acoplado a proteína G; nAChR, receptor nicotínico de acetilcolina; sst, somatostatina.

Capítulo 4. Discusión

Los resultados de este trabajo pertenecen a cuatro especies de cónidos (*C. californicus*, *C. mahogani*, *C. regularis*, y *C. ximenes*) que habitan dos áreas muy distintivas de la península de Baja California (México). Con la excepción de *C. californicus*, del que existen reportadas dos bibliotecas de cDNA y algunas conotoxinas a nivel de proteína, el veneno de las restantes especies ha sido poco o nada explorado. Es la primera vez que se secuencian el transcriptoma de estas especies por lo que la información obtenida aquí no únicamente servirá como referencia en estudios posteriores dentro del laboratorio de Inmunología Molecular y Biotoxinas de CICESE; sino que será valiosa para análisis comparativos del veneno y de la toxínología de la familia Conidae en general, ya que la mayoría de los estudios se han visto sesgados hacia especies que habitan el área del IndoPacífico (Elliger et al., 2011; Kaas et al., 2010; Neves et al., 2013).

Con respecto a la ausencia del 28S ribosomal en la verificación de la calidad del ARN, es un fenómeno que se ha presentado en extracciones de ARN de otras especies. Fue reportado por primera vez en insectos y se le denominó como el “hidden break” (rotura oculta) por Ishikawa y colaboradores en 1972 (Winnebeck et al., 2010). En determinadas especies, como son algunos mamíferos y protostomados, el 28S se compone de dos subunidades (α y β) que están unidas la una con la otra a través de puentes de hidrógeno ocultos. Cuando la muestra es calentada, como es en el caso de una extracción de ARN, ocurre una desnaturalización de las dos subunidades que ocasiona que dichos puentes de hidrógenos se rompan, liberando los dos fragmentos (similares en tamaño) que migran hacia la posición del 18S. No obstante, resulta curioso que la presencia de este fenómeno en los cónidos, fue reportado sólo hasta hace poco en extracciones de ARN de *C. episcopatus* (Lavergne et al., 2015).

Es necesario tener en cuenta que cuando se ensambla por primera vez el transcriptoma de una especie sin un genoma de referencia (ensamble *de novo*) presenta la limitante de no contar con un punto de comparación. El objetivo principal en el ensamble de un transcriptoma *de novo* es reconstruir un set de secuencias cortas y contiguas (*contigs*) que reflejen de manera precisa una porción grande de los ARNs realmente transcritos en las células del organismo en estudio. Esto no resulta sencillo porque la cantidad de información retenida de los transcritos del gen original es limitada por las lecturas cortas que produce el secuenciador. El proceso de ensamblado debe colocar las lecturas del mismo transcrito de manera contigua a pesar de las variantes que puedan ser introducidas por polimorfismo y errores de secuenciación (<http://rnaseq.uoregon.edu/#analysis>). Trinity (Grabherr et al., 2011) es un ensamblador que construye y analiza sets de gráficas de Bruijn para reconstruir los transcritos. Ensambla el transcriptoma en tres pasos que corresponden a los módulos llamados "Inchworm", "Chrysalis", y "Butterfly", por su analogía con el

desarrollo de una mariposa. El primer módulo descompone cada lectura en subcadenas cortas de longitud k (k -meros) que se superponen el uno con el otro, permitiéndole construir *contigs* iniciales. El módulo “Chrysalis” agrupa los *contigs* y construye gráficas de Bruijn para cada uno de los grupos, en el que las superposiciones son nodos y los k -meros son bordes de conexión. Finalmente, “Butterfly” simplifica las gráficas cuando le es posible y las reconcilia con las lecturas originales para emitir *contigs* individuales representativos de una variante única, así como de transcritos parálogos. En este caso, debido a que las conotoxinas son de tamaño pequeño se decidió utilizar un *contig* de 100 para poder aumentar la sensibilidad, sin embargo, incrementa el número de transcritos que no tienen traducción ni anotación.

La variación de una familia de secuencias puede ser descrita de manera estadística (Krogh, 1998). Los primeros trabajos enfocados en la elucidación de conopéptidos en un transcriptoma utilizaron BLAST como alineador (Hu et al., 2012; Lluisma et al., 2012); no obstante, como pudimos ver en este trabajo, la identificación de nuevas conotoxinas utilizando el alineador BLAST no es adecuado por dos motivos: en primer lugar las conotoxinas que conforman el veneno de un cónido son muy distintas entre ellas por lo que este tipo de alineador excluye las secuencias que están débilmente relacionadas y que por lo tanto no aprueban la significancia estadística suficiente para ser agrupadas; en segundo lugar únicamente un porcentaje muy bajo de conotoxinas de estas especies ha sido identificado y por lo tanto son pocas las que están reportadas en una base de datos. Las nuevas propuestas son utilizar pHMM (Peng et al., 2016; Robinson et al., 2014) ya que estos modelos tienen la capacidad de “aprender”. Los datos de entrenamiento que se les entrega, en este caso los reportados en ConoServer, permiten entender determinados parámetros que muchas veces están ocultos. Por lo tanto, si la secuencia problema presenta similitudes “débiles” con varios miembros de una familia a la que se le ha construido un pHMM, es probable que pertenezca a dicha familia.

Aunque ConoSorter trabaja con pHMM la cantidad y la calidad de secuencias que entregaba no pudieron ser clasificadas, ni agrupadas en nuevos conjuntos. Las secuencias eran muy largas y resultaba difícil analizarlas. Es probable que esto sea debido a que el e -value establecido por el programa, es muy alto (10) lo que hace que el número de falsos positivos sea muy grande. Por lo tanto, se decidieron construir pHMM propios en los que estuviéramos seguros de cuáles secuencias se estaban utilizando. No obstante, aunque la identificación de conotoxinas utilizando los pHMM fue mejor que utilizando BLAST no rindió el número de secuencias deseadas (alrededor de los cientos). Se sugiere que en futuros trabajos el ARN del ducto venenoso sea extraído en el momento en que son capturados, para de esta manera asegurarse que las toxinas que están siendo expresadas sean lo más cercanas posibles a lo que se encuentra en la naturaleza. También se sugiere continuar alimentando los pHMM con las nuevas conotoxinas que están siendo

reportadas, se debe de construir una base de datos propia en la que se tenga un control de las conotoxinas descritas para el laboratorio, así como las que día con día se reportan en UNIPROT, NCBI y Conoserver.

El veneno del caracol marino *C. regularis* ha sido poco estudiado, y hasta la fecha sólo la conotoxina RsXXIVA se ha informado en la literatura (Bernáldez et al., 2013). En este trabajo se identificaron 57 nuevas secuencias, ya que RsXXIVA no fue recuperada. El número de secuencias descritas para *C. regularis* es baja comparada con otros trabajos de transcriptómica que existen como lo son el caso de *C. gloriamaris* y *C. victoriae*, donde se encontraron alrededor de 100 secuencias diferentes por especie (Robinson et al., 2014; 2017). Por otro lado, aunque con una técnica diferente, en el análisis del transcriptoma de la glándula venenosa de *C. geographus* se identificaron únicamente 63 secuencias (Hu et al., 2012), un número bajo sobre todo si se toma en cuenta que es la especie más estudiada de los cónidos. Debe de tomarse en cuenta que el número de conotoxinas esperado para cada cónido es una estimación y por lo tanto debe de darse lugar a que exista variación en el número de péptidos de una especie a otra.

Para *C. regularis* se identificaron un total de 18 superfamilias de genes de conotoxinas, que junto con los 7 grupos que no pudieron ser clasificados dentro de superfamilias establecidas dan un total de 25 grupos distintos. Esto puede ser comparado con las 31 superfamilias identificadas en el transcriptoma de *C. gloriamaris* o las 20 superfamilias del transcriptoma de *C. victoriae* (Robinson et al., 2014; 2017).

Cabe señalar que varias de las conotoxinas recuperadas para *C. regularis* resultaron ser atípicas porque los marcos de cisteínas que presentaban no han sido nombrados de acuerdo a la nomenclatura establecida. No obstante, la mayoría pudieron ser validados por que fueron reportados en otros trabajos, principalmente por Phuong y colaboradores (2016). Esto no es de sorprenderse ya que el porcentaje de conotoxinas descubiertas hasta la fecha está muy por debajo del total estimado de secuencias. Por lo tanto, en este trabajo se confirmaron marcos de cisteína que habían sido descritos anteriormente, por lo que ahora lo que procede es que sean nombrados de acuerdo a la nomenclatura establecida.

Un gran problema a la hora de clasificar las secuencias es el hecho de que varios conopéptidos han sido reportados en el UNIPROT pero no llegan a ser curados en la base de datos de ConoServer. Como no tienen un nombre correctamente asignado, muchas secuencias se quedan perdidas y no pueden ser añadidas en la construcción de los pHMM. Además, en Conoserver muchas veces se señala que existe un número mayor de secuencias al que realmente está descrito en la página. Tal es el caso de la superfamilia B2, en donde se reportan que existen 21 secuencias pero sólo se tiene acceso a dos de ellas.

Los pHMM contruidos para esta tesis pueden ser mejorados si son enriquecidos con las conotoxinas identificadas para *C. regularis* y después pueden ser utilizados con otras especies. También se propone construir una base de datos propia del laboratorio donde se descarguen todas las secuencias de cónidos que estén reportadas en UNIPROT y en el NCBI, clasificarlas con el pHMM y ser utilizadas para recuperar un mayor número de secuencias.

Por último, varias secuencias de *C. regularis* pudieron ser validadas gracias a que presentaban una alta similitud con secuencias de *C. tribblei* y con secuencias de *C. betulinus* que no han sido propiamente nombradas y que sólo están reportadas en el UNIPROT, pero no en ConoServer. Tanto *C. regularis* como *C. betulinus* son especies vermívoras (Peng et al., 2016), no obstante, nos queda por establecer qué otros puntos tienen en común y por qué presentan conotoxinas tan similares.

Capítulo 5. Conclusiones

- Se confirma que el 28S ribosomal de los cónidos está conformado por dos subunidades. Se sugiere que en futuros trabajos el ARN sea extraída de manera inmediata a la captura de los organismos.
- Aunque el programa Conosorter trabaja con pHMMs, el nivel de e-value que maneja es muy alto y por lo tanto se obtienen falsos positivos que complican la identificación de las conotoxinas reales.
- Los pHMMs construidos para este trabajo permitieron identificar las conotoxinas de *C. regularis*, no obstante, en un futuro deben de ser alimentados con aquellas secuencias que no están reportadas en ConoServer.
- Este trabajo representa la primera aproximación para conocer la totalidad de los péptidos que componen el veneno de cuatro especies de cónidos pobremente estudiados. Las 100 a 200 secuencias de genes de conotoxinas que se espera encontrar en un veneno de cónido es tan sólo un número estimado y por lo tanto cabe la posibilidad de que los 57 conopéptidos encontrados en este trabajo para *C. regularis* no estén alejados del total.
- Se identificaron 18 superfamilias de genes de conotoxinas para *C. regularis*, así como conotoxinas que no pudieron ser clasificadas dentro de superfamilias establecidas. Se identificaron grupos de secuencias que probablemente sean nuevas superfamilias de genes.
- Varias de las conotoxinas de *C. regularis* pudieron ser validadas porque existían en el UNIPROT conotoxinas con las que compartían un alto porcentaje de identidad y que no habían podido ser nombradas bajo la nomenclatura de cónidos. La mayoría de estas conotoxinas pertenecen a *C. tribblei* y *C. betulinus*.
- Se confirmaron nuevos marcos de cisteína que aún no han sido nombrados de acuerdo a la nomenclatura de cónidos pero que si han sido reportados en otros trabajos. Además, se identificaron nuevos marcos de cisteínas.
- Los conopéptidos reportados en esta tesis tienen el potencial de ser desarrollados como posibles candidatos a fármacos.

Parte II. Las proteínas disulfuro isomerasas y su papel esencial en el plegamiento adecuado de las conotoxinas.

Capítulo 1. Introducción

El veneno de los cónidos se conforma principalmente por pequeños péptidos denominados conotoxinas o conopéptidos, que en su mayoría contienen una gran cantidad de residuos de cisteínas y por lo tanto de una elevada tasa de enlaces disulfuros (Norton y Olivera, 2006). La expresión y el plegamiento de las toxinas tiene lugar en el ducto venenoso, hasta la fecha no existe reportado ningún otro tipo de tejido que produzca una diversidad y densidad tan alta de péptidos ricos en residuos de cisteínas (Safavi-Hemami et al., 2016).

La formación correcta de los enlaces disulfuro es una MPT esencial para la estabilidad estructural e integridad funcional de las proteínas secretadas, donde las conotoxinas no son la excepción (Safavi-Hemami et al., 2016). A pesar de que las conotoxinas han sido ampliamente estudiadas debido al valor farmacológico que tienen como pequeñas moléculas, es poco lo que se conoce de la maquinaria requerida para lograr su correcto plegamiento. Se ha propuesto que las proteínas disulfuro isomerasas (PDI) juegan un papel sumamente importante a la hora de adquirir su conformación tridimensional activa (Bulaj et al., 2003).

En este trabajo se utilizaron lecturas de RNA-seq para ensamblar *de novo* los transcriptomas de los ductos venenosos de 4 especies de caracoles marinos de la Familia Conidae. Utilizando el motivo conservado APWCGHCK del dominio activo, se identificaron las posibles enzimas PDI expresadas en cada una de las glándulas. Se encontró una familia de PDI en cada ensamble, lo que indica que cada una de las cuatro especies contiene su propio grupo de proteínas involucradas en el plegamiento oxidativo de sus conotoxinas. Además, la co-localización de las conotoxinas con las enzimas de PDI confirma una dependencia entre ellas para formar los enlaces disulfuro y adquirir su conformación activa.

1.1 Antecedentes

1.1.1 PDI típica

La PDI de humanos, conocida como PDI “típica”, es uno de los miembros de la familia de PDIs que más ha sido estudiado y por lo tanto se utiliza como punto de comparación en el estudio de estas enzimas. Esta proteína de 55 kDa presenta una función dual: a través de su actividad oxidasa cataliza la formación de enlaces disulfuro en polipéptidos, mientras que a través de su actividad isomerasa facilita el ordenamiento de los enlaces disulfuros que estén incorrectos (Wilkinson y Gilbert, 2004). En la Figura 27 se puede observar el arreglo de los cinco dominios vinculados que presenta una PDI “típica”: a, b, b', a' y c (Darby et al., 1999). Los dominios “a” se caracterizan porque son enzimáticamente activos y se componen de alrededor de 100 residuos de aminoácidos que exhiben una similitud muy alta con la proteína procariota de tioredoxina. Además los dominios “a” comparten varias características estructurales y funcionales, incluyendo la presencia del motivo activo “CXXC”, crucial para la actividad redox (Kemink et al., 1997; Ramos et al., 2005); dicho motivo se ha demostrado que puede presentarse como una secuencia larga de 8 residuos (APWCGHCK) no obstante pocas veces ha sido utilizado (Knodler et al., 1999; Tachikawa et al., 1995). Por otro lado, los dominios “b” son enzimáticamente inactivos y a pesar de que son capaces de adoptar un patrón de plegamiento similar al de la tioredoxina, su secuencia de aminoácidos no guarda algún parecido (Wilkinson y Gilbert, 2004). Ambos dominios (a y b) contribuyen a la unión de la enzima con su sustrato, sin embargo es el dominio b' el que provee el sitio de unión principal (Klappa et al., 1998). Por otro lado el dominio “c”, localizado cerca del carboxilo terminal, es rico en residuos ácidos y tiene una alta capacidad de unión al Ca^{2+} (Lucero y Kaminer, 1999). Las PDIs son proteínas residentes del RE por lo que contienen un péptido señal en su amino terminal y una secuencia canónica en su carboxilo terminal (KDEL) que funciona como señal de retención a dicho organelo (Denecke et al., 1992).

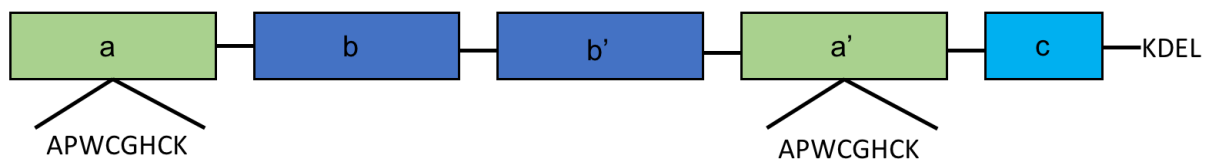


Figura 27. Organización de los dominios (a, b, b', a' y c) en una PDI típica (Modificado de Darby et al., 1999).

1.1.2 PDI de cónidos

La vía de síntesis de las conotoxinas ha sido un proceso del que se ha podido descifrar poco. Se sabe que todos los conopéptidos tienen una secuencia señal en el N-terminal sumamente conservada que dirige a la proteína a la membrana del RE de células del ducto venenoso. Esta vía de secreción transfiere a la conotoxina del citosol, un compartimiento bajo en Ca^{2+} y con un ambiente reductor, a la matriz del RE, que es rico en Ca^{2+} y con ambiente oxidante (Bulaj et al., 2003). De esta manera, en este organelo comienza la formación de enlaces disulfuros y probablemente también se originan otras MPT; aunque se piensa que la mayoría de estas ocurren en el aparato de Golgi. Sin embargo, los subsecuentes aspectos celulares y moleculares del proceso proteolítico del precursor así como cómo el empaquetamiento de las conotoxinas en sus vesículas secretorias son aspectos que se desconocen (Bulaj y Olivera, 2008).

Las conotoxinas pueden clasificarse de distintas maneras: dependiendo de su secuencia señal, del blanco molecular sobre el que estén actuando o de la manera en que estén dispuestos sus residuos de cisteínas. El arreglo de cisteínas es sumamente conservado entre grupos de conotoxinas, lo que les permite adquirir una configuración de sus enlaces disulfuros muy particular (“Marco de cisteínas”). Esto es sumamente determinante en la estructura terciaria del polipéptido porque, aunque existe una hipermutación en los aminoácidos que se encuentran entre los residuos de cisteínas, el marco de cisteínas permanece conservado y por lo tanto se genera un andamiaje característico (Bulaj et al., 2003).

Además de las conotoxinas, las PDI son uno de los grupos de proteínas más abundantes que han sido identificados en el ducto venenoso de los cónidos. Mediante la construcción de bibliotecas de cDNA; Bulaj y colaboradores (2013) identificaron la presencia de múltiples isoformas de PDI en extractos de ductos venenosos de diferentes cónidos. Posteriormente, Safavi-Hemami *et al.* (2011) identificaron (a nivel de proteína) un gran número de isoformas de PDI en las glándulas venenosas de los caracoles *Conus novaeollandiae* y *C. victoriae*. Además demostraron que las PDI juegan un papel muy importante tanto en la síntesis como la configuración de los residuos de cisteínas de las conotoxinas (Safavi-Hemami et al., 2012), un proceso que es altamente regulado y que requiere de la ayuda de una maquinaria de chaperonas sumamente eficiente. Por lo tanto, conocer cómo se lleva a cabo la modulación de la conformación de las conotoxinas nos permitirá entender la diversidad funcional y molecular de estas moléculas.

Finalmente, cabe mencionar que a la par en que los resultados de este capítulo de tesis de doctorado fueron publicados (Figueroa-Montiel et al., 2016) (en el mismo mes y año), Safavi-Hemami y colaboradores (2016) identificaron diversas secuencias de PDI de cónidos en los transcriptomas ensamblados de 17

especies de caracoles marinos diferentes. Además, identificaron una familia de PDI específica de cónidos, la cual sería la familia de oxidoreductasa más grande descrita para un solo género. Estas enzimas fueron denominadas como PDI específicas de cónidos (csPDI, por sus siglas en inglés) y, en promedio, encontraron 2.6 secuencias por especie.

1.2 Justificación

Gracias al valor farmacológico que poseen, el estudio del veneno de los cónidos se ha centrado en la caracterización de sus toxinas. No obstante, se ha dejado de lado al resto de los componentes que se encuentran en la glándula venenosa. Al igual que otros polipéptidos eucariotas con un patrón específico de enlaces disulfuro, las conotoxinas requieren de una maquinaria de chaperonas compleja que les permita adquirir su conformación tridimensional necesaria para desempeñar su función. Se sabe que las PDI son indispensables en el plegamiento y maduración estructural de las conotoxinas, no obstante, es poco lo que se conoce sobre ellas. Conocer las PDIs que conforman el veneno de cada uno de los cónidos es un factor indispensable para, en un futuro, entender cómo pueden ser sintetizadas y aprovechadas dentro de la industria farmacológica.

1.3 Hipótesis

Los transcriptomas del ducto venenoso de cónidos codifican para polipéptidos con homología a enzimas de la familia PDI.

1.4 Objetivos

1.4.1 Objetivo general

Identificar los genes codificantes para polipéptidos involucrados en la síntesis del veneno, en particular de las enzimas encargadas del correcto plegamiento de las conotoxinas, mediante la búsqueda en el transcriptoma ensamblado *de novo* de cuatro especies de cónidos de la península de Baja California

1.4.2 Objetivos específicos

- Caracterizar las familias de PDI que conforman el veneno de cuatro especies de la Familia Conidae, mediante un análisis comparativo de los tres primeros niveles de estructuración.

Capítulo 2. Metodología

La metodología correspondiente a la colecta de especímenes, extracción de ARN y secuenciación del transcriptoma de la glándula venenosa corresponden a las secciones 2.1 y 2.2 de la Parte I de esta tesis.

2.1 Ensamble del transcriptoma

Se realizó un ensamble *de novo* que permitió agrupar las lecturas cortas en *contigs*. Se utilizó el protocolo estándar (Grabherr et al., 2011) ejecutando el parámetro específico de cadena y normalizando las lecturas y, a diferencia de otras búsquedas (p.ej. de toxinas), se utilizó el tamaño mínimo de *contig* estándar (200). El ensamble del transcriptoma se llevó a cabo utilizando los siguientes parámetros: *Trinity --seqType fq --max_memory 60G --min_contig_lenght 200 --normalize_max_read_cov 30 --left R1.fastq --right R2.fastq --SS_lib_type RF --CPU 27 --no_bowtie* (Grabherr et al., 2011). Para medir la calidad del ensamble, se realizó una estadística del número de genes e isoformas, así como de su contigüidad, corriendo el script *TrinityStats.pl*.

2.2 Análisis Bioinformático

La identificación de las secuencias de PDI presentes en el transcriptoma de cada una de las especies se llevó a cabo mediante un análisis *in silico* que consistió de varios pasos. En primer lugar, el script de Transdecoder (incluido en la paquetería de Trinity) permitió identificar todos los ORFs que cumplieran con la mínima longitud de aminoácidos requerida para codificar una secuencia de proteína. Posteriormente, el motivo estructural “APWCGHCK” fue utilizado como sonda para identificar aquellas secuencias codificantes de proteínas similares a PDIs. Estas secuencias fueron nombradas utilizando las iniciales del organismo en cuestión seguidas de un número consecutivo (por ejemplo, para la primera secuencia de *C. californicus* se designó Cc01).

Cada una de las secuencias fueron analizadas de manera individual, la primera aproximación fue utilizar el alineador BLAST contra la base de datos no redundante del NCBI (<https://www.ncbi.nlm.nih.gov/>) que permitió determinar cuál era la identidad de estas secuencias. Este primer ejercicio reveló que todas las

secuencias obtenidas eran significativamente similares a PDIs reportadas (incluidas PDIs de *Conus*) ya que en promedio presentaban una similitud del 65.4 % (oscilando desde un 44 a un 98 %). Los dominios de tiorredoxina funcionales fueron predichos mediante homología utilizando la herramienta de búsqueda Conserved Domain Database (CDD) que pertenece al servidor del NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). El arreglo de dichos dominios se obtuvo utilizando los servidores SMART (<http://smart.embl-heidelberg.de>) y Pfam (<http://pfam.xfam.org>). El alineamiento múltiple de las secuencias se obtuvo utilizando el programa Clustal Omega que se encuentra en el servidor EMBL-EBI (<http://www.ebi.ac.uk/Tools/msa/clustalo>) para posteriormente ser editadas en la paquetería Genious v. 4.8.5 (Kearse et al., 2012). La identificación del sitio de hidrólisis del péptido señal fue realizada con ayuda del servidor SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP>) y el peso molecular fue calculado utilizando la plataforma EMBOSS Pepstats (http://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats).

2.3 Predicción de la estructura 3D de las PDI de cónidos

Las PDI presentan un arreglo tridimensional (3D) típico en el que sus 4 dominios (a, b, a' y b') son capaces de adquirir una conformación similar al de la tiorredoxina. Se decidió buscar, en cada una de las familias de PDI encontradas, una secuencia que presentara un arreglo tridimensional al esperado en una PDI típica. En primer lugar, se realizó un alineamiento de las PDI encontradas utilizando como templatado a PDI 4EKZ (PDI de humano completamente caracterizada). Las secuencias con mayor similitud al templatado (Cc03_i1, Cm05, Cr09 y Cx11) fueron las seleccionadas para predecir su estructura tridimensional. Una vez seleccionada la secuencia representativa de cada especie se realizó un BLAST de cada una de ellas y, en base al porcentaje de similitud, se seleccionaron 6 secuencias (4EKZ, 1MEK, 3UEM, 4JU5, 2KI8, 1X5C) para ser utilizados como molde. Estos andamios estructurales fueron obtenidos del Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>). Finalmente, se realizó un modelado por homología con un método avanzado utilizando MODELLER v. 9.14 (Eswar et al., 2007).

El refinamiento del modelado de las estructuras moleculares se realizó por dinámica molecular a través del método “simulated annealing”, utilizando el software NAMD Scalable Molecular Dynamics v.2.10 del grupo de Biofísica Teórica y Computacional de la Universidad de Illinois (Urbana-Champaign). Los parámetros utilizados fueron 16 ns, presión estándar y una temperatura constante de 300 °K. Este trabajo fue realizado en MIZTLI, la supercomputadora de la UNAM. De cada una de las secuencias seleccionadas

se agruparon las diferentes conformaciones de las estructuras dependiendo de la estabilidad energética y los cambios estructurales que suscitaron durante la simulación. Esto tuvo la finalidad de escoger la mejor estructura conformacional que tuviera el mayor tiempo de existencia. Posteriormente, para cada una de las estructuras seleccionadas, se realizó un análisis por gráficas de Ramachadran utilizando el servidor interactivo de PROCHECK de la Universidad de California, Los Angeles (Wang y Zhang, 2013). Estos gráficos permitieron conocer la calidad del modelo con respecto a su estructura; ya que comparan los mejores resultados de cada uno de los modelos con una PDI de humanos descrita mediante refracción de rayos-X.

2.4 Construcción de árbol filogenético de las PDI de cónidos reportadas a la fecha

Se construyó un árbol filogenético con las secuencias completas encontradas en cada transcriptoma junto con las reportadas por Safavi-Hemami et al (2016). Se utilizó como raíz una PDI de *Crassostrea gigas* (Gen Bank: XM_011447406.2) y una PDI de *Haliotis discus* (Gen Bank: EF103409.1). El programa utilizado fue MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 (Kumar et al., 2016). Se realizó un alineamiento de las secuencias utilizando MUSCLE con los parámetros predeterminados. A continuación, se construyó un árbol filogenético utilizando el método estadístico Neighbor-joining. La prueba de filogenia utilizada fue mediante el método Bootstrap con 1000 iteraciones. Una vez obtenido el árbol se fijó como raíz a las secuencias mencionadas anteriormente.

Capítulo 3. Resultados

Los resultados de la extracción de ARN y de los análisis de calidad de las secuencias están reportados en las secciones 3.1 y 3.2 de la Parte I de esta tesis.

3.1 Ensamble del transcriptoma con el programa de cómputo Trinity

El resumen de las estadísticas obtenidas con Trinity, utilizando un *contig* mínimo de 200, se encuentra en la Tabla 11. Los ensambles de los transcriptomas mostraron un valor de N50 que oscila entre 490 y 746 (Tabla 11). Lavergne *et al.* (2015) utilizaron el mismo ensamblador para el transcriptoma de *C. episcopatus* y encontraron un valor similar al nuestro (N50=478). Esto sugiere que la longitud del transcrito promedio, estimado para un transcriptoma del ducto venenoso de un cónido (utilizando un *contig* de 200) en Trinity es de 605.6 ± 116.9 .

Tabla 11. Resultados de las estadísticas obtenidas de los 4 ensambles realizados con Trinity

	<i>C. californicus</i>	<i>C. mahogani</i>	<i>C. regularis</i>	<i>C. ximenes</i>
Transcritos totales	24, 074	84,677	61,896	92, 433
<i>Contig</i> promedio	459.35	553.36	552.31	587.86
N10	1,477	2,439	2,254	2,411
N50	490	661	653	746

3.2 Análisis estructural de las familias de PDI de los cónidos

En la Tabla 12 se puede observar la nomenclatura que le corresponde a cada una de las secuencias de PDI obtenidas para cada especie; así como el resultado comparativo de cada alineamiento obtenido después de un BLAST. Se realizó una distinción entre las secuencias identificadas si estas se correspondían con una secuencia completa o parcial. Una vez realizado el alineamiento con el programa BLAST se identificó cuál proteína de la base de datos del NCBI era la que mayor porcentaje de identidad tenía con la pregunta problema y cuáles eran las características del alineamiento.

Tabla 12. Análisis de las secuencias obtenidas para cada una de las especies estudiadas.

Especie	Nomenclatura para el ORF		Hit top en Blast	
	Nombre	Estado	ACCN (Proteína, organismo)	ID (E)
<i>C. californicus</i>	Cc01	Parcial	XP_011419975 (PDI A5-like, <i>Crassostrea gigas</i>)	54 % (8e-70)
	Cc02	Parcial	ADZ76591 (PDI, <i>C. eburneus</i>)	90 % (0.0)
	Cc03_i1	Completa	ADZ76591 (PDI, <i>C. eburneus</i>)	68 % (0.0)
	Cc03_i4	Completa	ADZ76591 (PDI, <i>C. eburneus</i>)	68 % (0.0)
	Cc04	Completa	ADZ76591 (PDI, <i>C. eburneus</i>)	65 % (0.0)
	Cc05	Completa	XP_005100750 (PDI A3-like, <i>Aplysia californica</i>)	63 % (0.0)
	Cc06	Parcial	ADZ76590 (PDI, <i>C. virgo</i>)	80 % (1e-50)
<i>C. mahogani</i>	Cm01	Parcial	EKC22564 (PDI A4, <i>C. gigas</i>)	64 % (6e-60)
	Cm02	Completa	AEE36485 (PDI 1, <i>Fenneropenaeus chinensis</i>)	51 % (9e-134)
	Cm03	Parcial	ADZ76591 (PDI, <i>C. eburneus</i>)	68 % (0.0)
	Cm04_i1	Completa	XP_011453191 (PDI A3-like, <i>C. gigas</i>)	65 % (0.0)
	Cm04_i2	Parcial	XP_011453191 (PDI A3-like, <i>C. gigas</i>)	62 % (2e-64)
	Cm05	Completa	ADZ76593 (PDI, <i>C. betulinus</i>)	86 % (0.0)
	Cm06	Parcial	XP_002739058 (PDI A5-like, <i>Saccoglossus kowalevskii</i>)	51 % (9e-33)
	Cm07	Completa	XP_005108921 (PDI A6-like, <i>A. californica</i>)	71 % (0.0)
	Cm08	Parcial	XP_002732815 (PDI A4-like, <i>S. kowalevskii</i>)	44 % (3e-14)
	Cm09	Parcial	ABF48564 (PDI, <i>C. marmoreus</i>)	71 % (7e-42)
<i>C. regularis</i>	Cr01	Parcial	GAA48005 (PDI A1, parcial, <i>Clonorchis sinensis</i>)	60 % (4e-85)
	Cr02	Completa	XP_001602967 (probable PDI A6, <i>Nasonia vitripennis</i>)	64 % (0.0)
	Cr03	Parcial	XP_002732815 (PDI A4-like, <i>S. kowalevskii</i>)	50 % (7e-35)
	Cr04	Parcial	EKC22564 (PDI A4, <i>C. gigas</i>)	71 % (3e-53)
	Cr05	Parcial	AEE36485 (PDI 1, <i>Fenneropenaeus chinensis</i>)	57 % (9e-41)
	Cr06	Parcial	XP_011429706 (P5-like, <i>C. gigas</i>)	48 % (5e-80)

	Cr07_i1	Completa	ADZ76591 (PDI, <i>C. eburneus</i>)	65 % (0.0)
	Cr07_i2	Completa	ADZ76590 (PDI, <i>C. virgo</i>)	67 % (0.0)
	Cr07_i3	Parcial	ADZ76590 (PDI, <i>C. virgo</i>)	66 % (0.0)
	Cr08	Completa	XP_005100750 (PDI A3-like, <i>A. californica</i>)	64 % (0.0)
	Cr09	Completa	ADZ76591 (PDI, <i>C. eburneus</i>)	95 % (0.0)
	Cr10	Parcial	ADZ76593 (PDI, <i>C. betulinus</i>)	98 % (4e-76)
	Cr11	Parcial	XP_011679321 (PDI A5, <i>Strongylocentrotus purpuratus</i>)	51 % (7e-30)
<i>C. ximenes</i>	Cx01	Parcial	GAA48005 (PDI A1, parcial, <i>C. sinensis</i>)	62 % (2e-111)
	Cx02	Completa	ADZ76591 (PDI, <i>C. eburneus</i>)	68 % (0.0)
	Cx03	Completa	XP_011453191 (PDI A3-like, <i>C. gigas</i>)	62 % (0.0)
	Cx04	Parcial	XP_002739058 (PDI A5-like, <i>S. kowalevskii</i>)	46 % (1e-51)
	Cx05	Parcial	XP_008200981 (P5, <i>Tribolium castaneum</i>)	50 % (4e-82)
	Cx06	Parcial	GAA48005 (PDI A1, parcial, <i>C. sinensis</i>)	52 % (5e-77)
	Cx07	Parcial	XP_005108921 (PDI A6-like, <i>A. californica</i>)	73 % (5e-136)
	Cx08	Parcial	ADZ76591 (PDI, <i>C. eburneus</i>)	75 % (1e-41)
	Cx09	Parcial	XP_005099456 (PDI A4-like, <i>A. californica</i>)	58 % (0.0)
	Cx10	Parcial	ABJ89816 (PDI ER-60, parcial, <i>C. sinensis</i>)	58 % (2e-113)
	Cx11	Completa	ADZ76593 (PDI, <i>C. betulinus</i>)	86 % (0.0)

ACCN, número de acceso; ID, porcentaje de identidad; E, e-value.

Debido a que las proteínas de PDI comparten características estructurales muy similares como son los dominios similares a la tiorredoxina, resulta lógico analizar las PDI encontradas en este trabajo mediante acercamientos comparativos con las PDI ya establecidas. Por lo general, una estructura de PDI se compone de 1 a 3 dominios activos similares a la tiorredoxina (Ellgaard y Ruddock, 2005). En este estudio se encontró el arreglo de dominios de una PDI típica (a-b-b'-a') lo que indica que en las glándulas venenosas de los cónidos está ocurriendo una actividad oxidoreductiva involucrada en el plegamiento de las toxinas de estos animales.

3.2.1 PDIs de *C. ximenes*

11 secuencias fueron encontradas en el transcriptoma del ducto venenoso de *C. ximenes*; 3 de estas secuencias presentaron un ORF completo (Cx02, Cx03, and Cx11) y el resto presentaron regiones codificantes parciales (CDS, del inglés *coding DNA sequence*) (Tabla 13). La búsqueda con BLAST reveló que existe un alto grado de identidad con PDI homólogas que ya han sido anotadas, incluidas algunas de cónidos. El mayor porcentaje de identidad de todas las secuencias se observa entre Cx11 y una PDI de *C. betulinus* (86 %). Las PDI completas exhiben características estructurales de una PDI típica: peso molecular de 55 a 60 kDa, un arreglo de dominios a-b-b'-a', el motivo conservado (CXXC) en el dominio activo "a", y un tetrapéptido en el carboxilo terminal similar a la señal de retención en el RE (KDEL). El resto de las secuencias que se encontraron de manera incompleta también mostraron características que vale la pena mencionar. Todas presentaron por lo menos un dominio "a", con su sitio activo característico, y aquellos que presentaron carboxilo terminal (Cx01, Cx04, and Cx08-Cx10) exhibieron el tetrapéptido de señal de retención al ER.

Tabla 13. Análisis estructural de los diferentes miembros de la familia de PDI de *C. ximenes*.

Nombre	PM (kDa)	Organización de los dominios	Motivo del Sitio Activo	Secuencia de retención del RE
Cx01	ND	(*)-b-a	CGHC	RDEL
Cx02	56.21	a-b-b'-a'	CGHC, CGHC	RDEL
Cx03	57.94	a-b-b'-a'	CGHC, CGHC	KSEL
Cx04	ND	(*)-a	CGHC	KDEL
Cx05	ND	(*)-a-b-b'-a'-(*)	CGHC, CGHC	ND
Cx06	ND	(*)-a-a'-(*)	CGHC, CGHC	ND
Cx07	ND	(*)-a-b-(*)	CGHC	ND
Cx08	ND	(*)-a	CGHC	KDEL
Cx09	ND	(*)-a	CGHC	RDEL
Cx10	ND	(*)-b-a	CGHC	KVDL
Cx11	56.06	a-b-b'-a'	CGHC, CGHC	RDEL

ND: no determinado

3.2.2 PDIs de *C. regularis*

Se recuperaron 13 secuencias de PDI del transcriptoma ensamblado de *C. regularis*. Las secuencias que presentaron ORF completos fueron: Cr02, Cr07_i1, Cr07_i2, Cr08, y Cr09; mientras que el resto presentaron CDS parciales (Tabla 14). Después de realizar un BLAST, se encontró que las secuencias presentaban un alto grado de identidad con PDIs que ya han sido anotadas. Cabe mencionar que se encontraron tres isoformas de Cr07 con un alto grado de identidad (65–67 %) a PDI de *Conus* previamente reportadas (*C. eburneus* y *C. virgo*). Además, el mayor porcentaje de identidad se observó entre la secuencia Cr09 y una PDI de *C. eburneus* (95 %). Los polipéptidos Cr07_i1, Cr07_i2, Cr08, y Cr09 presentaron un arreglo de sus dominios característico al de una PDI “típica”. Por otro lado, el polipéptido completo Cr02 presenta características a las de P5/ERP5 (Tabla 14): un peso molecular alrededor de los 48 kDa, dos dominios “a” (ambos con el motivo conservado CXXC) y un dominio “b”, organizados como a-a'-b; así como una secuencia rica en residuos ácidos antes de la secuencia de retención del RE (KDEL) del carboxilo terminal. Para el resto de las secuencias incompletas, su análisis estructural se muestra en la Tabla 14: todas presentan por lo menos un dominio “a”, con su motivo CXXC, y aquellas que presentan un C-terminal (Cr06 and Cr10) contienen un tetrapéptido parecido a la señal de retención del RE.

Tabla 14. Análisis estructural de los diferentes miembros de la familia de PDI de *C. regularis*

Nombre	PM (kDa)	Organización de los dominios	Motivo del Sitio Activo	Secuencia de retención del RE
Cr01	ND	(*)-b-a-(*)	CGHC	ND
Cr02	47.47	a-a'-b	CGHC, CGHC	KDEL
Cr03	ND	(*)-a-(*)	CGHC	ND
Cr04	ND	(*)-a-(*)	CGHC	ND
Cr05	ND	(*)-a-(*)	CGHC	ND
Cr06	ND	(*)-a-a'	CGHC, CGHC	HTEL
Cr07_i1	55.35	a-b-b'-a'	CGHC, CKYC	KDEL
Cr07_i2	55.50	a-b-b'-a'	CGHC, CGHC	KDEL
Cr07_i3	ND	a-b-b'-a'-(*)	CGHC, CGHC	ND
Cr08	57.48	a-b-b'-a'	CGHC, CGHC	KTEL
Cr09	56.50	a-b-b'-a'	CGHC, CGHC	KDEL
Cr10	ND	(*)-a	CGHC	KDEL
Cr11	ND	(*)-a-(*)	CGHC	ND

ND: no determinado.

3.2.3 PDIs de *C. mahogani*

La familia de PDI de *C. mahogani* se compone de 10 secuencias. Únicamente Cm02, Cm04_i1, Cm05, y Cm07 presentan ORFs completos (Tabla 15). Todas las secuencias presentaron un alto grado de identidad con PDIs anotadas en el NCBI, siendo Cm05 la que presentó mayor similitud con una PDI de *Conus* (85 % con *C. betulinus*). Para Cm06 se encontraron dos secuencias idénticas (diferiendo únicamente en el residuo 37), sólo una de las secuencias fue utilizada para el análisis estructural. Además, se encontraron dos isoformas de Cm04 que compartían un alto grado de similitud (hasta del 65 %) con una PDI de *C. gigas*. Las secuencias completas Cm04_i1 y Cm05 exhiben características estructurales de una PDI “típica”; mientras que el polipéptido completo Cm07 presenta características similares a las de P5/ERP5 (Tabla 15). Por otro lado, la secuencia completa Cm02 presenta características similares a las de la PDI TXNDC5/ERP46: un peso molecular alrededor de 46 kDa, 3 dominios “a” (todos con su motivo conservado CXXC), y un tetrapéptido en el C-terminal parecido a la secuencia de retención del RE (KDEL). Aunque existen 6 secuencias con CDS parciales, el análisis estructural nos muestra ciertas características que valen la pena señalar: todas tienen por lo menos un dominio “a” y aquellas que tienen un carboxilo terminal completo (Cm01 y Cm06) presentan la etiqueta de retención del RE (KDEL).

Tabla 15. Análisis estructural de los diferentes miembros de la familia de PDI de *C. mahogani*.

Nombre	PM (kDa)	Organización de los dominios	Motivo del Sitio Activo	Secuencia de retención del RE
Cm01	ND	(*)-a	CGHC	KDEL
Cm02	43.90	a'-a-a'	CGHC, CGHC, CGHC	HTEL
Cm03	ND	a-b-b'-a'-(*)	CGHC, CGHC	ND
Cm04_i1	57.86	a-b-b'-a'	CGHC, CGHC	KSEL
Cm04_i2	ND	a-(*)	CGHC	ND
Cm05	56.03	a-b-b'-a'	CGHC, CGHC	RDEL
Cm06	ND	(*)-a	CGHC	KDEL
Cm07	47.86	a-a'-b	CGHC, CGHC	KDEL
Cm08	ND	a-(*)	CGHC	ND
Cm09	ND	(*)-a-(*)	CGHC	ND

ND, no determinado.

3.2.4 PDIs de *C. californicus*

Se recuperaron 7 secuencias de PDI en el transcriptoma del ducto venenoso de *C. californicus*. 4 de las secuencias presentaron ORF completos (Cc03_i1, Cc03_i4, Cc04 y Cc05), mientras que el resto presentó CDS parciales (Tabla 16). Al igual que en las otras especies, después de realizar un análisis con BLAST, se determinó que todas las secuencias presentaban un alto grado de identidad con PDI homólogas (incluidas algunas de *Conus*). Debido a que se encontraron 4 variantes de Cc03 (tres alelos Cc03_i1, _i2, y _i3, difiriendo en los residuos 288 y 387; y una isoforma Cc03_i4), solamente Cc03_i1 y Cc03_i4 fueron tomados en cuenta para el análisis estructural. El mayor grado de similitud fue observado entre las variantes de Cc03 y una PDI de *C. eburneus* (68 %). Los polipéptidos Cc03_i1, Cc03_i4, y Cc05, que codificaban ORF completos, presentan características de una PDI “típica”; mientras que Cc04, también una secuencia completa, presenta características similares a P5/ERP5 (Tabla 16). El análisis estructural de 3 secuencias con un CDS incompleto muestra las mismas características que las secuencias incompletas del resto de los transcriptomas (Tabla 16); siendo Cc01 y Cc02 las que muestran el tetrapéptido de la señal de retención del RE.

Tabla 16. Análisis estructural de los diferentes miembros de la familia de PDI de *C. californicus*.

Nombre	PM (kDa)	Organización de los dominios	Motivo del Sitio Activo	Secuencia de retención del RE
Cc01	ND	(*)-a-a'	CGHC, CGHC	KEEL
Cc02	ND	(*)-a-b-b'-a	CGHC	RDEL
Cc03_i1	56.27	a-b-b'-a'	CGHC, CGHC	KDEL
Cc03_i4	57.45	a-b-b'-a'	CGHC, CGHC	KDEL
Cc04	48.22	a-a'-b	CGHC, CGHC	KEEL
Cc05	55.73	a-b-b'-a	CGHC, CGHC	KTEL
Cc06	ND	(*)-a-(*)	CGHC	ND

ND, no determinado

3.3 Alineamiento de las secuencias de PDI

Se realizaron alineamientos de los dominios catalíticos de cada una de las especies. Se analizó particularmente cómo se presentaba el motivo conservado “CXXC”. Para *C. californicus* (Figura 28a) se

puede establecer como **(V/I)(E/M)FYAPWCGHC(K/Q)**; ya que para la secuencia CC04 se presenta una Q en lugar de una K. Para esta especie podemos ver que existe otro grupo de aminoácidos que se conserva (GYPT). Las letras resaltadas en negritas dentro de los dominios son los aminoácidos que aparecieron para la mayoría de las especies. Para *C. ximenes* (Figura 28b) se representa como **(V/I)(E/M/K)(F/L)(F/Y/H)APWCGHCK**; para *C. regularis* (Figura 28c) el motivo sería **(V/I)(E/M/K)(F/L)(F/Y)APWCGHC(K/Q)** y finalmente, para *C. mahogani* (Figura 28d) se representa como **(V/I)(E/M/K)(F/L)(F/Y)APWCGHCK(K/Q)**.

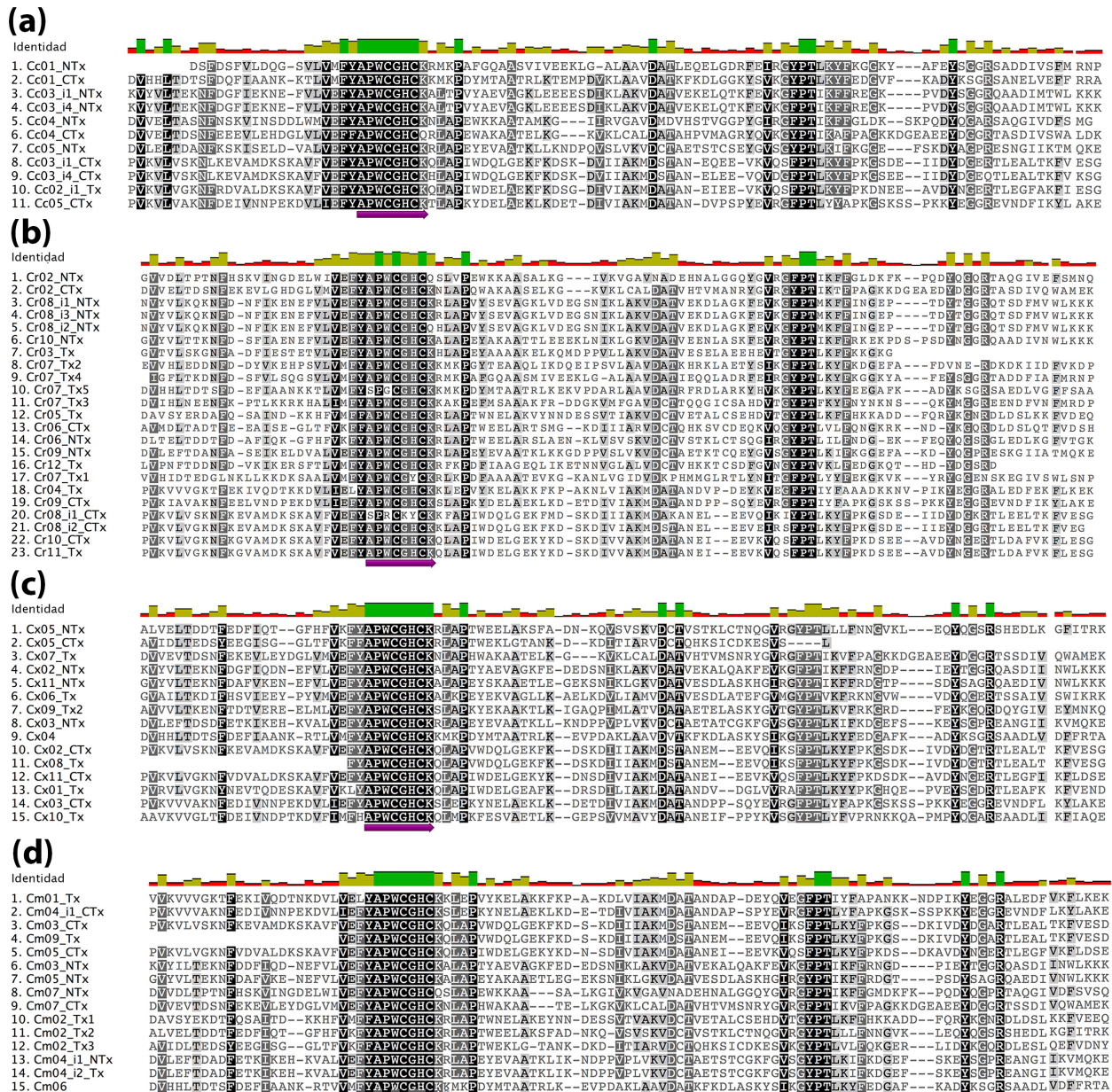


Figura 28. Comparación de los dominios catalíticos de las secuencias de PDIs de (a) *C. californicus*, (b) *C. regularis*, (c) *C. ximenes* y (d) *C. mahogani*.

3.4 Predicción de la estructura tridimensional de las PDI de los cónidos.

De cada uno de los transcriptomas, se seleccionó una secuencia que presentara la similitud más alta con la PDI de humano PDB 4EKZ, la cual ha sido resuelta por cristalografía de rayos X. Estas PDI seleccionadas fueron utilizadas como representativas de cada una de las familias de PDI para predecir un modelo tridimensional. Para *C. californicus* se seleccionó Cc03_i1, para *C. mahogani* fue Cm05, para *C. regularis* se utilizó Cr09 y finalmente, para *C. ximenes* fue Cx11. Después de un refinamiento del modelo, se obtuvieron las diferentes conformaciones de las estructuras y se pudo demostrar que las diferentes secuencias adquirirían una conformación tridimensional al esperado para una PDI típica (Figura 29). Su estructura sugiere que presentan una función relacionada con el proceso oxidativo del plegamiento de las proteínas nacientes. También se puede observar que los 4 dominios de las PDI de los cónidos (a-b-a' y b') adquieren una conformación similar al de la tioredoxina. Los modelos 3D refinados de las PDI de cónidos muestran tener similitud a las estructuras obtenidas mediante métodos experimentales. El porcentaje de residuo favorece satisfactoriamente las regiones (A, B, L) para Cc03_i1 (80.2 %), Cm05 (76.9 %), Cr09 (81.2 %) y Cx11 (82.9 %). Esto es similar al porcentaje obtenido en una PDI de humano obtenida mediante difracción de rayos X (93.2 %).

3.5 Árbol filogenético

La construcción del árbol filogenético de las secuencias de csPDI y PDI permitió realizar una primera aproximación de cómo se relacionaban las PDI encontradas en este trabajo con aquellas reportadas por Safavi-Hemami y colaboradores (2016) (Figura 30). Cómo era de esperarse se formaron dos grupos principales: las secuencias de csPDI y las secuencias PDI. Sólo una de las secuencias de csPDI (csPDI GH/GH de *C. bocki*) se agrupó con las secuencias de PDI (resaltada en gris). A continuación, se describe como se agruparon las secuencias descritas en este trabajo

- Cr07i1 y Cr07i2 (resaltadas en azul) se agruparon con las secuencias de csPDI a pesar de que no presentaban las características descritas para este tipo de secuencias. Sólo una secuencia de csPDI se agrupó con las secuencias regulares de PDI.
- En un grupo dentro de las csPDI pero más cercano a las PDI regulares, resaltadas en color amarillo, se situaron las secuencias Cx02, CC03i1 y CC03i4.

- En un grupo dentro de las PDI regulares y cercano a las csPDI (resaltadas en lila) se situaron las secuencias Cm05 y Cx11.
- La secuencia Cr09 (resaltada en verde) se agrupó perfectamente con todas las secuencias de PDI.
- Finalmente, se pudo observar un grupo externo (cercano a las PDI, resaltado en rosa) con las secuencias Cm07, Cr02, CC04, Cm02, CC05, Cr08, Cm04i1 y Cx03).

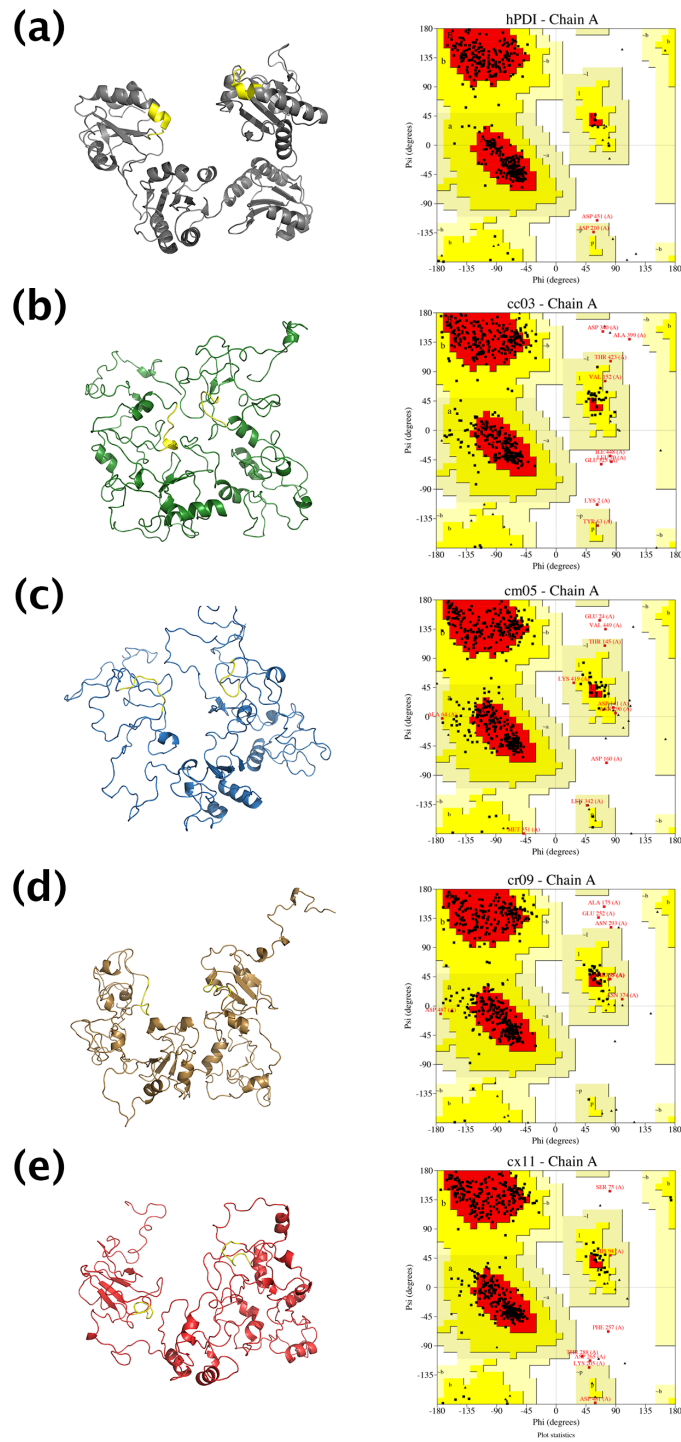


Figura 29. Patrón de plegamiento de las PDIs de los conos. A) Puede observarse en gris la estructura cristalina con su respectiva gráfica de Ramachadran para la hPDI (4EKZ) reducida (www.PDB.org). Los modelos refinados y su respectiva gráfica de Ramachadran se pueden observar en B) verde para Cc03_i1, C) en azul para Cm05, D) en café para Cr09 y E) en rojo para Cx11. Los motivos activos (APWCGCHCK) se muestran en amarillo para todas las moléculas.

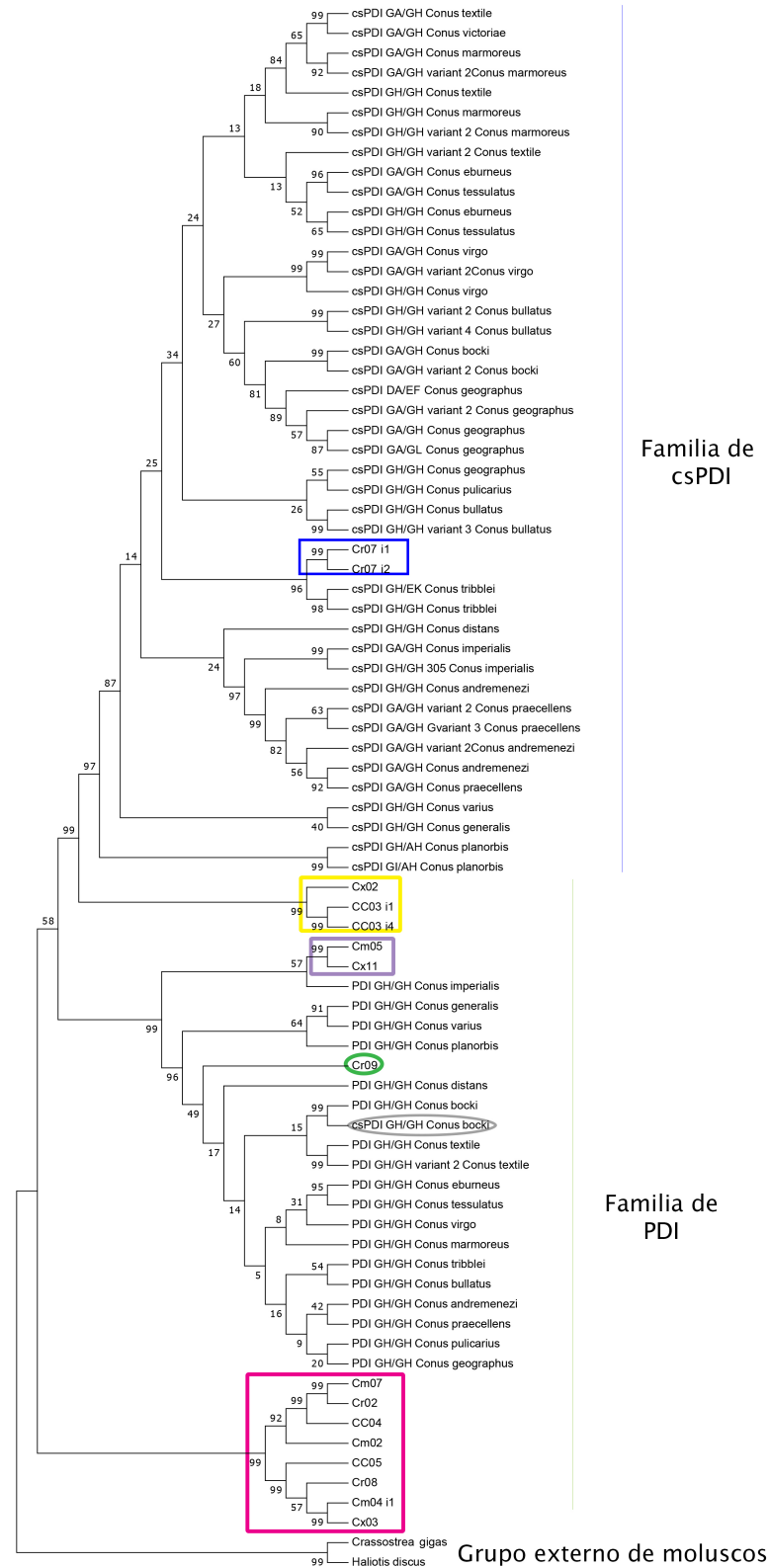


Figura 30. Árbol filogenético obtenido con las secuencias de Safavi-Hemmai y colaboradores (2016) y las obtenidas en este trabajo.

Capítulo 4. Discusión

El plegamiento de proteínas con residuos de cisteína requiere modificaciones bioquímicas adicionales a las de una proteína sin estos residuos: en primer lugar que el enlace disulfuro pueda formarse (oxidación), y que posteriormente mediante su reducción o rearreglos (isomerización) puedan corregirse aquellos que hayan sido incorrectamente formados (Ellgaard y Ruddock, 2005). Las conotoxinas no son ninguna excepción, y como se mencionó anteriormente, las cisteínas juegan un papel esencial en su conformación. Evidencia de esto es que las conotoxinas presentan arreglos de cisteína sumamente conservados (hasta la fecha existen reportados 25 marcos de cisteínas diferentes (Robinson et al., 2014); esto quiere decir que aunque existen miles de secuencias de diferentes conopéptidos estos tienden a utilizar el mismo tipo de andamio (Bulaj y Olivera, 2008). Además se estima que hasta el 50 % de los residuos de aminoácidos que componen una conotoxina pueden estar involucrados en la creación de enlaces disulfuro (Bulaj et al., 2003). A esto debe sumarse que la glándula venenosa donde son sintetizadas es el tejido que produce la mayor diversidad y densidad de péptidos ricos en residuos de cisteínas (de cientos a miles de secuencias) (Safavi-Hemami et al., 2016), lo que provoca que el reto para sintetizarlas sea aún más complejo. Por lo tanto, es de esperarse que las enzimas responsables de la formación de los enlaces disulfuro sean sofisticadas y variadas para poder mantener esta alta demanda. La alta diversificación de los genes de conotoxinas (Puillandre et al., 2012) debe de ir acompañada de una alta diversificación de las proteínas encargadas de su síntesis.

De esta manera, la diversidad de las familias de PDIs encontradas en este trabajo revela lo complejo que puede llegar a ser la maquinaria de chaperonas necesaria para generar la diversidad de conotoxinas contenidas en el veneno de los cónidos. Esta complejidad ha sido observada en organismos superiores; por ejemplo existen reportados 21 miembros para la familia de PDI de humanos, los cuales varían en masa molecular, composición de los dominios, tejido en el que son expresados y el tipo de procesamiento celular que reciben (Galligan y Petersen, 2012). Además, como es el caso de otras familias de PDI, es justificado suponer que sólo un porcentaje de la totalidad de las secuencias de PDI están involucradas en catalizar la isomerización de enlaces disulfuro de manera eficiente y que el resto no está directamente involucrada en la formación de enlaces sino que desempeña otra función (Ellgaard y Ruddock, 2005). Cabe mencionar que se ha visto que las PDI de cónidos más que trabajar de manera sinérgica, tienen funciones aditivas lo que sugiere que pueden trabajar de manera independiente unas de otras (Safavi-Hemami et al., 2012).

Se ha demostrado, mediante análisis filogenéticos, que las PDIs eucariotas presentan un ancestro en común cuya diversificación incluye la duplicación de los dominios de tiorredoxina; este concepto está

respaldado por la similitud de ciertas secuencias y la distribución de sus intrones (Kanai et al., 1998; Sahrawy et al., 1996). El motivo estructural “CXXC”, conservado en los dominios activos de las PDIs eucariotas, es indispensable para realizar la actividad oxidorreductasa. Se sabe que los dos aminoácidos que se encuentran dentro de los residuos de cisteínas juegan un papel muy importante en determinar el potencial redox de la enzima, y por lo tanto si presentará una función reductasa, oxidasa o isomerasa (Ellgaard y Ruddock, 2005). Los aminoácidos que flanquean dicho motivo están altamente conservados y existe una versión extendida de dicho motivo (APWCGHCK) que ha sido menos mencionada en la literatura pero que se ha visto se conserva dentro de las enzimas de PDI (Knodler et al., 1999; Mazzarella et al., 1990; Tachikawa et al., 1995). Por otro lado, se ha establecido que las secuencias de 7 residuos de longitud pueden ser utilizadas para identificar de manera inequívoca a grupos de proteínas (Gowd et al., 2007); el motivo “APWCGHCK” confirma esta aseveración y pudo ser utilizado por primera vez para recuperar de manera exitosa las secuencias de PDIs dentro de los transcriptomas de cónidos. Además, se pudo comprobar que este motivo puede ser extendido a una secuencia más larga de 10 aminoácidos FYAPWCGHCK, la cual coincide con la que Knodler *et al* (1999) habían identificado previamente como F(Y/F)APWCGHCK, e incluso se observa que en algunos casos puede ser extendida a una secuencia de 12 residuos (VEFYAPWCGHCK). Cabe mencionar que ligeras variaciones del motivo activo extendido fueron encontradas dentro de las diversas especies. Finalmente, de cada transcriptoma se seleccionó una secuencia para poder ser modelada en base a una secuencia de PDI típica. Esto nos permitió confirmar que por lo menos en cada especie existía una secuencia capaz de adquirir la conformación tridimensional esperada al de una PDI típica.

Por otro lado, nuestros resultados son consistentes con lo que se ha encontrado hasta la fecha en otros trabajos relacionados con PDIs: existen múltiples secuencias dentro de un solo veneno (Bulaj et al., 2003; Safavi-Hemami et al., 2012, 2016). No obstante, a diferencia de Safavi-Hemami et al (2016) no se encontraron lo que ellos identifican como secuencias específicas de cónidos (csPDIs) dentro del transcriptoma. Al igual que el resto de las PDIs, las csPDIs se conforman de 4 dominios similares a la tiorredoxina: dos dominios catalíticos (a y a') junto con dos dominios no catalíticos (b y b'). Sin embargo, las csPDIs presentan una alta variabilidad en los aminoácidos que se encuentran entre los dos residuos de cisteína sumamente conservados del motivo del sitio activo (CXX). La diferencia más notoria es el remplazo de la histidina por una alanina en los dominios “a” ya que el motivo “CAHC” es encontrado en 13 de las 18 especies estudiadas. Otras variantes encontradas en lugar del típico dominio “CGHC” fueron “CGIC”, “CEKC”, “CEFC” y “CRPC” (Safavi-Hemami et al., 2016). Esto plantea las siguientes preguntas: ¿estas especies carecen de csPDI? ¿Estas secuencias específicas deben de buscarse de una manera diferente? Cabe mencionar que a pesar de que *C. californicus* es considerado una especie atípica de la Familia Conidae

(Biggs et al., 2010), nuestros resultados demuestran que sus secuencias de PDI no presentan una divergencia notoria con el resto, sino que pueden agruparse con las secuencias de PDI encontradas. Otro punto que vale la pena resaltar es que Safavi-Hemami y colaboradores (2016) encontraron en promedio 2.6 secuencias de csPDI por especie, pero solamente una secuencia de PDI típica (con excepción de *Conus textile* de la que se encontraron 2). Este número de secuencias es mucho menor al número de secuencias de PDI encontradas por especie en este trabajo ¿Las especies de este trabajo tienen un mayor número de secuencias? ¿Safavi-Hemami y colaboradores están encontrando todas las secuencias? El árbol filogenético nos permitió observar los dos grupos principales que se esperaban que se formaran, no obstante, dos secuencias de *C. regularis* se agruparon con las secuencias de csPDI, esto a pesar de que no compartían las modificaciones en el dominio típico. Sólo la secuencia Cr09 (de la misma especie) se agrupó claramente con las secuencias de PDI y el resto formaron subgrupos ya sea dentro de las csPDI o dentro de las PDI.

Finalmente, se sabe que la producción de conotoxinas *in vitro* es un proceso complejo, por lo general las tasas de rendimiento son bajas y suelen generarse acumulación de productos que no fueron plegados correctamente (Bulaj y Olivera, 2008). El entendimiento de las PDI de la familia Conidae es necesario para en un futuro poder producirlas de manera eficiente. Además, como señalan Bulaj y Olivera (2008) estudiar estas enzimas no sólo nos permitirá entender cómo se lleva a cabo el plegamiento de conotoxinas sino que será un conocimiento que podrá ser expandido a la síntesis de pequeñas moléculas en general.

Capítulo 5. Conclusiones

- Cada especie presentó, en promedio, 4 secuencias codificantes para posibles proteínas PDI.
- El dominio “APWCGHCK” es suficiente para poder recuperar las secuencias esperadas, y que es posible extenderlo a la secuencia “FYAPWCGHCK.
- Por lo menos una de las secuencias de cada transcriptoma es capaz de adquirir una estructura tridimensional de una PDI típica, esto sugiere que los cónidos requieren de las PDI para realizar un plegamiento oxidativo correcto.
- No se encontraron secuencias de csPDI, no obstante, algunas de las secuencias identificadas se agruparon con este tipo de secuencias en el árbol filogenético. Se propone hacer un análisis similar al de Safavi-Hemami y colaboradores (2016) con el objeto de poder recuperar este tipo de secuencias.

Parte III. Combinación de estrategias de identificación en la secuenciación de conotoxinas

Capítulo 1. Introducción

Los venenos de *Conasprella ximenes* y de *Conus regularis* han sido poco explorados, de la primera especie no existe en la literatura conotoxina alguna reportada y de la segunda sólo existe una secuencia descrita hasta la fecha (Bernáldez et al., 2013).

Por otro lado, a pesar de que la tuberculosis es una enfermedad que causa estragos a nivel mundial, los tratamientos para combatirla son ineficientes. La bioprospección de los venenos de cónidos se ha realizado de manera sistemática durante los últimos 50 años (Akondi et al., 2014), además se ha demostrado que existen productos marinos con efecto antifímico (Kang et al., 2015); sin embargo, no se ha probado que las conotoxinas puedan tener actividad antimicrobiana.

Son varios los obstáculos que se encuentran a la hora de elucidar la estructura primaria de una conotoxina. En primer lugar, la manera tradicional de descubrimiento está basada en un fraccionamiento del veneno guiado por su actividad, lo cual requiere tiempo y cantidades de material que, por lo general, resulta difícil de obtener (Prashanth et al., 2012). En segundo lugar, las secuencias oscilan entre los 10 a 80 aminoácidos, incluyen enlaces disulfuros y en muchas ocasiones, múltiples MPT. La complejidad del veneno y la falta de bases de datos de cADN, ocasionan que hasta hace poco la tasa de descubrimiento de estas moléculas fuese de 1 secuencia/especie/año (Ueberheide et al., 2009). Afortunadamente, los avances en la transcriptómica y proteómica han incidido en que la identificación de conotoxinas se haya acelerado ostensiblemente (Dutertre et al., 2013). Sin embargo, la importancia farmacológica que tienen estas moléculas hace que no se abandone el método tradicional de identificación, sino que se emplee como técnica complementaria a las nuevas propuestas.

En este trabajo se describe la manera en que fue secuenciada parcialmente e identificada en una base de datos del transcriptoma de la glándula venenosa la primera conotoxina de *C. ximenes*: un polipéptido de 37 aminoácidos, con 8 residuos de cisteínas, y una masa molecular de 4109.69 Da. Esta conotoxina es capaz de inhibir el crecimiento de *Mycobacterium tuberculosis* (Mtb) a concentraciones molares similares a los de los fármacos utilizados actualmente en la terapia contra dicho patógeno.

Por otro lado, se reporta una conotoxina de *C. regularis* que hasta la fecha no presenta actividad ni se le ha podido determinar el blanco molecular sobre el que actúa. Según nuestro punto de vista, resulta interesante mostrar cómo una secuencia de 37 aminoácidos pueden poseer varias MPT: 8 cisteínas enlazadas formando cuatro enlaces de disulfuro, 3 residuos de ácido glutámico gammacarboxilados y dos hidroxiprolinas. Esta asignación solo pudo realizarse gracias a que se disponía de una base de datos de secuencias provenientes del transcriptoma de la glándula venenosa de dicha especie de *Conus*, demostrando la importancia de la estrategia utilizada: facilita y hace más eficiente la identificación de las conotoxinas de las especies que están siendo estudiadas en este trabajo de tesis doctoral.

1.1 Antecedentes

1.1.1 *Mycobacterium tuberculosis*

Mycobacterium tuberculosis (Mtb) es el agente causante de la tuberculosis (TB), una enfermedad infecciosa que está ocasionando severas afectaciones en la salud mundial, incluida la muerte. A pesar de que se han desarrollado nuevas técnicas para eliminar a este patógeno, de acuerdo con la Organización Mundial de la Salud, en el año 2015 ocurrieron 10.4 millones de nuevos casos y 1.4 millones de muertes relacionadas con esta enfermedad (“Global Tuberculosis Report,” 2016). En la actualidad, el tratamiento de primera línea para combatir la TB combina cuatro medicamentos que son administrados durante un periodo de 6 a 8 meses: rifampicina, isoniazida (INH), pirazinamida y el etambutol (EMB). Debido a la duración y complejidad del tratamiento, no es de sorprenderse que el desarrollo de nuevos fármacos está enfocado en moléculas con nuevos blancos moleculares que ayuden a acortar y mejorar los tratamientos actualmente utilizados. El crecimiento y los requerimientos nutricionales de las micobacterias han sido intensamente estudiados, y al igual que el resto de las bacterias, están sumamente relacionados con la permeabilidad que tiene la pared celular. Es por esta razón que las vías de señalización ligadas directamente con la pared celular, han sido consideradas como blancos moleculares para desestabilizar e inhibir el crecimiento de las micobacterias (Niederweis, 2008).

1.1.2 Los venenos como fuentes de fármacos

Existen varios estudios reportados de péptidos provenientes de fuentes marinas que presentan diversas actividades anti-infecciosas como lo pueden ser: antimicrobianas, antimicóticas, antimalarias o antimicobacteriales (Kang et al., 2015). Por otro lado, se ha demostrado que los venenos son una de las fuentes más grandes y diversas de moléculas químicas que pueden ser utilizadas de manera exitosa en la industria farmacéutica (Lewis et al., 2012). Dentro de los animales venenosos, los caracoles marinos de la familia Conidae son conocidos por la gran diversidad de conotoxinas o conopéptidos que producen. Estas pequeñas moléculas se caracterizan por la alta selectividad que presentan hacia sus blancos moleculares, dentro de los que destacan receptores y canales iónicos de membrana (Terlau y Olivera, 2004), convirtiéndolos en posibles candidatos para los tratamientos antimicrobianos, entre ellos las micobacterias por su relevancia para la salud humana, y es sin lugar a dudas una aplicación a considerar. Existen alrededor de 800 especies de cónidos cada uno de los cuáles produce un veneno único con alrededor de 1000 componentes (se calcula que sólo ~5 % de las toxinas se traslapa entre especies (Davis et al., 2009; Dutertre et al., 2013)). De la totalidad de conopéptidos, únicamente una pequeña fracción (~0.1 %) ha sido caracterizada farmacológicamente, (Lewis et al., 2012) y ninguna ha sido reportada con efecto anti-micobacteriano.

1.1.3 Elucidación de la estructura primaria de una conotoxina.

Mediante cromatografía líquida, el veneno de los cónidos puede ser fraccionado y evaluado en diferentes tipos de ensayos *in vitro* para llevar a cabo una bioprospección. No obstante, un paso clave en la identificación del blanco molecular y caracterización de una conotoxina, es conocer su estructura primaria que puede tener varias MPTs dentro de las que destacan la presencia de múltiples cisteínas. La espectrometría de masas se ha destacado por ser una herramienta analítica imprescindible para la caracterización de la estructura primaria de una conotoxina: su elevada sensibilidad y su fácil acoplamiento a la cromatografía líquida (LC-MS/MS) resulta ideal tanto en la secuenciación *de novo* como en la identificación de secuencias en bases de datos y la caracterización de las MPTs (Jakubowski et al., 2006).

La secuenciación *de novo* de un conopéptido mediante espectrometría de masas es un reto muy grande debido a que la mayoría de las veces su masa excede el tamaño óptimo ($MW < 3\,500$ Da) para realizar una fragmentación por CID (del inglés Collision Induced Dissociation) (Mikesh et al., 2006; Zubarev et al., 2008).

A esto debe sumarse que los conopéptidos frecuentemente contienen varios aminoácidos básicos en su secuencia, limitando la disponibilidad de los protones móviles y por lo tanto de la eficiencia de fragmentación mediante dicho método (Wysocki et al., 2000). Por lo tanto, en lugar de obtener una secuencia completa, resulta más realista obtener una secuencia parcial que pueda funcionar como *sequence tag* para identificar en una base de datos a la secuencia completa de la proteína que la contiene.

No obstante, la identificación de conopéptidos en bases de datos presenta tres grandes limitantes: 1) no existe un genoma reportado para los cónidos y el número de secuencias de cDNA es limitado (Ueberheide et al., 2009); 2) aún cuando varias conotoxinas estuvieran disponibles, el hecho de que existe muy poca superposición de las toxinas entre especies diferentes de cónidos, representa una limitante si no se cuenta con transcriptomas propios de la especie; 3) el software que identifique las secuencias debe ser capaz de reconocer eficientemente el número abrumador de variantes generadas por las múltiples MPTs que pudieran estar presentes de manera simultánea en un conopéptido, cuya presencia no puede predecirse con antelación.

La identificación de proteínas y la secuenciación *de novo* no son estrategias excluyentes (Ma et al., 2003) y pueden ser combinadas para identificar el conopéptido de interés si primero se obtiene una secuencia confiable *de novo*, que pueda ser utilizada como *sequence tag* (Mann y Wilm, 1994).

Finalmente, se sabe que el transcriptoma incluye todos aquellos genes que están siendo transcritos activamente en un órgano en específico (Rendón-Anaya et al., 2013). En los cónidos, la producción de conotoxinas está limitada al ducto venenoso, por lo que la secuenciación de su transcriptoma resulta ideal si se quiere construir una base de datos de las conotoxinas presentes en el veneno. Dutertre y colaboradores (2013) ya demostraron la utilidad que una base de datos específica de la especie puede tener en la elucidación de conotoxinas.

Conasprella ximenes y *Conus regularis* son cónidos de la costa del Mar de Cortez (México) cuyos conopéptidos no han sido estudiados. En este trabajo se combinaron diferentes estrategias para facilitar la secuenciación de una conotoxina que presentaba actividad anti-Mtb y otra conotoxina con múltiples MPT. Por lo tanto, se reporta la primera toxina de *C. ximenes* (I1_xm11a) que además es la primera conotoxina que tiene actividad anti-Mtb; así como la segunda conotoxina descrita para *C. regularis* (O2_rs15a).

1.2 Justificación

Aunque los nuevos avances en el estudio de transcriptómica y proteómica de los venenos de cónidos apuntan hacia un descubrimiento masivo de secuencias, la integración de métodos tradicionales permite un mejor acercamiento a la caracterización de las conotoxinas. Las *sequence tags* resultan sumamente útiles para buscar secuencias completas en bases de datos complejas, como son los conopeptidos, donde abundan las MPTs, múltiples residuos de cisteínas y que además son difíciles de fragmentar por métodos comunes como lo es el CID.

1.3 Hipótesis

- La utilización de *sequence tags* permite obtener de manera confiable la secuencia completa de una conotoxina en una base de datos de transcriptómica y dilucidar las modificaciones postraduccionales que presenta.

1.4 Objetivos

1.4.1 Objetivo general

Conocer la secuencia primaria de dos conotoxinas de cuyas especies se conoce el transcriptoma de la glándula venenosa.

1.4.2 Objetivos específicos

- Obtener *de novo* una *sequence tag* que permita identificar dentro del transcriptoma del ducto venenoso la secuencia completa de dos conotoxinas, una de las cuales presenta actividad anti-Mtb.
- Confirmar la asignación realizada mediante la reducción y alquilación de la conotoxina que presenta actividad anti-Mtb y la secuenciación mediante ESI-MS/MS de los péptidos tripticos resultantes para la verificación de la secuencia asignada.

Capítulo 2. Metodología

Bernáldez-Sarabia, J. (2013) identificó en su tesis de doctorado una conotoxina de *C. ximenes* con actividad antimicrobiana. Esta toxina, reportada como F11, fue evaluada nuevamente en ensayos de proliferación celular y fue capaz de inhibir el crecimiento de Mtb a una concentración mínima inhibitoria (MIC, por sus siglas en inglés) de 3.0 μ M. Este valor de concentración molar es similar al de dos fármacos utilizados en la actualidad para combatir la tuberculosis: la isoniazida (MIC 0.45 μ M) y el etambutol (MIC 9.8 μ M). En aquel momento la secuencia de esta conotoxina no pudo ser elucidada debido a dos razones 1) la metodología tradicional requiere que de una gran cantidad de material con la que no se contaba; 2) El transcriptoma de la glándula venenosa del cónido no estaba disponible. En este trabajo se describe la manera en que se secuenció dicha molécula, a la que se nombró como I1_xm11a.

Por otro lado, se purificó uno de los picos mayoritarios del veneno de *C. regularis*, se determinó la secuencia de la toxina nombrada como O2_rs15a, y se demostró la presencia de múltiples MPT.

2.1 Colecta de especímenes y extracción del veneno

Los especímenes de *C. regularis* y *C. ximenes*, presuntamente especies vermívoras y moluscívoras respectivamente (Lugo et al., 2016), fueron colectados en la zona del sublitoral de Bahía de Los Ángeles (Golfo de California, México). *C. regularis* habita zonas lodosas, mientras que *C. ximenes* prefiere zonas arenosas. Los ductos venenosos de 20 especímenes de cada especie fueron disecados e inmediatamente homogenizados por separado en 1 mL de una mezcla 40 % (v/v) de acetonitrilo (ACN, Fermont) y 0.1 % (v/v) de ácido trifluoroacético (TFA, Fluka). El homogenizado se mantuvo todo el tiempo en hielo, y una vez procesados todos los especímenes, fue centrifugado a 10,000 x g durante 10 min a 4 °C. Finalmente, el sobrenadante de cada especie fue liofilizado y almacenado a -80 °C hasta ser procesado para la purificación del péptido de interés.

2.2 Purificación de los péptidos de interés.

El veneno soluble fue fraccionado mediante cromatografía líquida de alta resolución en fase reversa (RP-HPLC) (Agilent 1220 Series LC System), con una columna analítica C₁₈ Zorbax 300SB (4.6 × 250 mm, tamaño de partícula 5 µm) y una pre-columna Zorbax 300SB C₁₈ (4.6 × 12.5 mm, tamaño de partícula 5 µm). Ambas fueron previamente equilibradas en una solución de TFA al 0.12 % (v/v) a la que se denominó como solución A. El veneno total fue desalado en la pre-columna con la misma solución de equilibrio a un flujo de 1.0 mL/min durante 5 minutos. Las fracciones de los componentes del veneno fueron colectadas cada 5 minutos y eluidas utilizando un gradiente lineal de 0 a 60 % (v/v) de una solución denominada como B (ACN con 0.10 % (v/v) de TFA) durante 65 min. El flujo se mantuvo constante a 1.0 mL/min.

Para *C. ximenes*, cada una de las fracciones fue evaluada en un ensayo de susceptibilidad antimicrobiana. La fracción cromatográfica que contenía el péptido de interés (para *C. ximenes* era el que presentaba actividad antimicrobiana, para *C. regularis* un pico mayoritario de la fracción 3), fue repurificada mediante RP-HPLC utilizando un gradiente lineal de 22 a 45 % (v/v) de solución B, en un periodo de 60 minutos a flujo constante de 1.0 mL/min. Todos los pasos de purificación fueron realizados a temperatura ambiente (TA) y la absorbancia fue monitoreada a 230 nm. El péptido purificado fue liofilizado, cuantificado por diferencia de peso seco y nombrado como I1_xm11a para *C. ximenes* y O2_rs15a para *C. regularis* (la nomenclatura será explicada en la sección de Resultados). Se utilizó en todo momento agua desionizada purificada mediante el sistema Mili-Q (Pure Lab Flex, Elga de Ion Torrent, Life Technologies).

2.3 Reducción y S-alquilación de las conotoxinas.

Se disolvió aproximadamente 1 µg de cada uno de los péptidos (I1_xm11a y O2_rs15a) en 10 µL de buffer de bicarbonato de amonio 50 mM (con 0.5 mol/L de cloruro de guanidio a pH 8.3). Se añadió ditioneol (DTT) para obtener una concentración final de 5 mM y se incubó durante 1 hr a 37 °C. Después de este período y una vez que la reacción alcanzó la temperatura ambiente (TA), se añadió una solución de iodoacetamida (IAM) para obtener una concentración final de 10 mM. La reacción de alquilación de los residuos libres de cisteínas tuvo lugar durante 30 min en la oscuridad a TA y se acidificó añadiendo 1 µL de

ácido fórmico (AF) al 10 % (v/v) para inmediatamente después ser analizada mediante LC-MS/MS. Todos los reactivos utilizados fueron de Sigma-Aldrich.

2.4 Análisis mediante LC-MS/MS

El péptido nativo y el péptido reducido y S-alquilado de cada especie fueron cargados en una precolumna de enriquecimiento de 40 nL empaquetada con partículas 5 μm (300SB-C₁₈) y con un chip de 300 Å (ProtID-Chip 43 II, Agilent Technologies), previamente equilibrado en una solución de TFA al 0.1 % (v/v). Las muestras fueron desaladas pasando 5 volúmenes de la precolumna en el mismo buffer de equilibrio a 2 $\mu\text{L}/\text{min}$. Posteriormente, la precolumna fue acoplada a una columna analítica de fase reversa (75 μm x 150 mm) empaquetada con el mismo material mencionado previamente. Los péptidos fueron eluidos a un flujo constante de 400 nL/min, durante 15 minutos, en un gradiente lineal de 3 – 70 % de una solución al 90 % de ACN con 0.1 % de AF.

A continuación, la solución en que eluyeron los péptidos fue nebulizada mediante ionización por electrospray en un sistema QTOF-LC/MS (Accurate Mass quadrupole time-of-flight, G6530AA de Agilent Technologies.) utilizando 1900 V de voltaje del capilar, 65 V de voltaje “skimmer” y 175 V para el “fragmentor”. Para detectar la masa molecular de los péptidos de interés, se obtuvo un espectro de masa de alta resolución ($m/\Delta m \sim 20,000$) para las relaciones masa/carga (m/z) de 300 – 2000 Da en modo MS. Las energías de colisión (CE) fueron ajustadas para obtener una fragmentación eficiente de los iones seleccionados de los péptidos de interés después de la reacción de reducción y S-alquilación. Los iones seleccionados para *C. ximenes* fueron los correspondientes a las cargas 4⁺, 5⁺ y 6⁺, mientras que para *C. regularis* fueron los correspondientes a las cargas 5⁺, 6⁺ y 7⁺. Los espectros de masas fueron adquiridos en modo positivo y analizados utilizando el software MassHunter Qualitative, Agilent Technologies (B.06.00). Los resultados de los espectros ESI-MS/MS fueron interpretados manualmente para extraer una secuencia parcial.

2.5 Transcriptómica de la glándula venenosa

La metodología correspondiente a la colecta de especímenes, extracción de ARN y su secuenciación, y el ensamblaje para la identificación de conotoxinas se encuentra explicada en las secciones 2.1, 2.2 y 2.3 respectivamente de la Parte I de esta tesis.

2.6 Búsqueda en la base de datos y análisis de la secuencia

Los espectros MS/MS de las dos conotoxinas fueron exportados en un archivo mgf y cargados en tres herramientas de búsqueda en bases de datos: MASCOT (Perkins et al., 1999), Protein Pilot (Shilov et al., 2007) y Peaks (Ma et al., 2003). La finalidad era identificar las conotoxinas de interés en cada una de las bases de datos de los transcriptomas ensamblados, a partir de los espectros MS/MS de los iones multicargados de los dos péptidos de interés. En todos los softwares la tolerancia de la masa para el ión precursor y los iones fragmentados fue de 0.1 y 0.05 Da, respectivamente. No se consideró ninguna restricción para la especificidad de enzima proteolítica y los residuos de cisteínas se consideraron carboamidometiladas y fueron definidas como una modificación fija en todos los experimentos. Tanto en el software Protein Pilot como en Peaks, se consideró que la búsqueda en la base de datos debía de considerar todas las MPT frecuentemente encontradas en los conopéptidos (Buczek et al., 2005). En el caso de MASCOT se utilizaron diferentes combinaciones de sólo 4 variantes de MPT con la finalidad de no saturar al software.

La interpretación manual de los espectros MS/MS permitió extraer una *sequence tag* confiable de cada conotoxina que fue confirmada tanto como por las series b_n como por las series y''_n . La precisión de la masa (0.01 – 0.02 Da), así como otros aspectos utilizados en la secuenciación *de novo* de péptidos, fueron considerados en la extracción de las dos secuencias parciales (Bernáldez et al., 2016; Cabrera et al., 2017). Con la ayuda del software MASCOT (Perkins et al., 1999), utilizando la opción de *sequence query*, se identificó las conotoxinas de interés en la base de datos de cada uno de los transcriptomas. Las *sequence tags* fueron buscadas en ambos sentidos, considerando la posibilidad de que hubieran sido extraídas ya sea a través de los iones de las series b_n o los iones de las series y''_n .

La identidad y similitud de las secuencias con otras identificadas en la literatura se realizó utilizando el programa BLAST contra la base de datos no redundante del NCBI, del UNIPROT y las secuencias reportadas

en ConoServer. La identificación del péptido señal se calculó con ayuda del servidor de SignalP 4.1, mientras que la región madura de la conotoxina se determinó al considerar la masa molecular de la secuencia.

2.7 Verificación de la secuencia de la toxina I1_xm11a mediante la digestión tríptica y análisis LC-MS/MS

1 µg de la conotoxina I1_xm11a reducida y S-alquilada fue disuelta en 10 µL de buffer de bicarbonato de amonio 50 mM. Para facilitar la solubilización de la toxina reducida y alquilada, se añadió cloruro de guanidinio 0.5 M al bicarbonato de amonio. Manteniendo un pH 8.3 se añadió 0.5 µg de tripsina grado secuencia (Promega). La digestión proteolítica se llevó a cabo durante 4 hrs a una temperatura de 37 °C. Para detener la reacción, la muestra fue inmediatamente congelada a -80 °C y posteriormente liofilizada. Los péptidos trípticos resultantes fueron disueltos en 10 µL de una solución de AF al 0.10 % (v/v) y analizadas mediante LC-MS/MS. Posteriormente los péptidos trípticos fueron analizados de la misma manera que el péptido nativo y el péptido reducido y S-alquilado (Ver apartado 2.4, Parte III). Para obtener una fragmentación eficiente se utilizaron rampas de energía de colisión predefinidas para los diferentes estados de carga de péptidos trípticos (2⁺, 3⁺, y 4⁺). Los datos fueron adquiridos y analizados de la misma manera que en el apartado 2.4 de la Parte III.

Para la verificación de la secuencia parcial de I1_xm11a, se consideró los posibles cortes generados por la tripsina. La cisteína carbamidometilada fue definida como una modificación fija en todos los experimentos. Tanto en el software Protein Pilot como en Peaks, se consideró que la búsqueda en la base de datos debía de considerar todas las MPT frecuentemente encontradas en los conopéptidos (Buczek et al., 2005). En el caso de MASCOT se utilizaron diferentes combinaciones de sólo 4 variantes de MPT esto con la finalidad de no saturar al software.

Capítulo 3. Resultados

3.1 Purificación de los péptidos de interés.

El perfil obtenido en la cromatografía de fase reversa del veneno crudo extraído, tanto de *C. ximenes* como de *C. regularis*, indica la presencia de varios componentes peptídicos (Figura 31a y Figura 32a), demostrando la gran diversidad esperada para los venenos de la familia Conidae.

Bernáldez-Sarabia (2013) identificó que la actividad biológica se encontraba en un péptido que eluyó en un pico específico en el minuto 39 (señalado con una flecha en la Figura 31a). Por esta razón se colectó la fracción del veneno que se encuentra entre los minutos 35 al 40 del cromatograma (sombreado en gris en la Figura 31a); para después realizar una repurificación utilizando un gradiente más lento (Figura 31b).

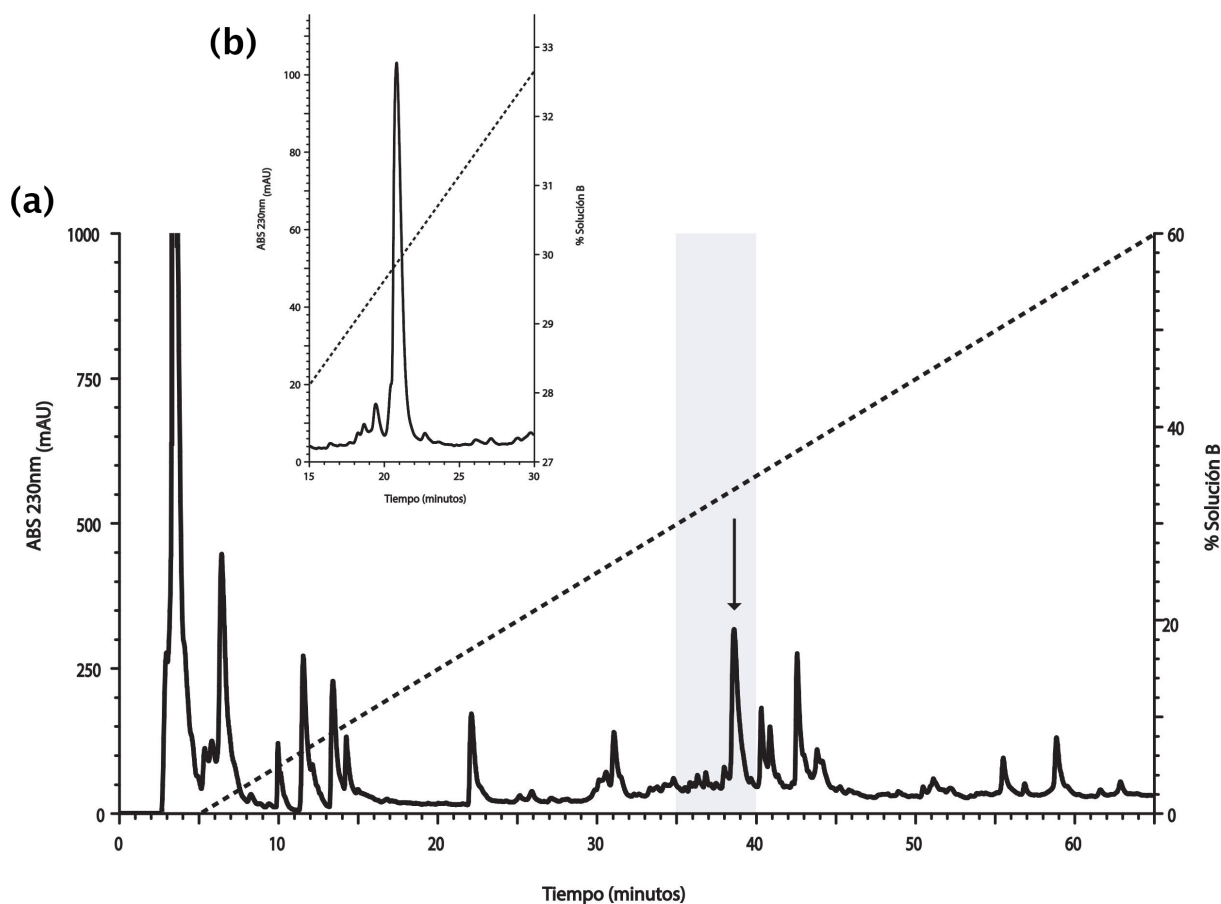


Figura 31. Fraccionamiento mediante RP-HPLC del veneno de *C. ximenes*. (a) Cromatograma correspondiente al fraccionamiento del veneno total utilizando un gradiente lineal de 0 a 60 % de solución B. Los primeros cinco minutos corresponden a un proceso de desalado de la muestra, la región en gris señala la fracción con actividad anti-MTb y la flecha indica el péptido purificado. (b) Cromatograma parcial de la repurificación del péptido con un gradiente lineal de 22 a 45 % de solución B. En ambas corridas de RP-HPLC las líneas discontinuas indican el gradiente lineal de solución B.

En el caso de *C. regularis* se colectó la fracción que corresponde del minuto 15 al 20 del cromatograma del veneno total (Figura 21a), para después realizar una repurificación del pico que correspondía al minuto 17 (Figura 21b).

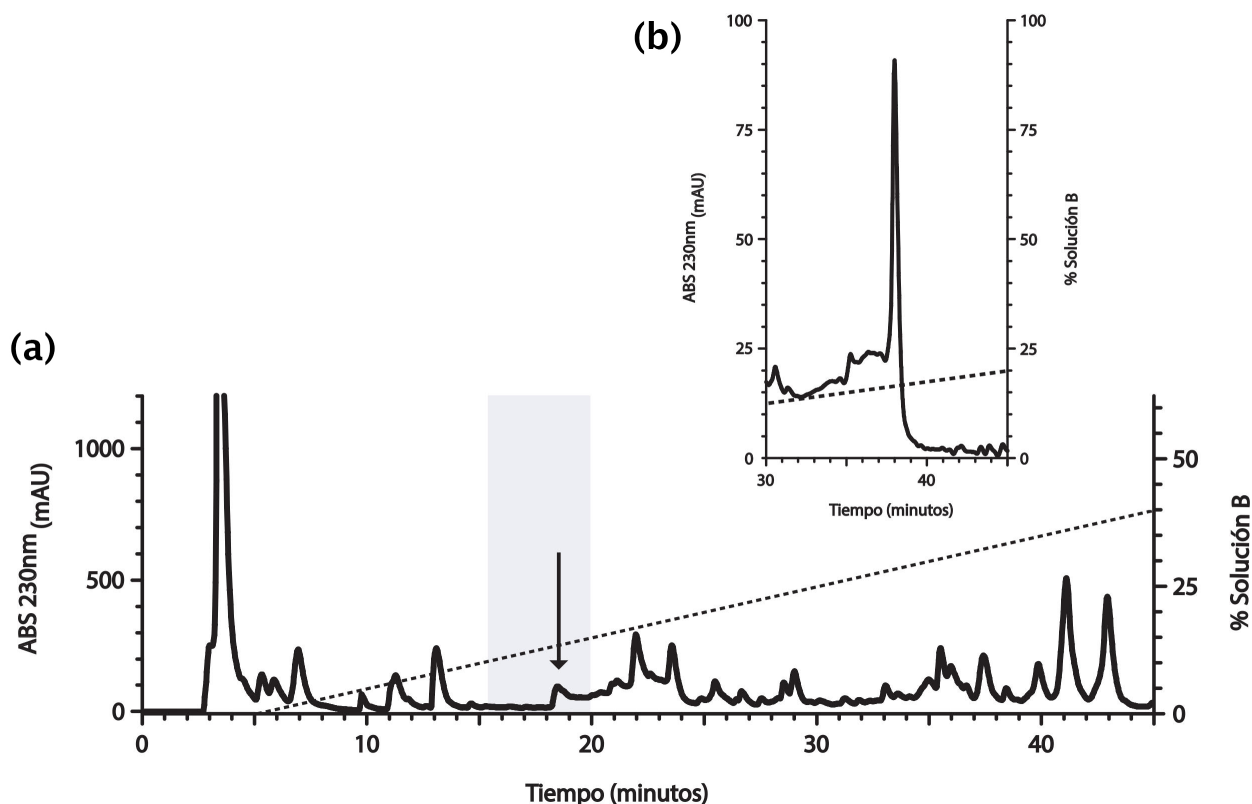


Figura 32. Fraccionamiento mediante RP-HPLC del veneno de *C. regularis*. (a) Cromatograma del veneno total utilizando un gradiente lineal de 0 a 60 % de solución B. Los primeros cinco minutos corresponden a un proceso de desalado de la muestra, la región en gris señala la fracción a partir de la cual se realizó una repurificación del pico que contenía al péptido de interés. (b) Cromatograma parcial de la repurificación del péptido con un gradiente lineal de 22 a 45 % de solución B. En ambas corridas de RP-HPLC las líneas discontinuas indican el gradiente lineal de solución B.

3.2 Identificación de la conotoxina de *C. ximenes*.

3.2.1 Cálculo del número de cisteínas.

En el espectro ESI-MS del péptido nativo de *C. ximenes* se detectaron en m/z de 1028.92, 823.54 y 686.28 tres señales intensas asignadas a los iones multicargados con 4^+ , 5^+ y 6^+ , respectivamente (Figura 33). Al

analizar detalladamente la separación de los picos de la distribución isotópica en la Figura 33, puede observarse la distribución isotópica experimental de los iones $[M+6H]^{6+}$ tanto del péptido nativo (inciso a) como del péptido carboamidometilado (inciso b). La diferencia en masa entre las señales de los picos monoisotópicos de los iones multicargados 6^+ fue de 464.22 Da según se muestra a continuación:

$$(\Delta m) * 6 = (763.32 - 685.95) * 6 = 464.22 \text{ Da}$$

Este incremento de masa después de la reducción y carboamidometilación puede ser asignada a la presencia de 8 residuos de cisteína (PM=58) enlazados por cuatro enlaces disulfuros presentes en el péptido l1_xm11a.

$$464.22/58 = 8 \text{ cisteínas}$$

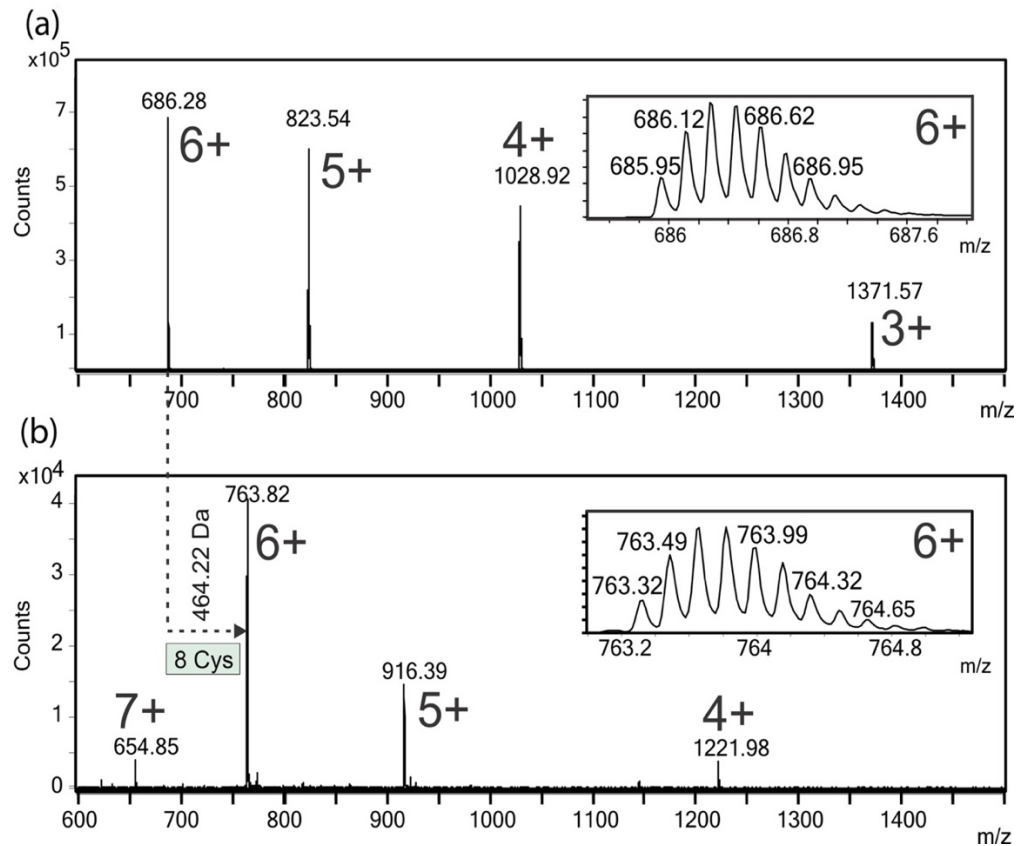


Figura 33. Espectro ESI-MS del péptido nativo (a) y reducido S-carboamidometilado (b) de la conotoxina l1_xm11a. En ambos espectros de masas, los recuadros interiores muestran la distribución isotópica de los iones correspondientes a $[M+6H]^{6+}$.

3.2.2 Obtención de la *sequence tag* e identificación de la secuencia.

Para poder identificar la conotoxina de interés en la base de datos del transcriptoma de la glándula venenosa, los iones multicargados (Figura 34) del péptido reducido y S-alquilado fueron fragmentados por CID. Los espectros MS/MS fueron exportados en formato mgf y evaluado por tres softwares distintos de identificación de proteínas en base de datos de secuencias (MASCOT (Perkins et al., 1999), Protein Pilot (Shilov et al., 2007) y el Peaks (Ma et al., 2003)). Ninguno de los 3 softwares pudo identificar una secuencia con un valor significativo de su puntuación. Probablemente, la masa molecular que posee la conotoxina (> 3500 Da) y la posible presencia de múltiples aminoácidos básicos, impiden una fragmentación eficiente del péptido analizado por la poca disponibilidad de protones móviles. Teniendo en cuenta estos antecedentes se decidió interpretar manualmente los espectros MS/MS para extraer una secuencia parcial confiable o *sequence tag*.

En el caso de *C. ximenes* se utilizó el espectro MS/MS del ion precursor $[M+6H]^{6+}$ (Figura 34a, m/z 763.32) y se extrajo una *sequence tag* confiable de 10 aminoácidos ($y_2^+-y_{12}^+$), utilizando los iones simple carga de la serie y_n^+ (333.19, PC*A-L/I-VT-L/I-VC*T, 1447.74, donde C* son las cisteínas carboamidometiladas). Esta secuencia parcial también pudo ser comprobada en los espectros MS/MS de los iones precursores con carga 3^+ , 4^+ y 5^+ del mismo péptido. La *sequence tag* (Mann y Wilm, 1994) permitió identificar en la base de datos del transcriptoma una única conotoxina que satisfacía las características necesarias para formar parte del péptido identificado (masa molecular, 8 cisteínas y los aminoácidos secuenciados manualmente). La masa calculada (4109.69 Da) y la masa experimental (4109.70 Da) para la conotoxina nativa de 37 aminoácidos coincidía con un margen de error de 2.4 ppm.

La secuencia del péptido tríptico permitió confirmar 6 de los 8 residuos de cisteína esperados para la conotoxina identificada como I1_xm11a, así como la secuencia parcial utilizada como *sequence tag* que se resalta en rojo (Figura 35). La distribución isotópica del péptido tríptico muestra una pérdida de la intensidad para el ion monoisotópico detectado para m/z 863.02, probablemente esto sea evidencia de un proceso de deamidación parcial de la Asn²³ durante la digestión, lo cual es posible porque existe un sitio probable Asn-Gly que fue incubado durante 4 h de digestión a 37°C en un pH básico (Li, 2008; Robinson y Robinson, 2001a; 2001b). Cabe mencionar que otros péptidos trípticos localizados en la región del N-terminal (1-14, Figura 35b) no fueron detectados en el análisis LC-MS/MS probablemente debido a que sus tamaños eran muy pequeños y a que su hidrofiliidad era incompatible con ser retenidos en la precolumna durante el proceso de desalado que es requerido en el análisis.

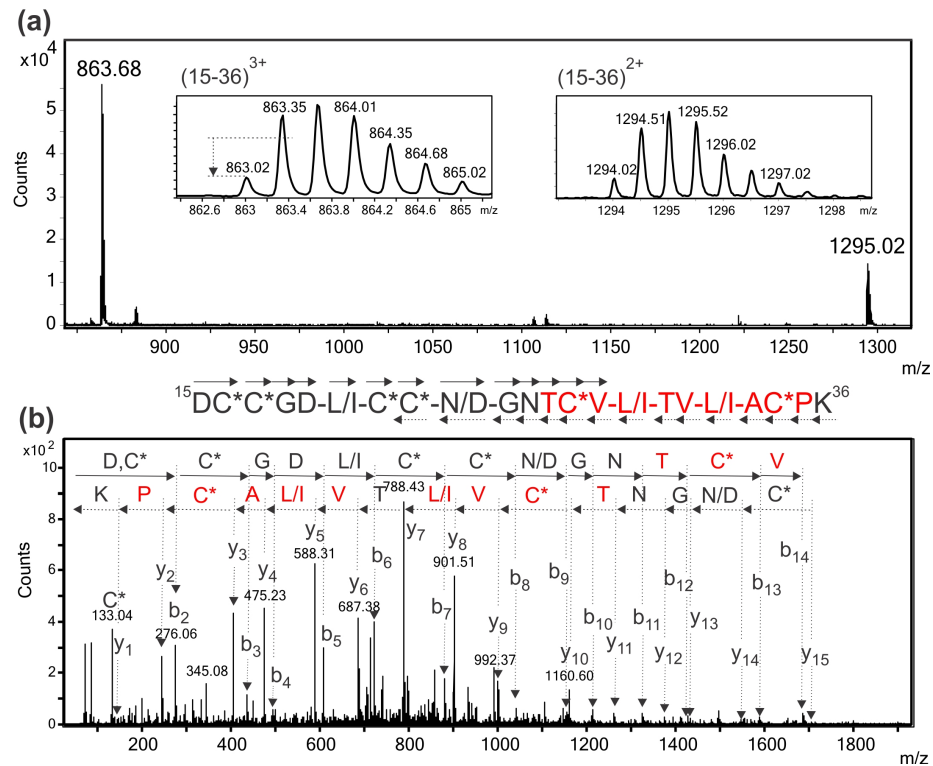


Figura 35. Espectro ESI-MS del péptido tríptico (15-36) derivado de la reducción y S-alquilación de la conotoxina I1_xm11a. (a) regiones ampliadas para el espectro ESI-MS mostrando la distribución isotópica de los iones $(15-36)^{2+}$ y $(15-36)^{3+}$. (b) corresponde al espectro ESI-MS/MS del ion doblemente cargados detectados en (a) con m/z 863.02.

3.2.4 Análisis de la conotoxina I1_xm11a.

Después de eliminar secuencias repetidas, la base de datos de *C. ximenes* quedó conformada por un total de 24, 284 secuencias. La identificación de la conotoxina en la base de datos permitió recuperar la secuencia completa (región pre, pro y toxina madura), que como se ha mencionado en el transcurso de esta tesis es sumamente importante para su clasificación.

La búsqueda a través de BLAST nos permitió determinar que el conopéptido aislado de *C. ximenes* pertenece a la superfamilia I1 (Buczek et al., 2005) y sus residuos de cisteína se encontraban acomodados de acuerdo al marco de cisteínas XI (C-C-CC-CC-C-C) reportado por Jimenez et al., (2003). Por lo tanto, de acuerdo a la nomenclatura establecida para los péptidos aislados de caracoles marinos venenosos (Walker et al., 1999), se nombró a esta conotoxina como I1_xm11a.

Recordando brevemente, en caso de que la conotoxina haya sido aislada a nivel de proteína, las letras minúsculas (xm) representan la especie a la que pertenece la secuencia; los números arábigos indican el marco de cisteínas que contiene (11) y el orden de descubrimiento está indicado por una letra minúscula comenzando con la letra “a”. Además, como sugieren Robinson y colaboradores (2014), se incorpora al nombre de la superfamilia (I1) a la que dicha conotoxina pertenece. El alineamiento de I1_xm11a con otros péptidos pertenecientes a la superfamilia I1 con un marco de cisteínas XI puede verse en la Figura 36.

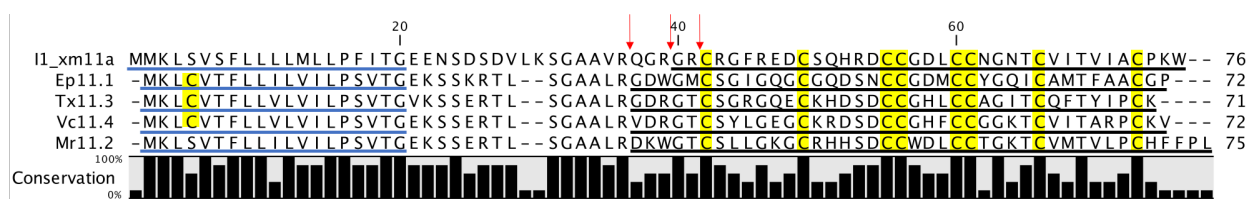


Figura 36. Alineamiento de I1_xm11a con otros miembros de la superfamilia I1 con marco de cisteínas XI. Los péptidos señal y los péptidos maduros están subrayados en azul y en negro, respectivamente. Los residuos de cisteína están resaltados en color amarillo y una barra de porcentaje de conservación está representada para cada aminoácido. Las flechas en rojo indican los tres posibles sitios de procesamiento para el N-terminal de acuerdo al mecanismo propuesto por Dutertre et al., 2013.

3.3 Identificación de la conotoxina de *C. regularis*.

3.3.1 Cálculo del número de cisteínas.

En el espectro ESI-MS del péptido nativo de *C. regularis* se detectaron tres señales intensas para las m/z de 856.51, 713.93 y 611.79 correspondientes a los iones multicargados de 5^+ , 6^+ y 7^+ , respectivamente. En la Figura 37 puede observarse la distribución isotópica experimental de los iones $[M+5H]^{5+}$ tanto del péptido nativo (Figura 37a) como del péptido carboxamidometilado (Figura 37b). Al igual que en el caso de l1_xm11a, la diferencia de masa de las señales de los picos monoisotópicos nos permitió calcular el incremento de masa, la cual puede ser asignada nuevamente a la presencia de 8 residuos de cisteína enlazados por cuatro enlaces disulfuros.

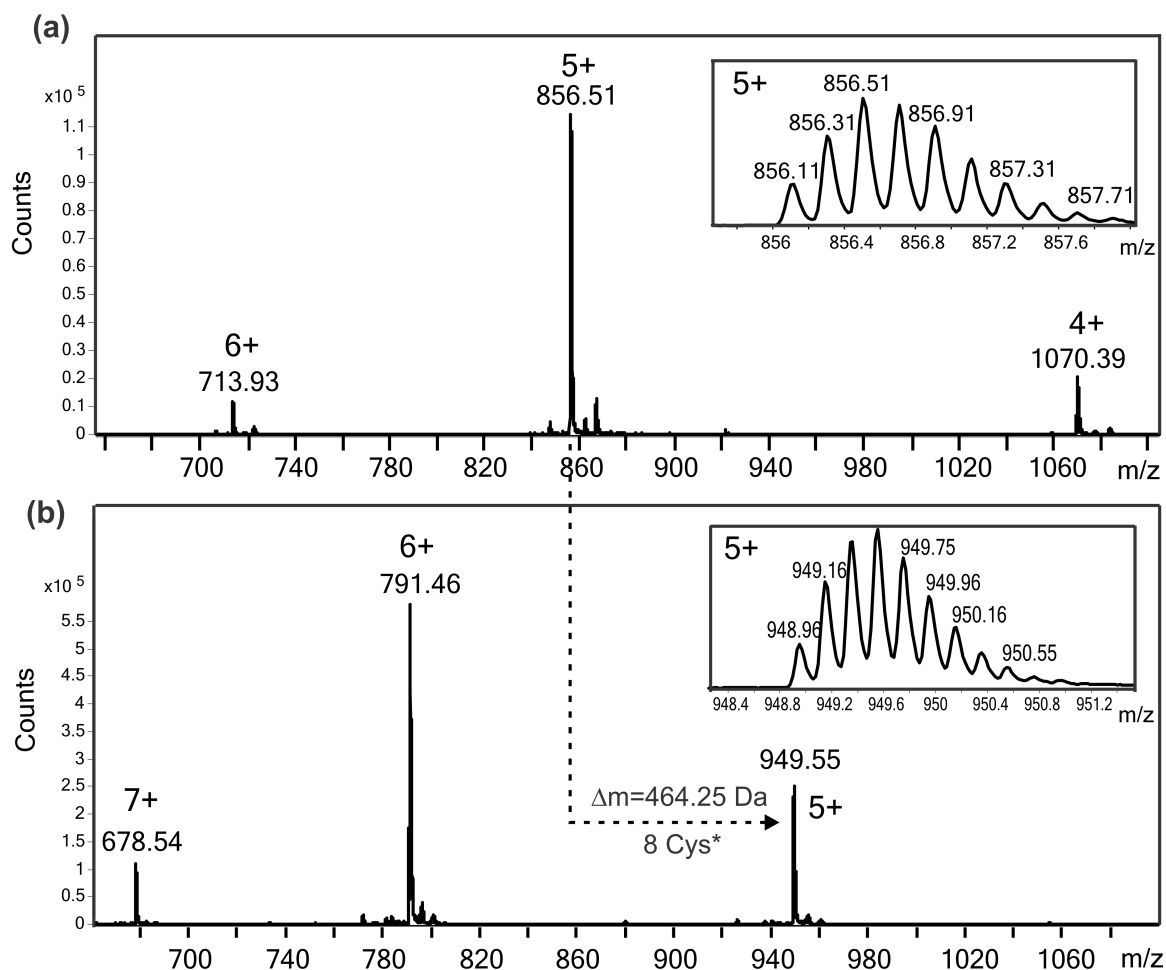


Figura 37. Espectro ESI-MS de la conotoxina O2_rs15a nativa (a) y reducida S-carboxamidometilada (b). En ambos espectros de masas, los recuadros interiores muestran la distribución isotópica de los iones correspondientes a los iones multicargados $[M+5H]^{5+}$. La diferencia de masa entre el péptido nativo y el péptido reducido y S-carboxamidometilado corresponde a 8 residuos de cisteína.

3.3.2 Obtención de la *sequence tag* e identificación de la secuencia

Los iones multicargados (Figura 39) del péptido reducido y S-alquilado fueron fragmentados por CID. Al igual que en el análisis de la conotoxina I1_xm11a, los espectros MS/MS fueron exportados en formato mgf y evaluado por tres softwares distintos de identificación de proteínas en base de datos de secuencias (MASCOT (Perkins et al., 1999), Protein Pilot (Shilov et al., 2007) y el Peaks (Ma et al., 2003)). Ninguno de los 3 softwares pudo identificar una secuencia con un valor significativo posiblemente por las mismas razones que en el caso anterior (masa molecular > 3500 Da, posible presencia de múltiples aminoácidos básicos). Teniendo en cuenta estos antecedentes se decidió interpretar manualmente los espectros MS/MS para extraer una secuencia parcial confiable o *sequence tag*.

La interpretación manual del espectro ESI-MS/MS del ion precursor $[M+6H]^{6+}$ (Figura 39a, m/z 856.51) permitió extraer una *sequence tag* confiable de 10 aminoácidos (1520.65, AGQC*P*QTDSK, 432.20, donde P* corresponde a una hidroxiprolina y C* a una cisteína carboamidometilada). Utilizando los iones doble carga de la serie y_n^{2+} se pudo confirmar el fragmento de secuencia LEC*KC*QA(G, Q)C*P*Q, mientras que los iones triple carga de la serie y_n^{3+} permitieron identificar el fragmento C*C*PNLEC*K (Figura 39a). La secuencia parcial también pudo ser comprobada en los espectros MS/MS de los iones precursores con carga 5^+ y 7^+ del mismo péptido (Figura 39).

En el espectro MS/MS se observó tres señales intensas separadas del ion precursor por tres pérdidas consecutivas de masas de 44 Da ($-CO_2$), lo que sugería la presencia de tres residuos de ácido gamma carboxiglutámico (Figura 38 y Figura 39b) por ser una modificación lábil en las condiciones de CID. En el espectro MS/MS estos residuos se detectan como un residuo de glutámico sin modificar.

La diferenciación entre los residuos de "Leu/Ile, 113.08 Da" y "P*, 113.04 Da" en las posiciones 17 y 29 de la secuencia se hicieron basándonos en la exactitud de la medición al calcular las diferencias en masas entre series y_n consecutivas (Figuras 39c y 39d). La presencia de P* presentes en la secuencia fue confirmada adicionalmente al ser asignados los iones imonios detectados en las m/z 86.09 y 86.04 a los aminoácidos (L/I) y (P*), respectivamente (Figura 39 a, c y d).

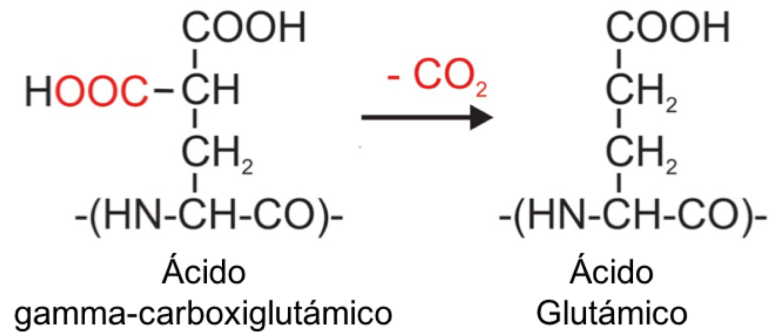


Figura 38. Estructura del ácido glutámico carboxilado y su transformación en ácido glutámico en las condiciones de fragmentación por CID debido a la pérdida parcial de su cadena lateral al eliminar CO_2 .

La *sequence tag* (Mann y Wilm, 1994) permitió identificar en la base de datos del transcriptoma una única conotoxina de 37 aminoácidos que contenía dicha secuencia parcial con 8 cisteínas y al menos dos prolínas con 3 residuos de ácidos glutámicos gamma carboxilados. La masa experimental y la calculada coincidían (con un error de 2ppm) sólo al considerar que tanto el carboxilo como el amino terminal estaban libres.

Alguna de las secuencias parciales antes mencionadas pudieron ser corroboradas al interpretar los espectros MS/MS de los otros estados de cargas de la toxina analizada como se muestra en la Figura 40.

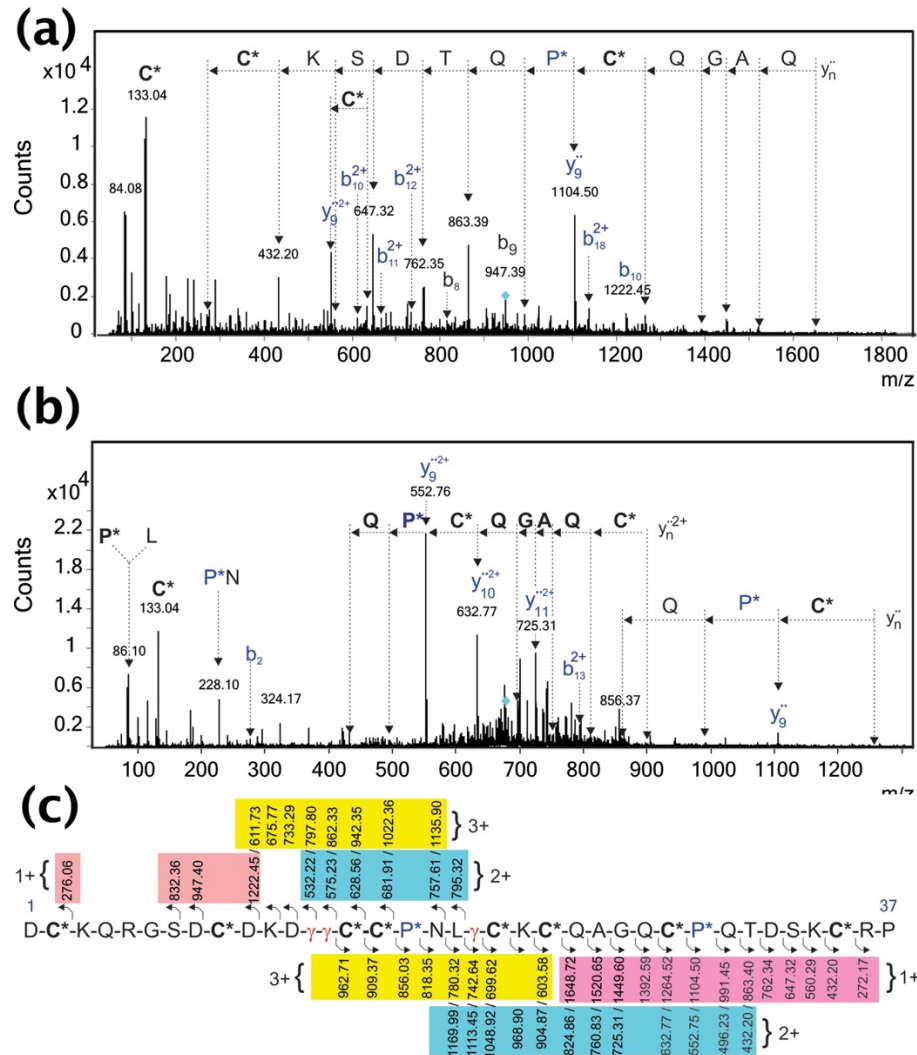


Figura 40. El espectro ES-MS/MS que se muestra en (a) y en (b) corresponde a los iones 5^+ y 7^+ del péptido rs15a. El rombo cian en ambos espectros MS/MS representa el ion precursor fragmentado en el experimento de CID. (c) Asignación de la estructura primaria asignada para los iones fragmentados de las series y_n y b_n observadas en el espectro MS/MS. Los 8 residuos de cisteínas están carboamidometilados y se representan por una C*. P* y γ representan residuos de hidroxiprolina y ácido glutámico carboxilado, respectivamente. El grupo amino y carboxilo en la toxina identificada se encuentran libres.

3.3.3 Análisis de la secuencia identificada

Utilizando la *sequence tag* se recuperó la conotoxina completa en la base de datos de *C. regularis*, que tenía un total de 163, 872 secuencias. La secuencia fue clasificada como miembro de la superfamilia O2 (Zhangsun

et al., 2006), dicha clasificación se realizó con ayuda del programa de pHMM de la Parte I de esta tesis. Sus residuos de cisteínas presentaban un acomodo idéntico a el marco de cisteínas XV (Zhangsun et al., 2006) y por lo tanto al ser la primera conotoxina de esta especie con dicho acomodo de cisteínas fue nombrada como O2_rs15a. Se realizó un alineamiento de O2_rs15a con otras conotoxinas de la superfamilia O2 con el mismo marco de cisteínas (Figura 41).

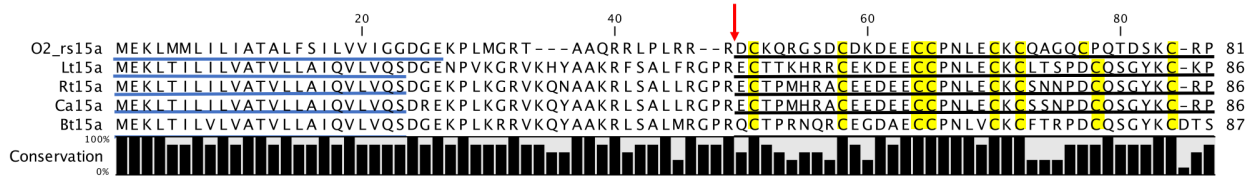


Figura 41. Alineamiento del péptido O2_rs15a con otros miembros de la superfamilia O2 con marco de cisteínas XV. Los péptidos señal y los péptidos maduros están subrayados en azul y en negro, respectivamente. Los residuos de cisteína están resaltados en color amarillo y una barra de porcentaje de conservación está representada para cada aminoácido. La flecha en rojo indica el sitio de corte de la conotoxina.

Capítulo 4. Discusión

Se ha demostrado que los componentes del veneno de los cónidos tienen un alto potencial como agentes terapéuticos y farmacológicos. El método tradicional para identificar conotoxinas se ha basado en el fraccionamiento del veneno crudo y repurificación de sus componentes generando dos pasos sumamente importantes: primero es necesario evaluar el veneno en ensayos de bioactividad y segundo caracterizar por completo aquella conotoxina que haya tenido resultados favorables. Estos procesos requieren de tiempo, mucho esfuerzo y de regresar continuamente al veneno puro que es difícil de conseguir. Las nuevas propuestas de identificación se enfocan en métodos (transcriptómica y proteómica) que permitan obtener secuencias de manera masiva, sin embargo ¿qué sucede cuando tenemos múltiples secuencias, pero desconocemos cómo pueden ser aprovechadas? En este trabajo se propone que es necesario combinar la metodología tradicional que se ha realizado durante los últimos 50 años con los nuevos avances que permiten conocer secuencias completas en un tiempo mucho más breve al anterior como es el caso de la secuencia del transcriptoma.

I1_xm11a es una conotoxina que presenta actividad antimicobacteriana, efecto nunca antes reportado para la familia Conidae, con un valor de MIC (3.0 μ M) comparable con los fármacos antifímicos utilizados en la actualidad. I1_xm11a es tres veces más potente que EMB (MIC 9.8 μ M) utilizado en la terapia actual y se mantiene en un rango similar al de INH (MIC 0.45 μ M). Existen otros trabajos que se han enfocado en la identificación de extractos biológicos de fuentes marinas con actividad anti-tuberculosis (Niederweis, 2008); no obstante solo muy pocos logran conseguir aislar una sola molécula y mucho menos secuenciarla. I1_xm11a plantea la pregunta si existen otras conotoxinas de otros miembros de la familia Conidae que presenten la actividad Mtb.

La identificación de I1_xm11a y de O2_rs15a no hubiera sido posible sin la presencia de sus transcriptomas. No fue posible recuperar la secuencia completa de la toxinas mediante la interpretación del espectro MS/MS: ni tratando de reducir y alquilar el péptido, ni por sus fragmentos trípticos. En ocasiones, en el caso de las conotoxinas, sólo es posible secuenciarlas parcialmente por espectrometría de masas y es necesario destacar que esto no resulta suficiente para la identificación en las bases de datos utilizando los programas de búsquedas que existen hasta la fecha. Estos programas están optimizados para la identificación de péptidos trípticos y no para las conotoxinas que difieren en cuanto a su longitud, el número de residuos básicos y el número de protones móviles.

En este trabajo se confirma la utilización de *sequence tags*, combinado con la transcriptómica, como una manera eficiente y relativamente rápida para la identificación de conopéptidos. Estas moléculas son secuencias no indicadas para ser fragmentadas con CID pues presentan una alta tasa de MPT incluyendo algunas que son lábiles en las condiciones como es el caso del ácido gammacaboxiglutámico y el material es difícil de obtener. En las últimas 3 décadas, alrededor de 1400 secuencias de conopéptidos han sido aisladas de 92 especies diferentes de cónidos, pero sólo 210 péptidos han sido validados a nivel de proteína (Dutertre et al., 2013). En este capítulo se presentan dos secuencias: una con actividad determinada y otra de la que se desconoce cuál es su función o sobre qué blanco molecular podría estar actuando. Esto pone en evidencia que la bioprospección o el método tradicional de descubrimiento de conotoxinas no puede ser dejado totalmente de lado.

Aquellos miembros de la superfamilia I1 que han sido completamente caracterizados, son conocidos por su actividad excitatoria (Jimenez et al., 2003); específicamente por su afinidad a los canales de Na^+ dependientes de voltaje (Buczek et al., 2007; Fiedler et al., 2008). Aunque el alineamiento con BLAST permitió clasificar a I1_xm11a dentro de la superfamilia I1, vale la pena mencionar que la región de la toxina únicamente presenta similitud con aquellos miembros de esa superfamilia que aún no han sido caracterizados (Figura 36). Las secuencias utilizadas para realizar el alineamiento también presentaban el marco de cisteínas XI, y fueron las siguientes: Ep11.1 que corresponde a una conotoxina de *Conus episcopatus*, Tx11.3 que pertenece a *Conus textile*, Vc11.4 a *C. victoriae* y Mr11.2 a *Conus marmoreus*. La divergencia de la secuencia de I1_xm11a con aquellos miembros de la superfamilia I1 descritos hasta la fecha, sugiere que la conotoxina probablemente presente un blanco molecular distinto al de los canales de Na^+ dependientes de voltaje en las membranas celulares. Además, como se pudo ver en la Figura 36, la longitud de la toxina madura I1_xm11a difiere en la longitud que fue predicha para el resto de las toxinas (Ep11.1, Tx11.3, Vc11.4 y Mr11.2). Estas conotoxinas homólogas sólo han sido identificadas a nivel de cDNA, por lo tanto, la región de la toxina madura fue predicha utilizando los sitios de corte R/K que presentan otros conopéptidos. El mecanismo llamado Procesamiento Variable del Péptido (VPP, por sus siglas en inglés), que ayuda a entender la diversidad tan excepcional de conopéptidos en el veneno de los cónidos, propone que los conopéptidos pueden ser procesados de manera diferente tanto en su amino como en su carboxilo terminal. Estas posibles variantes corresponden a procesamiento alternativo en los sitios de corte R/K que se presentan en la secuencia. En el caso de I1_xm11a existen tres posibles sitios de corte en las argininas ubicadas hacia el extremo del amino (flechas rojas, Figura 36), por lo que identificamos una de las 3 posibles variantes de procesamiento de dicha conotoxina. El VPP, así como las MPT, son un indicador de la economía celular que posiblemente estén realizando los cónidos a la hora de sintetizar los miles de conotoxinas que componen su veneno a partir de un mismo mRNA (Vázquez-Laslop y Mankin, 2014).

Por otro lado, la mayoría de las conotoxinas de la superfamilia O2 presentan un arreglo de cisteínas VI/VII o XV, no obstante sólo las conotoxinas que presentan el arreglo VI/VII han sido caracterizadas a nivel de proteína y algunas han mostrado actividad específica en canales neuronales de moluscos (no presentan efecto en otros organismos como mamíferos o larvas) (Robinson y Norton, 2014). Las conotoxinas caracterizadas de la superfamilia O2 se han clasificado como γ -conotoxinas porque presentan ácidos gamma-carboxiglutámicos. También se han reportado conotoxinas de esta superfamilia con hidroxiprolinas y bromotriptófanos en la misma secuencia. O2_rs15a es probablemente la primera conotoxina de esta superfamilia con un marco de cisteínas XV en ser identificada a partir del veneno total. O2_rs15a, al igual que otros miembros de la superfamilia O2, presenta γ -carboxiglutámicos e hidroxiprolinas. O2_rs15a ha sido probada sobre canales “pacemakers” pero no se obtuvieron resultados positivos por lo que hasta la fecha se desconoce el blanco molecular sobre el que pueda estar actuando. No obstante, cabe la posibilidad de que al igual que otros miembros de la superfamilia O2, sólo actúe sobre canales específicos de moluscos.

Capítulo 5. Conclusiones

- La utilización de *sequence tags* para obtener una secuencia completa en una base de datos, creada a partir de un transcriptoma, es un método confiable y eficiente; permitiéndonos aislar dos conotoxinas a partir del veneno total de dos especies de cónidos.
- l1_xm11a, una conotoxina con actividad antifímica, es la primera secuencia de *Conasprella ximenes* reportada. La secuencia pudo ser verificada tanto con la secuencia parcial como con los péptidos trípticos.
- O2_rs15a, una conotoxina con un elevado número de MPT, es la segunda conotoxina de *Conus regularis* en ser identificada. Presenta ocho residuos de cisteínas, dos hidroxiprolinas y tres residuos de ácido gamma carboxiglutámico. Además, es la primera conotoxina con el marco de cisteínas XV identificada para la superfamilia O2 a nivel de proteína.

Literatura citada

- Akondi, K. B., Muttenthaler, M., Dutertre, S., Kaas, Q., Craik, D. J., Lewis, R. J., & Alewood, P. F. (2014). Discovery , Synthesis , and Structure – Activity Relationships of Conotoxins. *Chemical Reviews*, 114(11), 5815–5847. <http://doi.org/10.1021/cr400401e>
- Azam, L., & McIntosh, J. M. (2009). α -conotoxins as pharmacological probes of nicotinic acetylcholine receptors. *Acta Pharmacol*, 30(6), 771–783. <http://doi.org/10.1038/aps.2009.47>
- Bandyopadhyay, P. K., Colledge, C. J., Walker, C. S., Zhou, L. M., Hillyard, D. R., & Olivera, B. M. (1998). Conantokin-G Precursor and Its Role in Gamma-Carboxylation by a Vitamin K-dependent Carboxylase from a *Conus* Snail. *Journal of Biological Chemistry*, 273(10), 5447–5450. <http://doi.org/10.1074/jbc.273.10.5447>
- Bandyopadhyay, P. K., Garrett, J. E., Shetty, R. P., Keate, T., Walker, C. S., & Olivera, B. M. (2002). Gamma-Glutamyl carboxylation: An extracellular posttranslational modification that antedates the divergence of molluscs, arthropods, and chordates. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3), 1264–1269. <http://doi.org/10.1073/pnas.022637099>
- Bernaldez, J., Jiménez, S., González, L., Ferro, J., Soto, E., Salceda, E., ... Licea-Navarro, A. (2016). A New Member of Gamma-Conotoxin Family Isolated from *Conus princeps* Displays a Novel Molecular Target. *Toxins*, 8(2), 39. <http://doi.org/10.3390/toxins8020039>
- Bernaldez, J., Román-González, S. A., Martínez, O., Jiménez, S., Vivas, O., Arenas, I., ... Licea, A. (2013). A *Conus regularis* conotoxin with a novel eight-cysteine framework inhibits CaV2.2 channels and displays an anti-nociceptive activity. *Marine Drugs*, 11(4), 1188–1202. <http://doi.org/10.3390/md11041188>
- Biggs, J. S., Watkins, M., Puillandre, N., Ownby, J.-P., Lopez-Vera, E., Christensen, S., ... Olivera, B. M. (2010). Evolution of *Conus* peptide toxins: analysis of *Conus californicus* Reeve, 1844. *Molecular Phylogenetics and Evolution*, 56(1), 1–12. <http://doi.org/10.1016/j.ympev.2010.03.029>
- Buczek, O., Bulaj, G., & Olivera, B. M. (2005). Conotoxins and the posttranslational modification of secreted gene products. *Cellular and Molecular Life Sciences: CMLS*, 62(24), 3067–79. <http://doi.org/10.1007/s00018-005-5283-0>
- Buczek, O., Wei, D., Babon, J. J., Yang, X., Fiedler, B., Chen, P., ... Norton, R. S. (2007). Structure and Sodium Channel Activity of an Excitatory I1-Superfamily Conotoxin. *Biochemistry*, 46(35), 9929–9940. <http://doi.org/10.1021/bi700797f>
- Buczek, O., Yoshikami, D., Watkins, M., Bulaj, G., Jimenez, E. C., & Olivera, B. M. (2005). Characterization of D-amino-acid-containing excitatory conotoxins and redefinition of the I-conotoxin superfamily. *FEBS Journal*, 272(16), 4178–4188. <http://doi.org/10.1111/j.1742-4658.2005.04830.x>
- Bulaj, G., Buczek, O., Goodsell, I., Jimenez, E. C., Kranski, J., Nielsen, J. S., ... Olivera, B. M. (2003). Efficient oxidative folding of conotoxins and the radiation of venomous cone snails. *Proceedings of the National Academy of Sciences of the United States of America*, 100 Suppl, 14562–14568. <http://doi.org/10.1073/pnas.2335845100>
- Bulaj, G., & Olivera, B. M. (2008). Folding of conotoxins: formation of the native disulfide bridges during chemical synthesis and biosynthesis of *Conus* peptides. *Antioxidants & Redox Signaling*, 10(1), 141–155. <http://doi.org/10.1089/ars.2007.1856>

- Cabrera, G., Lundberg, U., Rodríguez-Ulloa, A., Herrera, M., Machado, W., Portela, M., ... González, L. J. (2017). Protein content of the *Hylesia metabus* egg nest setae (Cramer [1775]) (*Lepidoptera: Saturniidae*) and its association with the parental investment for the reproductive success and lepidopterism. *Journal of Proteomics*, 150, 183–200. <http://doi.org/10.1016/j.jprot.2016.08.010>
- Callaghan, B., Haythornthwaite, A., Berecki, G., Clark, R. J., Craik, D. J., & Adams, D. J. (2008a). Analgesic α -Conotoxins Vc1.1 and Rs1A Inhibit N-Type Calcium Channels in Rat Sensory Neurons via GABAB Receptor Activation. *The Journal of Neuroscience*, 28(43), 10943–10951. <http://doi.org/10.1523/JNEUROSCI.3594-08.2008>
- Craig, A. G., Norberg, T., Griffin, D., Hoeger, C., Akhtar, M., Schmidt, K., ... Olivera, B. M. (1999). Contulakin-G, an O-Glycosylated Invertebrate Neurotensin*. *The Journal of Biological Chemistry*, 274(20), 13752–13759.
- Darby, N. J., Van Straaten, M., Penka, E., Vincentelli, R., & Kemmink, J. (1999). Identifying and characterizing a second structural domain of protein disulfide isomerase. *FEBS Letters*, 448(1), 167–172. [http://doi.org/10.1016/S0014-5793\(99\)00374-9](http://doi.org/10.1016/S0014-5793(99)00374-9)
- Davis, J., Jones, A., & Lewis, R. J. (2009). Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides*, 30(7), 1222–7. <http://doi.org/10.1016/j.peptides.2009.03.019>
- Denecke, J., De Rycke, R., & Botterman, J. (1992). Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope. *The EMBO Journal*, 11(6), 2345–2355.
- Duda, T. F., & Kohn, A. J. (2005). Species-level phylogeography and evolutionary history of the hyperdiverse marine gastropod genus *Conus*. *Molecular Phylogenetics and Evolution*, 34(2), 257–272. <http://doi.org/10.1016/j.ympev.2004.09.012>
- Dutertre, S., Jin, A., Kaas, Q., Jones, A., Alewood, P. F., & Lewis, R. J. (2013). Deep Venomics Reveals the Mechanism for Expanded Peptide Diversity in Cone Snail Venom. *Molecular & Cellular Proteomics*, 12(2), 312–329. <http://doi.org/10.1074/mcp.M112.021469>
- Ellgaard, L., & Ruddock, L. W. (2005). The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Reports*, 6(1), 28–32. <http://doi.org/10.1038/sj.embor.7400311>
- Elliger, C. a, Richmond, T. a, Lebaric, Z. N., Pierce, N. T., Sweedler, J. V., & Gilly, W. F. (2011). Diversity of conotoxin types from *Conus californicus* reflects a diversity of prey types and a novel evolutionary history. *Toxicon: Official Journal of the International Society on Toxinology*, 57(2), 311–22. <http://doi.org/10.1016/j.toxicon.2010.12.008>
- England, L. J., Imperial, J., Jacobsen, R., Craig, A. G., Gulyas, J., Akhtar, M., ... Olivera, B. M. (1998). Inactivation of a Serotonin-Gated Ion Channel by a Polypeptide Toxin from Marine Snails, 281(July), 575–579.
- Eswar, N., Webb, B., Marti-Renom, M. a, Madhusudhan, M. S., Eramian, D., Shen, M.-Y., ... Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Current Protocols in Protein Science/Editorial Board, John E. Coligan et al., Chapter 2, Unit 2.9*. <http://doi.org/10.1002/0471140864.ps0209s50>
- Fainzilber, M., Gordon, D., Hasson, A., Spirazs, M. E., & Zlotkin, E. (1991). Mollusc-specific toxins from the venom of *Conus textile neovicarius*, 595, 589–595.

- Fainzilber, M., Nakamura, T., Lodder, J. C., Zlotkin, E., Kits, K. S., & Burlingame, A. L. (1998). γ -Conotoxin-PnVIIA, A γ -Carboxyglutamate-Containing Peptide Agonist of Neuronal Pacemaker Cation Currents $\dagger$. *Biochemistry*, 2960(97), 1470–1477.
- Favreau, P., & Stöcklin, R. (2009). Marine snail venoms: use and trends in receptor and channel neuropharmacology. *Current Opinion in Pharmacology*, 9(5), 594–601. <http://doi.org/10.1016/j.coph.2009.05.006>
- Fiedler, B., Zhang, M., Buczek, O., Azam, L., Bulaj, G., Norton, R. S., ... Yoshikami, D. (2008). Specificity, affinity and efficacy of iota-conotoxin RXIA, an agonist of voltage-gated sodium channels NaV1.2, 1.6 and 1.7. *Biochemical Pharmacology*, 75(12), 2334–2344. <http://doi.org/10.1016/j.bcp.2008.03.019>
- Figueroa-Montiel, A., Ramos, M. A., Mares, R. E., Dueñas, S., Pimienta, G., Ortiz, E., ... Licea-Navarro, A. F. (2016). In Silico Identification of Protein Disulfide Isomerase Gene Families in the De Novo Assembled Transcriptomes of Four Different Species of the Genus *Conus*. *Plos One*, 11(2), e0148390. <http://doi.org/10.1371/journal.pone.0148390>
- Fry, B. G., Scheib, H., Junqueira, I. D. L. M., Azevedo, D., Andrade, D., & Casewell, N. R. (2012). Toxicon Novel transcripts in the maxillary venom glands of advanced snakes. *Toxicon*, 59(7–8), 696–708. <http://doi.org/10.1016/j.toxicon.2012.03.005>
- Galligan, J. J., & Petersen, D. R. (2012). The human protein disulfide isomerase gene family. *Human Genomics*, 6(1), 6. <http://doi.org/10.1186/1479-7364-6-6>
- Global Tuberculosis Report. (2016). *World Health Organization*. <http://doi.org/WHO/HTM/TB/2016.13>
- Gowd, K. H., Krishnan, K. S., & Balaram, P. (2007). Identification of *Conus amadis* disulfide isomerase: minimum sequence length of peptide fragments necessary for protein annotation. *Molecular bioSystems*, 3, 554–566. <http://doi.org/10.1039/b705382g>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. a, Amit, I., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <http://doi.org/10.1038/nbt.1883>
- Hu, H., Bandyopadhyay, P. K., Olivera, B. M., & Yandell, M. (2011). Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genomics*, 12(1), 60. <http://doi.org/10.1186/1471-2164-12-60>
- Hu, H., Bandyopadhyay, P. K., Olivera, B. M., & Yandell, M. (2012). Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC Genomics*, 13(1), 284. <http://doi.org/10.1186/1471-2164-13-284>
- Jakubowski, J. a, Kelley, W. P., & Sweedler, J. V. (2006). Screening for post-translational modifications in conotoxins using liquid chromatography/mass spectrometry: an important component of conotoxin discovery. *Toxicon: Official Journal of the International Society on Toxinology*, 47(6), 688–99. <http://doi.org/10.1016/j.toxicon.2006.01.021>
- Jimenez, E. C., Olivera, B. M., & Teichert, R. W. (2007). α C-Conotoxin PrXA: A New Family of Nicotinic Acetylcholine Receptor Antagonists, 46(1), 8717–8724. <http://doi.org/10.1021/bi700582m>
- Jimenez, E. C., Shetty, R. P., Lirazan, M., Rivier, J., Walker, C., Abogadie, F. C., ... Olivera, B. M. (2003). Novel excitatory *Conus* peptides define a new conotoxin superfamily. *Journal of Neurochemistry*, 85(3), 610–621. <http://doi.org/10.1046/j.1471-4159.2003.01685.x>

- Jin, A., Vetter, I., Hamilton, B., Sunagar, K., Lavergne, V., Dutertre, V., ... Lewis, R. J. (2014). Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. <http://doi.org/10.1038/ncomms4521>
- Jin, A., Vetter, I., Himaya, S. W. A., Paul, F., & Lewis, R. J. (2015). Transcriptome and proteome of *Conus planorbis* identify the nicotinic receptors as primary target for the defensive venom, 4030–4040. <http://doi.org/10.1002/pmic.201500220>
- Kaas, Q., Westermann, J.-C., & Craik, D. J. (2010). Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon : Official Journal of the International Society on Toxinology*, 55(8), 1491–509. <http://doi.org/10.1016/j.toxicon.2010.03.002>
- Kanai, S., Toh, H., Hayano, T., & Kikuchi, M. (1998). Molecular evolution of the domain structures of protein disulfide isomerases. *Journal of Molecular Evolution*, 47(2), 200–210. <http://doi.org/10.1007/PL00006377>
- Kang, H., Seo, C., & Park, Y. (2015). Marine Peptides and Their Anti-Infective Activities. *Marine Drugs*, 13(1), 618–654. <http://doi.org/10.3390/md13010618>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <http://doi.org/10.1093/molbev/mst010>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. <http://doi.org/10.1093/bioinformatics/bts199>
- Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M., & Creighton, T. E. (1997). The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. *Current Biology : CB*, 7(4), 239–245. [http://doi.org/10.1016/S0960-9822\(06\)00119-9](http://doi.org/10.1016/S0960-9822(06)00119-9)
- Ken Gayler, David Sandall, David Greening, David Keays, Megan Polidano, Bruce Livett, John Down, Narmatha Satkunanathan, Z. K. (2005). Molecular Prospecting for Drugs from the Sea. Isolating Therapeutic Peptides and Proteins from Cone snail Venom. *IEEE Eng Med Biol Mag.*, (April), 79–84.
- Klappa, P., Ruddock, L. W., Darby, N. J., & Freedman, R. B. (1998). The b' domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. *EMBO Journal*, 17(4), 927–935. <http://doi.org/10.1093/emboj/17.4.927>
- Knodler, L. a., Noiva, R., Mehta, K., McCaffery, J. M., Aley, S. B., Svärd, S. G., ... Gillin, F. D. (1999). Novel protein-disulfide isomerases from the early-diverging protist *Giardia lamblia*. *Journal of Biological Chemistry*, 274(42), 29805–29811. <http://doi.org/10.1074/jbc.274.42.29805>
- Krogh, A. (1998). An Introduction to Hidden Markov Models for Biological Sequences (Chapter 4). *Computational Methods in Molecular Biology*, 32, 45–63. [http://doi.org/10.1016/S0167-7306\(08\)60461-5](http://doi.org/10.1016/S0167-7306(08)60461-5)
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <http://doi.org/10.1093/molbev/msw054>

- Lavergne, V., Dutertre, S., Jin, A., Lewis, R. J., Taft, R. J., & Alewood, P. F. (2013). Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. *BMC Genomics*, 14, 708. <http://doi.org/10.1186/1471-2164-14-708>
- Lavergne, V., Harliwong, I., Jones, A., Miller, D., Taft, R. J., & Alewood, P. F. (2015). Optimized deep-targeted proteotranscriptomic profiling reveals unexplored *Conus* toxin diversity and novel cysteine frameworks. *Proceedings of the National Academy of Sciences*, 201501334. <http://doi.org/10.1073/pnas.1501334112>
- Lewis, R. J., Dutertre, S., Vetter, I., & Christie, M. J. (2012). *Conus* Venom Peptide Pharmacology. *Pharmacological Reviews*, 64(2), 259–298. <http://doi.org/10.1124/pr.111.005322>
- Lewis, R. J., & Garcia, M. L. (2003). Therapeutic potential of venom peptides. *Nature Reviews Drug Discovery*, 2(10), 790–802. <http://doi.org/10.1038/nrd1197>
- Li, W. (2008). Molecular Clocks. *eLS*, 2000. <http://doi.org/10.1002/9780470015902.a0005111.pub2>
- Lirazan, M. B., Hooper, D., Corpuz, G. P., Ramilo, C. A., Bandyopadhyay, P., Cruz, L. J., & Olivera, B. M. (2000). The Spasmodic Peptide Defines a New Conotoxin Superfamily †. *Biochemistry*, 39(7). <http://doi.org/10.1021/bi9923712>
- Liu, J., Wu, Q., Pi, C., Zhao, Y., Zhou, M., Wang, L., ... Xu, A. (2007). Isolation and characterization of a T-superfamily conotoxin from *Conus litteratus* with targeting tetrodotoxin-sensitive sodium channels. *Peptides*, 28, 2313–2319. <http://doi.org/10.1016/j.peptides.2007.09.006>
- Lluisma, A. O., Milash, B. A., Moore, B., Olivera, B. M., & Bandyopadhyay, P. K. (2012). Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. *Marine Genomics*, 5, 43–51. <http://doi.org/10.1016/j.margen.2011.09.002>
- Loughnan, M., Nicke, A., Jones, A., Schroeder, C. I., Nevin, S. T., Adams, D. J., ... Lewis, R. J. (2006). Identification of a Novel Class of Nicotinic Receptor Antagonists Dimeric Conotoxins VxXIIA, VXXIIB, and VXXIIC from *Conus Vexillum**. *The Journal of Biological Chemistry*, 281(34), 24745–24755. <http://doi.org/10.1074/jbc.M603703200>
- Lucero, H. a, & Kaminer, B. (1999). The Role of Calcium on the Activity of ERcalcistorin / Protein- disulfide Isomerase and the Significance of the C-terminal and Its, 274(5), 3243–3251.
- Lugo, P., Díaz, F., Re, A. D., Olivares, F., González, R., Dueñas, S., & Licea, A. (2016). Thermoregulatory behaviour and thermal tolerance of three species of Conidae in the Eastern Pacific and Gulf of California coasts of Baja California, Mexico. *Molluscan Research*, (July), 1–8. <http://doi.org/10.1080/13235818.2016.1172545>
- Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., & Lajoie, G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry : RCM*, 17(20), 2337–2342. <http://doi.org/10.1002/rcm.1196>
- Mann, M., & Wilm, M. (1994). ErrorTolerant identification of Peptides in Sequence Databases by Peptide Sequence Tags, 66(24), 4390–4399.
- Marshall, J., Kelley, W. P., Rubakhin, S. S., Bingham, J. P., Sweedler, J. V., & Gilly, W. F. (2002). Anatomical correlates of venom production in *Conus californicus*. *Biological Bulletin*, 203(August), 27–41.

- Mazzarella, R. A., Srinivasan, M., Haugejorden, S. M., & Green, M. (1990). ERp72, an abundant luminal endoplasmic reticulum protein, contains three copies of the active site sequences of protein disulfide isomerase. *Journal of Biological Chemistry*, 265(2), 1094–1101.
- Mcdougal, O. M., Turner, M. W., Ormond, A. J., & Poulter, C. D. (2008). Three-Dimensional Structure of Conotoxin tx3a : An m-1 Branch Peptide of the M-superfamily. *Biochemistry*, 47, 2826–2832.
- Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E. P., Shabanowitz, J., & Hunt, D. F. (2006). The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1764(12), 1811–1822. <http://doi.org/10.1016/j.bbapap.2006.10.003>
- Miles, L. A., Dy, C. Y., Nielsen, J., Barnham, K. J., Hinds, M. G., Olivera, B. M., ... Norton, R. S. (2002). Structure of a Novel P-superfamily Spasmodic Conotoxin Reveals an Inhibitory Cystine Knot Motif *, 277(45), 43033–43040. <http://doi.org/10.1074/jbc.M206690200>
- Moller, C., & Marí, F. (2011). 9.3 kDa Components of the Injected Venom of *Conus purpurascens* Define a New 5-disulfide Conotoxin Framework. *Biopolymers*, 96(2), 158–165. <http://doi.org/10.1002/bip.21406.9.3>
- Neves, J., Campos, A., Osório, H., Antunes, A., & Vasconcelos, V. (2013). Conopeptides from Cape Verde *Conus crotchii*. *Marine Drugs*, 11(6), 2203–15. <http://doi.org/10.3390/md11062203>
- Niederweis, M. (2008). Nutrient acquisition by mycobacteria. *Microbiology*, 154(3), 679–692. <http://doi.org/10.1099/mic.0.2007/012872-0>
- Norton, R. S., & Olivera, B. M. (2006). Conotoxins down under. *Toxicon : Official Journal of the International Society on Toxinology*, 48(7), 780–98. <http://doi.org/10.1016/j.toxicon.2006.07.022>
- Oers, N. S. C. Van, Killeen, N., Weiss, A., Wiest, D. L., Ashe, J. M., Abe, R., ... Olivera, B. M. (1998). Inactivation of a Serotonin-Gated Ion Channel by a Polypeptide Toxin from Marine Snails, 281(July), 575–579.
- Olivera, B. M. (2002). CONUS VENOM PEPTIDES: Reflections from the Biology of Clades and Species. *Annual Review of Ecology and Systematics*, 33(1), 25–47. <http://doi.org/10.1146/annurev.ecolsys.33.010802.150424>
- Olivera, B. M. (2006). *Conus* peptides: biodiversity-based discovery and exogenomics. *The Journal of Biological Chemistry*, 281(42), 31173–7. <http://doi.org/10.1074/jbc.R600020200>
- Olivera, B. M., & Cruz, L. J. (2001). Conotoxins, in retrospect. *Toxicon : Official Journal of the International Society on Toxinology*, 39(1), 7–14. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10936619>
- Peng, C., Tang, S., Pi, C., Liu, J., Wang, F., Wang, L., ... Xu, A. (2006). Discovery of a novel class of conotoxin from *Conus litteratus* , lt14a , with a unique cysteine pattern. *Peptides*, 27, 2174–2181. <http://doi.org/10.1016/j.peptides.2006.04.016>
- Peng, C., Yao, G., Gao, B.-M., Fan, C.-X., Bian, C., Wang, J., ... Shi, Q. (2016). High-throughput identification of novel conotoxins from the Chinese tubular cone snail (*Conus betulinus*) by multi-transcriptome sequencing. *GigaScience*, 5(1), 17. <http://doi.org/10.1186/s13742-016-0122-9>
- Perkins, D. N., Pappin, D. J. C., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18), 3551–3567. [http://doi.org/10.1002/\(SICI\)1522-2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](http://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2)

- Petrel, C., Hocking, H., Reynaud, M., Upert, G., Favreau, P., Biass, D., ... Servent, D. (2013). Identification, structural and pharmacological characterization of t-CnVA, a conopeptide that selectively interacts with somatostatin sst3 receptor. *Biochemical Pharmacology*, 85(11), 1663–1671. <http://doi.org/10.1016/j.bcp.2013.03.019>
- Phuong, M. A., Mahardika, G. N., & Alfaro, M. E. (2016). Dietary breadth is positively correlated with venom complexity in cone snails. *BMC Genomics*, 17, 401. <http://doi.org/10.1186/s12864-016-2755-6>
- Pi, C., Liu, J., Peng, C., Liu, Y., Jiang, X., Zhao, Y., ... Xu, A. (2006). Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus* ☆, 88, 809–819. <http://doi.org/10.1016/j.ygeno.2006.06.014>
- Prashanth, J. R., Lewis, R. J., & Dutertre, S. (2012). Towards an integrated venomomics approach for accelerated conopeptide discovery. *Toxicon: Official Journal of the International Society on Toxinology*, 60(4), 470–7. <http://doi.org/10.1016/j.toxicon.2012.04.340>
- Puillandre, N., Duda, T. F., Meyer, C., Olivera, B. M., & Bouchet, P. (2015). One, four or 100 genera? A new classification of the cone snails. *Journal of Molluscan Studies*, 81(1), 1–23. <http://doi.org/10.1093/mollus/eyu055>
- Puillandre, N., Koua, D., Favreau, P., Olivera, B. M., & Stöcklin, R. (2012). Molecular phylogeny, classification and evolution of conopeptides. *Journal of Molecular Evolution*, 74(5–6), 297–309. <http://doi.org/10.1007/s00239-012-9507-2>
- Ramos, M. A., Sánchez-Lopez, R., Mares, R. E., Olvera, F., & Alagón, A. (2005). Identification of an *Entamoeba histolytica* gene encoding a protein disulfide isomerase that functionally complements the dsbA mutation in Escherichia coli. *Molecular and Biochemical Parasitology*, 143, 236–240. <http://doi.org/10.1016/j.molbiopara.2005.04.013>
- Rendón-Anaya, M., Camargos, T. S., & Ortiz, E. (2013). Scorpion Venom Gland Transcriptomics. In *Toxinology* (pp. 1–14). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-6647-1_27-1
- Rigny, A. C., Lucas-Meunier, E., Kalume, D. E., Czerwicz, E., Hambe, B., Dahlqvist, I., ... Stenflo, J. (1999). A conotoxin from *Conus textile* with unusual posttranslational modifications reduces presynaptic Ca^{2+} influx. *Proceedings of the National Academy of Sciences*, 96(May), 5758–5763.
- RNA-Seq De novo Assembly Using Trinity. Recuperado en abril del 2015 de <http://trinityrnaseq.github.io/>
- Robinson, N. E., & Robinson, A. B. (2001a). Prediction of protein deamidation rates from primary and three-dimensional structure. *Proceedings of the National Academy of Sciences*, 98(8), 4367–4372. <http://doi.org/10.1073/pnas.071066498>
- Robinson, N. E., & Robinson, a B. (2001b). Deamidation of human proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 98(22), 12409–12413. <http://doi.org/10.1073/pnas.221463198>
- Robinson, S. D., Li, Q., Lu, A., Bandyopadhyay, P. K., Yandell, M., Olivera, B. M., & Safavi-Hemami, H. (2017). The Venom Repertoire of *Conus gloriamaris* (Chemnitz, 1777), the Glory of the Sea. *Marine Drugs*, 15(6), 145. <http://doi.org/10.3390/md15050145>
- Robinson, S. D., & Norton, R. S. (2014). Conotoxin gene superfamilies. *Marine Drugs*, 12(12), 6058–6101. <http://doi.org/10.3390/md12126058>

- Robinson, S. D., Safavi-Hemami, H., McIntosh, L. D., Purcell, A. W., Norton, R. S., & Papenfuss, A. T. (2014). Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. *PLoS ONE*, 9(2). <http://doi.org/10.1371/journal.pone.0087648>
- Safavi-Hemami, H., Gorasia, D. G., Steiner, A. M., Williamson, N. a., Karas, J. a., Gajewiak, J., ... Purcell, A. W. (2012). Modulation of conotoxin structure and function is achieved through a multienzyme complex in the venom glands of cone snails. *Journal of Biological Chemistry*, 287(41), 34288–34303. <http://doi.org/10.1074/jbc.M112.366781>
- Safavi-Hemami, H., Hu, H., Gorasia, D. G., Bandyopadhyay, P. K., Veith, P. D., Young, N. D., ... Purcell, A. W. (2014). Combined proteomic and transcriptomic interrogation of the venom gland of *Conus geographus* uncovers novel components and functional compartmentalization. *Molecular & Cellular Proteomics : MCP*, 13(4), 938–53. <http://doi.org/10.1074/mcp.M113.031351>
- Safavi-Hemami, H., Li, Q., Jackson, R. L., Song, A. S., Boomsma, W., Bandyopadhyay, P. K., ... Ellgaard, L. (2016). Rapid expansion of the protein disulfide isomerase gene family facilitates the folding of venom peptides. *Proceedings of the National Academy of Sciences*, 113(12), 3227–3232. <http://doi.org/10.1073/pnas.1525790113>
- Sahrawy, M., Hecht, V., Lopez-Jaramillo, J., Chueca, A., Chartier, Y., & Meyer, Y. (1996). Intron position as an evolutionary marker of thioredoxins and thioredoxin domains. *Journal of Molecular Evolution*, 42(4), 422–431. <http://doi.org/10.1007/BF02498636>
- Sharpe, I. A., Gehrmann, J., Loughnan, M. L., Thomas, L., Adams, D. A., Atkins, A., ... Lewis, R. J. (2001). Two new classes of conopeptides inhibit the α 1-adrenoceptor and noradrenaline transporter. *Nature Neuroscience*, 4(9), 902–907.
- Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., ... Schaeffer, D. a. (2007). The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Molecular & Cellular Proteomics : MCP*, 6(9), 1638–1655. <http://doi.org/10.1074/mcp.T600050-MCP200>
- Tachikawa, H., Takeuchi, Y., Funahashi, W., Miura, T., Gao, X. D., Fujimoto, D., ... Onodera, K. (1995). Isolation and characterization of a yeast gene, MPD1, the overexpression of which suppresses inviability caused by protein disulfide isomerase depletion. *FEBS Letters*, 369(2–3), 212–216. [http://doi.org/10.1016/0014-5793\(95\)00750-4](http://doi.org/10.1016/0014-5793(95)00750-4)
- Teichert, R. W., Jimenez, E. C., & Olivera, B. M. (2005). α S-Conotoxin RVIIIA : A Structurally Unique Conotoxin That Broadly Targets Nicotinic Acetylcholine Receptors †. *Biochemistry*, 44, 7897–7902.
- Terlau, H., & Olivera, B. M. (2004). *Conus* Venoms: A Rich Source of Novel Ion Channel-Targeted Peptides. *Physiological Reviews*, 84(1), 41–68. <http://doi.org/10.1152/physrev.00020.2003>
- Ueberheide, B. M., Fenyő, D., Alewood, P. F., & Chait, B. T. (2009). Rapid sensitive analysis of cysteine rich peptide venom components. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17), 6910–5. <http://doi.org/10.1073/pnas.0900745106>
- Vázquez-Laslop, N., & Mankin, A. S. (2014). Protein Accounting in the Cellular Economy. *Cell*, 157(3), 529–531. <http://doi.org/10.1016/j.cell.2014.04.002>

- Violette, A., Blass, D., Dutertre, S., Koua, D., Piquemal, D., Pierrat, F., ... Favreau, P. (2012). Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. *Journal of Proteomics*, 75(17), 5215–25. <http://doi.org/10.1016/j.jprot.2012.06.001>
- Voelker, R., Small, C., Bassham, S., Catchen, J., Sydes, J. y Cresko B. RNA-seqlopedia. University of Oregon. Recuperado en abril del 2015 de <http://rnaseq.uoregon.edu/#library-prep>
- Walker, C. S., Jensen, S., Ellison, M., Matta, J. A., Lee, W. Y., Imperial, J. S., ... Maricq, A. V. (2010). A Novel *Conus* snail Polypeptide Causes Excitotoxicity by Blocking Desensitization of AMPA Receptors. *Current Biology : CB*, 19(11), 900–908. <http://doi.org/10.1016/j.cub.2009.05.017>
- Walker, C. S., Steel, D., Jacobsen, R. B., Lirazan, M. B., Cruz, L. J., Hooper, D., ... Olivera, B. M. (1999). The T-superfamily of Conotoxins. *Journal of Biological Chemistry*, 274(43), 30664–30671. <http://doi.org/10.1074/jbc.274.43.30664>
- Wang, X., & Zhang, B. (2013). customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics (Oxford, England)*, 29(24), 3235–7. <http://doi.org/10.1093/bioinformatics/btt543>
- Wilkinson, B., & Gilbert, H. F. (2004). Protein disulfide isomerase. *Biochimica et Biophysica Acta*, 1699(1–2), 35–44. <http://doi.org/10.1016/j.bbapap.2004.02.017>
- Winnebeck, E. C., Millar, C. D., & Warman, G. R. (2010). Why does insect RNA look degraded? *Journal of Insect Science (Online)*, 10(159), 159. <http://doi.org/10.1673/031.010.14119>
- Wysocki, V. H., Tsaprailis, G., Smith, L. L., & Brei, L. A. (2000). Mobile and localized protons: A framework for understanding peptide dissociation. *Journal of Mass Spectrometry*, 35(12), 1399–1406. [http://doi.org/10.1002/1096-9888\(200012\)35:12<1399::AID-JMS86>3.0.CO;2-R](http://doi.org/10.1002/1096-9888(200012)35:12<1399::AID-JMS86>3.0.CO;2-R)
- Zhangsun, D., Luo, S., & Wu, Y. (2006). Novel O-superfamily Conotoxins Identified by cDNA Cloning From Three Vermivorous *Conus* Species, (1), 256–265. <http://doi.org/10.1111/j.1747-0285.2006.00443.x>
- Zubarev, R. A., Zubarev, A. R., & Savitski, M. M. (2008). Electron Capture/Transfer versus Collisionally Activated/Induced Dissociations: Solo or Duet? *Journal of the American Society for Mass Spectrometry*, 19(6), 753–761. <http://doi.org/10.1016/j.jasms.2008.03.007>

Anexo 1. Construcción de los pHMM

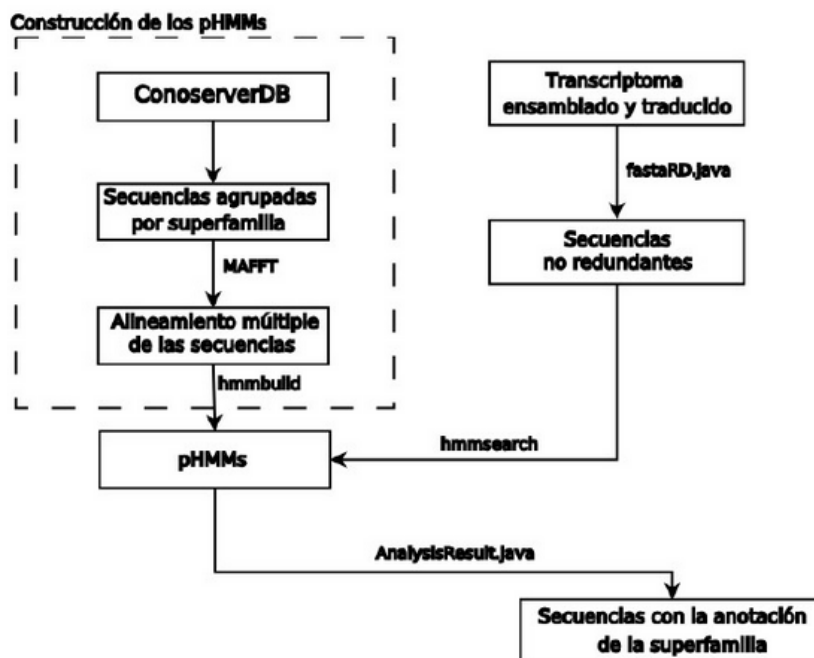


Figura 42. Flujo de trabajo para la anotación de las conotoxinas en las diversas superfamilias de genes utilizando pHMMs de un transcriptoma de la glándula venenosa de un miembro de la familia Conidae. Creado por M.C. Jesús Armando Beltrán Verdugo.

Este Anexo está basado en el reporte generado por el M.C. Jesús Armando Beltrán Verdugo. En primer lugar, se construyó una base de datos que permitiera fabricar los perfiles. Se recuperaron las secuencias de conopéptidos de la base de datos ConoServer y sólo se utilizaron aquellas que estuvieran completas (secuencias que contenían el péptido pre, pro y maduro). Las secuencias fueron agrupadas por superfamilias de genes (clasificación proporcionada por ConoServer y se les sumaron 6 secuencias de la superfamilia J de Biggs (Biggs et al., 2010). En total quedaron 1961 secuencias (Tabla 17). Posteriormente se realizó un alineamiento múltiple de las secuencias por cada superfamilia de genes. Se utilizó el programa MAFFT (Katoh & Standley, 2013) versión 7 con la estrategia L-INS-i con un máximo de 1000 iteraciones. Con el objetivo de automatizar el alineamiento múltiple de las secuencias por superfamilia se creó un *script* que recibe como entrada la dirección de los archivos fasta y la dirección donde se almacenarán los alineamientos múltiples de salida. El *script* da como salida un archivo con extensión *.aln por cada archivo fasta de entrada. A continuación, se crearon los pHMM para cada superfamilia de genes. En este paso se utilizó la herramienta *hmmbuild* del programa HMMER versión 3.0, que permite construir el pHMM. De la misma manera que en el paso anterior se creó un *script* que creaba de manera automática un pHMMs por

cada superfamilia de genes. Posteriormente se creó un *script* para concatenar todos los pHMMs en un sólo archivo. Después se realizó la anotación de la superfamilia de genes con los pHMMs creados. Los pHMMs son usados para buscar y anotar los conopéptidos a partir de un conjunto de secuencias de aminoácidos, que en este caso son las bases de datos creados a partir del transcriptoma. A estas bases de datos se le eliminaron las secuencias no redundantes con un programa llamado fastaRD.java (creado también por M.C. Jesús Armando Beltrán Verdugo). La herramienta *hmmsearch* del programa HMMER (versión 3.0) fue utilizada en las secuencias no redundantes del transcriptoma. Esta herramienta entrega aquellas secuencias que empataron con al menos un pHMM (Figura 42). Se estableció un e-value con valor de 0.011 como un umbral para clasificar las conotoxinas dentro de una superfamilia.

Tabla 17. Resumen de las secuencias de conopéptidos recuperadas de ConoServer.

Superfamilia	Nro. secuencias	Superfamilia	Nro. secuencias	Superfamilia	Nro. secuencias
A	247	JBiggs	6	Divergent_M---L-LTVA	9
B1	18	K	4	Divergent_MKFPLLFI SL	1
B2	2	L	15	Divergent_MKLCVVIVLL	3
B3	1	M	440	Divergent_MKVAVVLLVS	1
C	4	N	33	Divergent_MRCLSIFVLL	2
D	29	O1	566	Divergent_MRFLHFLIVA	1
E	1	O2	136	Divergent_MRFYIGLMAA	3
F	2	O3	42	Divergent_MSKLVILAVL	1
G	1	P	12	Divergent_MSTLGMTLL-	6
H	10	Q	22	Divergent_MTAKATLLVL	1
I1	25	S	19	Divergent_MTFLLLLVSV	1
I2	61	T	224	Divergent_MTLTFLLVVA	1
I3	9	V	2		
J	29	Y	1		

Anexo 2. Artículo publicado

Figuerola-Montiel, A., Ramos, M. A., Mares, R. E., Dueñas, S., Pimienta, G., Ortiz, E., ... Licea-Navarro, A. F. (2016). In Silico Identification of Protein Disulfide Isomerase Gene Families in the De Novo Assembled Transcriptomes of Four Different Species of the Genus *Conus*. *Plos One*, 11(2), e0148390. <http://doi.org/10.1371/journal.pone.0148390>

Anexo 3. Gráficas de la calidad de las secuencias

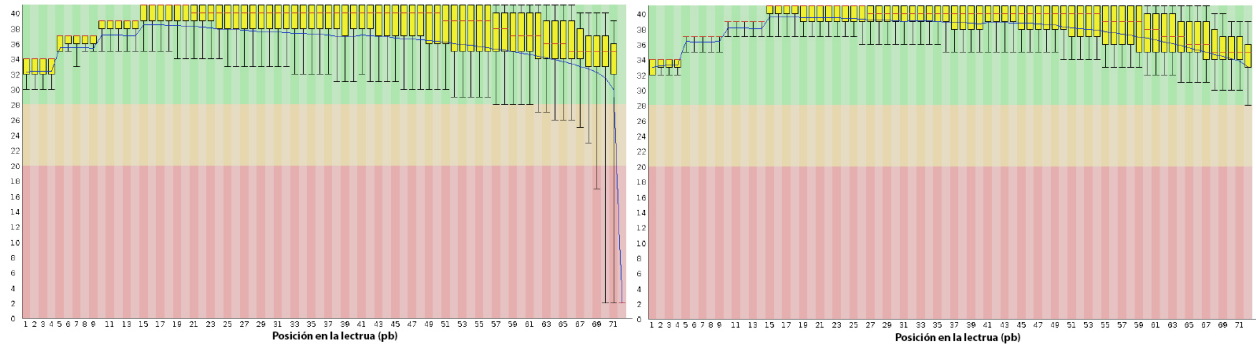


Figura 43. Gráfica de calidad por base para la secuenciación con el cebador R1 (izquierda) y con el cebador R2 (derecha) de *C. californicus*.

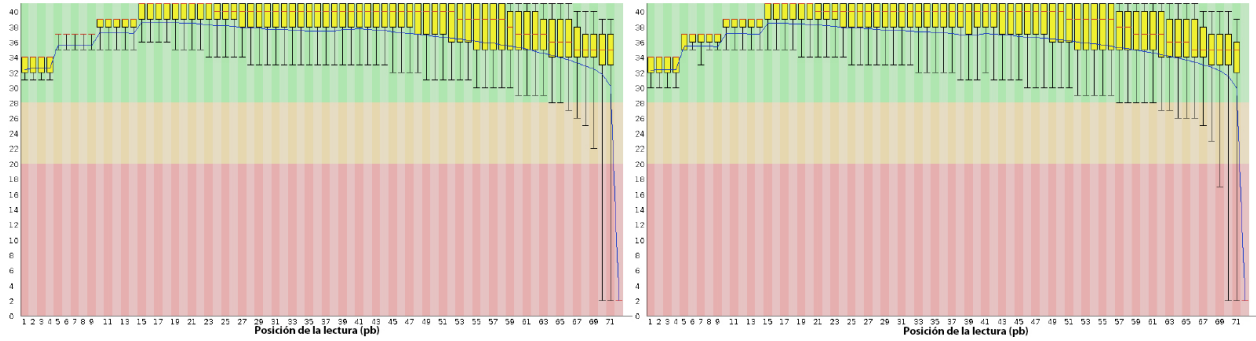


Figura 44. Gráfica de calidad por base para la secuenciación con el cebador R1 (izquierda) y con el cebador R2 (derecha) de *C. regularis*.

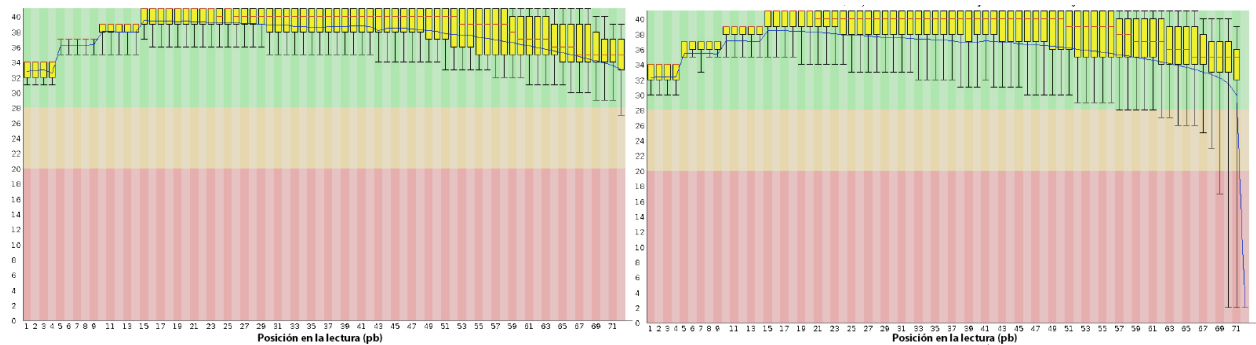


Figura 45. Gráfica de calidad por base para la secuenciación con el cebador R1 (izquierda) y con el cebador R2 (derecha) de *C. mahogani*.

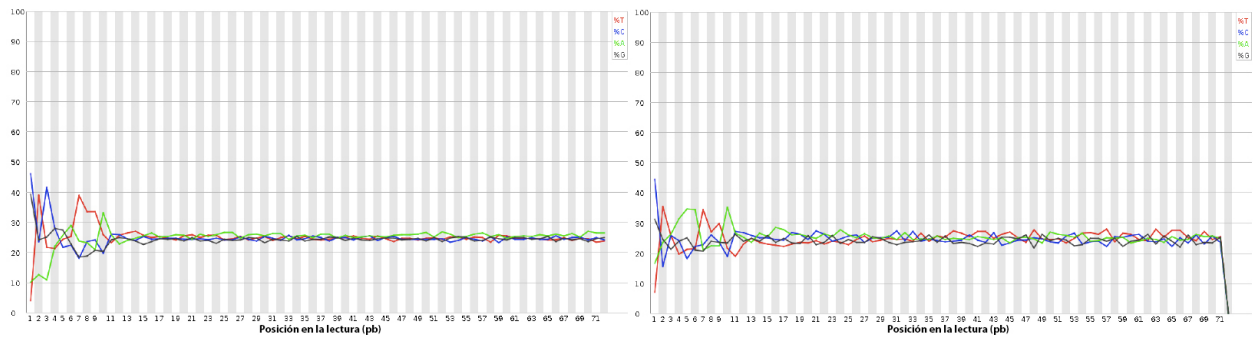


Figura 46. Contenido de bases a lo largo de la secuencia correspondiente al cebador R1 y R2 de *C. mahogani*.

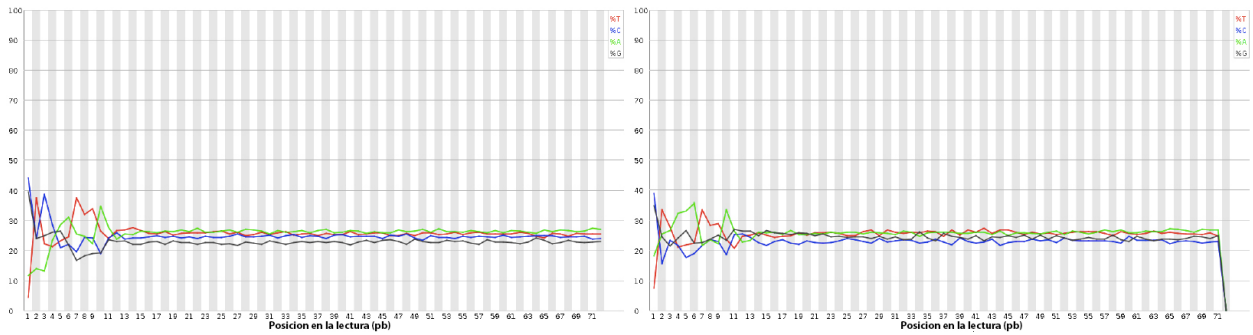


Figura 47. Contenido de bases a lo largo de la secuencia correspondiente al cebador R1 y R2 de *C. ximenes*.

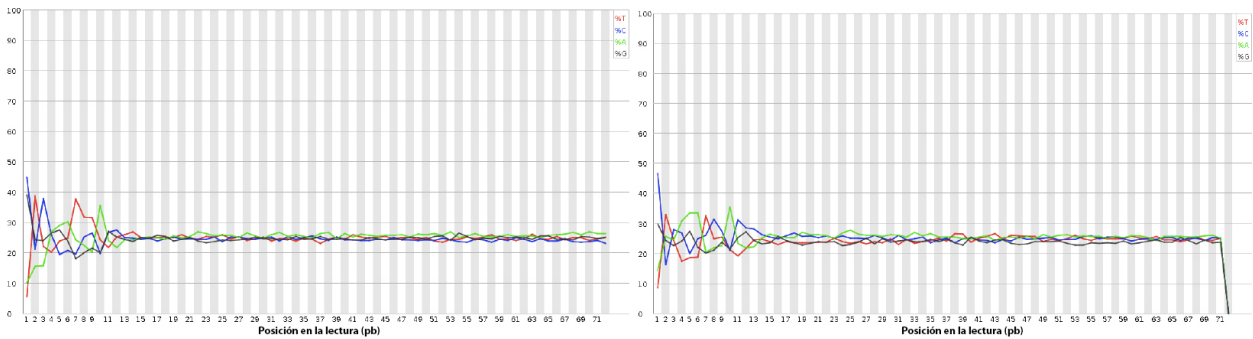


Figura 48. Contenido de bases a lo largo de la secuencia correspondiente al cebador R1 y R2 de *C. californicus*.

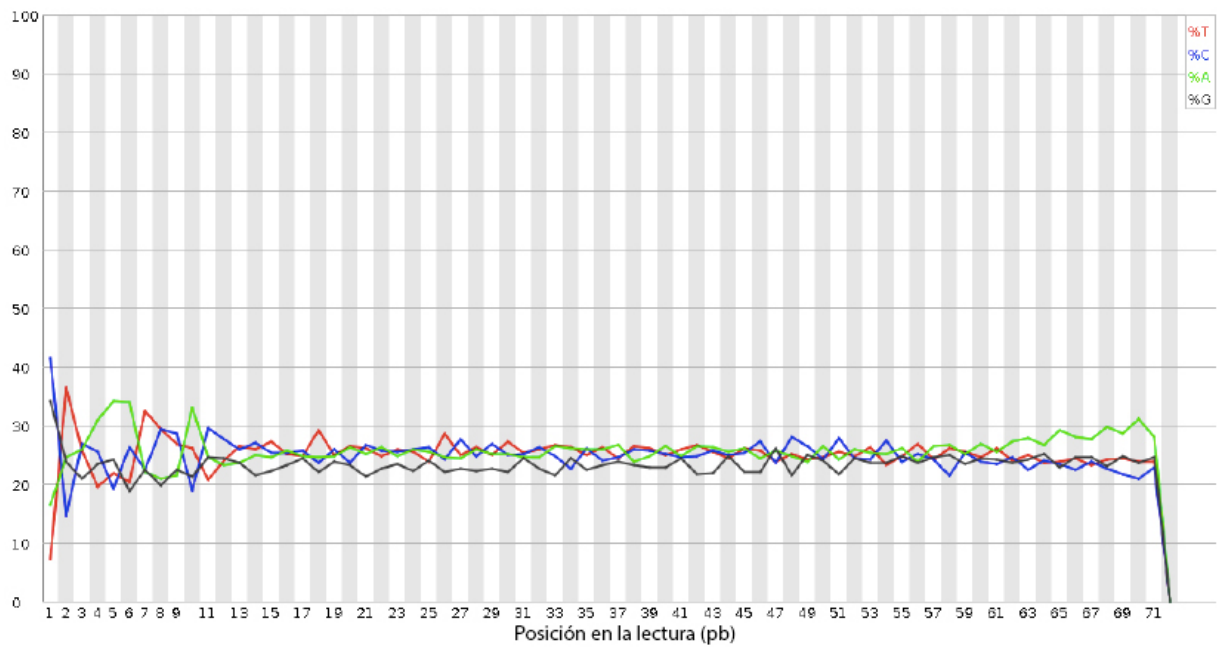


Figura 49. Contenido de bases a lo largo de la secuencia correspondiente al cebador R2 de *C. regularis*.

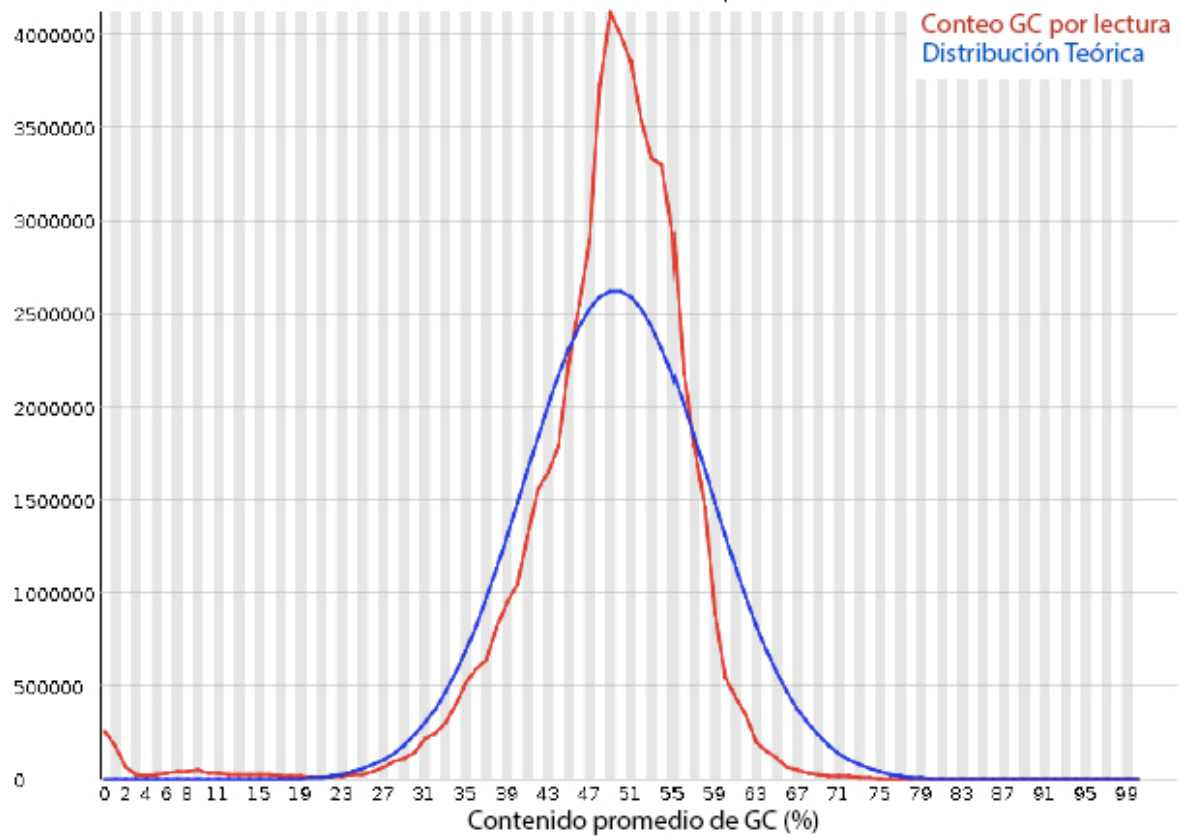


Figura 50. Distribución GC de la secuenciación con el cebador R2 para la especie *C. californicus*.

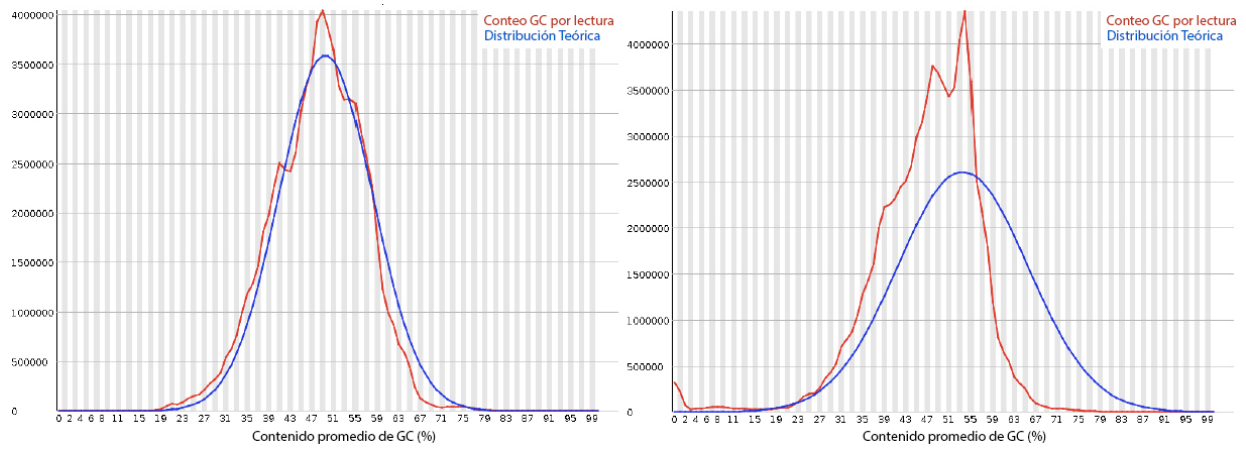


Figura 51. Distribución GC de la secuenciación con los cebadores R1 y R2 para la especie *C. ximenes*.

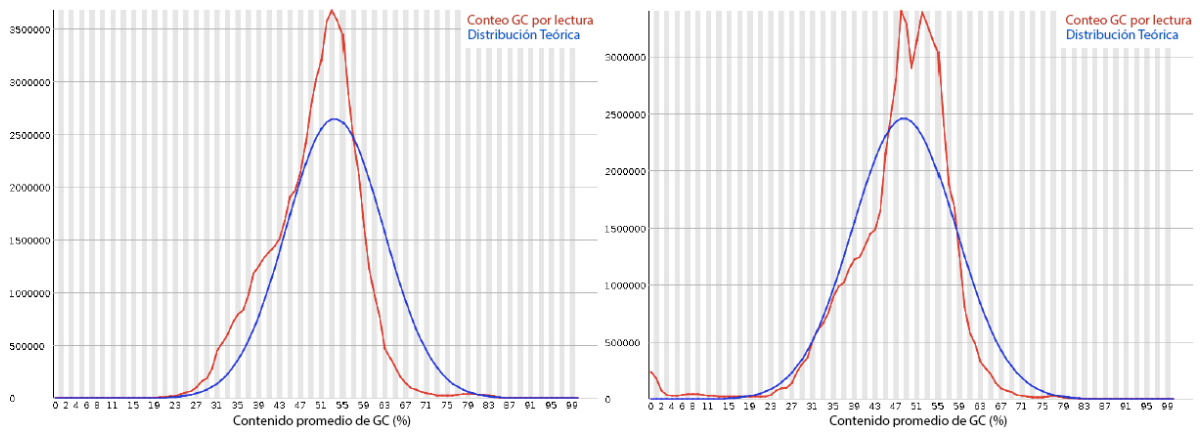


Figura 52. Distribución GC de la secuenciación con los cebadores R1 y R2 para la especie *C. mahogani*.

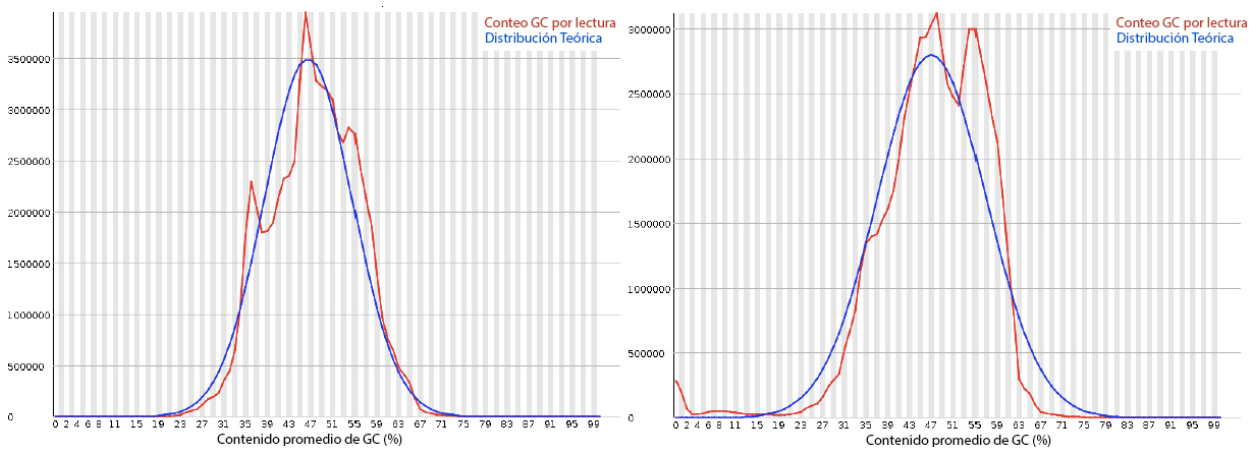


Figura 53. Distribución GC de la secuenciación con los cebadores R1 y R2 para la especie *C. regularis*.