VNIVERSITAT DE VALÈNCIA

Doctorado en Bioquímica y Biomedicina

Departament de Bioquímica i Biologia Molecular

Molecular and functional approaches to understand the natural history of snake short disintegrins

Raquel Sanz Soler

Tesis Doctoral 2016

Directores: Dr. Juan José Calvete Chornet

Dra. Libia Sanz Sanz

VNIVERSITAT DE VALÈNCIA

INSTITUTO DE BIOMEDICINA DE VALENCIA CSIC

Doctorado en Bioquímica y Biomedicina

Departament de Bioquímica i Biologia Molecular

Molecular and functional approaches to understand the natural history of snake short disintegrins

Raquel Sanz Soler

Tesis Doctoral 2016

Memoria presentada por Raquel Sanz Soler para optar al grado de Doctora Internacional por la Universidad de Valencia

Directores: Dr. Juan José Calvete Chornet

Dra. Libia Sanz Sanz

MINISTERIO
DE ECONOMÍA
Y COMPETITIVIDAD

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

JUAN JOSÉ CALVETE CHORNET, Doctor en Ciencias Biológicas y Profesor de Investigación en el Instituto de Biomedicina de Valencia del Consejo Superior de Investigaciones Científicas

LIBIA SANZ SANZ, Doctora en Ciencias Químicas y Técnico Superior Especializado en el Instituto de Biomedicina de Valencia del Consejo Superior de Investigaciones Científicas

INFORMAN: que Raquel Sanz Soler, Licenciada en Bioquímica por la Universitat de València, ha realizado bajo su dirección el trabajo que con el título "Molecular and functional approaches to understand the natural history of the snake short disintegrins" presenta para optar al grado de Doctora por la Universitat de València.

Valencia, Abril de 2016

Dr. Juan José Calvete Chornet

Dra. Libia Sanz Sanz

Este trabajo se ha realizado con el soporte económico de los proyectos de investigación que se enumeran a continuación:

BFU2010-17373: Venómica Traslacional. Ministerio de Ciencia e Innovación (IP. Juan José Calvete Chornet)

BFU2013-42833-P: Venómica de última generación. Ministerio de Economía y Competitividad. (IP. Juan José Calvete Chornet)

Raquel Sanz Soler ha disfrutado de una beca predoctoral FPI otorgada por el Ministerio de Ciencia e Innovación asociada al proyecto BFU2010-17373 y dos becas asociadas al mismo proyecto, concedidas para la realización de estancias breves con referencias EEBB-I-12-05509 y EEBB-I-2013-07728.

A mi familia, amigos y especialmente a Sise

Agradecimientos / Acknowledgements

Después de este largo camino y todo el apoyo recibido en él, va a ser difícil dar las gracias a todo el mundo como se merece, porque he tenido la suerte de rodearme de mucha y buena gente.

En primer lugar, gracias a mis directores de tesis. A Juanjo por abrirme las puertas de su laboratorio, por la libertad a elegir, y por curtirme como científica, además, gracias por su apoyo para participar en congresos y realización de estancias en el extranjero. Gracias a Libia por su apoyo, paciencia, consejos científicos y por su saber escuchar en cualquier momento aunque le borrase el nombre al cabo del día.

Gracias a todos los "calvetes" con los que he compartido laboratorio estos años.

A Jordi, ese hermano que me ha dado la carrera científica, porque sin él esta tesis no hubiera sido posible de principio a fin. A Davi, por estar siempre ahí, apoyándome científica y personalmente. Sin olvidar a las *disintegrin girls*: a Betty, porque aprendí mucho enseñándote, a Carol por esos mutantes *alemanokens* y por mucho más, a Gemi, por dejarlo todo para ayudarme en el laboratorio, por las discusiones científicas a las tantas, de camino a casa o en la biblioteca, y porque aunque las disintegrinas nos hayan dado muchos dolores de cabeza, hemos podido con ellas ☺.

A Alicia por sus consejos y ayuda en el laboratorio. A Yania, por echar una mano, con alegría, cuando veía que yo no llegaba a todo. A todos los estudiantes en prácticas, que te hacen descubrir que enseñando se aprende mucho. A todos los estudiantes y post-docs que han pasado por nuestro laboratorio desde lugares muy diversos del mundo (brasileñas, costarricenses, alemanes, argentina, venezolana, etc.), cada uno de ellos ha aportado una pizquita a esta tesis, a nivel profesional y/o personal.

No se como agradecer a todas las personas del IBV que en conversaciones de pasillo, cuartitos comunes, o en sus despachos, me han dado ideas o consejos para solucionar algún que otro problema de laboratorio. Gracias a todas las personas del laboratorio del Dr. Pascual Sanz, por dejarme un trocito de bancada cuando la necesité, además de vuestra ayuda, consejo y aceptación. Tampoco quiero dejar de agradecer la

ayuda y disponibilidad a las personas de administración, limpieza, mantenimiento, informática, seguridad, y biblioteca.

Gracias a la Cdf, por todos esos momentos de risas, discusiones científicas, congresos, y por sus eventos que han hecho más llevaderas las rachas de malos resultados. Además, gracias por aceptarme siempre después de volver de mis múltiples estancias, y a pesar de ello nombrarme becaria de honor.

Gracias a todos los científicos y veterinarios que han aportado muestras para llevar a cabo gran parte de la investigación de esta tesis, ya que sin ellas no hubiera sido posible.

Thanks to Professor Eble for receiving me in his laboratory and inviting me a second time. Also thanks for being so patient, teaching me so many aspects of science new to me. Thanks to everybody in the lab who taught me new technics and another point of view of science and life (Stephan, Alletta, Augusto, Dessi, Bartek, María, Flavia, etc.).

Thanks to Professor Mackessy for accepting me in his lab, teaching me how to handle a rattlesnake in my first day in Greeley, allowing me to use his lab and thanks for broaden my view of the snakes and venom world, in the lab and in the wild. Thanks to Cassie for sharing with me her new protocol and molecular biology discussions and thanks a lot to Anthony for going ahead with the behavior experiments and all his help in the last months. Also thanks a lot to everybody who made my life easier and happier in Colorado (Kaye, Tom, Kyle, Cara, Ana, Brenda…).

Gracias a Lola por aguantar siempre mis quejas y por las discusiones científicas en el sofá a la hora de cenar.

Gracias también a mis compañeros y amigos del departamento de B&BM, porque allí fue donde comencé a hacer ciencia, y a pesar de que aprendí que no todo eran resultados positivos, con personas como vosotros todo se puede sobrellevar. Gracias a tod@s por seguir apoyándome durante estos años de tesis. Además, gracias Inma por enseñarme tanto y tan bien, y a Carlos por su apoyo incondicional y estar siempre ahí.

Gracias a mis *xiquetes*, porque nuestros caminos casi paralelos han hecho que me hayáis entendido siempre. Gracias por escucharme, animarme siempre, por sacarme del laboratorio alguna tarde-noche, y por vuestro apoyo a diario en los últimos meses.

Gracias a mis cellanas, por hacer el esfuerzo de entender qué es lo que hago y estar más orgullosas de mí que yo misma. Gracias por vuestra paciencia, apoyo y cariño.

Gracias a *mis osos panda*, por su apoyo en todas mis decisiones a lo largo de la tesis. Por seguir ahí a pesar de todos los plantones que os he dado cuando el laboratorio/tesis me requería. Por vuestros ánimos desde cualquier sitio del mundo. Y especialmente a Marta, gracias por creer en mí y darme el primer empujón para hacer ciencia, y a Juanra por ayudarme a hacer perfecta una de mis figuras favoritas de la tesis.

Gracias a Javi, por estar ahí, por las comidas al sol de vez en cuando hablando de ciencia, experimentos, estancias, tesis, más tesis y de la vida.

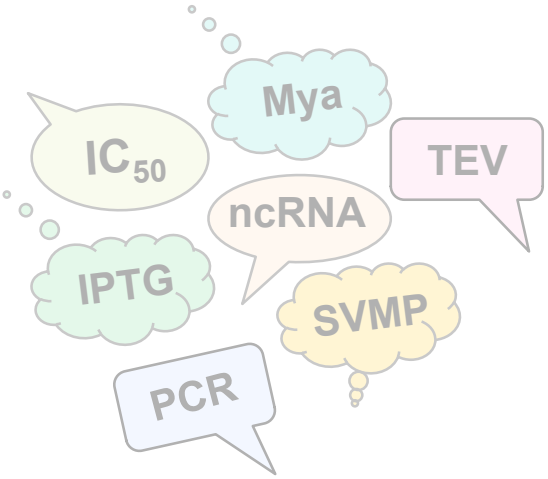
Thanks Ingo for being so patient, for your support, for believing in me even when it is pretty difficult for you to understand how the science/PhD world works. Dankeschön!

Y por supuesto, a mi familia, porque sin ellos nada de esto hubiera sido posible, porque gracias a ellos soy quien soy, y he llegado hasta aquí. Gracias por su apoyo incondicional en todos los sentidos, porque aunque no entiendan muy bien lo que hago, intentan entenderlo. Especialmente, gracias papás, tete, yaya y cuñada, gracias por estar ahí siempre. Y gracias a mi tete por su ayuda con el diseño de la portada.

¡GRACIAS! **GRÀCIES!** **THANK YOU!** **DANKE!**

Index

Abbreviations

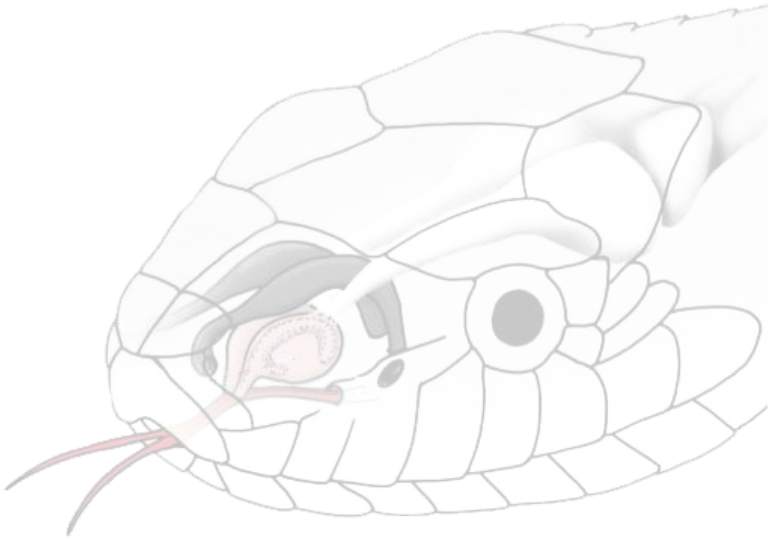


Abbreviations

aa: amino acid/s
ACD: Acid Citrate Dextrose solution
Amp: Ampicillin
AP: Adapter Primer
AUAP: Abridged Universal Amplification Primer
BCA: BiCinchoninic Acid (quantification protein method)
BLASTn: Basic Local Alignment Search Tool nucleotide
bp: base pair
BSA: Bovine Serum Albumin
cDNA: complementary DNA
Col I and Col IV: Collagen I and IV
Ct: Carboxyl-terminus
del/Δ: deletion
df: degrees of freedom (n-1)
DTT: Dithiothreitol
E: Envenomated
EDTA: Ethylenediaminetetraacetic acid
ESI-MS: ElectroSpray Ionization-Mass spectrometry
FPLC: Fast Protein Liquid Chromatography
Fw: Forward
g: gravitational force
gDNA: genomic DNA
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HisTRAP: Nickel Sepharose affinity column
HPLC: High Performance Liquid Chromatography
IC_{50}: half maximal inhibitory concentration
IPTG: isopropyl-D-thiogalactosidase
KTS: Lysine-Threonine-Serine
LB: Lauria-Broth
lncRNA: long non-coding RNA
LTA: Light Transmission Aggregometry
NCBI: National Center of Biotechnology Information
ncRNA: non-coding RNA
NE: Non-envenomated
nESI-MS/MS: nanoElectroSpray Ionization-tandem mass spectrometry
NMR: Nuclear Magnetic Resonace
nt: nucleotide
Nt: amino-terminus

NTC: Non-Template Control
MA: Maximum Aggregation
MALDI-TOF: Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight
mRNA: messenger RNA
Mya: Million Year Ago
oc: ocellatusin
OD_{600}: Optical Density (600nm)
PBS: Phosphate-Buffered Saline
PCR: Polymerase Chain Reaction
PEG: Prostaglandin E
PRP: Platelet-rich-plasma
QTrap: quadrupole linear-ion trap
r- : recombinant
RACE: Rapid Amplification of cDNA Ends
RGD: Arginine-Glycine-aspartic acid
RTS: Arginine-Threonine-Serine
Rv: Reverse
rpm: revolutions per minute
rRNA: ribosomal RNA
RACE: Rapid Amplification of cDNA Ends
RT- : Reverse transcriptase minus
RT-PCR: Real Time-PCR
s.e.m : Standard Error of the Mean
SDS: Sodium dodecyl sulfate
SDS-PAGE: SDS-Polyacrylamide gel electrophoresis
SICS: Strike-Induced Chemosensory Searching
sp.: specie
SVMPs: Snake Venom MetalloProteinases
TBS: Tris-Buffered Saline
TEV: Tobacco Etch Virus
TFA: Trifluoroacetic acid
tRNA: transfer RNA
trxA : thioredoxin A
v/v: Volume/Volume
w/v: Weight/Volume
WHO: World Health Organization
w/w: Weight/Weight
X (Xaa): Any amino acid

Introduction



1 Reptile cladogenesis

Amniota is a remarkably diverse clade of tetrapod vertebrates consisting of mammals, non-avian reptiles, and birds. The most recent common amniote ancestor likely lived approximately 325 million years ago (Mya) (Burghardt 1970; Shedlock and Edwards 2009) (Figure 1).

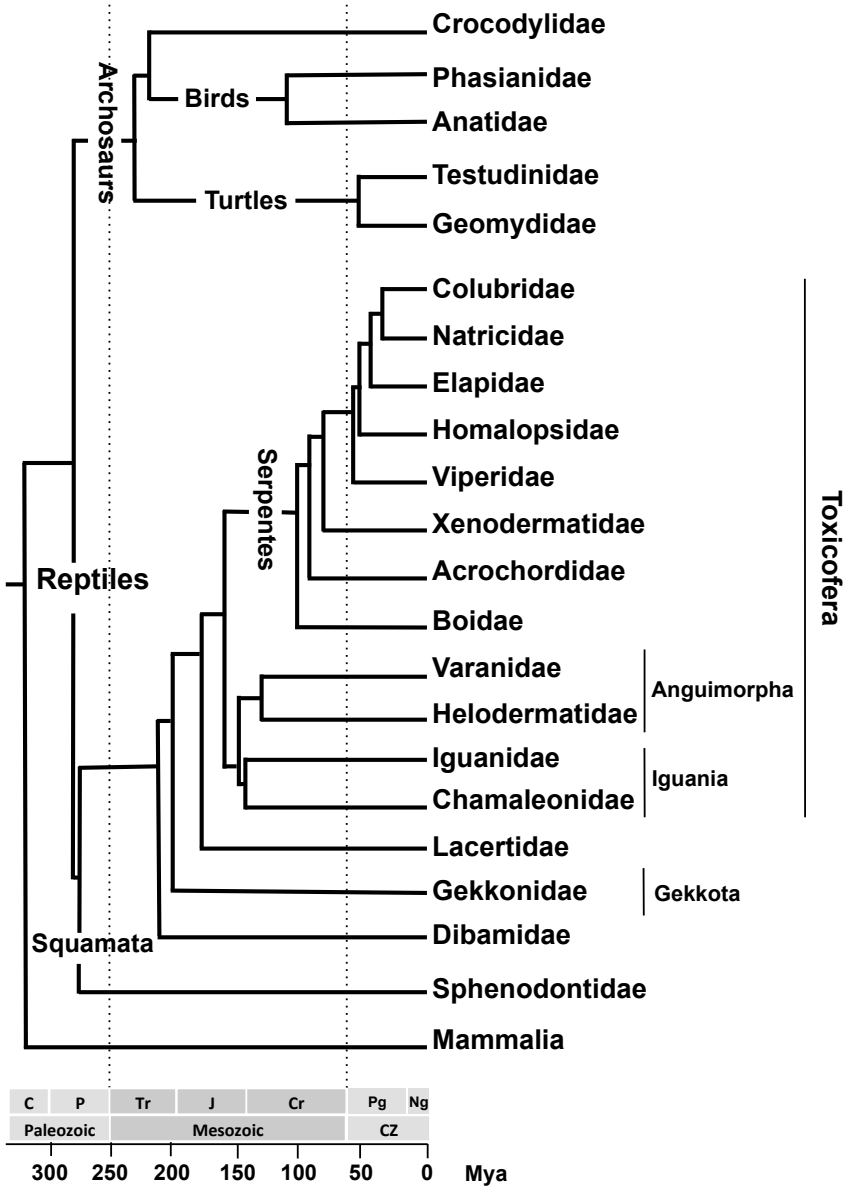


Figure 1. Simplified Reptilia cladogenesis. Adapted from (Hedges and Vidal 2009; Pereiraa and Bakera 2009; Shaffer 2009; Shedlock and Edwards 2009; Vidal et al. 2009). Abbreviations: C (Carboniferous), P (Permian), Tr (Triassic), J (Jurassic), Cr (Cretaceous), CZ (Cenozoic), Pg (Paleogene), and Ng (Neogene). Mya (Million years ago).

Since some reptiles (crocodiles) are more closely related to birds than they are to other reptiles (lizards), many modern day scientists prefer to make Reptilia a monophyletic group, which also includes birds. Molecular data and fossil records suggest that the major lineages of extant reptiles arose in the Permian and Triassic periods (299–200 Ma), when land areas were coalesced into the single supercontinent known as Pangaea. Squamates diverged from other reptiles approximately (~) 275 Mya (Shedlock and Edwards 2009) (Figure 1).

Nevertheless, there is an extensive debate focused on reptilian evolution. For example, disagreeing morphological and molecular results on the origin of turtles has been challenging to resolve (Hedges 2012). Historically, turtles are placed as the closest relatives of all other living reptiles, with embryological data supporting this view too (Benton 1990; Werneburg and Sánchez-Villagra 2009) (Figure 2B). Nonetheless, most recent morphological analyses (Schoch and Sues 2015) and additional molecular genetic and genomic data (Hedges 1999; Cao et al. 2000; Rieppel 2001; Zardoya and Meyer 2001; Rest et al. 2003; Iwabe et al. 2005; Shedlock et al. 2007; Shedlock and Edwards 2009; Janes et al. 2010; Carroll 2013), place turtles with birds (Aves) and crocodilians (Figure 2A). Presumably, turtles separated from the crocodilians and birds clade approximately 231 Mya, and crocodilians separated from birds ~219 Mya (Shedlock and Edwards 2009) (Figure 1).

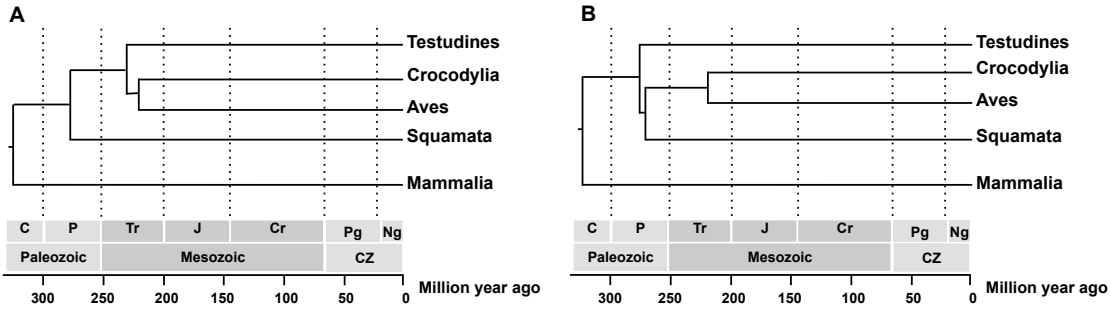


Figure 2. Uncertain position of turtles in the amniotes tree of life. A. Turtles as sister group of Crocodylia and Aves (Shedlock and Edwards 2009). B. Turtles in the base of the Reptilia tree of life. Abbreviations: C (Carboniferous), P (Permian), Tr (Triassic), J (Jurassic), Cr (Cretaceous), CZ (Cenozoic), Pg (Paleogene), and Ng (Neogene). Adapted from Shedlock and Edwards 2009.

Reptiles are one of the most ecologically and evolutionary remarkable groups of living organisms, having successfully colonized most of the planet. The evolutionary history of reptiles has given considerable rise to species richness among phylogenetic groups. Most reptile diversity is concentrated in the Squamata clade, where lizards and snakes represent a 96,4% of the diversity. In contrast, turtles (3.4%) and crocodilians (0.2%) are far less diverse (Pincheira-Donoso et al. 2013; Uetz and Hošek 2015). Most of the Squamata groups diversified in the Jurassic and Cretaceous roughly 200-66 Mya, possibly in response to the breakup of supercontinents. Gekkota clade separated from the rest of reptiles ~198 Mya, and Lacertata is said to have split from "Toxicofera" (Serpentes, Anguimorpha, and Iguania) reptiles ~180 Mya (Hedges and Vidal 2009). Currently, integrated molecular, morphological, and fossil record analyses, propose that Iguania is placed with snakes and anguimorphs in the tree of life (Fry 2005; Hedges and Vidal 2009; Reeder et al. 2015). Nonetheless, their relative position is still unresolved (Figure 3). Higher-level Squamate phylogeny is considered controversial due to conflicts between hypotheses based on analyses of morphological (Gauthier et al. 2012) and molecular datasets (Townsend et al. 2011; Pyron et al. 2013). However, integrated phylogenetic studies combining fossil records, morphological and molecular analysis have attempted to resolve these incongruences. Recently, Jones et al. 2013 supports the phylogeny suggested by the morphological studies, which considers Iguania as a sister group to Serpentes+Anguimorpha (Figure 3A). If this is confirmed, it would mean a split ~166 Mya (Jurassic) (Hedges and Vidal 2009). Nevertheless, Reeder et al. 2015 results of the combined-data analyses, suggest that Serpentes are a sister group to Iguania+Anguimorpha, supporting the phylogeny recovered in the molecular analysis (Pyron et al. 2013) (Figure 3B).

Technically, Reptilia denotes birds and nonavian reptiles, as mentioned. However, throughout this dissertation, Reptilia will be used to refer to non-avian reptiles.

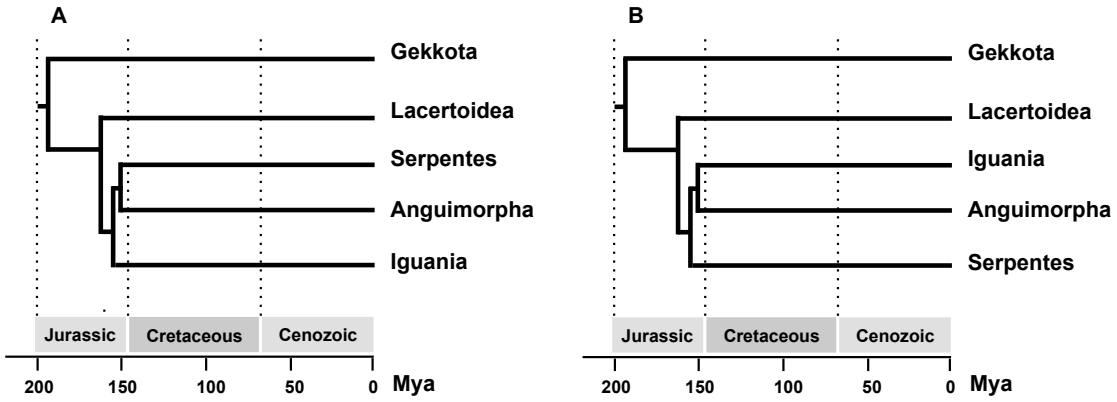


Figure 3. Unresolved Toxicoferan time tree of evolution. Adapted and simplified phylogenies from integrated studies by A. (Jones et al. 2013) Serpentes+Anguimopha are a sister group to Iguania and B. (Reeder et al. 2015) Serpentes are in the base of the "Toxicofera" tree.

2 Snakes (Serpentes)

Snakes belong to the suborder Serpentes, which represent a large radiation of terrestrial vertebrates comprising of approximately 3567 extant species distributed across every continent except Antarctica (*reptile-database.org*) (Uetz and Hošek 2015). Although snakes are considered as non-model organisms in biological research, the tremendous phenotypic and molecular plasticity observed in these animals have made them attractive subjects in an array of biological fields ranging from evolution, to ecology and behavior, physiology, and medicine.

Recent phylogenetic analyses of Serpentes have been generated by integration of morphological and molecular characteristics (Kelly et al. 2003; Lawson et al. 2005; Vidal et al. 2007; Castoe et al. 2009; Pyron et al. 2011; 2013; Reeder et al. 2015) (Figure 4A). However these phylogenies are still under significant debate. Boas, pythons and Caenophidians (Colubroidea superfamily) make up most of the extant species of snakes, with the Boidae and Pythoniedae families splitting from Caenophidia ~104 Mya. Snakes belonging to the Boidae and Pythoniedae families are non-venomous and subdue prey using a combination of constriction and/or jaw-holding behaviors. The vast majority of extant snakes (>2/3 *sp.*), including venomous snakes, belong to the Colubroidea super family (Vidal et al. 2009; Pyron et al. 2011; Pyron and Burbrink 2012) and are primarily distributed into the three major families of Viperidae, Elapidae, and Colubridae. It has

been suggested that Viperidae split from the Elapidae and Colubridae families ~54 Mya, and Colubridae separated from Elapidae ~46 Mya (Figure 4A). Although the Viperidae family only comprises roughly 9% of the total diversity of colubroid snakes, their distribution is widespread across all continents except Australia and Antarctica, as well as displaying tremendous morphological diversity, occupying a wide variety of niches (Greene 1992; Greene et al. 1997). Viperidae is further divided into two major subfamilies, the Viperinae (Old World or pitless vipers) and the Crotalinae (pitvipers and rattlesnakes). Only Crotalinae species have been found on the American continent, and are therefore referred to as New World pitvipers. (Figure 4B).

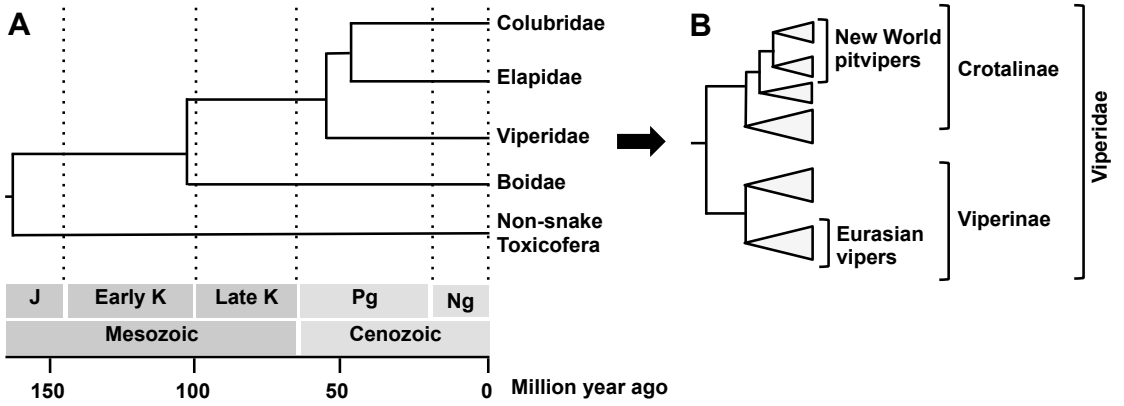


Figure 4. Snake families distribution along evolution. A. Timetree of snakes, adapted from Vidal et al. 2009. B. Diagram of the distribution of Viperidae taxa adapted and simplified from Pyron et al. 2013. The distances in B do not correspond to evolutionary time. Abbreviations: J (Jurassic), K (Cretaceous), Pg (Paleogene) and Ng (Neogene).

2.1 Vipers foraging strategy and behavior

Vipers employ a sit-and-wait foraging strategy, where these venomous predators strike, envenomate and release prey. Subsequently, chemoreception is used to trail and relocate the envenomated prey that may wander several meters or more from the attack site (Chiszar et al. 1992). This chemical recognition is characteristically mediated by rapid tongue flicking, activated by the detection of volatile chemical cues by the nasal olfactory system (Burghardt 1970; Saviola et al. 2010) or by visual or thermal stimulation (Chiszar et al. 1981; Saviola et al. 2011; 2012). Consequently, tongue flicking delivers volatile and non-volatile stimuli to the vomeronasal organ located in the roof of the upper jaw (Halpern 1992; Schwenk 1995). The vomeronasal organ of squamates is

associated with a sophisticated tongue delivery system. When the tongue protrudes from the mouth of the snake it collects volatile and non-volatiles molecules, which adhere to the moist surface of the tongue. When the tongue is retracted, these molecules are carried into the mouth and pass by a disputed mechanism into the openings of the vomeronasal ducts in the roof of the mouth, traveling through the ducts to the chemosensory epithelium of the vomeronasal organ (Gillingham and Clark 1981; Halpern 1992; Young 1993). This ultimately allows for definitive analysis of chemical information (Cowles and Phelan 1958; Halpern 1992) (Figure 5). Tongue-flicking is a unique synapomorphy among squamate reptiles that has been used in numerous studies of squamate behavior as an index of chemosensory response (Cooper 1994; Schwenk 1995; Desfilis et al. 2002).

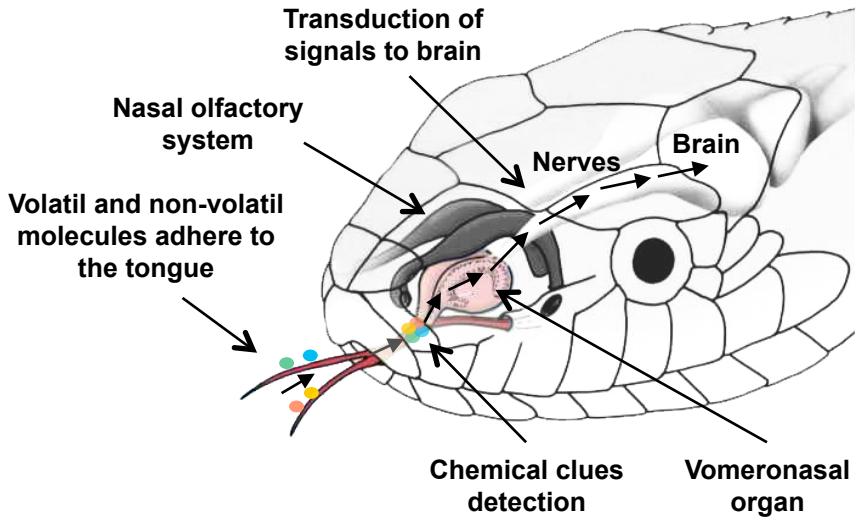


Figure 5. Vomeronasal chemoreception mediated by tongue flicking and discrimination by the vomeronasal organ. Adapted from the Dan Erickson Illustration in (Grzimek 2003).

Following a viperid predatory strike, prey is released and escapes wandering from the attack site. After a few minutes, the snake searches for its prey while rapidly tongue flicking, which allows the animal to pick up chemicals in the air and sniff out the trail left behind by the prey. Several studies of chemoreception by tongue flicking have demonstrated that venomous snakes prefer envenomed to non-envenomed prey (Chiszar et al. 1999; Greenbaum et al. 2003; Clark 2004; Chiszar et al. 2008; Saviola et al. 2010; 2011; 2012), and that prey preference is common in viperids. These studies

suggest that after striking the prey, vipers use chemical cues from envenomated tissue to a larger extent than chemical cues produced by the prey (e.g., prey scent, urine, or alarm pheromones) to locate their prey (Chiszar et al. 1992).

Recent studies from the laboratory of Professor Stephen Mackessy (University of Northern Colorado) tried to identify the compounds in the venom that may assist the snake to relocate the prey. *Crotalus atrox* venom protein fractions were used for independent chemoreception studies by tongue flicking. *Crotalus atrox* responded significantly only to the venom fraction, containing disintegrins (Saviola et al. 2013).

3 Valuable applicability of the snake venom study: from a neglected disease to drug discovery

Venomous Squamates are a particularly significant group for humans since they cause tens of thousands of deaths every year (Gutiérrez et al. 2006; 2010; 2015; Habib et al. 2015) and yet their venom toxins are a unique resource for diversity of medicines and even applied physics (Geim et al. 2003).

3.1 Snakebites

In 2015, West Africa was in the media spotlight, and rightly so after nearly 11.000 people died in the largest Ebola outbreak ever recorded. However, another deadly killer often fails to exist in the public conscience: snakebites (Shultz 2015). Snakebite is a major, yet seriously neglected public health issue, not only in Africa, but also in Asia, Latin America and parts of Oceania (Gutiérrez et al. 2006). Traditional data suggest that between 1.2 and 5.5 million people suffer from snakebites every year, resulting in 25,000 to 125,000 deaths and leaving approximately 400,000 victims with permanent sequela (Kasturiratne et al. 2008; Williams et al. 2010) (http://www.who.int/mediacentre/news/notes/2010/antivenoms_20100504/en/) (Figure 6).

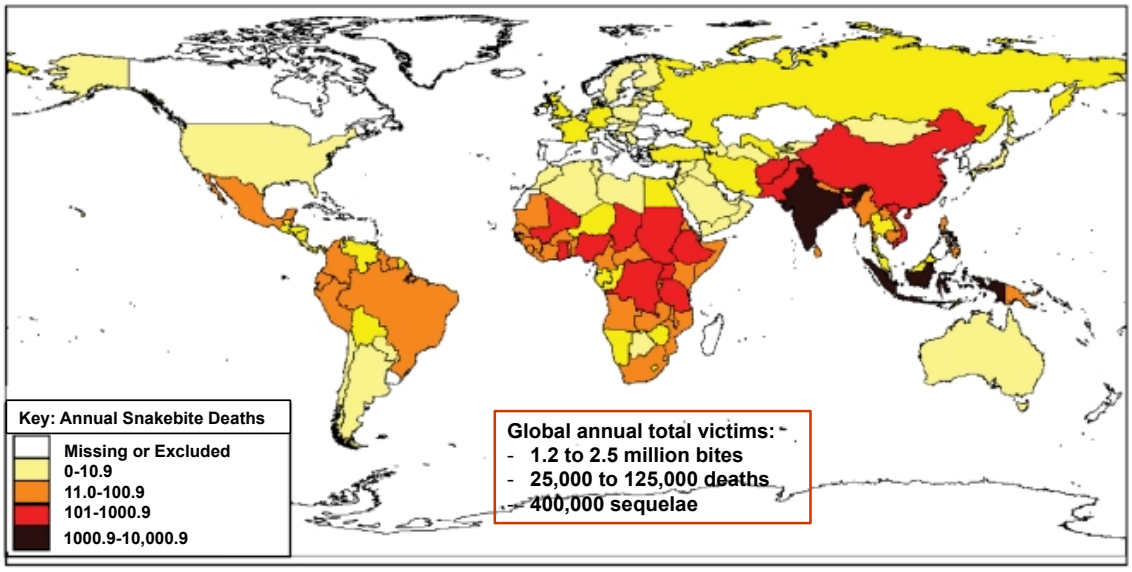


Figure 6. Annual snakebite mortality. Annual estimates of snakebite-induced deaths for 138 countries were obtained from the data published by Kasturiratne et al. 2008 and depicted on a world map using Epi-info; the darker a country's colour the greater the estimated snakebite mortality – see key for details. Adapted from Harrison et al. 2009.

These victims are often farmers living in poor rural communities, where the consequence of envenomation, such as limb loss, may significantly impact their quality of life, and can be devastating for their the entire family. Yet, despite its significant impact on human health, snakebite remains largely neglected (Gutiérrez et al. 2006; Williams et al. 2010; Gutiérrez et al. 2015; Habib et al. 2015), and in fact, it is currently not considered as problematic, or as a neglected tropical disease (NTD) by the World Health Organization (WHO) (http://www.who.int/neglected_diseases/diseases/en/). Nevertheless, the WHO highlights the critical need for life-saving antivenoms, and are included on as an essential medicine on the 19th WHO Model List of Essential Medicines (April2015) (http://www.who.int/medicines/publications/essentialmedicines/EML2015_8-May-15.pdf?ua=1). However, there is still significant work necessary to improve the prevention, first aid and treatment of snakebites, and generating knowledge on snake venom composition is essential to produce effective antivenoms (Williams et al. 2010; 2011; Gutiérrez et al. 2013; 2014; 2015) (Figure 7).

3.2 Applications in biomedicine. Medicines from nature's superstore

Historically, snakes venoms have been used in medicinal practices dating back to ancient Egypt and Greece, as well as in traditional Chinese and Indian therapies, where Cobra venoms were used to treat arthritis for thousands of years. Similarly, Hippocrates used bee venom to treat joint pain and arthritis (Seifert 1954; Scarborough J. 2010). Venom derived toxins have had several applications in drug discovery where they have been used as therapeutic agents (botulinum toxin), or molecular scaffolds in early stages of drug design (captopril®), or used as reagents for the identification of therapeutic targets (Harvey et al. 1998; Harvey 2001; 2002; Kapoor 2010; Harvey 2014). The first clinically successful Angiotensin Converting Enzyme inhibitor, captopril®, used to treat hypertension, was developed from peptides isolated from the venom of *Bothrops jararaca* (Ferreira 1965; Cushman and Ondetti 1999), reviewed by Harvey 2014. Other drugs derived from animal venoms that are currently clinically approved include Aggrastat® (Tirofiban) and Integrilin® (Eptifibatidae) (Granada and Kleiman 2004), which are both used to treat acute coronary syndrome, by inhibiting platelet aggregation. Additionally, Prialt® (Zicotonida) is a painkiller that blocks calcium ion channels, and Byetta® (exenatide), isolated from the venom of the Gila Monster, is used as treatment of type-2 diabetes (Scheen 2014). Finally, Batroxobin® (Baquting) is a treatment used for perioperative bleeding (King 2011).

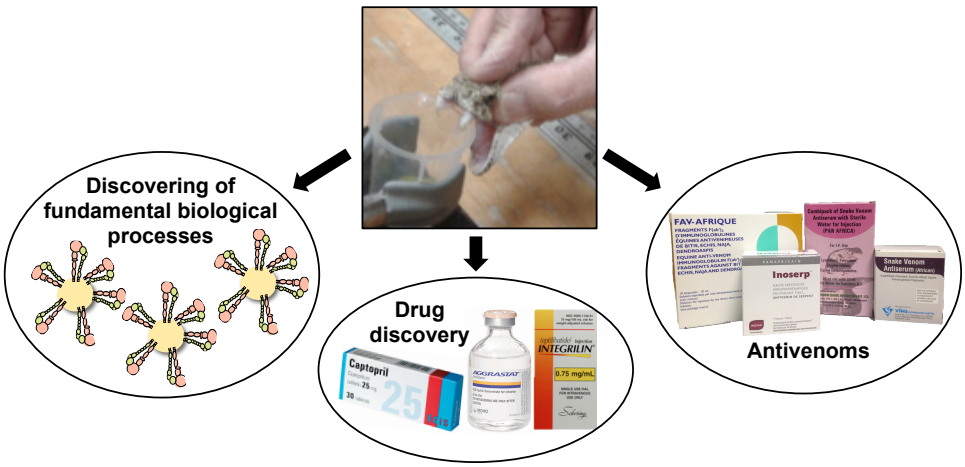


Figure 7. Biomedical applications from the study of snake venom. Few examples of antivenoms and commercial drugs derived from snake venom proteins. Picture of venom extraction by Professor Mackessy at the University of Northern Colorado.

In addition, several additional venom peptides are in clinical trials for the treatment and/or diagnosis of chronic pain, autoimmunity, cardiovascular syndromes, cancer (glioma) or infection (Marsh 2001; Lewis and Garcia 2003; King 2011; Cossins 2013). For example, pre-clinical data on the peptides known as mambalgins, isolated from the venom of the Black Mamba, suggest an analgesic effect as strong as morphine (Diochot et al. 2012), and represents a novel pain treatment if it passes clinical analyses.

In addition to the development of several major human drugs derived from snake venom proteins, an array of critical biological processes have also been revealed using toxins in cell-receptor studies in mammals (Figure 7).

For all these reasons, snake venom research is crucial in many aspects of biological and biomedical related research. The knowledge generated based on venom composition and individual toxin studies, may provide a foundation for new drug development and also help generate improved antivenoms. Furthermore, new biological and molecular mechanism might be discovered, which might be conserved in mammals (Figure 7).

4 Snake venom

Snake venoms are mixtures of biologically active proteins (90% dry weight), salts, and organic molecules such as polyamines, amino acids, and neurotransmitters (Fry 2005; Fox and Serrano 2008a; Casewell et al. 2013). This complex cocktail of bioactive compounds ("*toxins*") is produced in a postorbital venom gland (Vonk et al. 2008) that is present in elapid and viperid snakes. These snakes also contain a small downstream accessory gland, however the specific function of this gland is currently unknown (Vonk et al. 2013). The most recent definition of venom is "a secretion, produced in a specialized tissue (generally encapsulated in a gland) in one animal and delivered to a target animal through the infliction of a wound. A venom must further contain molecules that disrupt normal physiological or biochemical processes to facilitate feeding and/or defense for the animal" (Fry et al. 2009a; 2009b; 2012). However, this definition of "venom" is under much debate, and specifically what constitutes some reptiles as venomous or non-venomous species. Snake venom proteins play a number of

adaptive roles: immobilizing, paralyzing, killing, and digesting prey, in addition to deterring competitors. However, venoms may serve other possible functions, besides feeding or defense, are suggested to include in the venom definition (Fox and Serrano 2005; Kini 2006; Fry et al. 2012; Jackson et al. 2012; Kardong 2012; Weinstein et al. 2012). Although much debate surrounds the evolution of venomous systems, recent evidence suggest that some toxin types appear to be co-expressed in the venom gland and other tissues (Casewell et al. 2012; Makran et al. 2012; Junqueira-de-Azevedo et al. 2014; Reyes-Velasco et al. 2014), complicating the definition and evolution of toxins and non-toxins.

4.1 Snake venom composition

Venom composition significantly varies, not only between family, genus, and species, but also between individuals. Elapid venoms are generally more rapidly acting than other Caenophidian venoms, due to the high content of neurotoxic peptides. On the other hand, colubrid venoms are generally considered to be much simpler in venom composition than Elapidae and Viperidae venoms.

Although Viperidae venoms may contain an abundance of proteins and protein isoforms (Vonk et al. 2011; Fahmi et al. 2012; Casewell et al. 2013), the majority of the venom compounds can be grouped into a several major families such as the enzymatic; (serine proteinases, Zn^{2+}-metalloproteases, L-amino acid oxidase, PLA_2 (fosfolipases A_2), nucleotidases...) and non-enzyimatic activity (C-type lectins, natriuretic peptides, ohanin, myotoxins, CRISP (cysteine-rich secretory protein) toxins, nerve and vascular endothelium growth factors, cystatin, Kunitz-type protease inhibitors and disintegrins) (Calvete et al. 2007a; Gutiérrez et al. 2009; Fry et al. 2009a; Casewell et al. 2012; 2013; Pincheira-Donoso et al. 2013; Junqueira-de-Azevedo et al. 2014; Uetz and Hošek 2015). Every Viperidae species produces in their venoms a variable number of these components (Markland 1998; Menez 2002; Juárez et al. 2004; Lomonte et al. 2014), thought different species venoms exhibit a distinct toxin family distributional profile (Figure 8). The functional and biological activities observed in venoms depends on the concentration and interaction of these individual components (Aird 2002; Mackessy 2010; Aird et al. 2015), which exhibit a diversity of biological activities in the context of

predator-prey interactions. Specifically, some proteins in the viperid venoms interfere with the coagulation cascade, homeostatic system and tissue repair, leading to the clotting disorders, hypofibrinogenemia and local tissue necrosis commonly observed in humans following envenomation (Markland 1998; Fox and Serrano 2005; Kini 2006).

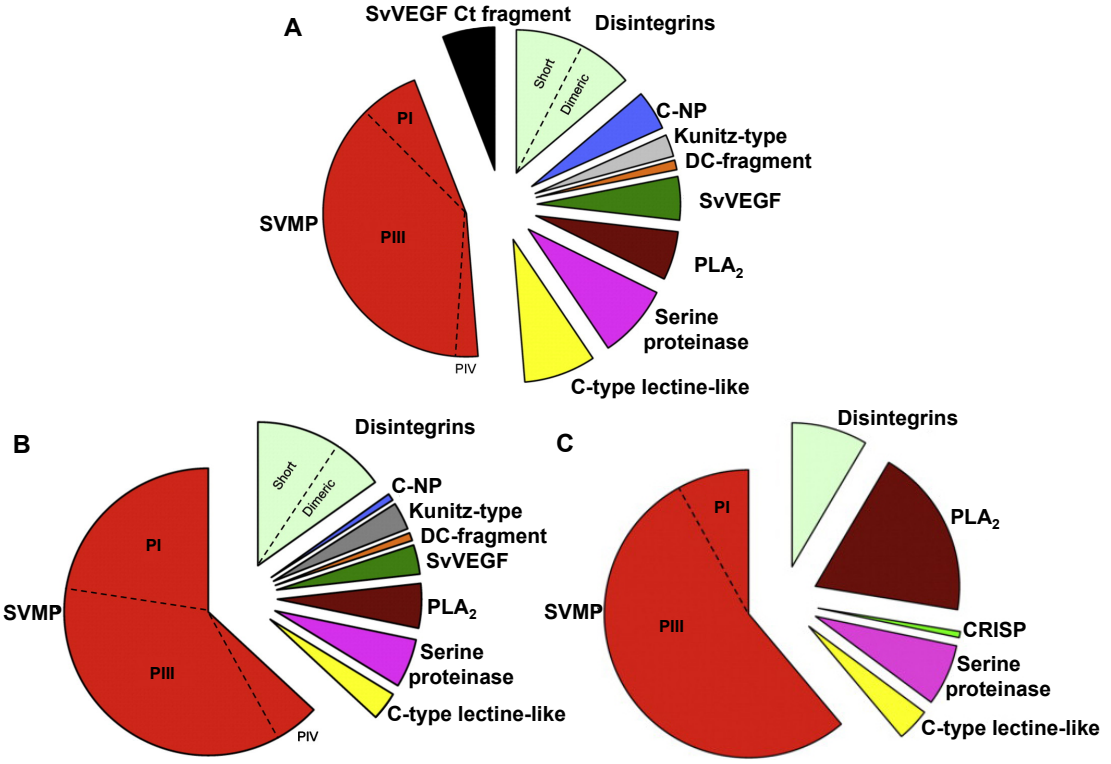


Figure 8. Modified pie charts comparing relative occurrence of proteins from different toxin families in (A) *Macrovipera mauritanica* and (B) *Macrovipera lebetina* by (Makran et al. 2012) and (C) *Cerastes cerastes* by (Fahmi et al. 2012). PI- and PIII-SVMP, snake venom Zn^{2+}-metalloproteinase (SVMPs) of classes I and III, respectively; PLA_2, phospholipase A_2; svVEGF, snake venom vascular endothelial growth factor; C-NP, C-type natriuretic peptide; DC-fragment, disintegrin-like/cysteine-rich domain. Each toxin family encloses different protein isoforms.

4.2 Snake venom evolution

It is hypothesized that venoms evolved from genes that code for ancestral endogenous proteins with normal physiological functions that have been recruited into the venom proteome (Fry et al. 2004; Fry et al. 2009a; Casewell et al. 2012; 2013; Junqueira-de-Azevedo et al. 2014). These genes were shown to be expressed in a variety of tissue types as diverse as the venom proteins themselves (Fry 2005; Fry et al. 2012). Toxin variability is driven by positive accelerated evolution (Moura-da-Silva et al. 1996;

Fry et al. 2003; Ogawa et al. 2005; Juárez et al. 2008; Casewell et al. 2011). When a particular gene has been recruited into the venom gland, additional gene duplication, combined with protein neo and/or subfunctionalization, are involved as the key regulatory processes driving bioactivity. This evolutionary path results in large multigene families that encode toxins exhibiting a variety of functional activities and potencies in addition to non-functional or pseudogenes (Kordis and Gubensek 2000; Fry et al. 2003; Hedges and Vidal 2009; Casewell et al. 2011). That form of multigene family evolution, in which new genes are created by gene duplication, where some are maintained in the genome, and others are deleted or inactivated is termed as the birth-and-death evolution model (Tan and Saifuddin 1990; Nei et al. 1997; Nei and Rooney 2005). This model is used to explain the evolutionary origin of many venom toxins (Casewell et al. 2011; 2012; Peichoto et al. 2012; Vonk et al. 2013).

Despite the increasing number of studies that have conducted transcriptomic and proteomic analyses on venoms (reviewed by Calvete 2013a and Ducancel et al. 2014), the evolution of venomous systems still remains highly controversial. Some researches believe that venom originated only once in the course of the reptilian evolution, giving rise to the Toxicofera hypothesis (Vidal and Hedges 2005; Fry et al. 2009a; 2009b; 2012; 2013), whereas others suggest that the venom has originated independently several times in the course of venom evolution. This later hypothesis suggest that venoms evolved once at the base of the advanced snakes, once in *Heloderma* linage, and an additional time in monitor lizards (*Varanus*) (Hargreaves et al. 2014a; 2014b).

5 ADAMs and Snake Venom Metalloproteinases

Snake Venom Metalloproteinases (SVMP) comprise a large multigene family that are one of the most studied venom protein families due to their significant contribution to the hemotoxicity commonly observed in venoms across most extant Viperidae taxa (Escalante et al. 2011; Markland and Swenson 2013; Uetz and Hošek 2015). SVMPs are zinc-dependent enzymes that belong to the M12 reprolysin family of metalloproteinases, and are classified based on domain structure into the PI, PII and PIII classes. PI-SVMPs contain only the metalloproteinase domain in the mature protein, whereas PII-SVMPs

exhibit an additional disintegrin domain carboxy to the spacer region, which, in many PII-SVMPs, is proteolytically cleaved from its precursor and is released in the venom (Moura-da-Silva et al. 1996; Calvete et al. 2005; Fox and Serrano 2005; Fry et al. 2008; Casewell et al. 2011; Casewell 2012; Casewell et al. 2014). PIII-SVMPs contain in addition to the metalloproteinase domain, a disintegrin-like (Dis-like) and a cysteine-rich (Cys-rich) domain. Some PIII-SVMPs are also post-translational processed, releasing the Dis-like and Cys-rich domain (DC domain) into the venom (Fox and Serrano 2008b; Takeda et al. 2012) (Figure 9).

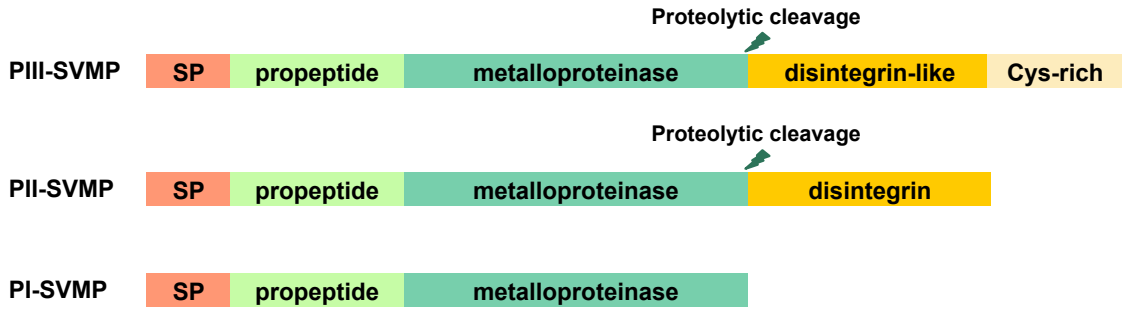


Figure 9. Schematic of snake venom metalloproteinases (SVMPs) present in Viperidae venoms. SP, signal peptide. PII-SVMP and PIII-SVMP cleavage site of the mature protein is indicated.

The ancestral PIII-SVMPs evolved from a gene encoding the extracellular region of a membrane-associated ADAM (a disintegrin and metalloproteinase) molecule, which was possibly recruited into the snake venom proteome after the divergence of squamate reptiles (Fry et al. 2006; 2012) in the Jurassic, approximately 170–150 Mya (Hedges and Vidal 2009). This hypothesis is supported by the presence of PIII-SVMPs in the venoms of Viperidae (Fox and Serrano 2005; Lu et al. 2005), Elapidae (Tan and Saifuddin 1990), Colubridae (Peichoto et al. 2012) and Atractaspididae (Ovadia 1987) snakes, whereas, PII-SVMPs occur only in Viperidae venoms (McLane et al. 1998; Calvete et al. 2003; Juárez et al. 2008; Calvete et al. 2009b). Therefore, PII-SVMPs may represent a derivation from ancestral PIII-SVMP genes, by a single loss of the cysteine-rich domain (Calvete et al. 2003; Juárez et al. 2006a; Casewell et al. 2011), subsequent to the emergence of Viperidae as a distinct taxonomical group of advanced snakes, approximately 37 Mya, in the Cenozoic era (Pyron and Burbrink 2012).

6 Disintegrins function and evolution

Disintegrins are a broad group of small (40–100 amino acids), cysteine-rich polypeptides (Calvete 2010; 2013b; Carbajo et al. 2015), synthesized from short-coding mRNAs (Okuda et al. 2002) or released into the venom of Viperinae (vipers) and Crotalinae (pitvipers) snakes by the proteolytic processing of PII-SVMP precursors (Kini and Evans 1992) (Figure 10).

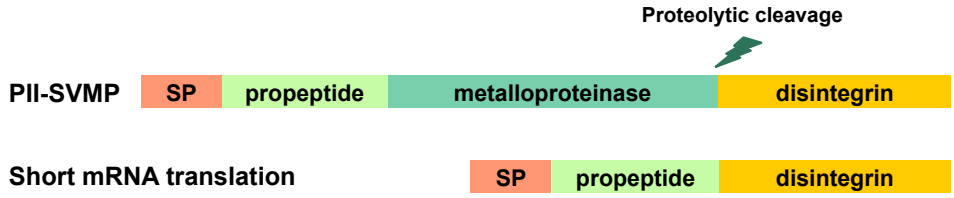


Figure 10. Cartoon of the disintegrin precursors released in Viperidae venoms. SP, signal peptide. The cleavage site is indicated in the snake venom metalloproteinase (PII-SVMP).

6.1 Canonical-disintegrin evolution

Canonical disintegrins have evolved by positive Darwinian evolution guided by the adaptation of a conformational epitope (the integrin recognition loop and the C-terminal tail) to the active site of the targeted integrin receptors (Juárez et al. 2008). The RGD integrin-binding tripeptide has been inferred to represent the ancestral integrin recognition motif, which emerged during the Paleogene period of the Cenozoic Era (approximately 54–64 Mya) from a subgroup of PIII-SVMP bearing the RDECD sequence (Juárez et al. 2008), which originated by recruitment, duplication, and neofunctionalization of a cellular ADAM-7 or 28 ancestor gene, (*Anolis carolinensis* lizard genome) (Moura-da-Silva et al. 1996; Fry 2005; Fry et al. 2006; Alföldi et al. 2011; Casewell 2012). After the deletion of the PIII-SVMP-lineage-specific cysteine residue (RDE**C**D →RDED), conversion of RDE into RGD can be accomplished with a minimum of two mutations (Calvete et al. 2009b; Calvete 2010). Furthermore, the lost of the cysteine-rich domain along SVMP evolution is supported by BA-5A transcript. BA-5A was found in *Bitis arietans* venom gland, and encodes a SVMP with structure similar to those SVMPs of the PIII classification, excluding the cysteine-rich domain, although BA-5A is not translated. Structural features of BA-5A have never been

reported in snake venoms. BA-5A may thus represent a relic of evolution, being an intermediate step in SVMP evolution, between PIII and PII-SVMPs (Juárez et al. 2006a; Carbajo et al. 2015) (Figure 11).

6.2 Disintegrin classification by size and number of cysteine residues

The disintegrin family comprises four subfamilies that are classified according to their polypeptide length and number and pattern of disulfide bonds. Long-chain disintegrins contain ~84-residues cross-linked by 7 intramolecular disulfide linkages, and medium-sized disintegrins consist of ~70 amino acids and 6 intramolecular cysteine sulfide bridge. Homo- and heterodimeric disintegrins consist of subunits of about 67 residues with 10 cysteines involved in the formation of 4 intra-chain disulfides and 2 inter-chain cystine linkages, and finally, short disintegrins are composed of 41-51 residues crosslinked by 4 disulfide bonds (Juárez et al. 2008). Disintegrin activity critically depends on the appropriate pairing of cysteines (Juárez et al. 2008).

The structural diversity of disintegrins, ranging from the ancestral long disintegrins to the more recently evolved short disintegrins (Figure 11), occurred after the emergence of Viperidae as a distinct taxonomical group of advanced snakes ~37 million years ago, in the Eocene epoch of the Cenozoic era. It has been suggested that short disintegrins evolved as the result of a reduction in the size of the disintegrin fold, including the stepwise loss of a pair of cysteine linkages and processing of the N-terminal region (Calvete et al. 2009b; Calvete 2010). Recent structural analyses of two proteoforms of the long disintegrin bitistatin, from *Bitis arietans* venom, confirm the hypothesis that long disintegrins represent the first step in the evolutionary history of the disintegrin family through a disulfide bond engineering process (Carbajo et al. 2015). Particularly, canonical short disintegrins emerged from short-coding RGD dimeric disintegrin precursors and/or short-disintegrin specific PII-SVMP (Sanz et al. 2006; Bazaa et al. 2007; Carbajo et al. 2015) (Figure 11).

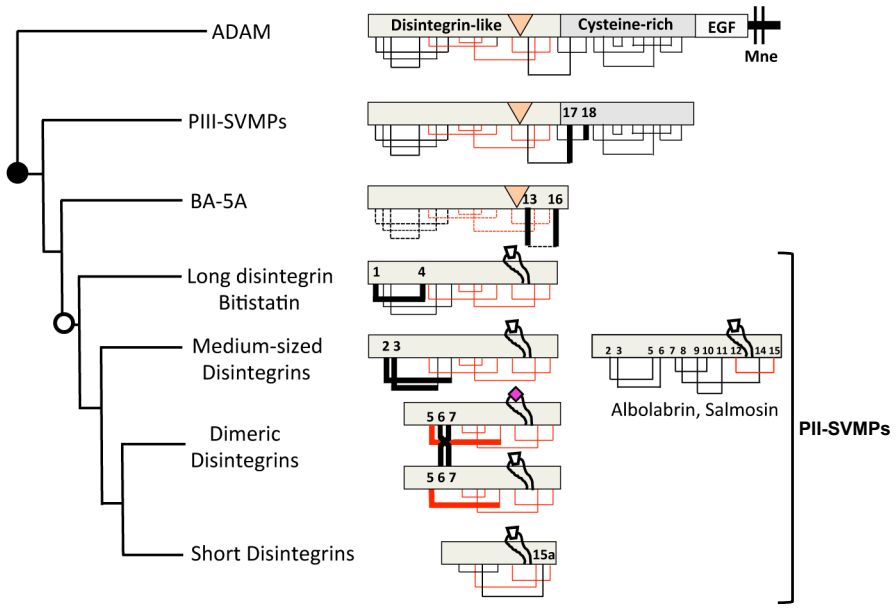


Figure 11. Cartoon of the evolution of the canonical disintegrin family. Scheme of the evolutionary path of disintegrins, highlighting the ancestral recruitment of an ADAM gene and its transformation into a PIII-SVMP toxin into the venom gland (black-filled circle) and the emergence of the disintegrin family (white circle) through the successive loss of disulfide linkages. Cysteine residues that are lost along the proposed evolutionary pathway are depicted as thick bars and numbered as in ADAM molecules. Conserved disulfide bonds between the disintegrin-like domains and disintegrins proper of ADAM/PIII-SVMP are colored red. The divergent disulfide-bonding pattern of albolabrin and salmosin is depicted to the right of the canonical S-S pattern of medium-sized disintegrins. The topology of the ancestral RSECD motif of ADAM and PIII-SVMP molecules, the integrin-binding RGD tripeptide, and the derived KGD, MGD, VGD, WGD or MLD tripeptide motifs are indicated, respectively, by an orange triangle, a white trapeze and a magenta rhomb. The emergence of the integrin-binding loop in PII-disintegrins is highlighted by a loop structure in long, medium-sized, dimeric and short disintegrins. EGF, epidermal growth factor domain; BA-5A represents a nontranslated transcript found in B. arietans venom gland cDNA library (Juárez et al. 2006a). Mne, membrane. Adapted from Carbajo et al. 2015.

6.3 Disintegrin functional classification

Due to the presence of disintegrins, and other integrin antagonist, Viperinae and Crotalinae venoms have evolved a restricted panel of β_1 and β_3 integrin inhibitory motifs, which have emerged via positive Darwinian evolution (Calvete 2010; Carbajo et al. 2015). However, it should be noted that integrin $\alpha_2\beta_1$ does not appear to be a binding site for disintegrins, and this receptor is only targeted by C-type lectin-like proteins (Ogawa et al. 2005; Arlinghaus and Eble 2012) (Figure 12).

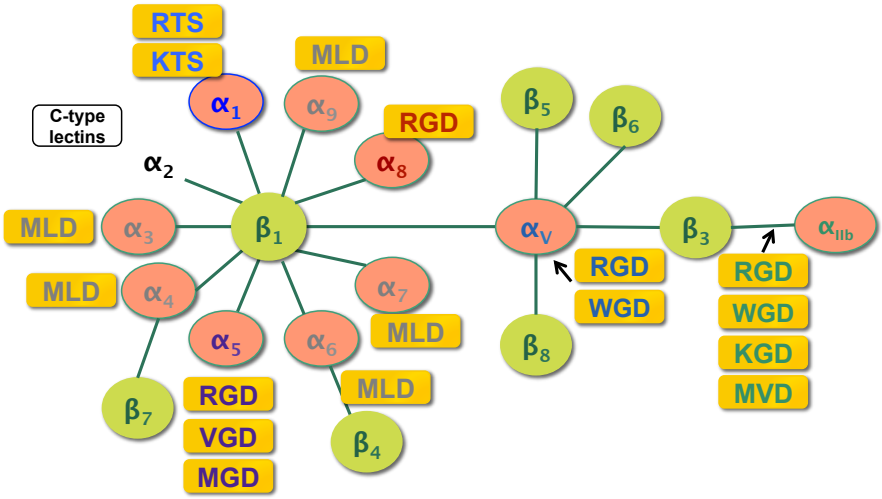


Figure 12. Evolutionary adaptation of disintegrin inhibitory motifs to the ligand-binding sites of integrins. Diagram of the integrin family and the different disintegrin tripeptide motifs (in yellow boxes) that block specific integrin-ligand interactions. Integrin heterodimers antagonized by snake venom disintegrins are encircled. Adapted from Calvete et al. 2007b.

Integrins are a evolutionary old family of heterodimeric transmembrane receptors which are the bridge for cell-cell and cell-extracellular matrix (ECM) interactions (Johnson et al. 2009; Sebé-Pedrós et al. 2010), and are composed of α and β subunits. Currently, 18 α and 8 β subunits have been identified in mammals cells, whose association is limited only to 24 heterodimers that strictly dictate ligand specificity (Hynes 2002; Humphries et al. 2006). Integrins play vital roles in many physiological functions, participating in developmental and pathological processes (reviewed by Melissa Millard 2011). Therefore, selectively blocking specific integrins is a desirable therapeutic goal for a number of pathological conditions, including acute coronary ischemia and thrombosis ($\alpha_{IIb}\beta_3$), tumor metastasis, osteoporosis, restenosis and rheumatoid arthritis ($\alpha_V\beta_3$), bacterial infections and vascular diseases ($\alpha_5\beta_1$), inflammation and autoimmune diseases ($\alpha_4\beta_1$, $\alpha_7\beta_1$, $\alpha_9\beta_1$), and tumor angiogenesis ($\alpha_1\beta_1$, $\alpha_V\beta_3$).

Disintegrin-integrin interactions have been extensively studied (Calvete 2013b), and the therapeutic potential of these compounds warrants continued investigation. Indeed, two drugs currently available on the market as antiplatelet agents (reviewed by Koh and Kini 2012) have been designed based on disintegrin structure, Tirofiban (Aggrastat®) and Eptifibatide (Integrillin®). Platelet aggregation is promoted mainly by the $\alpha_{IIb}\beta_3$ integrin (or GPIIb-IIIa receptor) being mediated by adhesive interaction

between $\alpha_{IIb}\beta_3$ and fibrinogen. Upon integrin-ligand interaction, integrins are capable of transmitting "outside-in" as well as "inside-out" signals across the cell membrane. Particularly, "inside-out" signaling is very important for activation of $\alpha_{IIb}\beta_3$ integrin, which following stimulation of platelets by several agonists undergoes a conformational change to bind fibrinogen. "Outside-in" signaling generates signal transductions upon activation by ligand binding (Hynes 2002; Harburger and Calderwood 2009; Lin et al. 2016). Disintegrins that exhibit an Arg-Gly-Asp (RGD) binding motif are able to inhibit platelet aggregation by binding the active $\alpha_{IIb}\beta_3$ integrin, and thereby blocking the fibrinogen-$\alpha_{IIb}\beta_3$ integrin interaction (Calvete et al. 1994) (Figure 13).

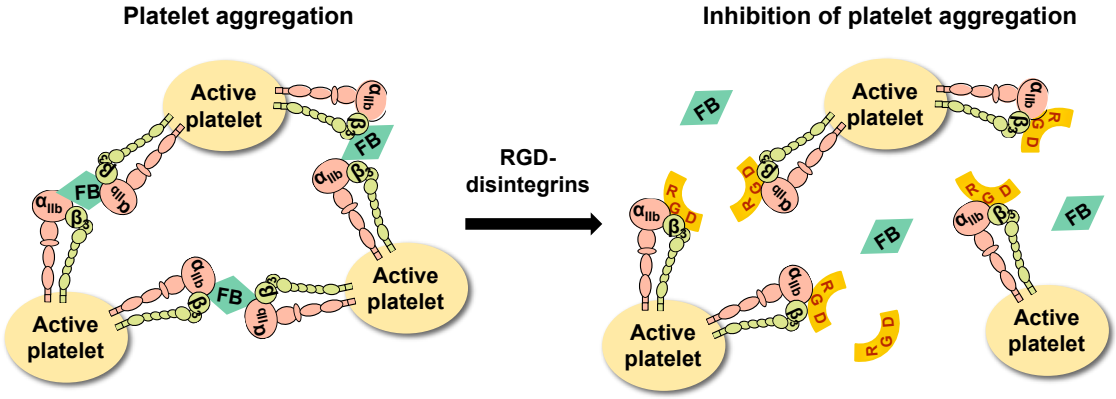


Figure 13. ***In vivo*** **platelet aggregation and platelet aggregation inhibition mediated by RGD-disintegrin o peptides based on RGD disintegrins (RGD).** α_{IIb} **and** β_3 **denote the integrin subunits and FB, fibrinogen**

Similarly, other disintegrins have been shown to inhibit integrin binding to their natural ligands, affecting normal or pathogenic processes.

Disintegrins are also classified according to the integrin-binding motifs found in their active site. The vast majority of single-chain (long, medium-sized and short) disintegrins express the canonical RGD sequence, however, a few medium-sized disintegrins bear a KGD sequence. Dimeric disintegrins exhibit the largest variability in their integrin recognition motifs, including another common motif of MLD, in addition to RGD, KGD, VGD, MGD, and WGD (Calvete et al. 2009b; Calvete 2010). On the other hand, the KTS and RTS binding-motifs are only observed in short disintegrins (Figure 12).

Except for the KTS/RTS short disintegrins (Calvete et al. 2007b), the most parsimonious nucleotide substitutions events required a minimum of three mutations for the emergence of all currently known integrin-recognition motifs from the ancestral

RGD sequence (Figure 14).

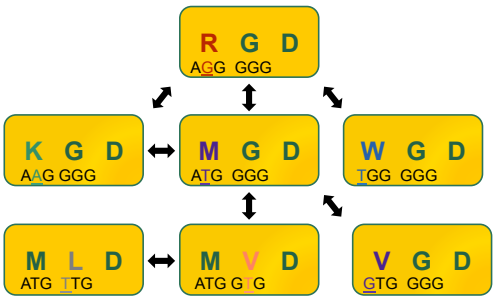


Figure 14. Evolution of the integrin-binding site. Most parsimonious nucleotide substitution events required for the emergence of all known disintegrin's integrin-recognition motifs from an ancestral RGD sequence. Arrows indicate mutational transitions at the underlined sites. Adapted from Juárez et al. 2008.

Calvete et al. 2007 suggest that disintegrins evolved in a restricted panel of integrin blocking sequences, indicating an evolutionary adaptation of venom disintegrins to the specific ligand-binding sites of integrin expressed in their prey items (Figure 15).

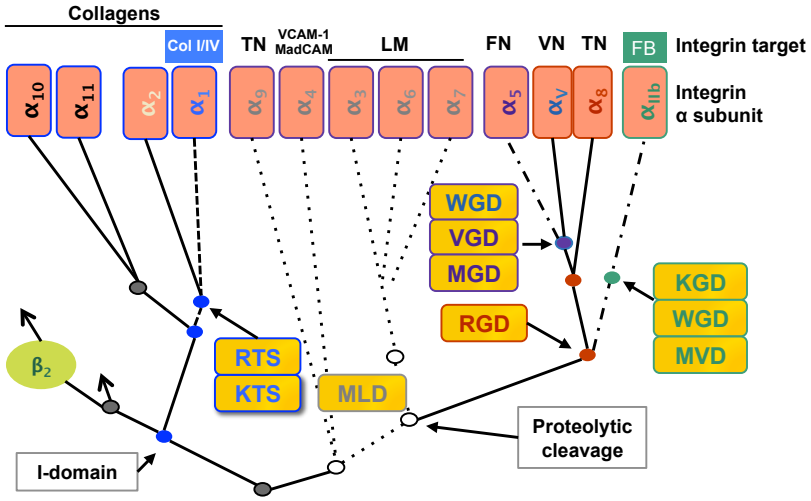


Figure 15. Segregation of disintegrin binding motifs (displayed in yellow boxes), across the phylogenetic tree of the integrin α-subunits (displayed in coral boxes). Branches are not scaled according to evolutionary distance. The disintegrin motifs and their specific α subunit targets are display in the same color. Integrin major ligands are indicated. $\alpha_1\beta_1$ integrin is a receptor for collagen I and IV (Col I/IV); $\alpha_2\beta_1$ binds collagen I; $\alpha_4\beta_1$ interacts with fibronectin and VCAM-1; $\alpha_4\beta_7$ bind the same ligands as $\alpha_4\beta_1$ and in addition is a receptor for MdCAM; $\alpha_5\beta_1$ represents the major fibronectin (FN) receptor; integrins $\alpha_3\beta_1$, $\alpha_6\beta_1$, and $\alpha_7\beta_1$ represent major laminin (LM) receptors; $\alpha_8\beta_1$ and $\alpha_9\beta_1$ bind tenascin (TN); $\alpha_V\beta_1$ and $\alpha_V\beta_3$ are major vitronectin (VN) receptors; and $\alpha_{IIb}\beta_3$ is the platelet fibrinogen (FB) receptor involved in platelet aggregation. Branch points linked with the emergence of the proteolytic cleavage of integrin α-subunits, and the acquisition of an I-domain, are indicated. Adapted from Calvete et al. 2007b.

6.4 Disintegrin structure-function studies

Structural studies on a number of short (echistatin, 1RO3), medium-sized (kistrin, 1N4Y; flavoridin, 1FVL; albolabrin (Smith et al. 1996); salmosin, 1IQ2, 1L3X; rhodostomin, 1Q7J, 1Q7I, 2PJI, 2PJF, 1JYP; trimestatin, 1J2L) and dimeric disintegrins (1Z1X, 1RMR, 1TEJ, 3CO5) have revealed that their RGD/KGD integrin inhibitory motifs have evolved at the apex of an 11 amino acid mobile loop, protruding 14–17 Å from the disintegrin protein scaffold and maintained in the active conformation by the appropriate pairing of cysteine residues. Currently known integrin-blocking motifs include RGD, which blocks integrins $\alpha_8\beta_1$, $\alpha_5\beta_1$, $\alpha_V\beta_1$, $\alpha_V\beta_3$, and $\alpha_{IIb}\beta_3$; MLD targets the $\alpha_4\beta_1$, $\alpha_4\beta_7$, $\alpha_3\beta_1$, $\alpha_6\beta_1$, $\alpha_7\beta_1$ and $\alpha_9\beta_1$ integrins; VGD and MGD impair the function of the $\alpha_5\beta_1$ integrin; KGD inhibits the $\alpha_{IIb}\beta_3$ integrin with a high degree of selectivity; WGD has been reported to be a potent inhibitor of the RGD-dependent integrins $\alpha_5\beta_1$, $\alpha_V\beta_3$, and $\alpha_{IIb}\beta_3$; the adhesive function of the latter integrin is also blocked by MVD (Shimokawa et al. 1998; Sanz et al. 2006; Calvete et al. 2009b; Calvete 2010; Carey et al. 2012) (Figure 12).

The crystal structure of the extracellular segment of integrin $\alpha_V\beta_3$ in complex with an RGD ligand (Xiong 2002) revealed that the peptide fits into a crevice between the α_V propeller and the β_3 A-domain. The Arg side-chain is held in place by interactions with α_V carboxylates, the Gly residue makes several hydrophobic interactions with α_V, and the Asp ligand interacts primarily with βA residues. Thus, the conserved aspartate residue might be responsible for the binding of disintegrins to integrin receptors which share a β subunit, while the two other residues of the integrin-binding motif (RG, KG, MG, WG, ML, MV, VG) may dictate the primary integrin-recognition specificity, with residues flanking the active tripeptide finely tuning the potency and integrin receptor selectivity of disintegrins (McLane et al. 1996; Wierzbicka-Patynowski et al. 1999), reviewed by (Calvete et al. 2005; Calvete 2005). High-resolution NMR studies (Monleón et al. 2005) provided a structural ground for the biochemically defined functional synergy between the RGD loop and the C-terminal region of echistatin (Marcinkiewicz et al. 1997).

Short RTS/KTS disintegrins have only been found in Eurasian vipers, selectively target the integrin $\alpha_1\beta_1$ (Calvete et al. 2007b), and form a distinct clade within the short disintegrin subfamily (Figure 16).

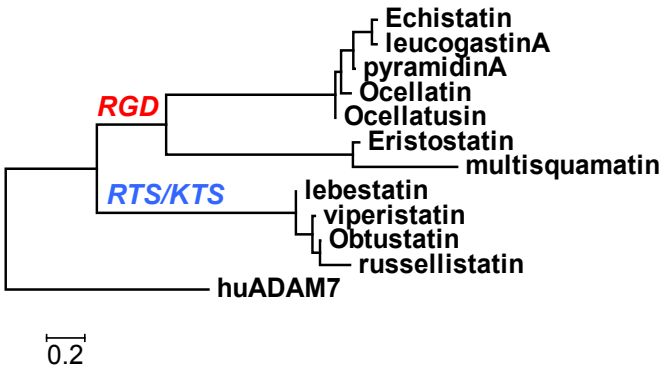


Figure 16. Inferred phylogeny within the short disintegrin subfamily. Cladistic relationships between the RGD and the KTS/RTS short disintegrins were inferred through Neighbor-Joining using maximum-likelihood distances using the PHYML program (Guindon and Gascuel 2003), with the disintegrin-like domain of human ADAM-7 serving as out-group. This distribution parallels the phylogenetic tree of the species in whose venoms short disintegrins have been characterized. Expression of RGD-disintegrins appear to be restricted to *Echis* and *Eristicophis* taxa, whereas RTS/KTS-disintegrins have been only reported in venoms from Eurasian vipers, genera *Vipera, Macrovipera*, and *Daboia* (Sanz-Soler et al. 2012)

Compared to all known disintegrin structures, in which the RGD motif is located at the apex of an eleven residue hairpin loop, the active RTS/KTS tripeptide is oriented towards a side of a nine residue integrin-binding loop (Moreno-Murciano et al. 2003b) (Figure 17). Structure-function correlation studies have shown that the selectivity of KTS-disintegrins for the $\alpha_1\beta_1$ integrin resides within a conformational epitope encompassing the integrin-binding loop and the C-terminal tail (Monleón et al. 2003; Moreno-Murciano et al. 2003a; Kisiel et al. 2004; Carbajo et al. 2011). The potency of recombinant KTS-disintegrin obtustatin also depends on the residues found carboxy to the active motif (Brown et al. 2009) (Figure 17).

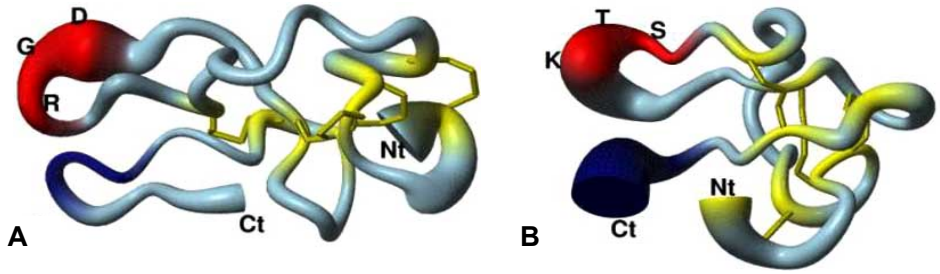


Figure 17. Comparison of the NMR structures of echistatin and obtustatin. A, echistatin (PDB code 1RO3) and, B, obtustatin (PDB code 1MPZ, bottom) are rendered in 'sausage' representation, where the thickness of the backbone is proportional to the structural dispersion (RMSD value) of the backbone atoms coordinates. Disulfide bonds and the active tripeptides (RGD in echistatin; KTS in obtustatin) are colored yellow and red, respectively. The C-terminal tail residues (43-46 in echistatin and 40-41 in obtustatin), represent another functional epitope, are colored in dark blue. Adapted from Calvete 2005.

In the present dissertation we focus on different aspects of the disintegrin family of snake venom proteins. Specifically, we studied the disintegrins crotatroxin, a medium size RGD-disintegrin, as well as, two short disintegrins, ocellatusin, an RGD-disintegrin and jerdostatin, an RTS-disintegrin.

7 Short disintegrins

Short disintegrins can be classified as: i) the classical RGD-disintegrins, which evolved as previously described in (Figure 12), and might have emerged after the split of Asian and New World (NW) Crotalinae and African and Eurasian Viperinae (Juárez et al. 2008; Calvete 2013a; Ducancel et al. 2014; Gonçalves-Machado et al. 2015) (Figure 3) and ii) RTS/KTS-disintegrins, whose evolution is not clearly understood (Calvete et al. 2007b).

7.1 Ocellatusin, an RGD-disintegrin

Ocellatusin is a short RGD-containing monomeric disintegrin, present in the venom of the African saw-scaled viper, *Echis ocellatus*.

Ocellatusin, is encoded by two different transcripts in the venom gland: a 1485 base pair (bp) mRNA encoding the PII-SVMP Q14FJ4, and a short coding message (384 bp), that lacks the metalloproteinase domain and most of the propeptide sequence [Q3BER1] (Juárez et al. 2006b; Wagstaff et al. 2009) (Figure 18).

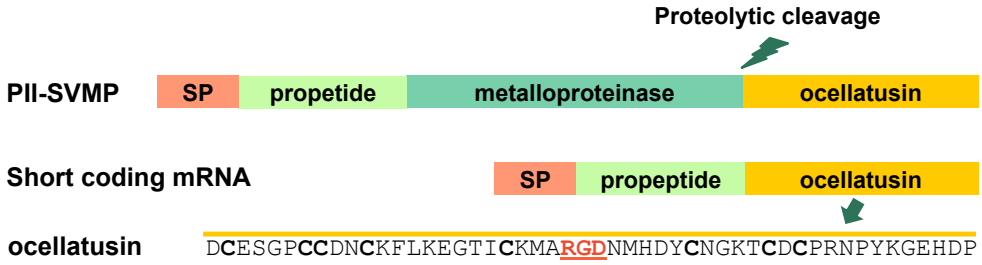


Figure 18. Cartoon of ocellatusin precursors found in *Echis ocellatus* venom gland coding DNA described by (Juárez et al. 2006b). SP, signal peptide. The cleavage site is indicated in the snake venom metalloproteinase (SVMP-PII). Amino acid sequence of the ocellatusin disintegrin domain; Cysteines are in bold and the active tripeptide in red and underlined.

Ocellatusin express the RGD tripeptide at the apex of the 11 amino acid integrin-binding loop and inhibits RGD-dependent integrins, including the integrins $\alpha_5\beta_1$, $\alpha_V\beta_3$, and $\alpha_{IIb}\beta_3$ (Smith et al. 2002). These three integrins are involved in many crucial physiological processes such as homeostasis and thrombosis, angiogenesis, inflammation, bone resorption and wound healing, cell growth and organ development (Smith et al. 2002).

In vitro and *in vivo* analyses indicate that ocellatusin inhibits ADP-induced platelet aggregation, and exhibits a similar potency to those reported for other short monomeric disintegrins form the genus *Echis* (Smith et al. 2002). In addition, it is a potent inhibitor of $\alpha_5\beta_1$ integrin.

7.2 Jerdostatin, an RTS-disintegrin

Reverse transcription of mRNA isolated from the venom gland of *Protobothrops jerdonii* led to the amplification of cDNA coding for jerdostatin, an RTS disintegrin belonging to the KTS/RTS disintegrin family. As suggested by the name, jerdostatin was first identified in *P. jerdonii* venom gland cDNA. Particularly, its nucleotide sequence comprises 369 bp (GenBank accession code AY262730) coding for an open reading frame of 333 bp, including a signal peptide sequence, propeptide, and short disintegrin domain (residues 69 –110) (Sanz et al. 2005) (Figure 19). Short coding mRNA sequence for jerdostatin has been cloned from several snake venom glands cDNA (Sanz et al. 2006; Bazaa et al. 2007) (Figure 19).

Despite the presence of transcripts of the gene coding for jerdostatin in several snake venom glands, this protein had not been identified in any venom proteome, to date.

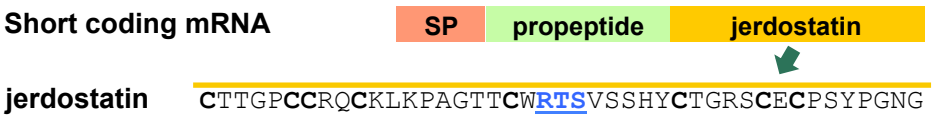


Figure 19. Diagram of jerdostatin unique precursor found in *Protobothrops jerdonii* venom gland coding DNA described by (Sanz et al. 2005). SP, signal peptide. Amino acid sequence of the mature disintegrin domain, jerdostatin, is displayed. Cys are in bold and the active tripeptide in blue and underlined.

RTS-jerdostatin amino acid sequence displayed 80% identity with KTS-disintegrins obtustatin (Marcinkiewicz et al. 2003), lebestatin (Olfa et al. 2005) and viperistatin (Kisiel et al. 2004). Recombinant jerdostatin selectively targets the collagen IV binding integrin, $\alpha_1\beta_1$, (Sanz et al. 2005; Calvete et al. 2007b). This activity is shared by other short KTS-disintegrins: obtustatin (Marcinkiewicz et al. 2003; Brown et al. 2008), viperistatin (Staniszewska et al. 2009; Momic et al. 2011) and lebestatin (Olfa et al. 2005). Type IV collagen is an exclusive constituent of basement membranes, where it creates complex supramolecular networks that influences cell adhesion, migration and differentiation (Khoshnoodi et al. 2008). Specifically, the $\alpha_1\beta_1$ integrin is involved in tumor angiogenesis and metastasis (Marcinkiewicz et al. 2000; Jin and Varner 2004; Tucker 2006; Garmy-Susini and Varner 2008; Staniszewska et al. 2009; Ghazaryan et al. 2015); hence selective blocking this receptor is a desirable goal for cancer treatment. K/RTS-disintegrins block endothelial cell proliferation and angiogenesis (Olfa et al. 2005; Brown et al. 2008). Particularly, r-jerdostatin produced in both a mammalian cell system (Juárez et al. 2010) and in BL21 *Escherichia coli* (Bolás et al. 2014), inhibited the binding of soluble $\alpha_1\beta_1$ integrin to the CB3 fragment of collagen IV in a dose-dependent manner (Figure 20). Further, jerdostatin disrupted the adhesion of glioblastoma (RuGli) cells to collagen IV (Juárez et al. 2010) and also affects $\alpha_1\beta_1$-mediated adhesion, migration and proliferation of rat aortic smooth muscle cells and angiogenesis (Bolás et al. 2014).

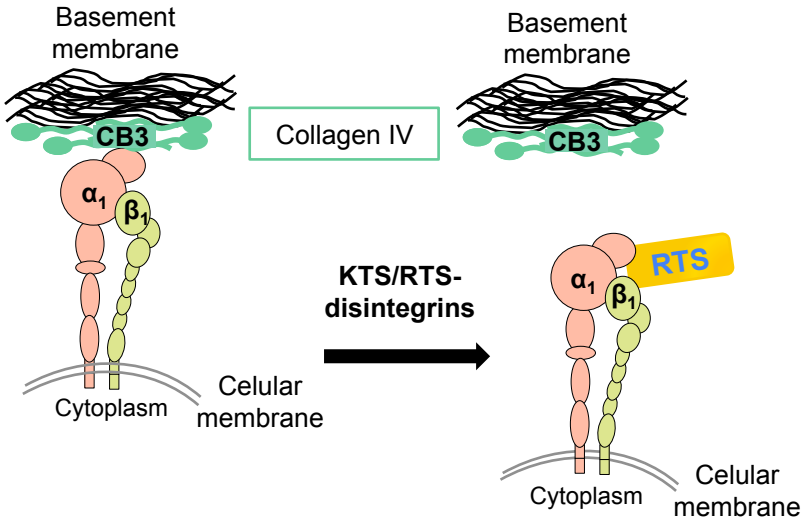


Figure 20. Cartoon of the inhibition of $\alpha_1\beta_1$ integrin-binding to the CB3 domain (CB3) of the collagen IV by KTS/RTS-disintegrins (RTS). α_1 and β_1 designs the subunits of the integrin.

NMR studies of a $\alpha_1\beta_1$-blocking short disintegrins KTS-obtustatin (from *Macrovipera lebetina obtusa*), mentioned previously (Brown et al. 2009) and the recombinant jerdostatin, wild-type (RTS) and mutants (KTS, RTSΔNG, KTSΔNG) (Carbajo et al. 2011) revealed that the integrin binding loop and the C-terminal tail of these disintegrins form a conformational functional epitope and display concerted motions (Figure 21). The shape and size of its 2 residue shortened integrin-binding loop (9 amino acid) respect RGD loop (11 amino acid), along with its composition, flexibility, and the distinct (lateral) orientation of the KTS/RTS tripeptide, may underlay the structural basis of their unique selectivity and specificity for integrin $\alpha_1\beta_1$ (Marcinkiewicz et al. 2003; Monleón et al. 2003; Moreno-Murciano et al. 2003a; Calvete et al. 2007b; Carbajo et al. 2011) (Figure 21). NMR characterization of jerdostatin molecules reported in Carbajo et al. in 2011 highlights the role played by dynamics in the integrin-inhibitor recognition process, suggesting that the conformation and dynamics of the integrin binding loop of jerdostatin are important for fast recognition of integrin $\alpha_1\beta_1$ (Carbajo et al. 2011) (Figure 21).

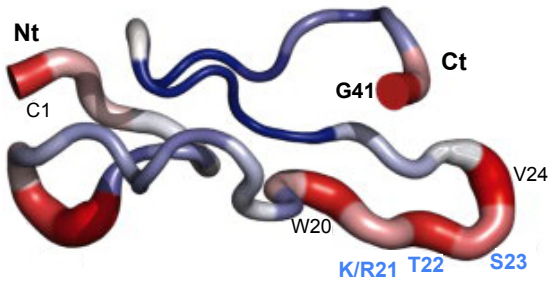


Figure 21. Representation of the backbone superposition of the average structures of r-RTSjerdostatin, r-RTSjerdostatinΔNG, r-KTSjerdostatin21, and r-KTSjerdostatinΔNG. Regions showing low dispersion are narrower and depicted in blue, whereas regions exhibiting higher structural dispersion are wider and colored in red. Adapted from Carbajo et al. 2011.

8 Medium-size disintegrin, crotatroxin

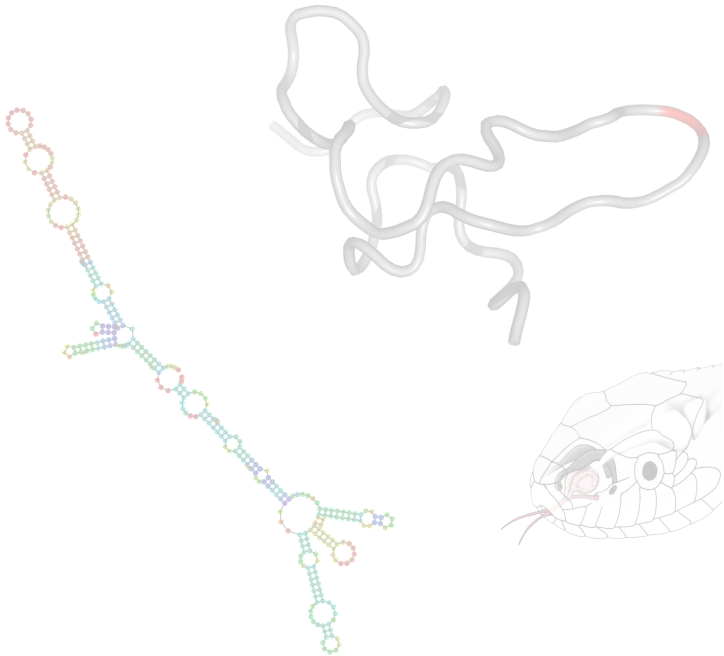
Crotatroxins are medium-sized monomeric disintegrins released in the venom of *Crotalus atrox* (Western diamondback rattlesnake). Crotatroxin 1 and 2 contain 72 to 71 amino acids, respectively, which conform their structure with six disulfide bonds. The only difference between Crotatroxin 1 (GenBank: AAB24808.1) and 2 is found with the Alanine (A) present at the N terminal end of crotatroxin 1 (Figure 22).

```
                1         10        20        30        40        50        60        70
crotatroxin1    AGEECDCGSPANPCCDAATCKLRPGAQCADGLCCDQCRFIKKGTVCRPARGDWNDDTCTGQSADCPRNGLYG  72
crotatroxin2     GEECDCGSPANPCCDAATCKLRPGAQCADGLCCDQCRFIKKGTVCRPARGDWNDDTCTGQSADCPRNGLYG  71
```

Figure 22. Medium size disintegrins expressed in *Crotalus atrox* venom. Crotatroxin 1 and 2 amino acid sequences; cysteine are displayed in bold, RGD active motif red and underlined.

Functionally, these disintegrins might be important at clinical level, given that crotatroxin inhibited human whole blood platelet aggregation with an IC_{50} of 17.5 nM, inhibit M21 melanoma cell adhesion to fibronectin (Scarborough et al. 1993; Fox and Serrano 2005; Kini 2006), and may be inhibitory against integrin $\alpha_5\beta_1$. Further, crotatroxin inhibited lung tumor colonization *in vivo* using a mouse model (Galán et al. 2008). On the other hand, in the prey-predator context, crotatroxin contributes to the envenomation with its hemorrhagic function.

Objectives



Objectives

In this dissertation we sought to explore the evolutionary relationship of two different groups of short disintegrins, RGD- and RTS/KTS-disintegrins, in addition to studying the *RPTLN* gene distribution and its possible function across Reptilia. Hence, the following objectives have been addressed in this work:

1- Investigate the evolutionary history of the RTS/KTS-disintegrin family, analyzing RGD- and RTS/KTS-disintegrins structure-function, by generating chimeric RTS-RGD-disintegrin mutants.

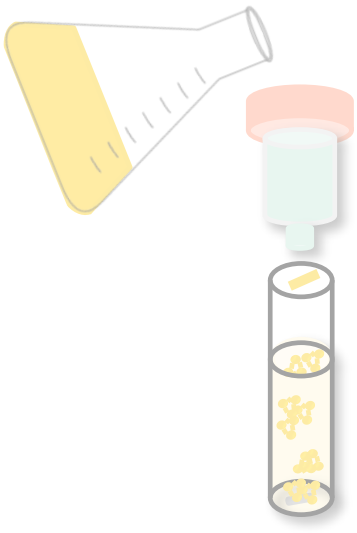
2- Define essential residues for ocellatusin natural functionality, studying the key structure-function determinants in this short disintegrin.

3- Characterize the complete *RPTLN* gene coding for RTS-disintegrins, in addition to exploring the distribution of this gene across Reptilia.

4- Study the distribution of *RPTLN* gene transcripts across Reptilia and possible translation of its disintegrin domain in organs other than the venom gland.

5- Assess a possible new crotatroxin disintegrin functionality in *Crotalus atrox* venom, as an element to relocate prey following the strike-and-release predatory strategy, and investigate the possibility that short RGD-disintegrins are recognized by *Crotalus atrox* as a chemoattractant element.

Methodology



1 Genomic DNA extraction, RNA isolation and complementary DNA synthesis

Blood and tissues samples for genomic DNA and RNA extraction and proteins isolation were provided or donated by following people and institutions listed in Table 1.

Table 1. Blood, tissues and fresh venom provided by different people and institutions

Material	Species	Provided by
Blood	*Ophiophagus Hannah*	José María López (SoHeVa)*
Blood	*Boa constrictor*	José María López (SoHeVa)*
Blood	*Timon lepidus*	José María López (SoHeVa)*
Blood	*Heloderma horridum*	José María López (SoHeVa)*
Blood	*Uroplatus ebenaui*	José María López (SoHeVa)*
Blood	*Chamaeleo calyptratus*	José María López (SoHeVa)*
Blood	*Testudo graeca*	José María López (SoHeVa)*
Blood	*Testudo Hermanni*	José María López (SoHeVa)*
Blood	*Stigmochelys pardalis*	José María López (SoHeVa)*
Blood	*Mauremys annamensis*	José María López (SoHeVa)*
Blood	*Mauremys sinensis*	José María López (SoHeVa)*
Blood	*Chelonoidis carbonaria*	José María López (SoHeVa)*
Blood	*Chelonoidis chilensis*	José María López (SoHeVa)*
Liver	*Alligator*	José María López (SoHeVa)*
Tail	*Lacerta hispanica*	caught in Valencia
Tail	*Tarentola mauritanica*	caught in Valencia
Tissue	*Naja haje haje*	César Olmos (Private zoological, Cullera)
gDNA	*Echis ocellatus*	Dr. Robert A. Harrison (Liverpool, UK)
gDNA	*Bothrops asper*	Dr. Bruno Lomonte (I. Clodomiro Picado, Costa Rica)
gDNA	*Bothriechis lateralis*	Dr. Bruno Lomonte (I. Clodomiro Picado, Costa Rica)
gDNA	*Atropoides picadoi*	Dr. Bruno Lomonte (I. Clodomiro Picado, Costa Rica)
Tissues	*Podarcis muralis*	Dr. Enrique Font (Universitat de Valencia, Valencia)
Tissues	*Podarcis hispanica*	Found in Valencia
Tissues	*Mus musculus*	Our laboratory
Tissues	*Rhinechis scalaris*	Found in IBV (Valencia)
gDNA	*Arabidopsis thaliana*	Anonymous laboratory (Valencia)
gDNA	*Homo sapiens sapiens*	Anonymous laboratory (Valencia)
Blood	*Gallus gallus*	Carlos Núñez (Valencia)
Blood	*Alectoris rufa*	Carlos Núñez (Valencia)
Blood	*Anas platyrhychos domesticus*	Carlos Núñez (Valencia)
Venom	*Crotalus atrox*	Dr. Stephen Mackessy (UNC**, Colorado)

***SoHeVa (Sociedad Herpetológica Valenciana). **UNC (University of Northern Colorado)**

1.1 Isolation of genomic DNA

Blood cells and tissues were incubated overnight at 55˚C in lysis buffer (100 mM Tris, 25 mM EDTA, 100 mM NaCl, 0.5% SDS) and 0.2 µg/µL proteinase K (Sigma-Aldrich®). Genomic DNA (gDNA) was isolated using phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) extraction and precipitated by adding 1/10 vol of 3M sodium acetate (pH 5.2) and 2 vol of 100% ethanol. gDNA was resuspended in TE buffer (10 mM Tris, pH 8, 1 mM EDTA) (modified from Longmire et al. 1997).

1.2 Tissues preparation, RNA isolation and reverse transcription.

Organs from *Rhinechis scalaris* (lung, heart, skeletal muscle, skin), *Podarcis muralis* (bladder, liver, lung, kidney, skeletal muscle, skin, stomach, heart), and *Podarcis hispanica* (liver, lung, skin, stomach, heart, brain) were dissected, minced manually, and stored in RNA*later*® (Sigma-Aldrich®). Total RNA extraction was performed using the TRIzol method following the manufacturer's (Life Technologies, NY, USA) recommended protocol. Total RNA was treated with RNAase-free DNAaseI following the manufacturer's (Thermo Scientific) protocol. DNAseI-treated total RNA integrity was assessed by electrophoresis in a 2% agarose gel. One µg of the RNA was reverse-transcribed to first strand cDNA, using oligo $(dT)_{18}$ and the RevertAid H Minus. First Strand complementary DNA (cDNA) Synthesis Kit (Thermo Scientific). The cDNA was stored at -80˚C until used. Reverse transcriptase minus (RT-) negative control was performed to verify the absence of gDNA in the RNA sample. The RT- control contained the same reaction mixture used for reverse transcription except for the RevertAid reverse transcriptase, which was substituted for 1 µl of RNAse-free water. Non-template control (NTC) was also included to discard reagent contaminations. The NTC reaction contained the same reagents than the reverse transcription reaction, excepting the RNA template.

1.3 Snake milking, RNA isolation from the venom, and reverse transcription

Treated material with RNase AWAY® (Sigma), was used for *Crotalus atrox* milking. Immediately, 100 µl of the venom was added to 1 mL TRIzol® solution, followed by the RNA isolation procedure described previously. 3´RACE System for Rapid Amplification of cDNA Ends (Invitrogen) manufacture protocol was followed for

Crotalus atrox cDNA synthesis. Adapter Primer (AP), containing oligo $(dT)_{16}$ and AUAP (Abridged Universal Amplification Primer) sequence was used for the first strand cDNA synthesis from total RNA (Modahl et al., *in press*).

2 PCR-amplification of DNA and RNA sequences

2.1 *RPTLN* and *RPTLN-like* sequences amplification

RPTLN sequences were PCR-amplified in a final volume of 25 µL containing 0.02 units of iProof™ High-Fidelity DNA Polymerase (Bio-Rad), 1.5 mM $MgCl_2$, 0.2 µM of each forward SP_jerdostatin 5′-ATGATCCAGGTTCTCTTGGTAACTATATG-3′ [MIQVLLVTI] and reverse 3′jerdostatin 5′-TAGCCATTCCCGGGATAACTGG-3′ [PSYPGNG] primers (Table 2), and 100 ng of gDNA, or 1 µL of cDNA, RT- or NTC, as template. PCR protocol included denaturation at 98˚C for 2 min, followed by 35 cycles of denaturation (10 s at 98˚C), annealing (20 s at 58˚C), extension (40 s at 72˚C), and a final extension step for 5 min at 72˚C. One µL of Mili-Q® water, without template, was used as negative control in every PCR-amplification. The PCR reaction mixture was run in 1% agarose gel, and the candidate 333 bp DNA band was excised.

2.2 PCR-amplification of the housekeeping gene 28S ribosomal RNA

Amplification of a partial sequence of the 28S ribosomal RNA (rRNA) gene was used as an internal control in every *RPTLN* gene expression study. The forward rRNA28S and reverse rRNA28S primers (Table 2) were designed based on the partial sequence of *Anolis carolinensis* 28S ribosomal RNA gene [AY859623]. The homologous 275 bp sequences of *Podarcis muralis* [KU556683], *Podarcis hispanica* and *Rhinechis scalaris* were amplified using the FirePol® DNA Polymerase (Solis BioDyne) protocol. PCR-amplification was performed using an initial denaturation step at 94˚C for 5 min, followed by 35 cycles of denaturation (20 s at 94˚C), annealing (20 s at 58˚C), extension (30s at 72˚C), and a final extension step for 5 min at 72˚C.

2.3 PCR-amplification of intron 7 of fibrinogen β-chain

PCR-amplification of a 296 bp fragment of intron 7 of fibrinogen β-chain was performed as a double check to confirm the absence of amplicons arising from contaminating gDNA. *Podarcis muralis* and *Podarcis hispanicus* cDNA were used as

templates in the FirePol® (Solis BioDyne) DNA Polymerase protocol, with Fw_Intron7FGB_*Podarcis* and Rv_Intron7FGB_*Podarcis* as forward and reverse primers, respectively (Table 2). These primers were designed from the sequence of intron 7 of the *Podarcis muralis* haplotype B80 β-fibrinogen (FGB) gene [EU269550]. PCR-amplification was performed using an initial denaturation step (94˚C for 5 min), followed by 35 cycles of denaturation (20 s at 94˚C), annealing (20 s at 60˚C), extension (30 s at 72˚C), and a final extension step for 5 min at 72˚C. *Rhinechis scalaris* cDNA was also subject to PCR-amplification of a 1519 bp sequence of intron 7 of fibrinogen β-chain [KU556682], using forward Fw_7IFGB and reverse Rv_7IFGB primers (Table 2), designed based on the β-fibrinogen intron 7 sequence of *Trimeresurus* species [AF517209]. FirePol® DNA Polymerase (Solis BioDyne) protocol consisted of an initial denaturation cycle at 94˚C for 5 min, followed by 35 cycles of denaturation (20 s at 94˚C), annealing (20 s at 56˚C), extension (90 s at 72˚C), and a final extension step for 5 min at 72˚C.

Table 2. *RPTLN*, housekeeping gene 28S ribosomal RNA, and intron 7 of fibrinogen β-chain primers

Primers	Sequence
SP_jerdostatin	ATGATCCAGGTTCTCTTGGTAACTATATG
3´jerdostatin	TAGCCATTCCCGGGATAACTGG
Forward rRNA28S	GTAACGCAGGTGTCCTAAGG
Reverse rRNA28S	CGCTTGGTGAATTCTGCTTC
Fw_Intron7FGB_*Podarcis*	GGATCATGCTGTCAGGCTGG
Rv_ Intron7FGB_*Podarcis*	CAGTGGTACCTTGGGTTAAGAAC
Fw_7IFGB	AGAGACAATGATGGATGGTAAG
Rv_7IFGB	AGAGACAATGATGGATGGTAAG

3 Purification, cloning and sequencing of PCR products

DNA candidate fragments, were isolated from the agarose gel and purified by GeneClean® TurboKit (MP BioMedicals, LLC) or Illustra GFX Gel Band Purification Kit (GE Healthcare, Buckinghamshire, UK) and cloned into a pGEM®-T (Promega, Madison, WI, USA), or a pJET1.2/blunt vector (Thermo Scientific). Ligation mixtures were transformed in DH5α *E. coli* strain cells (Novagen, Madison, WI) by electroporation using and Eppendorf 2510 electroporator. Positive clones were selected

by growing the transformed cells in Luria-Broth (LB) medium 2% agar, containing 100 µg/ml ampicillin, along with beta-galactosidase selection only in the case of pGEM®-T clones. Positive clones were confirmed by PCR amplification using vector-specific primers; M13, for pGEM®-T, and pJET1.2, for pJET1.2/blunt vector. The PCR-amplified positive clones were isolated by DNA Purification System Wizard® Plus SV Minipreps (Promega) and sequenced (using an Applied Biosystems model 377 DNA sequencer) or ABI PRISM-3130XL Genetic Analyzer (Applied Biosystems).

4 Semiquantitative PCR and Real-time PCR

4.1 Semiquantitative PCR

Semiquantitative PCR was performed by electrophoresis in 1% agarose gel following the conventional PCR-amplifications of the *RPTLN* gene and the 28S rRNA fragments (previously described), for 25, 30, and 35 cycles. PCR products were run arranged by cycle's number, Tube 1, 25 cycles; tube 2, 30 cycles and tube 3, 35 cycles.

4.2 Real-time PCR

Quantitative real-time PCR was performed in duplicate using Light Cycler FastStart DNA Master SYBR green I (Roche) in a Light Cycler 480 (Roche), following the manufacture's protocol using SP_jerdostatin and 3´jerdostatin primers (Table 2) for *RPTLN*-like amplification, and forward and reverse rRNA28S primers (Table 2) for the housekeeping fragment. 1 µg of total RNA from *Podarcis muralis* heart (in a final volume of 20 µl) was transcribed into cDNA, and 3µL (0.15 µg of RNA) of the reaction mixture were used as template. pMD18-T/*RPTLN* and pJET1.2/28SrRNA plasmids were used as positive controls and internal standard, respectively.

5 Sequence analyses

5.1 Sequence identification

Sequence similarity searches were done using BLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The DNA sequence chromatograms of the *RPTLN* genes and mutagenized disintegrins were analyzed by Chromass software

(Technelysium). The *translation* of a nucleotide (DNA/RNA) to a protein sequence was done using the translate ExPaSy tool (http://web.expasy.org/translate/).

5.2 Sequences alignment

Multiple sequence alignment was performed using MEGA (Molecular Evolutionary Genetic Analysis; http://www.megasoftware.net).

Note that the alignment of *RPTLN-like* sequences became complex in number and it included redundant information. Therefore, *RPTLN* alignment was simplified considering as *bona fide* changes, only the nucleotide substitution observed at least twice in different species. To avoid the inclusion of sequencing mistakes, the changes found only once were discarded, assuming we might lose some information.

5.3 RNA secondary structure prediction

Prediction of RNA secondary structure was performed using the RNAfold WebServer (http://rna.tbi.univie.ac.at) at the Institute for Theoretical Chemistry, University of Vienna (Mathews et al. 2004; Gruber et al. 2008; Lorenz et al. 2011).

6 Sequences accession codes

All nucleotide sequences gathered in this work have been deposited with the NCBI database under accession codes KU556682 (partial 1519 bp sequence of intron 7 of fibrinogen β-chain of *P. scalaris*), KU556683 (partial 275 bp sequence of 28S ribosomal RNA from *Podarcis muralis*) and KU563546-KU563619 (*RPTLN* 2-21 sequences from different organisms listed in table 1S (p.171) and table 2S (p.172).

7 Protein extraction and Western blotting

7.1 Protein extraction

For total protein extraction, 5 mm-thick portions of minced tissues or organs where homogenized in lysis buffer (10 mM Tris, pH 7.5, 1 mM EDTA, 1 mM $MgCl_2$, 10% glicerol, 5 μM β-mercaptoetanol, containing a tablet of EDTA-free protease inhibitor (Roche) per 50mL of lysis buffer and 0.4 mM Pefabloc SC (AEBSF) (Roche), using an Ultra-Turrax® (Ika® Werke) homogenizator. SDS was added to a final

concentration of 2% (w/v). The samples were vortexed for ~15 seconds, incubated for 10 min on ice. This process was repeated twice. Immediately, the samples were centrifuged at 14,000xg for 20 min at 4˚C, and the supernatants were transferred to clean tubes. The protein extracts were suspended in loading sample buffer in reducing conditions and heated 5 minutes at 95°C, prior loading in polyacrylamide gel.

7.2 Protein analysis and Western blotting

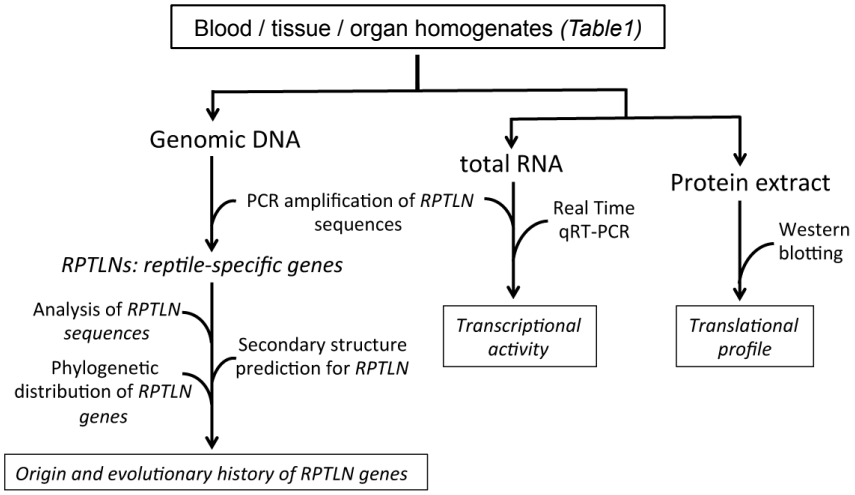
Aliquots of 40-60 μg of total proteins extracted from 200 μg of organ/tissue homogenates of *R. scalaris* (liver, skeletal muscle, kidney, heart, lung, skin), *P. muralis* (bladder, liver, skeletal muscle, stomach, skeletal muscle, kidney, lung, heart), and *P. hispanicus* (skin, stomach, skeletal muscle, brain, lung, heart) and 50 ng of r-jerdostatin were analyzed in a 10% Tris-tricine SDS-PAGE gel under reducing conditions. Replicate gels were i) stained with Coomasiee Blue R250 and ii) electrotrasferred to PVDF membrane (Hybond-P, GE Healthcare) using a semi-dry electrotransfer device. PVDF membranes were blocked in 5% (w/v) non-fat dried milk in PBS (20 mM Na_2HPO_4, 150 mM NaCl, pH 7.5) overnight at 4˚C, and incubated for 1 h at 25˚C with a 1:500 (v/v) dilution of anti-PEP160 polyclonal antibodies (against jerdostatin, see in detail below), in 5% non-fat dried milk/PBS-Tween (20 mM Na_2HPO_4 pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20). The membranes were then washed three times with PBS-Tween followed by incubation (1 h at 25˚C) with a 1:5000 (v/v) dilution of peroxidase-conjugated anti-rabbit IgG (Sigma) in PBS/5% non-fat dried milk. After 3 times washing with PBS-Tween, the membranes were developed using the chemiluminescence ECL Prime kit (GE Healthcare).

Another set of total protein extracts were fractionated into ≥ 10 kDa, 10-3 kDa, and ≤ 3 kDa fractions using Microcon® YM-10K (Amicon Bioseparations, Millipore) and Amicon Ultracel®-3K (Merck Millipore Ltd.) centrifugal filters. These fractions were run in a 10%Tris-tricine–SDS-PAGE gel, electroblotted onto PVDF membrane, and the blots developed as described above.

Recombinant soluble r-jerdostatin was produced in transformed *BL21 Escherichia coli* and isolated as described in Gema Bolás dissertation (and Sanz-Soler et al. 2012). Antiserum PEP160 against the C-terminal tail of jerdostatin (^{35}CKPSYPGNG^{43}) was

generated in rabbit by Abintek Biopharma, S.L. (Parque Tecnológico de Bizkaia, Derio, Bizkaia, Spain) using a standard immunization protocol and the synthetic peptide CKPSYPGNG conjugated to keyhole limpet hemocyanin (KLH) as immunogen. Anti-CKPSYPGNG antibodies were affinity-purified on a peptide-Sepharose column (Juárez et al. 2010).

Scheme 1 summarizes and introduces the various methodological approaches (described in detail earlier) employed in this study to address specific questions about the evolution and possible function(s) of the *RPTLN* genes (Chapter II).



Scheme 1. Summary of the various methodological approaches (described in detail in the Methodology section) employed in this study to address specific questions about the evolution and possible structure-function correlations of the reptile-specific *RPTLN* genes.

8 Generation of expression plasmids for recombinant disintegrin production.

8.1 Design and cloning of recombinant disintegrins

8.1.1 Design and cloning of recombinant ocellatusin

The short disintegrin wild-type RGD-ocellatusin was PCR-amplified from *E. ocellatus* venom gland cDNA (Juárez et al. 2006a) using forward, BglII_TEV_ocellatusin and reverse primer, HindIII_Ocellatusin, including respectively, BglII and HindIII restriction sites. The sequence coding the TEV protease cleavage site [ENLYFQG] was

included in the forward primer, and a stop codon in the reverse primer (Table 3). The PCR-amplification protocol included initial denaturation at 94 °C for 2 min, followed by 35 cycles of denaturation (94°C for 30 s), annealing (60°C for 30 s), extension (72°C for 30 s) and a final extension step for 7 min at 72°C. The resulting PCR product was purified, cloned in the vector pGEM®-T, (Promega, Madison, WI, USA), selected, confirmed and sequenced, as mentioned above.

Table 3. Primers for ocellatusin and crotatroxin cloning.

Primers	Sequence
Fw_*BglII*_**TEV**_ocellatusin	GG*AGATCT*C**GAGAATCTTTACTTCCAAGGA**GACTGTGAATCTGGACC
Rv_*HindIII*_ocellatusin	GT*AAGCTT*CTACGGATCATGTTCGCCTTTG
Fw_disintegrinUV*	GAGGTGGGAGAAGAWTGYGACTG
AUAP (3´RACE Invitrogen)	GGCCACGCGTCGACTAGTAC
Fw_*BglII*_**TEV**_crotatroxin2	GG*AGATCT*C**GAGAATCTTTACTTCCAAGGA**GGAGAAGAATGTGACTGTGG
Rv_*EcoRI*_crotatroxin2	GT*GAATTC*TTAGCCATAGAGGCCATTTCTGGG

***Modahl et al., in press**

8.2 Determination of the crotatroxin nucleotide sequence and cloning of crotatroxin for its recombinant expression

8.2.1 Cloning and sequence identification of crotatroxin-coding DNA

For the crotatroxin-coding DNA sequences determination, PCR-amplification was performed, using our *Crotalus atrox* venom cDNA library, and the forward degenerated primer, disintegrinUV, coding for the N-terminal disintegrin sequences and the reverse primer AUAP (Abridged Universal Amplification Primer) from the 3´RACE Invitrogen kit. AUAP primer hybridized onto the oligo(dT_{16}) containing the adapter primer, with which the cDNA library was generated (Modahl et al., in press). HiFidelity DNA polymerase master mix (Invitrogen) was used for that PCR-amplification. Touchdown 58-55 PCR protocol included an initial denaturation step at (94°C for 1 min 30sec) followed by 10 cycles of denaturation (30 s at 94°C), annealing (45 s at 58°C), and extension (1min at 68°C); 25 cycles starting with the above conditions, decreasing the annealing temperature by 0.12°C (reaching 55°C in cycle 25); and a final extension for 5 min at 68°C. PCR-amplification product was purified by PCR Clean-up system (Promega) and cloned in pGEM®-T vector, (Promega, Madison, WI, USA). Positive

clones were sequenced and compared with the known crotatroxin2 amino acid sequence [AAB24808.1] (Modahl et al., in press).

8.2.2 Design and cloning of recombinant crotatroxin sequence

Equal crotatroxin PCR-amplification mixture and protocol (Touchdown 58-55) was followed, using *Crotalus atrox* venom cDNA and Fw_BglII_TEV_crotatroxin2 and Rv_EcoRI_crotatroxin2 specific primers (Table 3), which were designed based in our crotatroxin2 nucleotide sequence, including BglII and EcoRI restriction sites, respectively, in addition to TEV cleavage site, similar to ocellatusin construct. PCR resulting product was cloned in pGEM®-T vector (Promega, Madison, WI, USA), as shown earlier in section 3 of Methodology.

8.3 Cloning TEVcleavage site and disintegrin in pET32a(+) expression vector

To construct an expression vector of disintegrin-thioredoxin-His_6 fusion protein the pGEM-T/TEV-disintegrin plasmids and the pET32a(+) vector (Novagen, Madison, WI) were respectively digested with BglII/HindIII (ocellatusin) and BglII/EcoRI (crotatroxin) restriction enzymes for 24 h at 37°C. The disintegrin fragments and the linear pET32a(+) vector were purified by agarose gel electrophoresis as mentioned previouly, and ligated with T4 DNA ligase (Invitrogen) overnight at 4°C. *E. coli* DH5a strain cells were transformed with this construct, and positive clones were confirmed and sequenced as described above (Figure 23).

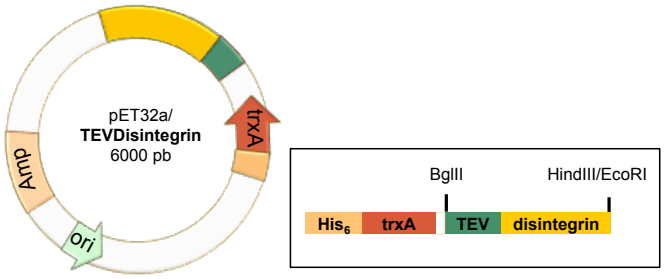


Figure 23. Expression plasmid pET32a(+)/TEV-disintegrin for expression of recombinant disintegrins and detail of the fusion protein: Histidine tail (His_6), thioredoxin A (trxA), sequence coding the TEV protease cleavage site [ENLYFQG] and sequence coding for the mature disintegrin domain. BglII, HindIII and EcoRI restriction sites are indicated. Amp denotes the sequence ampiciline restistance and ori, Origen of replication sequence.

8.4 Site-directed mutagenesis for RTS-ocellatusin generation

Hybrid ("*Frankenstein*") ocellatusin–jerdostatin constructions were generated by site-directed mutagenesis, using as template pET32a(+)/TEV-ocellatusin or new generated pET32a(+)/TEV-*Frankenstein*-disintegrin constructs. To this end, iProof™ High Fidelity Master Mix was used for PCR-amplification (denaturation at 94°C for 2 min, followed by 12 cycles of denaturation for 30 s at 94°C, annealing for 60 s at 55°C, extension for 12 min at 68°C, and a final extension for 10 min at 68°C), using different complementary primers for each mutant (Table 4). Additionally, to obtain the jerdloop-ocellatusin mutant an intermediate construct, pET32/K21W-M22R-A23T-R24S ΔG25D26-ocellatusin, was generated and served as template. Forward and reverse primers used are listed in Table 4. PCR products were treated with *DpnI* restriction enzyme, which is specific for methylated DNA, to digest the template plasmid in the PCR mix (Figure 24). These PCR products were used to transform electrocompetents *E. coli* DH5a cells. Plasmids were isolated and sequenced as described formerly, to confirm the mutations.

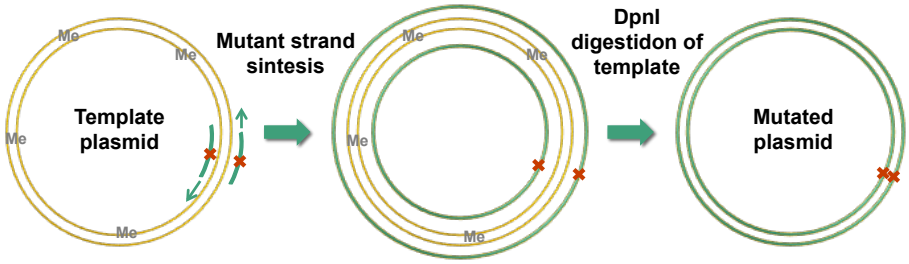


Figure 24. Simplified site-directed mutagenesis diagram. Primers and punctual mutations are display by green lines and red crosses, respectively. DNA synthesis direction is indicated. Me, show template plasmid metilation.

XGD-ocellatusin mutants (SGD, TGD and GGD) were generated using identical site-directed mutagenesis protocol (Figure 24). The primers for the generation of the mutants R24S-, R24T- and R24G-ocellatusin are noted in Table 4.

Table 4. Template plasmid and oligonucleotides used to generate *Frankenstein*- and XGD-ocellatusin mutated plasmids by site-directed mutagenesis

Mutant plasmid	Template plasmid	Oligonucleotides used for generate each mutant
pET32a(+)/TEV-ocellatusin	**pET32a(+)/TEV-ocellatusin**	
M22R-A23T-R24S	Wild-type	Fw-AGGAACAATATGCAAGAGGACAAGCGGTGATAACATGC Rv-GCATGTTATCACCGCTTGTCCTCTTGCATATTGTTCC
M22R-A23T-R24S-G25_D26del	M22R-A23T-R24S	Fw-CAATATGCAAGAGGACAAGCAACATGCATGATTACTGC Rv-GCAGTAATCATGCATGTTGCTTGTCCTCTTGCATATTG
A23R-R24T-G25S	Wild-type	Fw-GGAACAATATGCAAGATGAGAACGAGTGATAACATGCATG Rv-CATGCATGTTATCACTCGTTCTCATCTTGCATATTGTTCC
G25T-D26S	Wild-type	Fw-ATGCAAGATGGCAAGGACTAGTAACATGCATGATTACTGC Rv-TGCAGTAATCATGCATGTTACTAGTCCTTGCCATCTTGC
G25R-D26T-N27S	Wild-type	Fw-TGCAAGATGGCAAGGCGTACTAGCATGCATGATTACTG Rv-AGTAATCATGCATGCTAGTACGCCTTGCCATCTTGC
D26R-N27T-M28S	Wild-type	Fw-ATGGCAAGGGGTCGTACCAGCCATGATTACTG Rv-AGTAATCATGGCTGGTACGACCCCTTGCCATC
K21W-M22R-A23T-R24S ΔG25D26 *	M22R-A23T-R24S ΔG25D26	Fw-TGAAGGAAGGAACAATATGCTGGAGGACAAGCAACATGCATG Rv-CATGCATGTTGCTTGTCCTCCAGCATATTGTTCCTTCCTTCA
jerdloop	K21W-M22R-A23T-R24S ΔG25D26	Fw-TGCTGGAGGACAAGCGTCAGCAGTCATTACTGCAATGG Rv-CCATTGCAGTAATGACTGCTGACGCTTGTCCTCCAGCA
E47_P50del	Wild-type	Fw-CAGAAATCCTTACAAAGGCTAGAAGCTTGCGGCCGCACTC Rv-GAGTGCGGCCGCAAGCTTCTAGCCTTTGTAAGGATTTCTG
M22R-A23T-R24S-E47_P50del	M22R-A23T-R24S	Fw_CAGAAATCCTTACAAAGGCTAGAAGCTTGCGGCCGCACTC Rv-GAGTGCGGCCGCAAGCTTCTAGCCTTTGTAAGGATTTCTG
M22R-A23T-R24S-G25_D26del-E47_P50del	M22R-A23T-R24S ΔG25D26	Fw_CAGAAATCCTTACAAAGGCTAGAAGCTTGCGGCCGCACTC Rv-GAGTGCGGCCGCAAGCTTCTAGCCTTTGTAAGGATTTCTG
A23R-R24T-G25S-E47_P50del	A23R-R24T-G25S	Fw_CAGAAATCCTTACAAAGGCTAGAAGCTTGCGGCCGCACTC Rv-GAGTGCGGCCGCAAGCTTCTAGCCTTTGTAAGGATTTCTG
G25T-D26S-E47_P50del	G25T-D26S	Fw_CAGAAATCCTTACAAAGGCTAGAAGCTTGCGGCCGCACTC Rv-GAGTGCGGCCGCAAGCTTCTAGCCTTTGTAAGGATTTCTG

Mutant plasmid	Template plasmid	Oligonucleotides used for generate each mutant
pET32a(+)/TEV-ocellatusin	**pET32a(+)/TEV-ocellatusin**	
G25R-D26T-N27S-E47_P50del	G25R-D26T-N27S	Fw_CAGAAATCCTTACAAAGGCTAGAAGCTTGCGGCCGCACTC Rv-GAGTGCGGCCGCAAGCTTCTAGCCTTTGTAAGGATTTCTG
D26R-N27T-M28S-E47_P50del	D26R-N27T-M28S	Fw_CAGAAATCCTTACAAAGGCTAGAAGCTTGCGGCCGCACTC Rv-GAGTGCGGCCGCAAGCTTCTAGCCTTTGTAAGGATTTCTG
Jerdloop-E47_P50del	jerdloop	Fw_CAGAAATCCTTACAAAGGCTAGAAGCTTGCGGCCGCACTC Rv-GAGTGCGGCCGCAAGCTTCTAGCCTTTGTAAGGATTTCTG
jerdloop Y44G K45N-E47_P50del	Jerdloop-E47_P50del	Fw-GACTGTCCCAGAAATCCTGGCAATGGCTAGAAGCTTGCGG Rv-CCGCAAGCTTCTAGCCATTGCCAGGATTTCTGGGACAGTC
jerdloop R41S N42Y Y44G K45N-E47_P50del	Y44G K45N-E47_P50del	Fw-ACTTGTGACTGTCCCAGCTATCCTGGCAATGGCTAGA Rv-TCTAGCCATTGCCAGGATAGCTGGGACAGTCACAAGT
I19T jerdloop R41S N42Y Y44G K45N-E47_P50del	jerdloop R41S N42Y Y44G K45N-E47_P50del	Fw-CTGAAGGAAGGAACAACATGCTGGAGGACAAGCGTC Rv-GACGCTTGTCCTCCAGCATGTTGTTCCTTCCTTCAG
N33T jerdloop R41S N42Y Y44G K45N-E47_P50del	jerdloop R41S N42Y Y44G K45N-E47_P50del	Fw-GTCAGCAGTCATTACTGCACTGGCAAAACTTGTGACTG Rv-CAGTCACAAGTTTTGCCAGTGCAGTAATGACTGCTGAC
jerdloop-oc R41S N42Y Y44G-K45_P50del	jerdloop R41S N42Y Y44G K45N-E47_P50del	Fw-CAGTCACAAGTTTTGCCAGTGCAGTAATGACTGCTGAC Rv-GTGCGGCCGCAAGCTTCTAGCCAGGATAGCTGGGACAG
I19T jerdloop-oc R41S N42Y Y44G-K45_P50del	I19T jerdloop R41S N42Y Y44G K45N-E47_P50del	Fw- CTGAAGGAAGGAACAACATGCTGGAGGACAAGCGTC Rv-GACGCTTGTCCTCCAGCATGTTGTTCCTTCCTTCAG
R24S	Wild-type	Fw-CAATATGCAAGATGGCAAGCGGTGATAACATGCATG Rv-CATGCATGTTATCACCGCTTGCCATCTTGCATATTG
R24T	Wild-type	Fw-CAATATGCAAGATGGCAACGGGTGATAACATGCATG Rv-CATGCATGTTATCACCCGTTGCCATCTTGCATATTG
R24G	Wild-type	Fw-CAATATGCAAGATGGCAGGGGGTGATAACATGCATG Rv-CATGCATGTTATCACCCCCTGCCATCTTGCATATTG

***The intermediate construction, in grey was not recombinantly expressed.**

9 Expression and purification of recombinant disintegrins

For r-jerdostatin expression and purification was followed the protocol described by Bolas, G., in her dissertation, as mentioned previously. The procedure is comparable to the disintegrin expression and purification protocol of r-ocellatusin, r-crotatroxin, *Frankenstein* ocellatusin and XGD-ocellatusin mutants, which were expressed and purified by the procedure described below.

9.1 Expression of recombinant disintegrins in E.coli

E. coli BL21 strain cells (Novagen, Madison, WI) were transfected with the corresponding pET32a(+)/TEV-disintegrin construct (Figure 23) for each recombinant disintegrin. The presence of the disintegrin–thioredoxin fusion constructs in positive clones was checked by PCR using specific primers. Positive *E. coli* BL21 clones were grown overnight at 37°C in LB medium containing 100 µg/ml ampicillin, followed by a 1:50 (v/v) dilution in the same medium until an OD_{600} of around 0.8–1 was reached. Expression of the recombinant fusion proteins was then induced by addition of isopropyl-D-thiogalactosidase (IPTG) to a final concentration of 0.75-1 mM, and incubation of the cell suspensions for 4 h at 37°C or overnight at 21°C for some of the mutants. Cells were pelleted by centrifugation (4000 for 30 min), resuspended in the same volume of 20 mM sodium phosphate, 150 mM NaCl, pH 7.4 buffer, washed twice with this buffer, and resuspended in 50 mL of 20 mM sodium phosphate, 250 mM NaCl, pH 7.4 (Figure 25).

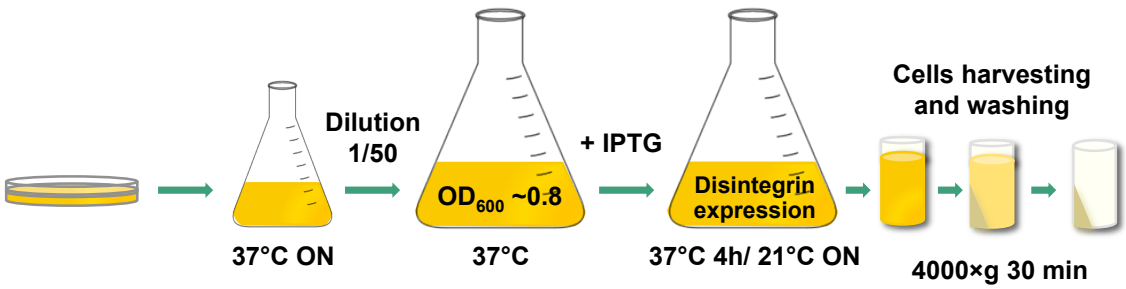


Figure 25. Simplified diagram of the recombinant disintegrins expression protocol.

9.2 Purification of recombinant disintegrins

Cells were lysed by sonication (15 cycles of 15 s sonication/1 min resting) in an ice bath. The lysates were centrifuged at 10,000×g for 30 min at 4°C twice. Soluble and insoluble fractions were analyzed by SDS-PAGE using 12% polyacrylamide gels under reducing conditions. Disintegrin-thioredoxin-His_6 fusion proteins were purified from the soluble fraction of the lysate by affinity chromatography using an ÄKTA Basic chromatograph equipped with a 5 ml HisTrap HP column (Amersham Biosciences) equilibrated in 20 mM sodium phosphate, 250 mM NaCl, pH 7.4. Bound protein was eluted with a linear gradient of 50–500 mM imidazole. Eluted fractions were checked by SDS-PAGE and those containing the fusion protein were pooled, dialyzed against 20 mM sodium phosphate, 250 mM NaCl, pH 7.4, and digested overnight at 4°C with TEV-His_6 protease (1:20, w/w) (Tropea et al. 2009). Recombinant disintegrins were separated from thioredoxin-His_6 and TEV-His_6 by chromatography on a HisTrap column (as above). The flow-through fractions were concentrated using Amicon filtration membranes with a pore size of 3000 Da (Millipore, MA, USA) and the r-disintegrins purified by reverse-phase HPLC on a C18 column (4.6×250 mm, 5 µm, SunFire™, Waters, MA, USA) equilibrated with 5% acetonitrile 0.05% TFA in water (solution A) and eluted with a 45 min linear gradient of 5–70% acetonitrile in 0.05% TFA (Figure 26). The purity of the isolated protein was assessed by SDS-PAGE and electrospray-ionization mass spectrometry. Disintegrin were lyofilized and resuspended in PBS buffer (20 mM Na_2HPO_4, 150 mM NaCl, pH 7.5). Soluble protein concentration was determined using bicinchoninic acid protein method (BCA™ Protein Assay, Pierce).

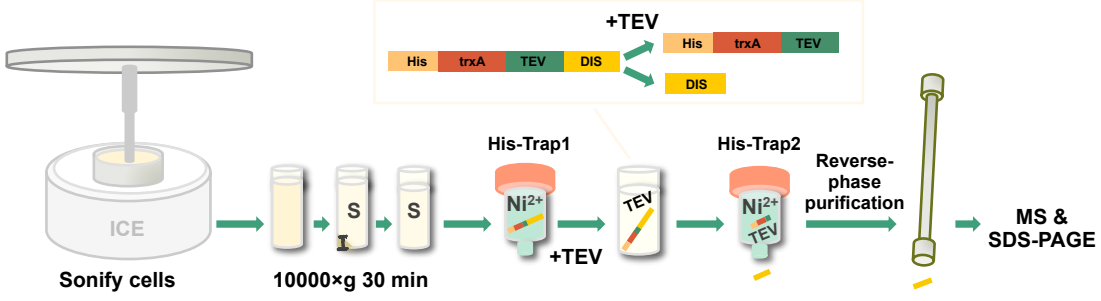


Figure 26. Simplified diagram of the recombinant disintegrin purification protocol. Soluble and insoluble fraction are displayed as S and I, respectively; TEV protease by TEV. His-Trap, denotes the purification steps of the affinity chromatography, in which His_6 tagged proteins bind Nickel Sepharose affinity column (HisTrap HP, Amersham Biosciences). MS, Mass spectrometric measurement and SDS-PAGE to test purify.

10 Isolation of snake venom proteins

10.1 Isolation of lebestatin from *Macrovipera lebetina* venom

Lebestatin was isolated from the venom of *Macrovipera lebetina* by HPLC. Two mg of crude venom were dissolved in 300μl water containing 0.1% TFA and applied to a reverse phase C18 column (250×4.6 mm, 5 μm, Europa® Teknokroma), as shown in (Makran et al. 2012). The purity of the disintegrin was assessed by SDS-PAGE and Mass spectrometry in native conditions.

10.2 Isolation of RTS-disintegrin from *Daboia russelli* venom

Two miligrams of crude venom samples were dissolve in 300μL water containing 0.1% TFA (solution A). Soluble venom proteins were separated by reverse-phase HPLC on a C18 column (250×4.6 mm, 5 μm, Europa® Teknokroma), equilibrated with 5% acetonitrile 0.05% TFA in water (solution A) and eluted with linear gradient of 5–20% (in 20min), 20%-45% (in 110min) and 45-70% (30min) acetonitrile in 0.05% TFA. Protein detection was performed at 215 nm and peaks were collected manually. Possible disintegrin peaks were used directly for mass spectrometry in native conditions and dried in a Speed-Vac (Savant), dissolved in water, and further loaded in SDS-PAGE separation in 12% gels, under reducing conditions.

11 Molecular mass determination and collision-induced fragmentation by nESI-MS/MS

11.1 Mass spectrometry (MS) in native conditions

The purity and identity of the isolated recombinant disintegrins, as well as, possible disintegrin peaks isolated from snake venom were assessed by Electrospray-ionization mass spectrometry using a QTrap 2000 instrument (Applied Biosystems) (Le Blanc et al. 2003) equipped with a nanoelectrospray source (Proxeon, Denmark). Electrospray-ionization mass spectrometric characterization was performed on the native (nonreduced) proteins, obtaining ESI-MS isotope-averaged molecular mass of HPLC-purified disintegrins (Figure 26). The protein molecular mass can be obtained determining the charge state of the peaks in the spectrum. For that, one may assume

that the two adjacent peaks are from the same compound and they differ by a single charge. If the two peaks are related and differ by one proton then more specifically we can write: *m+1=808,3×(z+1)* and *m=942,7×(z)* (example, Figure 27). Since *m* is assumed to be the same for both peaks we can set the two equations equal to each other *808,3z+807.3=942,7×z* and solve *z= 6.* Now that the charge (*z)* is known, we obtained the accurate protein mass and standard deviation (Calvete 2014) (Figure 27).

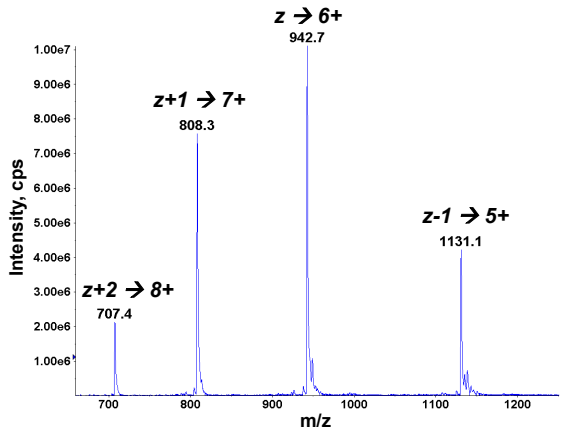


Figure 27. Electrospray-ionization mass spectra. ESI_MS m/z, mass-to-charge; M=molecular mass of the protein; Charge (z) Proton mass (H^+=1.00794u). The equation necessary to obtain the protein molecular mass is displayed.

Theoretical disintegrin monoisotopic mass with 4 disulfide bonds was calculated using Paws® software, considering native cysteine disulfide bridge formation.

11.2 In-gel enzymatic digestion and collision-induced fragmentation by nESI-MS/MS

Protein bands of interest were excised from a Coomassie Brilliant Blue-stained SDS-PAGE and subjected to automated reduction with DTT and alkylation with iodoacetamide, and in-gel digestion with sequencing grade bovine pancreas trypsin (Roche) using a ProGest digestor (Genomic Solutions) following the manufacturer's instructions.

For peptide sequencing, the protein-digested mixture was loaded in a nanospray capillary column and subjected to electrospray-ionization mass spectrometric analysis using a QTrap mass spectrometer (Applied Biosystems). Doubly- or triply- charged ions

of selected peptides from the MALDI-TOF mass fingerprint spectra were analyzed in Enhanced Resolution MS mode and the monoisotopic ions were fragmented using the Enhanced Product Ion mode with Q0 trapping and QTrap 2000 instrument (Applied Biosystems). Enhanced Resolution was performed at 250 amu/s across the entire mass range. Settings for MS/MS experiments were as follows: Q1-unit resolution; Q1-to-Q2 collision energy — 30–40 eV; Q3 entry barrier — 8 V; LIT (linear ion trap) Q3 fill time — 250 ms; and Q3 scan rate — 1000 amu/s. CID spectra were interpreted manually or using the online form of MASCOT program at www.matrixscience.com.

12 Inhibition of soluble $\alpha_1\beta_1$ integrin binding to CB3

12.1 Human $\alpha_1\beta_1$ integrin ectodomain expression and issolation

Recombinant soluble human $\alpha_1\beta_1$ integrin ectodomain was produce in transfected *Drosophila* Schneider cells. The established clone was grown in a biofermenter to a density of 12×10^6cells/ml, then, $\alpha_1\beta_1$ integrin expression was induced with 0.8 mM Cu^{2+} ions for five days. For $\alpha_1\beta_1$ integrin isolation, the cell supernatant was harvested, concentrated, and loaded onto a CB3 [IV] affinity column. Soluble $\alpha_1\beta_1$ integrin was eluted by 10 mM EDTA as previously described by Eble et al. 2006. Its concentration and purity were determined by BCA test (Pierce) and SDS-PAGE. The Collagen IV fragment CB3 was generated as described by Kern et al. 1993.

12.2 Inhibition of soluble $\alpha_1\beta_1$ integrin binding to CB3 by *Frankenstein* disintegrins

The collagen type IV fragment CB3 was immobilized on a 96-well plate overnight at 4°C in TBS/Mg^{2+} (20 mM Tris, 150 mM NaCl, pH 7.5 containing 2 mM $MgCl_2$) at a concentration of 5 µg/ml (100µL). The plate was washed three times with binding buffer (TBS/Mg^{2+}) and non-specific binding sites blocked with 1%(v/v) BSA in TBS/Mg^{2+} at room temperature for 1 h. Then, 3.5 µg/ml of soluble $\alpha_1\beta_1$ integrin, dissolved in 100µL TBS/Mg^{2+}, were mixed with 100µL of increasing concentrations of recombinant disintegrins. The mixtures were added to the plate and incubated for 2 h at room temperature. The plate was then washed twice with 50 mM HEPES (pH 7.5) containing 150 mM NaCl, 2 mM $MgCl_2$ and 1 mM $MnCl_2$, and the bound integrin was fixed with 2.5% (v/v) glutaraldehyde in the same buffer for 10 min at room temperature. For $\alpha_1\beta_1$

integrin detection a primary rabbit anti-β_1 antiserum (1:2000) was employed. After 3 times washing with PBS, goat anti-rabbit IgG conjugated with alkaline phosphatase (AP, 1:2000) was added. 4-nitrophenyl phosphate disodium salt hexahydrate (Sigma) was used as AP substrate, color was developed at room temperature and quantified in an ELISA plate reader at 405 nm. CB3 coated wells were incubated with the integrin either in the absence of disintegrin (positive control) or in the presence of 10 mM EDTA (negative control). Lebestatin purified from *Macrovipera lebetina transmediterranea* and r-jerdostatin were used as positive control for $\alpha_1\beta_1$ integrin binding inhibition.

13 Platelets aggregation assay. Inhibition of collagen I-induced platelets aggregation assay

13.1 Platelets isolation. Preparation of washed platelets suspensions

Human platelets were isolated from fresh blood from healthy volunteers not having taken any drug known to interfere with platelet responses in the previous 2 weeks. For preparation of washed platelets, blood was collected 6:1 (v/v) in ACD anticoagulant (117 mM trisodium citrate, 78 mM citric acid, 282 mM dextrose), and centrifuge at 200 ×g for 15 minutes at 37 °C obtaining platelet-rich-plasma (PRP). PRP was draw off, transfer to a clean tube and centrifuge again. Aliquots of PRP were distributes in 2 mL tubes with addition of 200 ng/mL PGE1 (Prostaglandin E1) final concentration, and centrifuge at 10,000 ×g for 60 seconds at room temperature (RT). Platelet pellets were resuspended in modified calcium-free Tyrode buffer A (134 mM NaCl, 2.9 mM KCl, 3 mM NaH_2PO_4, 1 mM $MgCl_2$, 10 mM HEPES, 5 mM dextrose, 50 ng/ml PGE1, pH 6.2) from Antunes et al. 2010. This procedure was repeated once more. Finally, platelet pellets were resuspended in Tyrode buffer B containing calcium (134 mM NaCl, 2.9 mM KCl, 3 mM NaH_2PO_4, 1 mM $MgCl_2$, 10 mM HEPES, 5 mM dextrose, 2 mM $CaCl_2$, pH 7.4) and the suspension was maintained at 37 °C. The platelet counts were determined using a Neubauer chamber and adjusted to 2 $\times 10^8$/mL for starting the assay. Modified from Antunes et al. 2010; Navdaev et al. 2011.

13.2 Platelet aggregation assay.

Human platelet aggregation was monitored, by light transmission aggregometry (LTA), in an platelet aggregometer profiler PAP-E8 (MöLab GMbH), with continuous stirring at 1100 rpm and constant temperature at 37°C in eight channels at the time. Platelets aggregation studies were performed at counts of 200,000 platelets/µL, 250 µL of those platelets in Tyrode buffer B, were pre-incubated stirring at 37 °C for 2 minutes till the signal is stable. Straightaway, platelets aggregation was induced by addition of 10 µL of 100 µg/mL Collagen I (Col I). Then, 1 minute after, disintegrins in PBS (20 mM Na_2HPO_4, 150 mM NaCl, pH 7.5), or negative control (PBS), were added. Platelets aggregation reaction was allowed to proceed for at least 5 min. r-ocellatusin wild-type was our positive control and PBS maximum aggregation (MA) was used as a reference of 100% platelets aggregation. The results were expressed in percentage of platelet aggregation through optical measurements of turbidity (Figure 28). Platelets in Tyrode buffer B was used as blank in the tests with washed platelets suspension.

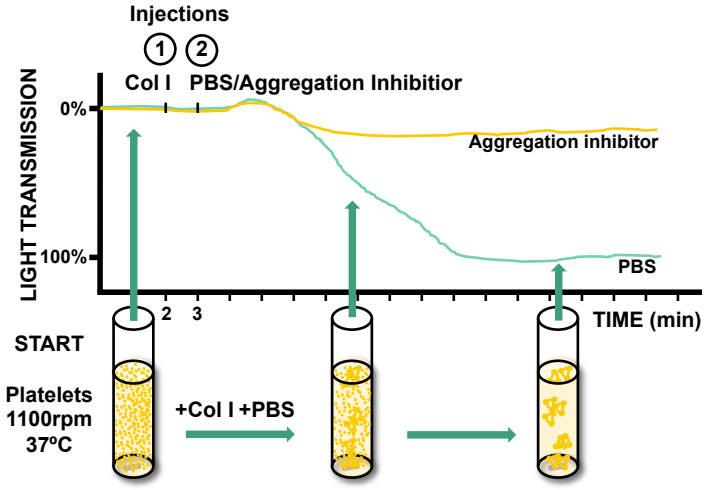


Figure 28. Platelets aggregation cartoon and its profile of aggregation, together with the profile of platelets aggregation inhibition by an aggregation inhibitor (e.g. RGD-ocellatusin). Platelet aggregation profile with (10µl ocellatusin) and without (10µl PBS buffer) inhibitor are indicated in yellow and green, respectively. The simplified cartoon presents the process of collagen I-induced platelets aggregation, and each tube indicates different points of the PBS platelet aggregation profile. Col I, denotes 10µL of 100µg/mL Collagen I. PBS buffer (20 mM Na_2HPO_4, 150 mM NaCl, pH7.5)

In our studies we used Maximum Aggregation to compare different concentrations and different disintegrins inhibitory potencies. The percentage of platelet aggregation was recorded and used to calculate the concentration necessary to reduce the induced platelet aggregation by 50% with respect to control (IC_{50} value).

14 Rattlesnakes behavior trials

14.1 Experimental animals. *Crotalus atrox* and *Mus musculus*

Behavioral trials were performed as approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Northern Colorado. Eight *Crotalus atrox*, all adult long-term captive snakes, were housed individually in glass aquaria (61.0 × 41.0 × 44.5 cm) containing a paper floor, water bowls and hide boxes, and maintained on a 12:12 L:D (Light:Dark) cycle at 26 ± 2°C. All snakes were healthy and fed bi-weekly on pre-killed inbred Swiss/Webster mice (*Mus musculus*). Snakes were never fed on the day of trials, which occurred at least 7 to 10 days after the last feeding session, and all behavioral trials were separated by at least 14 days. On testing days, similar size and sex mice were selected and euthanized by cervical dislocation just before r-disintegrin injection and subsequent testing. Although *M. musculus* is not a natural prey item of *C. atrox*, it has been shown that the magnitude of the snakes chemosensory response, termed strike-induced chemosensory searching (SICS), is no different towards natural rodent prey such as *Peromyscus maniculatus* (deer mice) when compared to lab mice (*M. musculus*) (Furry et al. 1991). Therefore, the strain of laboratory mice used will not influence the results.

14.2 *Crotalus atrox* behavior trials using recombinant disintegrins

For behavioral trials, strike-induced chemosensory searching was induced by allowing *C. atrox* to strike and envenomate a prey carcass suspended from long forceps (Chiszar et al. 1992). Since rattlesnakes release prey immediately after the strike, the envenomated mouse carcass was removed without ever touching the floor or walls of the snakes cage, and discarded. The testing apparatus consisted of a flat 4 × 10 cm metal base with two wire mesh baskets approximately 4.0 cm apart. One wire mesh basket

contained the "envenomated" mouse carcass which was injected with 1.5mg in 100 µL of reconstituted recombinant disintegrin in PBS, whereas the other basket held the "non-envenomated" (control) mouse, injected with 100 µL of PBS (20 mM Na_2HPO_4, 150 mM NaCl, pH 7.5) (Chiszar et al. 1999; 2008). We used the terminology "envenomated" (E) and "non-envenomated" (NE) mice according to previous behavioral studies (Saviola et al. 2013). In addition, the 100 µL volume of reconstituted venom is comparable to the volume of venom injected during a predatory strike (Hayes et al. 1992). Two injections, each containing 50 µL, were made in the thoracic region, dorsal and ventral to the shoulder blade, as these are the regions most commonly struck during rattlesnake predatory episodes (Kardong 1986). The testing apparatus containing both the E and NE mice was placed in the opposite end of the snake cage, and to account for human disturbance of opening and closing the cage, a 2-minute undisturbed acclimation period occurred before the start of the trial. It should be noted, that during this acclimation period, none of the snakes approached the testing apparatus containing the E and NE carcasses. Following this 2-minute acclimation period, 10 minute trials were recorded on a JVC Everio GZ-MG330 30GB HDD Camcorder with a 35x optical zoom lens. Videos were analyzed and data was recorded as the number of tongue flicks directed within 1 cm of either the E or the NE mouse. In squamates, tongue flicking represents a stimulus-seeking behavior and is the main process for delivering volatile and non-volatile cues to the vomeronasal organ (Schwenk 1995). Since tongue flicking is activated by the detection of volatile cues by the nasal olfactory system, or visual, thermal or vibratory stimuli, measuring the rate of tongue flicking is an accurate and convenient assay of nasal as well as vomeronasal chemoreception in snakes (Schwenk 1995; Saviola et al. 2012). Data was recorded with the observer blind to the conditions, therefore, the observer was unaware of which mouse carcass was injected with the control or recombinant disintegrin sample. Cages and test apparatus were thoroughly cleaned with Quatricide® (Pharmacal) and 70% ethanol between trials.

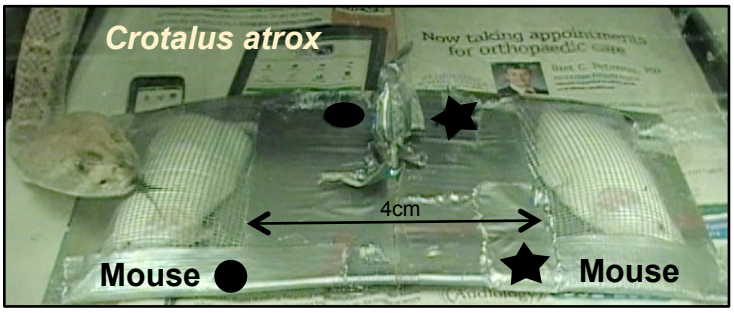
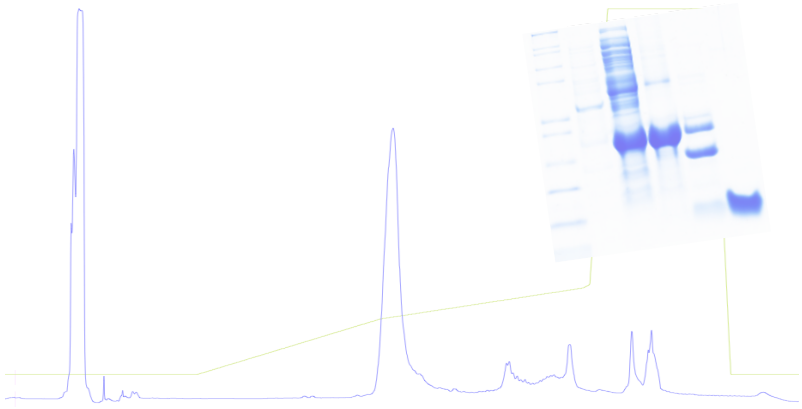


Figure 29. Testing apparatus containing envenomated (E) and non-envenomated (NE) mice. E and NE mice were injected with r-disintegrin and PBS (20 mM Na_2HPO_4, 150 mM NaCl, pH 7.5), respectively. The mice were denoted as circle and star as the data were recorded with the observer blind to the conditions. In the picture, *Crotalus atrox* flicks towards Mouse "circle". The picture was extracted from one of the videos analyzed in this dissertation.

14.3 Tongue flicks statistics analyses

The mean number of tongue flicks directed towards the E and NE mouse carcasses for the recombinant disintegrin and the PBS control were analyzed using a two-sample t-test. The number of tongue flicks was also converted to percentages (that is, percent of tongue flicks emitted to E and NE mice) by dividing the number of tongue flicks aimed at the E carcass by the total number of tongue flicks for both carcasses. These data were then analyzed by single sample t-tests in which the mean percent of tongue flicks directed toward E mice were compared to 50%, as 50% is the expected value of tongue flicks directed towards each of the E and NE mice under the null hypothesis. Rate of tongue flicking can be highly variable among snakes, so converting rate of tongue flicking to percentages places all snakes on the same scale (Chiszar et al., 2008; Saviola et al., 2013).

Results & Discussion



Chapter I. Recombinant expression and functionality of the *"Frankenstein"* [1] and XGD ocellatusin mutants

[1]The term *Frankenstein* used here as a synonym for "a chimeric protein made with pieces from different molecules", is a tribute to Mary W. Shelley's "*Frankenstein*, or, the Modern Prometheus", Lackington, Hughes, Harding, Mavor & Jones, Gradifco, Switzerland, 1818.

It is known that short RGD-disintegrins appear to be restricted to African and Asian *Echis* and *Eristicophis* species, and represent the most recent members of the disintegrin family (Juárez et al. 2008; Calvete 2010). This also suggest that short RGD-disintegrins may have evolved after the radiation of Viperinae during the late Oligocene or the early Miocene, between 22 and 24 Mya (Castoe et al. 2009), at a time when eastern North America and Eurasia were widely separated across the Atlantic, whereas northeastern Asia and Alaska remained connected via the Bering land bridge. The existence of two distinct messengers coding for the short disintegrin ocellatusin (Figure 18 p.30) suggest key events of the evolutionary emergence of the short disintegrin ocellatusin from a short-coding dimeric disintegrin precursor genes (Juárez et al. 2006b).

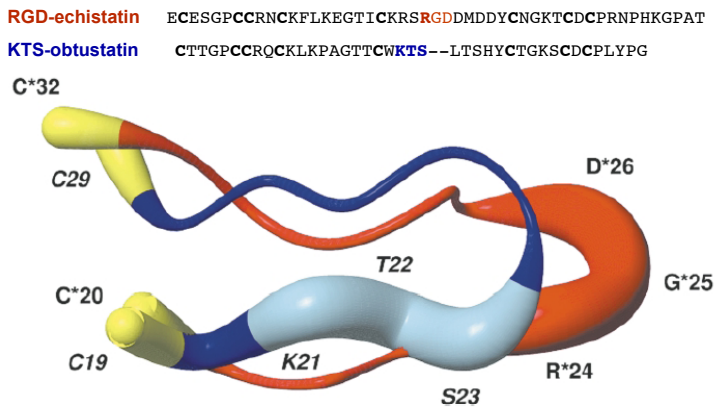


Figure 30. Superposition of the integrin-binding loops of echistatin (red) and obtustatin (blue), highlighting their different size and the different topology of their active motifs, RGD and KTS. Adapted from Moreno-Murciano et al. 2003a.

Furthermore, the most parsimonious nucleotide substitution model required for the emergence of all known XXD disintegrin's integrin inhibitory sequences from an ancestral RGD motif (Juárez et al. 2008) involves a minimum of three mutations (Calvete 2010) (Figure 14 p.26). However, circumstantial evidence suggests that KTS/RTS disintegrins may not follow this canonical evolutionary scenario. Hence, i) no putative dimeric disintegrin precursor has been found in the few species from which KTS/RTS-disintegrins have been isolated from their venoms or cloned from their cDNA venom glands in the species shown in Table 5, ii) whereas the integrin-inhibitory loops of XXD disintegrins are absolutely conserved in residue length (11 amino acids) and harbor the active tripeptide at the tip. However, the active tripeptides of jerdostatin (RTS, Carbajo et al. 2011) and obtustatin (KTS, Monleón et al. 2003; Moreno-Murciano et al. 2003b) are oriented towards the side of nine-residue integrin-binding loops. (Figure 30).

Table 5. Existing KTS/RTS-disintegrins released in snake venoms, and coding DNA cloned from their venom gland cDNA libraries.

Snake species	KTS/RTS-disintegrin	Accession code	References
Macrovipera lebetina M.mauritanica	**lebestatin (KTS)**	Q3BK14.1	(Olfa et al. 2005; Makran et al. 2012))
M. lebetina obtusa	**obtustatin (KTS)**	P83469.1	(Marcinkiewicz et al. 2003; Moreno-Murciano et al. 2003a)
Daboia paleastinae	**viperistatin (KTS)**	P0C6E2.1	(Kisiel et al. 2004)
Daboia russelii	**russellistatin (RTS)**	Q7ZZM2.1	**At this dissertation (Chapter II)** (Sanz-Soler et al. 2012)
Protobothrops jerdonii	**cDNA RTS**	Q7ZZM2.1	(Sanz et al. 2005)
Daboia russelii	**cDNA RTS**	FF277034.1	Unpublished data
Macrovipera mauritanica	**cDNA RTS**	AM261813.1	(Bazaa et al. 2007)
Echis ocellatus	**cDNA RTS**	AM286798.1	(Bazaa et al. 2007)
Cerastes vipera	**cDNA RTS**	AM114012.1	(Sanz et al. 2006)

Disintegrin´s KTS and RTS integrin-binding motifs are indicated. cDNA RTS denotes the cloned sequences from the snake venom glands.

The striking similar functional requirements (synergy between the integrin binding loop and the C-terminal tail) and structural differences between RGD and RTS/KTS short disintegrins prompted us to investigate a possible transformation route of RGD-ocellatusin into RTS-jerdostatin.

To elucidate the RGD-disintegrin and KTS/RTS-disintegrin evolutionary correlation, minimal requirements to transform RGD-disintegrin in KTS/RTS-disintegrin were studied. For that matter, RGD-ocellatusin and RTS-jerdostatin were selected, expressed and purified.

1 Expression, purification and identification of recombinant disintegrins

1.1 Cloning of RGD ocellatusin

Firstly, in order to express a short RGD disintegrin, the complete disintegrin domain, wild-type RGD-ocellatusin (Figure 31) nucleotide sequence, together with the sequence coding for the cleavage site for the tobacco etch virus (TEV) protease in 5´, was cloned in the pET32a(+) expression vector (Novagen, Madison, WI) (Figure 31 and Figure 23, p.50).

Figure 31. Cloned ocellatusin nucleotide and amino acid sequence and TEV (tobacco etch virus) cleaveage recognizing site in N-terminal. The exact cleavage site is displayed.

Once we created this pET32a(+)/TEV-ocellatusin expression construct, we designed ocellatusin mutants by site-directed mutagenesis (Figure 24, p.51), as it is explained below.

1.2 Expression of recombinant disintegrins

The different recombinant disintegrins were overexpressed in *BL21*(DE3) expression strain transformed with the pET32a(+)/TEV-disintegrin plasmids. Every individual *BL21*_pET32a(+)/TEV-disintegrin strain overexpressed soluble His_6-thioredoxin-TEV-disintegrin fusion protein (Figure 32A). Co-expressed thioredoxin in tandem with the disintegrins catalyze a proper cystine linkage, challenging the reducing conditions in the bacterial cytoplasm, and assisting a correct disintegrins folding.

Recombinant disintegrin purification method consists in two-steps of affinity chromatography, on a His-Trap column, including in-between, His_6-trxA-TEV-disintegrin digestion by the r-TEV-protease (Tropea et al. 2009) and followed by a reverse-phase chromatography (Figure 32). Purification yields of r-disintegrins were approximately 0.5-1 mg/L of cell culture.

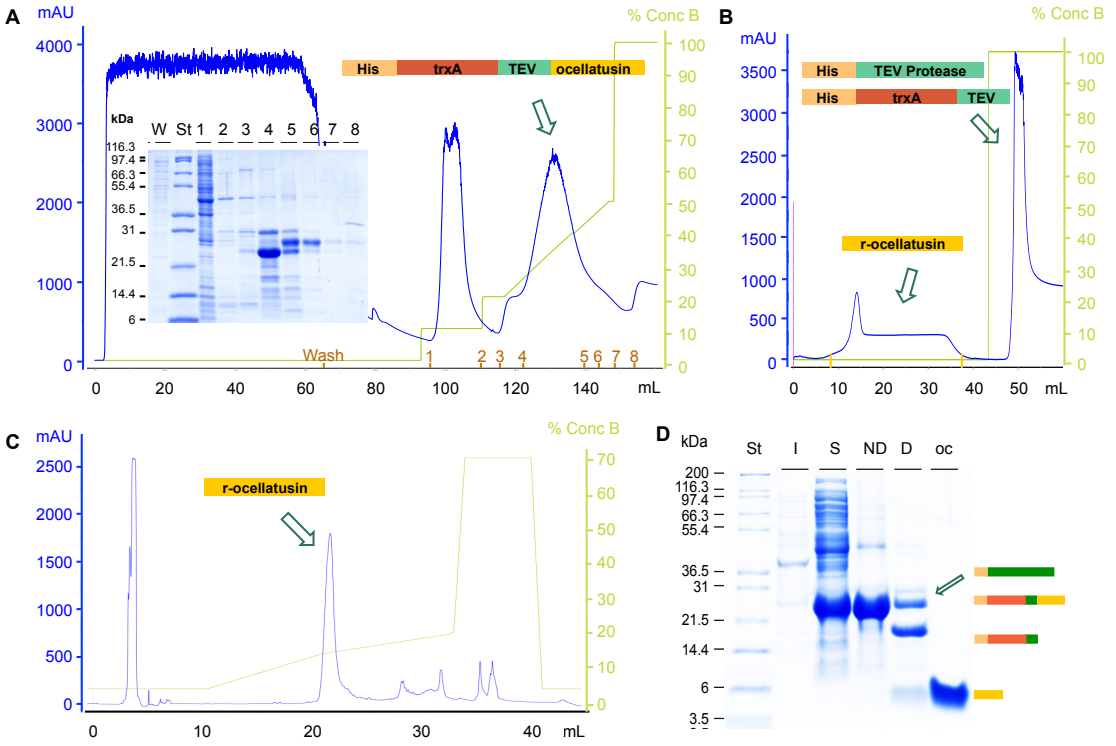


Figure 32. Purification of recombinant disintegrins. r-ocellatusin purification steps. A. Soluble fraction of lysated of *E.coli* BL21 cells expressing the His_6-thioredoxin-TEV-disintegrin fusion protein was loaded in the first step of His-Trap affinity chromatography. To identify the fraction containing our construct, eluted fractions were loaded in 12% SDS-PAGE. An arrow points the fusion protein fraction. B. Second step of His-Trap affinity chromatography after the digestion of the construct (fraction 4) by r-TEV-protease. The retained and non-retained protein identities are specified. The flow through fraction, containing the recombinant disintegrin was concentrated. C. The concentrate fraction was injected in a reverse-phase chromatography column. The purified r-disintegrin peak is denoted. D. Analysis by Tris–Tricine–(10%)SDS-PAGE of the overexpression and purification steps of ocellatusin. Lanes I and S, insoluble and soluble fractions, respectively, of lysates of *E. coli* BL21 cells expressing the disintegrin fusion protein. Lane ND (Non-digested), HisTrap affinity-purified disintegrin-thioredoxin-His_6 (24 kDa). Lane D, digestion products of ocellatusin-TEV-thioredoxin-His_6 fusion protein after incubation with TEV protease, plus His_6-TEV protease (28,6 kDa). Lane oc, r-ocellatusin purified by reverse-phase HPLC from the flow-through of the HisTrap affinity column of the protein mixture shown in lane D. Lanes S, molecular weight markers (Mark12™, Invitrogen), whose apparent molecular mass is indicated at the left side of the gels.

1.3 Disintegrin identification by MS spectrometry

The purity of the isolated recombinant disintegrins was assessed by Tris-tricine-SDS-PAGE (Figure 32 D). Then, recombinant disintegrins were identified by electrospray-ionization mass spectrometry (Figure 33).

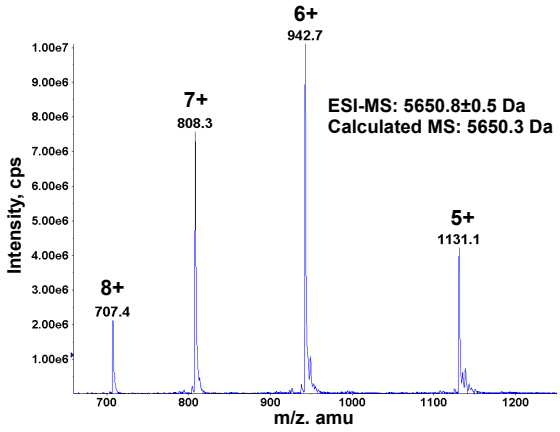


Figure 33. Electrospray-ionization mass spectrometry of reverse-phase HPLC-purified recombinant wild-type ocellatusin. ESI-MS is the experimental mass and calculated molecular mass corresponds to fully oxidized (4 disulfide bonds) monoisotopic species.

Mass spectrometry is perhaps the best-suited technique for counting cysteine residues, discriminating free cysteine residues (SH) and disulfide bonds (S-S) (Calvete et al. 2007a). Accordingly, electrospray-ionization mass spectrometry proved that our experimental molecular masses accurately matched the calculated masses for the recombinant disintegrins with fully oxidized cysteine (S-S) residues content. These data confirmed the correct r-disintegrins primary sequence (Figure 33) (Table 6 and 7).

1.4 Recombinant RGD-disintegrins functionality. Inhibition of collagen-induced platelet aggregation by recombinant ocellatusin

Short RGD disintegrins bind different kinds of integrins, blocking the binding of the $\alpha_5\beta_1$, $\alpha_V\beta_3$, $\alpha_{IIb}\beta_3$ integrins to their natural ligands. These major integrin ($\alpha_5\beta_1$, $\alpha_V\beta_3$ and $\alpha_{IIb}\beta_3$) ligands are fibronectin, vitronectin and platelet fibrinogen, respectively. Integrin $\alpha_{IIb}\beta_3$ is implicated in platelet aggregation, therefore, one common method to study and compare different activities of the RGD-disintegrins is measuring the inhibition of the platelet aggregation induced by ADP or Collagen I (McLane et al. 1996;

Sanchez et al. 2009; Yang et al. 2015). Accordingly, to test our recombinant ocellatusin functionality, inhibition of platelet aggregation induced by collagen I (Col I) was measured (Figure 34).

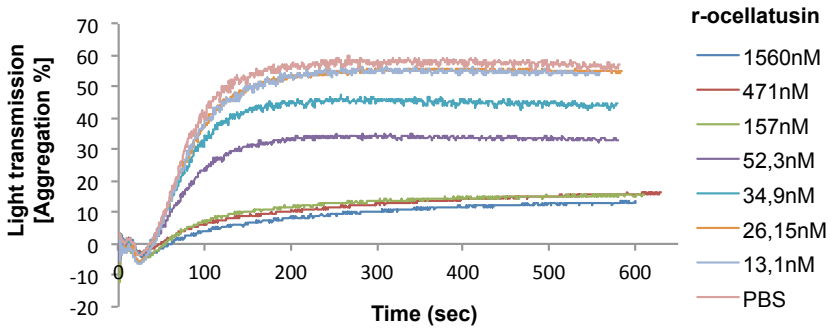


Figure 34. Inhibitory activity of recombinant ocellatusin. Concentration-dependent blocking of collagen-induced platelet aggregation by r-ocellatusin.

Recombinant ocellatusin inhibited the collagen I-induced aggregation of human washed platelets, in a dose dependent manner, with an IC_{50} of 3.6 $\pm 0.6\times10^{-8}$ M (Figure 34). Thus, evidences that r-ocellatusin inhibited platelet aggregation (collagen I-induced) with a potency similar to that reported for the natural short RGD-disintegrin echistatin (IC_{50} of 3 $\times10^{-8}$M) (Gan et al. 1988). Furthermore, r-ocellatusin has similar potency to natural ocellatusin, which has an ADP-induced platelet rich plasma aggregation IC_{50} of 16.8 $\pm2\times10^{-8}$ M, which is similar to the reported for echistatin, 13.6 $\pm$ 2.9×10^{-8} M (Smith et al. 2002). These result strongly suggests that the recombinant disintegrin may have been folded into the same biologically-active conformation as the venom-isolated ocellatusin and its homologue echistatin (1RO3, Monleón et al. 2005).

2 Design and generation of recombinant RTS-ocellatusin mutants

To approach the evolutionary path of XXD *versus* KTS/RTS-disintegrins, structure-function analyses were performed. Considering that it might exist a possible evolutionary pathway from XXD-disintegrins towards KTS/RTS-disintegrins, we explored the minimal requirements to transform the RGD-ocellatusin into the $\alpha_1\beta_1$-blocking RTS-jerdostatin.

For that aim, recombinant *"Frankenstein"* disintegrins (hybrid ocellatusin-jerdostatin constructs) were designed, generated and functionally tested. "*Frankenstein"* RTS-ocellatusin constructs were generated by site-directed mutagenesis using as a template the wild-type expression vector, pET32a(+)/TEV-ocellatusin, or some of the generated mutants (Table 4, p.52 and 53) (Figure 35).

2.1 Generation of recombinant loop-mutated RTS-ocellatusin disintegrins

The disintegrin active sites have been significantly studied (reviewed by Calvete 2013b), and the integrin-binding loop is the first determinant of the epitope that provides the disintegrin functionality. Therefore, the RGD integrin-binding motif was substituted for the RTS sequence (r-R24-G25T-D26S-ocellatusin) in the tip of the loop, (as is located in the ocellatusin). In addition, mutants r-D26R-N27T-M28S-ocellatusin, r-G25R-D26T-N27S-ocellatusin, r-A23R-R24T-G25S-ocellatusin and r-M22R-A23T-R24S-ocellatusin were designed to vary the topology of the engineered RTS motif from the tip to the loop towards lateral position, as occurs in the jerdostatin loop (Figures 30 and 35). The 11 amino acid length of the integrin-binding loop of ocellatusin was preserved in these five *Frankenstein* mutants, since our aim was to discover the minimal necessary substitutions to transform ocellatusin in to an active RTS-disintegrin. Besides, we generated the mutant, r-M22R-A23T-R24S-G25_D26del-ocellatusin, which was designed to shorten the length of the integrin-binding loop from 11 to 9 residues (as in jerdostatin) while maintaining the lateral topology of the RTS motif (Figures 30 and 35).

In addition, the full sequence of the inhibitory loop of ocellatusin was replaced by the complete amino acid sequence of jerdostatin´s integrin-binding loop in the r-jerdloop-ocellatusin mutant [^{20}**C***W*RTS--*VSSHY***C**^{32}] (Figure 35). r-ocellatusin mutants containing this sequence in the disintegrin loop are labeled as "jerdloop" to simplify terminology.

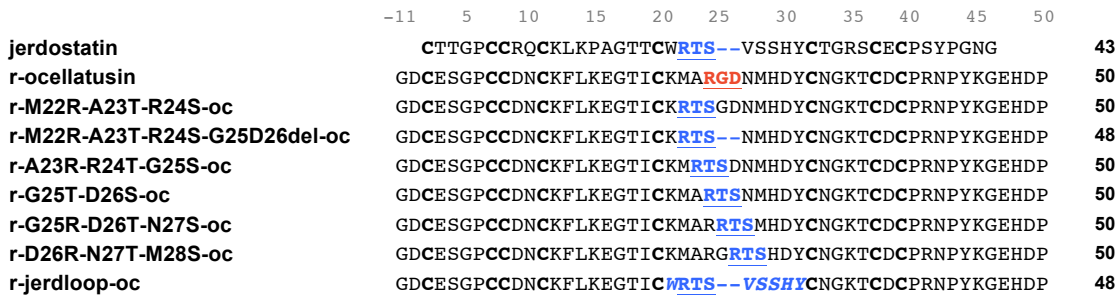


Figure 35. Amino acid sequences and nomenclature of wild-type jerdostatin, wild-type r-ocellatusin and the "*Frankenstein*" r-RTS-ocellatusin mutants. Cysteine residues are in bold, and the RTS and RGD motifs are underlined and denoted in blue and red, respectively, according to the figure 30. The substitutions in the mutants are in indicated in italics and colored in blue, according to jerdostatin assigned color. Residue labeled -1 corresponds to the last residue of TEV protease cleavage site, ENLYFQG, inserted between the His_6-thioredoxin tag and the disintegrin sequence. The length, in amino acids, of each disintegrin is indicated.

Each of these r-RTS-ocellatusin mutants were expressed, purified, and isolated in a soluble form, and identified by SDS-PAGE and mass spectrometry, similarly to the wild type r-ocellatusin (Table 6) (Figure 32 and Figures 25 and 26, p.54 and 55).

Table 6. Experimental (ESI-MS) and calculated molecular masses of the recombinant disintegrins listed in Figure 35. Calculated masses correspond to fully oxidized (4 disulfide bonds) monoisotopic species.

Recombinant disintegrins	ESI-MS (Da)	Calculated monoisotopic mass (Da) with 4 S-S bonds
r-jerdostatin	4764.7±0.5	4765.4
r-ocellatusin	5650.8±0.5	5650.3
r-M22R-A23T-R24S-ocellatusin	5636.7±1.5	5636.2
r-M22R-A23T-R24S ΔG25D26-ocellatusin	5465.0±0.7	5464.1
r-A23R-R24T-G25S-ocellatusin	5710.9±0.7	5710.3
r-G25T-D26S-ocellatusin	5666.7±1.3	5666.3
r-G25R-D26T-N27S-ocellatusin	5708.6±0.2	5708.4
r-D26R-N27T-M28S-ocellatusin	5634.9±0.3	5634.3
r-jerdloop-ocellatusin	5435.1±1.0	5435.0

3 Inhibition of soluble $\alpha_1\beta_1$ integrin-binding to CB3 by wild-type and r-RTS-ocellatusin mutants

In contrast to the short RGD disintegrins, KTS/RTS-disintegrins uniquely inhibit the binding of integrin $\alpha_1\beta_1$ to its natural ligands, the CB3 fragment of collagen IV, being the most specific ligand, and also collagen I (Figure 20, p.32). In agreement with previous studies (Juárez et al. 2010), both r-jerdostatin and lebestatin (Olfa et al. 2005) (our positive controls) blocked the high-affinity interaction between the soluble heterodimeric ectodomain of integrin $\alpha_1\beta_1$ and the CB3 fragment of collagen IV in a concentration-dependent and divalent ion-independent manner (Figure 36). In contrast, wild-type r-ocellatusin, our negative control, did not show any inhibitory activity, as expected. Similarly, none of the *Frankenstein* r-RTS-disintegrins blocked the binding of the $\alpha_1\beta_1$ integrin to the CB3 fragment (Figure 36).

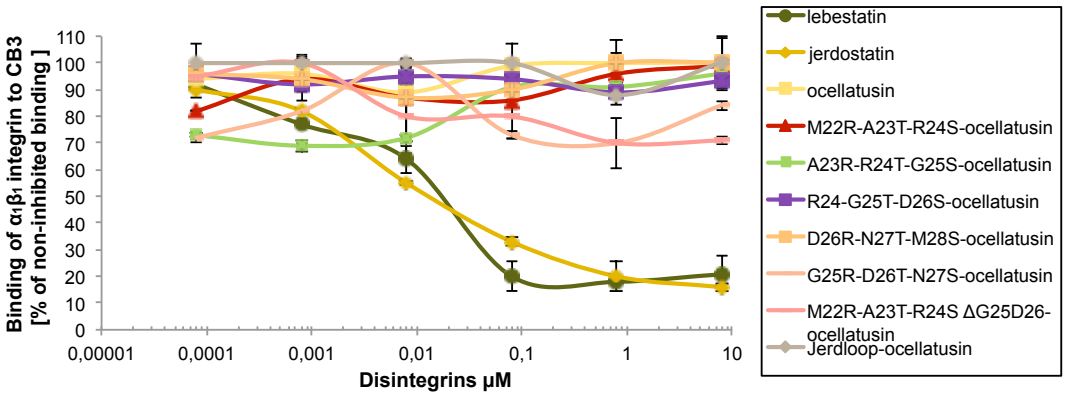


Figure 36. Inhibitory activity of recombinant disintegrins. Inhibition of the integrin $\alpha_1\beta_1$ binding to CB3 fragment of collagen IV by incubating soluble integrin $\alpha_1\beta_1$ (3.5 µg/mL) with increasing concentrations r-RTS-jerdostatin and KTS-lebestatin positive r-RGD-ocellatusin and r-RTS-ocellatusin mutants in 96-wells plates coated with 5 µg/mL of CB3. Bound integrin was detected by ELISA. Lebestatin/r-jerdostatin and wild-type r-ocellatusin were used as positive and negative inhibition controls, respectively.

The lack of inhibitory activity strongly suggests that neither the insertion of the RTS motif in different positions of the integrin-binding loop of an XXD short-disintegrin scaffold, nor the lateral position of RTS motif in a 9 or 11 amino acids loop, are sufficient for conferring $\alpha_1\beta_1$ integrin-binding specificity. Not even the replacement of the whole RGD-loop of ocellatusin [^{20}**C**KMARGDNMHDY**C**^{32}] by the RTS-loop of

jerdostatin [^{20}**C**WRTS--VSSHY**C**^{32}] is enough to provide $\alpha_1\beta_1$ integrin-binding specificity. Clearly, factors other than the integrin-binding loop sequence may modulate its active conformation and/or provide additional elements involved in determining the disintegrin' selectivity and specificity for integrin $\alpha_1\beta_1$.

4 Design and generation of *Frankenstein*-C-terminal-ocellatusin (r-RTS-C-terminal-ocellatusin) mutants

In this respect, it is worth mentioning that the integrin-binding loop and the C-terminal tail of both (K/R)TS-disintegrins (Monleón et al. 2003; Carbajo et al. 2011), and the RGD disintegrin echistatin (Monleón et al. 2005) display concerted motions. Replacement of echistatin's C-terminal end sequence ^{44}HKGPAT^{49} with that of the RGD-disintegrin eristostatin (WNG) decreased but did not abolish the inhibitory potential when examining ADP-induced platelet aggregation (Wierzbicka-Patynowski et al. 1999; Chen et al. 2012). As a whole, these data indicate that the C-terminal tail may act in synergy with the integrin-binding loop to modulate the high affinity and selectiveness of disintegrins for their target integrin receptors.

4.1 Generation of *Frankenstein*-C-terminal-ocellatusin mutants

Structure-function studies in our laboratory confirmed that RTS-disintegrin loop and the C-terminal tail of jerdostatin form an integrin-binding conformational epitope (Carbajo et al. 2011). Consequently, further mutants were designed to address the contribution of the C-terminal region for transforming a disintegrin scaffold from the RGD clade into another form the RTS/KTS clade. With this object, and given that the C-terminal end is longer in ocellatusin than in jerdostatin, we generated shortened C-terminal tail mutants. Initially, r-ocellatusin and each *Frankenstein* r-RTS- and r-jerdloop-ocellatusin mutants were shortened by 4 amino acids [^{47}EHDP^{50}] in the C-terminal end, generating r-E47_P50del-ocellatusin, r-RTS-E47_P50del-ocellatusin and r-jerdloop-E47_P50del-ocellatusin mutants (Figure 37).

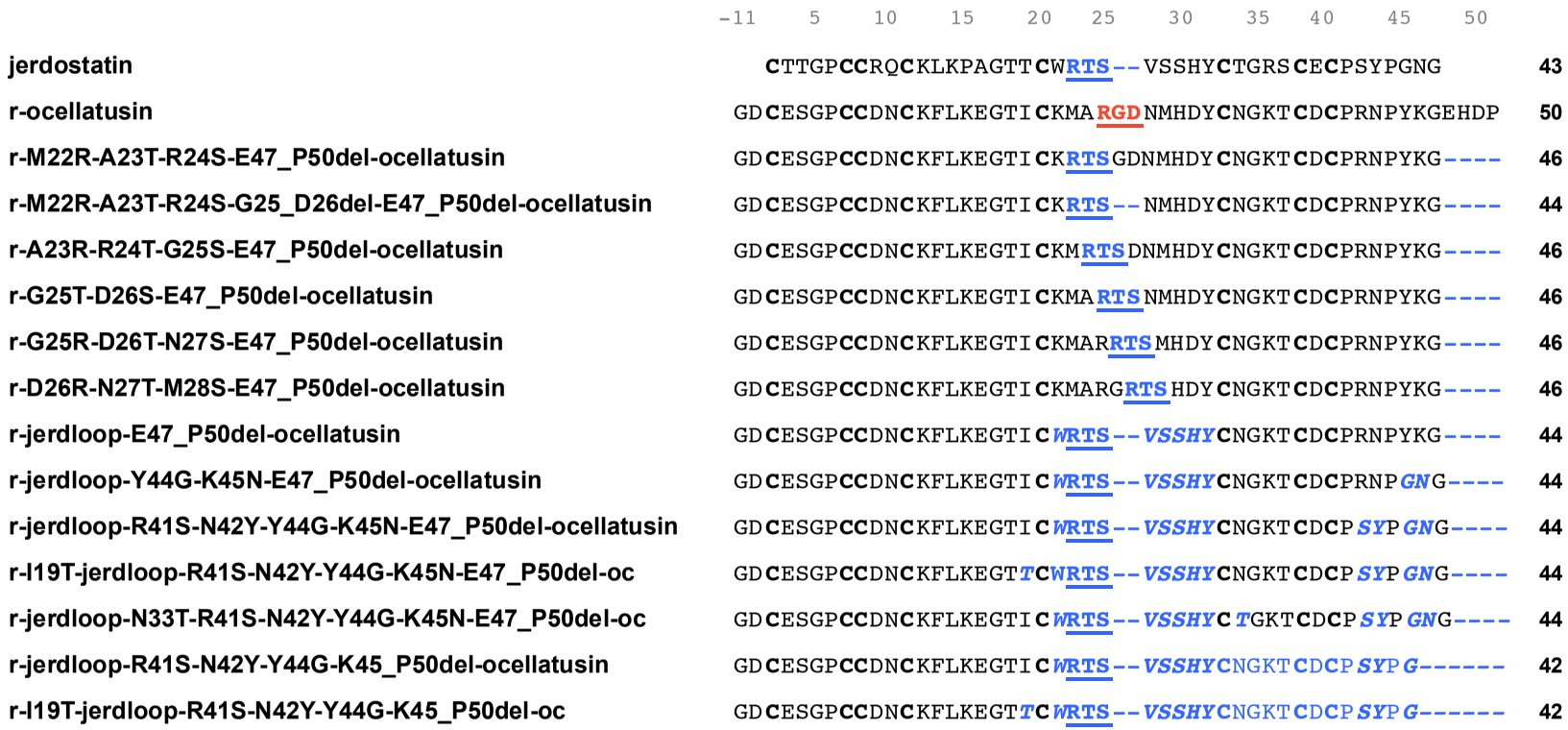


Figure 37. Amino acid sequences of wild-type jerdostatin, wild-type ocellatusin and *Frankenstein* rRTS-, jerdloop-ocellatusin Carboxyl terminal mutants. Cysteine residues are in bold, and RTS and RGD motifs are underlined and denoted in blue and red, respectively, according to the figure 30. The substitutions in the mutants are in indicated in italics and colored in blue, according to jerdostatin assigned color. Residue labeled -1 corresponds to a Glycine, the last residue of the TEV protease cleavage site, ENLYFQG, inserted between the His6/trxA tag and the disintegrin sequence.

In order to generate an active "r-*jerdostatinloop-C-terminal*" integrin-binding epitope, additional C-terminal tail mutants were generated using the jerdloop-E47_P50del-ocellatusin mutant as a template. Primarily, two mutants were generated; r-jerdloop-Y44G-K45N-E47_P50del-ocellatusin, replacing partially ocellatusin C-terminal end, [^{39}**C**PRNPYKG^{46}], for jerdostatin C-terminal tail, [^{39}**C**PRNP*GN*G^{46}], and, r-jerdloop-R41S-N42Y-Y44G-K45N-E47_P50del-ocellatusin, which includes the complete jerdostatin C-terminal end sequence [^{39}**C**P*SY*P*GN*G^{46}] (Figure 37). This sequence might be critical to conform the proper C-terminal tail synergy with the jerdostatin integrin-binding loop (Carbajo et al. 2011).

Further analyses (Carbajo et al. 2011) indicates that the r-jerdostatin-N45_G46del mutant, combined with a shortened C-terminal end sequence [^{39}CP*SY*P*G*^{44}] increases its activity in relation to the jerdostatin wild-type, with IC_{50}'s of 80nM and 180nM, respectively. Therefore, r-jerdloopR41S-N42Y-Y44G-K45_P50del-ocellatusin mutants with this shortened C-terminal tail [^{39}CP*SY*P*G*^{44}] were produced (Figure 37) (Table 7).

Considering the structural requirements and the RTS-, KTS- and RGD amino acid sequence conservation (Figure 38), other "*Frankenstein*" mutants were generated, including changes around the jerdostatin loop. First, N33T substitution was included in the mutant jerdostatin loop [^{19}I**C**WRTSVSSHY**C***T*^{33}], which contains the jerdostatin C-terminal sequence, [^{39}**C**P*SY*P*GN*G^{46}], resulting the r-N33T-jerdloop-R41S-N42Y-Y44G-K45N-ΔEHDP-ocellatusin mutant (Figure 37 and 40).

```
                  1    5    10   15   20   25   30   35   40   45   50
ocellatusin       DCESGPCCDNCKFLKEGTICKMARGDNMHDYCNGKTCDCPRNPYKGEHDP
jerdostatin        CTTGPCCRQCKLKPAGTTCWRTS--VSSHYCTGRSCECPSYPGNG
obtustatin         CTTGPCCRQCKLKPAGTTCWKTS--LTSHYCTGKSCDCPLYPG
viperestatin       CTTGPCCRQCKLKPAGTTCWKTS--RTSHYCTGKSCDCPVYQG
lebestatin         CTTGPCCRQCKLKPAGTTCWKTS--RTSHYCTGKSCDCPVYQG
```

Figure 38. RGD-ocellatusin and KTS/RTS-disintegrins. RGD non-conserved sites in KTS/RTS-disintegrins nucleotide sequence are double underlined in the ocellatusin sequence. Cysteine residues are in bold and RGD and (R/K)TS are in red and blue, respectively, according to the figure 30. Amino acids around the loop included in our mutants are highlighted in grey.

Beyond the sequence conservation, jerdostatin structure analyses indicate that the C-terminal tail of the RTS/KTS-jerdostatin is in close proximity to the loop, permitting structural linkages through hydrophobic interactions (Carbajo et al. 2011). The amino acid T19, in N-terminal of the jerdostatin loop, is one of those residues involved in hydrophobic interactions. This indicates that T19, together with additional amino acids in this loop (W21, H28, Y29) and the C-terminal tail (P38, P41, G42), are involved in the active conformation of jerdostatin (Carbajo et al. 2011) (Figure 39).

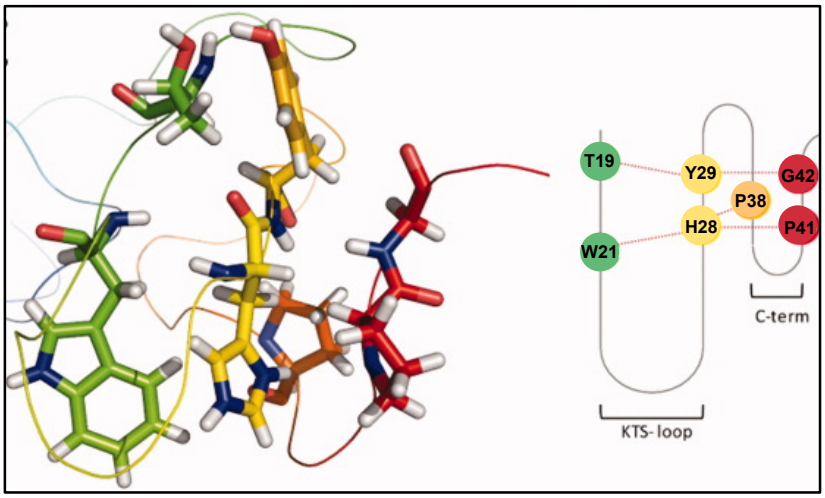


Figure 39. Detail of the KTS-jerdostatin structure, highlighting the network of hydrophobic interactions linking the active loop and the C-terminal tail of rJerK21. The side chains of T^{19} and W^{21} (in green) interact with residues H^{28} and Y^{29} (in yellow), which at the same time interact with P^{38} (orange) and the C-terminal residues P^{41} and G^{42} (red). Adapted from Carbajo et al. 2011.

Considering these interactions, I19T mutants were generated, containing the jerdloop sequence [^{19}*T*CWRTSVSSHY**C**N^{33}] and two different jerdostatin C-terminal sequences, [^{39}**C**P*SYPGNG*^{46}] or [^{39}**C**P*SYPG*^{44}] (Figure 40).

```
                          1    5    10   15   20   25   30   35   40   45   50
ocellatusin               DCESGPCCDNCKFLKEGTICKMARGDNMHDYCNGKTCDCPRNPYKGEHDP
r-I19T-jerdloop-Ct-oc     DCESGPCCDNCKFLKEGTTCWRTS--VSSHYCNGKTCDCPSYPG
r-N33T-jerdloop-Ct-oc     DCESGPCCDNCKFLKEGTICWRTS--VSSHYCTGKTCDCPSYPGNG
jerdostatin                CTTGPCCRQCKLKPAGTTCWRTS--VSSHYCTGRSCECPSYPGNG
lebestatin                 CTTGPCCRQCKLKPAGTTCWKTS--RTSHYCTGKSCDCPVYQG
```

Figure 40. Alignment of KTS/RTS-disintegrin, ocellatusin, I19T-jerdloop-R41S-N42Y-Y44G-K45_P50del-ocellatusin (r-I19T-jerdloop-Ct-oc) and N33T-jerdloop-R41S-N42Y-Y44G-K45N-E47_P50del-ocellatusin (r-N33T-jerdloop-Ct-oc) mutants. In the mutants, RGD- and (K/R)TS-disintegrin characteristic sequences are colored in red and blue, respectively, according to figure 30. Cysteine residues are in bold and the RGD and RTS motifs are underlined.

Similar to both the r-ocellatusin (Figure 32 and 33, p.70 and 71) and r-RTS-ocellatusin mutants, recombinant C-terminal ocellatusin mutants were obtained by overexpression in bacteria, purified by several affinity purification steps (Figure 25 and 26, p.54 and 55) and identified by SDS-PAGE and mass spectrometry (Table 7).

Table 7. Experimental (ESI-MS) and calculated molecular masses of the *Frankenstein* recombinant rRTS-, jerdloop-C-terminal ocellatusin mutants. Calculated masses correspond to fully oxidized (4 disulfide bonds) monoisotopic species.

Recombinant disintegrins	ESI-MS (Da)	Calculated monoisotopic mass (Da) with 4 S-S bonds
r-ocellatusin-E47_P50del	5174.6 ±2.4	5171.8
r-M22R-A23T-R24S- E47_P50del-ocellatusin	5158.3±0.7	5157.7
r-M22R-A23T-R24S ΔG25D26-E47_P50del-ocellatusin	4986.2±0.3	4985.6
r-A23R-R24T-G25S-E47_P50del-ocellatusin	5232.2±0.6	5231.8
r-G25T-D26S-E47_P50del-ocellatusin	5188.2±2	5187.8
r-G25R-D26T-N27S-E47_P50del-ocellatusin	5230.1±0.5	5229.9
r-D26R-N27T-M28S-E47_P50del-ocellatusin	5156.6±1	5155.8
r-jerdloop-E47_P50del-ocellatusin	4957.1±1	4956.5
r-jerdloop Y44G K45N-E47_P50del-ocellatusin	4836.9±0.5	4836.3
r-jerdloop R41S N42Y Y44G K45N-E47_P50del-ocellatusin	4816.8±0.6	4816.3
r-I19T-jerdloopR41S N42Y Y44G K45N-E47_P50del-oc	4803.7±1.1	4804.3
r-N33T jerdloop R41S N42Y Y44G K45N-E47_P50del-oc	4803.8±1.3	4803.3
r-jerdloop R41S N42Y Y44G-K45_P50del-ocellatusin	4644.7±0.9	4645.2
r-I19T jerdloop-R41S-N42Y-Y44G-K45_P50del-ocellatusin	4632.3±0.3	4633.1

4.2 Inhibition of soluble $\alpha_1\beta_1$ integrin-binding to CB3 by *Frankenstein* r-RTS- and r-jerdostatinloop-C-terminal-ocellatusin mutants

Similar to the previous r-RTS-ocellatusin and r-jerdloop-ocellatusin mutants, none of the jerdloop-C-terminal-ocellatusin mutants blocked $\alpha_1\beta_1$ binding to the CB3 collagen IV fragment (Figure 41) (Figure 20, p.32). The lack of binding inhibition imply that neither the complete jerdostatin C-terminal sequences, [^{39}**C**P*SYPGN*G^{46}] or the shortened sequence [^{39}**C**P*SYPG*^{44}], plus the jerdostatin loop [^{18}**C**WRTSVSSHY**C**^{32}], nor its combination with the improved jerdostatin loops, [^{19}**I****C**WRTSVSSHY**C***T*^{33}] or [^{19}*T***C**WRTSVSSHY**C**N^{33}], are sufficient enough to achieve functionality similar to the KTS/RTS disintegrins.

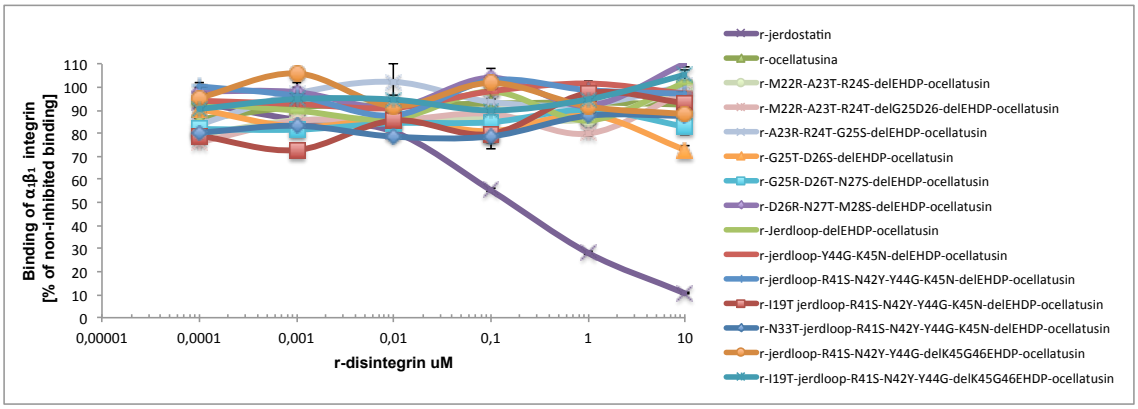


Figure 41. Inhibitory activity of recombinant disintegrins. Inhibition of the integrin $\alpha_1\beta_1$ binding to CB3 fragment of collagen IV by incubating soluble integrin $\alpha_1\beta_1$ (3.5 µg/mL) with increasing concentrations of r-RTS-ocellatusin-C-terminal mutants in 96-wells plates coated with 5 µg/mL of CB3. Bound integrin was detected by ELISA. r-jerdostatin and wild-type r-ocellatusin were used as positive and negative inhibition controls, respectively.

These results suggest that other elements beyond: i) the integrin-binding loop sequence, ii) the C-terminal tail and iii) the amino acids around the loop, might modulate the active conformation of jerdostatin. Comparison of the amino acid sequences of RGD-ocellatusin and KTS/RTS-disintegrins (Figure 37), combined with our results, indicate that the non-conserved amino acids located at the N-terminal region of the loop and/or the Serine36 are required to conform a functional (R/K)TS-disintegrin. Previous studies examining disintegrin structure fail to address these amino acid regions as a key element for selectivity and specificity for their integrin receptors. Nonetheless, it is worth mentioning that there is 100% conserved sequence identity in the N-terminal region of every KTS/RTS-disintegrins (Figure 38). These N-terminal sequences may be essential for a proper conformation of disintegrins involved in the $\alpha_1\beta_1$ integrin-binding inhibition.

In addition, our results may be suggestive that i) (K/R)TS-disintegrins may have diverged from a common precursor to the RGD-disintegrins in an unusually accelerated pace, or ii) (K/R)TS-disintegrins have been recruited independently of the canonical XXD disintegrins. The second hypothesis is supported by structural data of RTS- and RGD-disintegrins. When the structure of RTS-jerdostatin (Carbajo et al. 2011) is compared to RGD-echistatin (Chen et al. 1994; McLane et al. 1996; Marcinkiewicz et al. 1997), distinct chemical and physical features modulate the integrin inhibitory motif,

permitting the different integrin-binding strategies observed in each of these disintegrins. Therefore, it is likely that RGD- and RTS-disintegrin scaffolds evolved independently.

Furthermore, all disintegrins with exception of KTS/RTS-disintegrins bind integrins which emerged early in evolution, preceding first metazoans (Sebé-Pedrós et al. 2010), and none of these integrins contain the I-domain (or A-domain) "insertion" in the α subunit (αI) (Figure 13 A). The insertion of the αI domain occurred approximately 500 Mya in Chordate. The presence of the αI domain permits additional flexibility in regards to ligand recognition by integrins (Chouhan et al. 2014) (Figure 42B). Therefore, novel sources with epitopes that are targeted by αI-integrins (Figure 42B) may have been an adaptation in venoms to the presence of the αI domain in the prey of venomous snakes. This may be the case of RTS/KTS-disintegrins and C-type lectin, which inhibit the binding of integrins $\alpha I_1\beta_1$ and $\alpha I_2\beta_1$ (Figure 42B) to collagen, respectively (Ogawa et al. 2005; Calvete et al. 2007b). Interestingly, KTS/RTS-disintegrins, which we propose evolved independently from other disintegrins (XXD-type), are unique in the aspect that they inhibit $\alpha_1\beta_1$ integrin-collagen IV/I binding. Further, KTS/RTS-disintegrins are not able to block non-αI-integrin (Figure 42A) binding (Figure 12, p.24). This may support our hypothesis of an independent evolutionary origin of the RTS/KTS and the RGD short disintegrins clades.

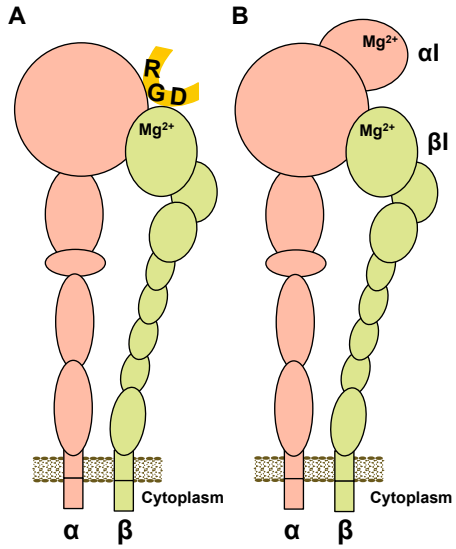


Figure 42. Schematic representation of integrin heterodimers. Integrins are large heterodimeric, bi-directionally signaling, cell surface receptors that consist of a large extracellular ectodomain, a transmembrane region and relatively short intracellular ''tails'' (right). Integrins with (B) and without (A) I-domain in α domain (αI). The αI domain contains the highly solvent-exposed MIDAS site (Mg^{2+}). Cartoon of the disintegrin RGD motif is presented in the integrin-binding site (A). Adapted from Chouhan et al. 2014.

5 Residues responsible for the functional activity of ocellatusin.

From these results, and the high RTS/KTS-disintegrin sequence conservation necessary to inhibit $\alpha_1\beta_1$ integrin-binding to its natural ligand, we decide to examine how r-ocellatusin function was depleted in our r-RTS-ocellatusin mutants. Hence and alternatively to the $\alpha_1\beta_1$ integrin-binding inhibition assay, RGD-disintegrin functionality of the *"Frankenstein"* mutants was studied.

Similar to the activity analysis of r-ocellatusin (Figure 34), we measured the ability of r-RTS-ocellatusin mutants (Figure 35) to inhibit platelet aggregation (Figure 43). Results show, that out of all the disintegrins tested, only r-M22R-A23T-R24S-ocellatusin [^{20}**C**K*RT**S***__**GD**__NMHDY**C**^{32}] inhibits the collagen I-induced aggregation of washed human platelets in a dose dependent manner, with an IC_{50} of 1.046×10^{-6} M (Figures 43 A and B). However, the potency of r-RTS-ocellatusin was significantly less when compared to that of the wild-type r-ocellatusin (IC_{50} of $3.9\times10^{-8}\pm0.5$ M). Conversely, none of the r-RTS-ocellatusin isoforms, including G25 and/or D26 substitutions were able to inhibit collagen I-induced platelet aggregation (Figures 43 C and D).

These results are consistent with the conserved nature of the aspartate (D26) residue among small disintegrins (Calvete et al. 2005), reviewed in Calvete 2013b. However, Glycine25 is not as conserved as the aspartate26 in XXD disintegrins motifs (MLD, MVD, KGD, RGD, WGD, VGD, MGD). Furthermore, G26 is not essential for integrin $\alpha_{IIb}\beta_3$ binding inhibition, given that disintegrins containing MVD motif (Shimokawa et al. 1998; Carey et al. 2012), RGD (Marcinkiewicz et al. 1997; Wierzbicka-Patynowski et al. 1999), KGD (Scarborough et al. 1991) and WGD (Calvete et al. 2002) motifs, are able to block integrin $\alpha_{IIb}\beta_3$. In addition, the mutant r-A23R-R24T-G25S-ocellatusin, which contains the loop [^{23}RTSD^{26}], failed to inhibit platelets aggregation, suggesting that the Gly25 substitution by a less hydrophobic residue, containing a bigger lateral chain (Serine) might interfere in the proper binding of the D26 to integrin $\alpha_{IIb}\beta_3$ (Figure 42). In this mutant, D26 is not enough to confer RGD-ocellatusin functionality. Our results suggest that for sufficient binding to $\alpha_{IIb}\beta_3$, a ligand must have a small amino acid in the second position of the integrin-binding motif (e.g. XGD, XVD).

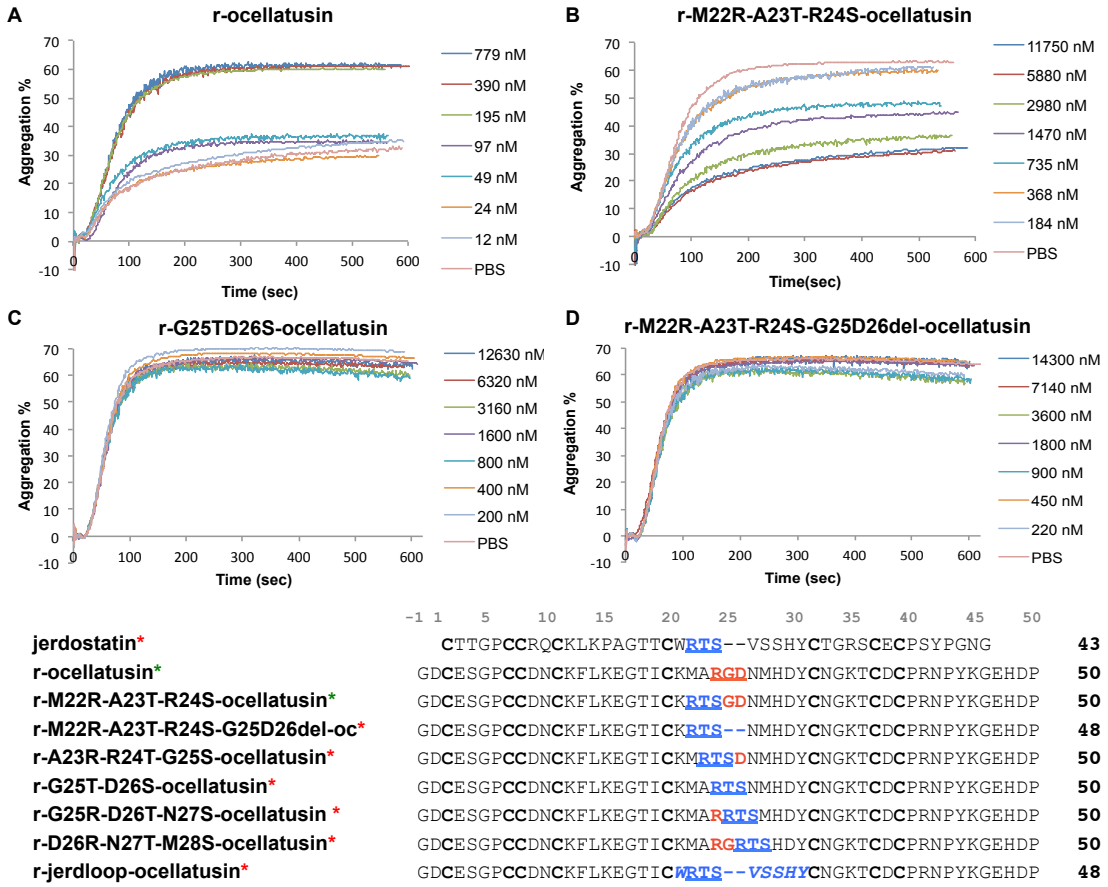


Figure 43. Recombinant r-RTS-ocellatusin mutants lost of r-RGD-ocellatusin "*natural*" functionality A. Concentration-dependent blocking of the collagen I-induced platelet aggregation by r-RTS-ocellatusin mutants. r-A23T-R24S-G25S-ocellatusin, r-G25R-D26T-N27S-ocellatuisn, r-D26R-N27T-M28S-ocellatusin and r-jerdloop-ocellatusin mutants present similar aggregation curves to the r-G25T-D26-ocellatusin and r-M22R-A23T-R24S-G25_D26del-ocellatusin. B. r-RTS-ocellatusin mutants alignment, in which the RTS and RGD motifs are underlined and colored in blue and red, respectively, according to the figure 30. Cysteine residues are in bold. Green and red asterisk (*) denote active and non-active r-disintegrin, respectively.

In addition, we tested for the ability of C-terminal mutants, specifically, r-E47_P50del-ocellatusin and r-M22R-A23T-R24S-E47_P50del-ocellatusin (Figure 37), to inhibit collagen-induced platelet aggregation (Figure 43). Both r-E47_P50del-ocellatusin and r-M22R-A23T-R24S-E47_P50del-ocellatusin inhibited collagen I-induced aggregation of human platelets in a dose dependent manner, with an IC_{50} of $3.6\pm0.1\times10^{-8}$M and IC_{50} of 2.97×10^{-6} M, respectively (Figure 44). Interestingly, the results of r-E47_P50del-ocellatusin IC_{50} were comparable to those of r-ocellatusin IC_{50} ($3.6\pm0.6\times10^{-8}$M). Even the C-terminal region of ocellatusin might be important for its activity, as our results show that the C-terminal amino acids, E47, H48, D49, P50, are

not crucial for $\alpha_{IIb}\beta_3$ integrin-binding potency. These results are supported by Chen et al. 2012, with the r-E47_P50del-ocellatusin being comparable in length, as well as biological activity when compared to wild type echistatin (Chen et al. 2012).

On the other hand, the IC_{50} of r-M22R-A23T-R24S-E47_P50del-ocellatusin is increased in respect to r-ocellatusin, decreasing its inhibitory potency, and showing activity similar to r-M22R-A23T-R24S-ocellatusin. Although a shortened C-terminal end does not affect to the activity of wild-type ocellatusin, the combination of the [^{20}**C**K***RTS***GDNMHDY**C**^{32}] loop and C-terminal deletion leads to an increased IC_{50} in r-M22R-A23T-R24S-ocellatusin (Figure 44). These data suggest that the amino acids E47, H48, D49, P50 in C-terminal tail may contribute to the stability of the integrin-binding site. However, further analyses are necessary to confirm this hypothesis.

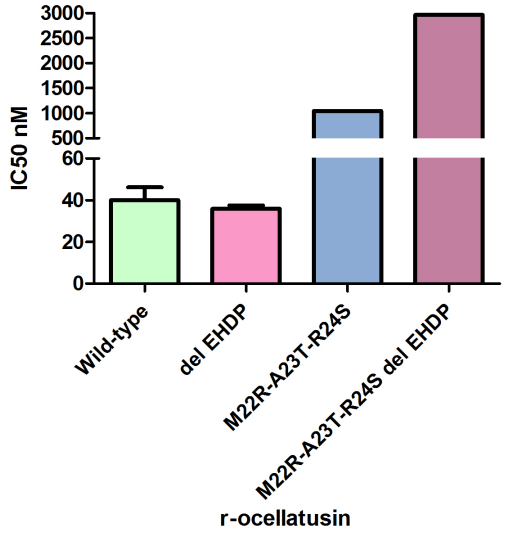


Figure 44. IC_{50} values of ocellatusin and *Frankenstein* r-RTS- and r-C-terminal-ocellatusin mutants, obtained from dose-dependent inhibition of collagen I-induced aggregation (MA) of human washed platelets.

The functionality of r-ocellatusin and r-*Frankenstein*-ocellatusin mutants agrees with previously published reports on RGD-disintegrin structure-function studies (Marcinkiewicz et al. 1997; Wierzbicka-Patynowski et al. 1999; Smith et al. 2002; Chen et al. 2012; Yang et al. 2015), which indicates that the recombinant disintegrins used throughout these studies have been properly folded as a "natural" short disintegrin by the appropriate cysteine-pairing [(2-11)(7-32)(8-37)(20-39)](Carbajo et al. 2015).

6 Generation and functionality of SGD, GGD and TGD mutants

Our results suggest that just mutants expressing the ^{25}GD^{26} or ^{25}VD^{26} (Carey et al. 2012) sequence at the tip of the inhibitory loop are able to block $\alpha_{IIb}\beta_3$ integrin-binding to fibrinogen. Interestingly, these results provide evidence for a new active disintegrin integrin-recognition motif consisting of the residues SGD. However, mutants expressing the loop ^{20}**C**K*RT*<u>**SGD**</u>NMHDY**C**^{32} are not as potent as wild-type r-ocellatusin. That might be because of the amino acids substitution in the N-terminal side of SGD motif (M22R and A23T), which may interfere with the potency of the integrin-binding loop.

6.1 Design and generation of the mutants SGD, GGD and TGD

Although SGD represents a novel tripeptide integrin-binding sequence, an endogenous disintegrin expressing this sequence has not yet been characterized. Juarez *et al.* (2008) suggested that only the most parsimonious nucleotide substitution events are required for the emergence of the different disintegrins´ integrin-recognition XXD motifs from the ancestral RGD sequence (Juárez et al. 2008) (Figure 14 p.26). We decided to explore the possible natural emergence of an SGD-disintegrin by a unique nucleotide substitution in the arginine codon, necessary to transform RGD to SGD (Figure 45).

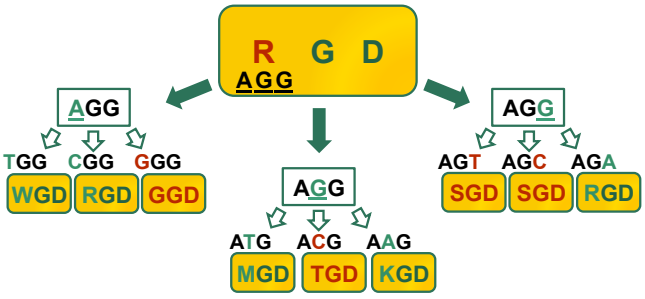


Figure 45. Most parsimonious nucleotide substitution events needed for the emergency of SGD and other integrin-recognition motifs from RGD ancestral tripeptide. Arrows indicate a single mutational transition at the underlined site.

We observed that the SGD motif could emerge twice in evolution by a single nucleotide substitution in the arginine (R) codon. Taking this into account, in addition to the potential effect of two additional mutations in r-M22R-A23T-R24S-ocellatusin mutant, the SGD-ocellatusin mutant was generated (Figure 46).

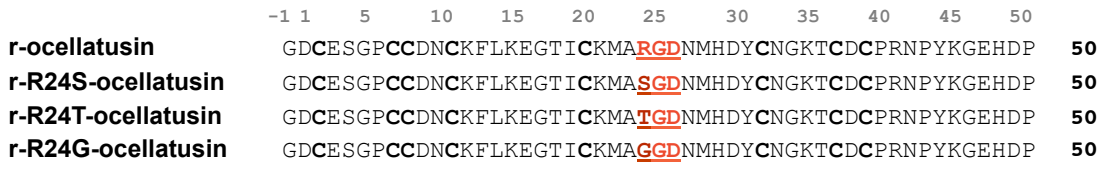


Figure 46. Alignment of new generated r-XGD-ocellatusin mutants and r-RGD-ocellatusin, containing disintegrin integrin-binding motifs. Cysteine residues are in bold, and XGD motifs are underlined. Residue labeled -1 corresponds to the last residue of the TEV protease cleavage site, ENLYFQG, inserted between the His_6-trxA tag and the disintegrin sequence. The length, in amino acids, of each disintegrin is indicated next to their sequence.

Additionally, we generated the following two ocellatusin mutants; r-R24T- and r-R24G-ocellatusin, both of which equally consider a single nucleotide substitution in the R codon (AGG) (Figure 45). Result mutants express TGD and GGD integrin-binding motifs (Figure 46). Those mutants as well as r-R24S-ocellatusin were recombinantly expressed, purified and identified, following similar methods used for recombinant wild type ocellatusin (Figures 25 and 26, p.54 and 55) (Figures 32 and 33, p.70 and 71).

Table 8. Experimental (ESI-MS) and calculated molecular masses of the recombinant r-XGD-disintegrin mutants. Calculated masses correspond to fully oxidized (4 disulfide bonds) monoisotopic species.

Recombinant XGD disintegrin mutants	ESI-MS (Da)	Calculated monoisotopic mass (Da) with 4 S-S bonds
r-R24S-ocellatusin	5582.7±1.3	5581.2
r-R24T-ocellatusin	5596.5±0.3	5595.2
r-R24G-ocellatusin	5552.8±1.5	5551.2

Mass spectrometry data indicated that the XGD mutants were fully oxidized; given that their ESI-MS values match with calculated monoisotopic mass with 4 disulfide bonds (Table 8).

6.2 Functionality of r-SGD, r-TGD- and r-GGD-disintegrins

Inhibition of collagen I-induced platelet aggregation by the r-SGD-, r-TGD, and r-GGD-ocellatusin mutants was measured and compared with the r-M22R-A23T-R24S-ocellatusin mutant. Wild-type r-ocellatusin and r-jerdostatin were used as the positive and negative control, respectively. Although r-R24S-ocellatusin, r-R24T-ocellatusin and

r-R24G-ocellatusin are capable of inhibiting collagen I-induced platelet aggregation in a dose dependent manner, with an IC_{50} of 1.8 $\times 10^{-6}$ M, 9.5×10^{-7} M and 4.3 $\times 10^{-6}$ M, respectively, these activities were not as potent as wild-type ocellatusin (IC_{50} of 3.9×10^{-8} ±0.5 M) (Figure 47).

r-SGD-ocellatusin does block $\alpha_{IIb}\beta_3$ integrin-fibrinogen binding with similar potency as seen with the r-M22R-A23T-R24S-ocellatusin mutant (IC_{50} of 1.05×10^{-6} M). This result suggests that the M22R and A23T substitutions, do not considerably affect the integrin-binding activity. It appears that the amino acids in the N-terminal region of the SGD motif are not critical for the r-RGD-disintegrin activity, whereas amino acids located carboxyl to the RGD motif significantly contribute to disintegrin functionality (Yang et al. 2015).

The TGD motif appears to have slightly higher affinity when compared SGD. However, the presence of a glycine in the first position of the r-GGD-ocellatusin tripeptide appears to decrease the disintegrin selectivity for the integrin-binding site (Figure 47).

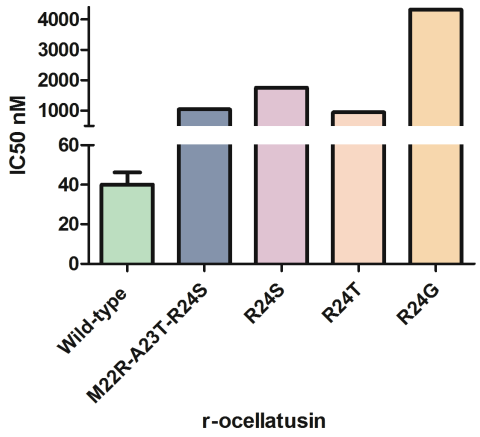


Figure 47. IC_{50} values of ocellatusin and XGD-ocellatusin mutants, obtained from dose-dependent inhibition of collagen-induced aggregation (MA) of human washed platelets.

Our results suggest that disintegrins expressing SGD, TGD, GGD integrin-binding motifs may not have been selected for in nature due to their low potency at blocking integrin binding. This is supported by the fact that natural disintegrins (RGD, WGD,

KGD, MGD) are able to block $\alpha_{IIb}\beta_3$ integrin with IC_{50}'s 100 times lower than observed by our r-(S/T/G)GD-ocellatusin mutants (Figure 47) (Scarborough et al. 1991; Marcinkiewicz et al. 1997; Shimokawa et al. 1998; Wierzbicka-Patynowski et al. 1999; Calvete et al. 2002; Carey et al. 2012; Angulo et al. 2014; Yang et al. 2015).

7 Concluding remarks, reflexions and perspectives

When analyzing the residues responsible for the functional activity of ocellatusin, only G25 and D26, in the RGD disintegrin motif, seem essential for the functional activity of ocellatusin, whereas, the R24 residue is important for ocellatusin potency. However, the amino acids M22 and A23, which are situated adjacent to RGD motif in N-terminal side, as well as the residues E47, H48, D49, P50, in C-terminal end, are not crutial for ocellatusin activity and potency.

The lack of inhibitory activity of the different *Frankenstein* disintegrins, even when the whole RGD-loop and C-terminal end of ocellatusin were replaced by the RTS-loop and jerdostatin C-terminal tail, suggest that RTS/KTS short disintegrins could have been recruited independently of the canonical RGD disintegrins. This independent evolutionary origin is supported by structure-function data, which show significant differences in the RGD- and KTS/RTS-disintegrins integrin-binding scaffolds (Figure 17, p.29). Furthermore, this is supported by integrin evolution, since natural ligand binding to $\alpha_1\beta_1$ is inhibited only by KTS/RTS-disintegrins, including integrins expressing the αI-domain. Moreover, KTS/RTS-disintegrins are not able to inhibit (non-αI)-integrins-ligand binding. Therefore, the emergence of αI-integrin and the structural change around the possible binding site supports the independent emergence of RTS/KTS-disintegrins in Eurasian viper venoms.

KTS-disintegrins appear to be uniquely translated in the venom proteomes of the relatively recent clade of Eurasian vipers, including *Macrovipera* and *Daboia* (Lenk 2001) (Table 5). In addition, the RTS-disintegrin coding sequence has been identified in several viper venom glands (Table 5), however, to date, none of the studied vipers produce an RTS-disintegrin in their venoms. Therefore, other Eurasian viper venom proteome analyses are needed to confirm the expression of RTS-disintegrin(s) in venom.

However, the evolutionary pressure to express integrin $\alpha_1\beta_1$–specific inhibitors in these taxa appears to be difficult to rationalize in the context of a predator-prey arms race. Hence, another functionality of KTS/RTS-disintegrins in regards to prey-predator evolution warrants further exploration. In addition, the hypothesis that the molecular machinery operating on the neofunctionalization of disintegrin scaffolds evolved only in Viperinae deserves future detailed investigations. Therefore, jerdostatin-coding gene analyses are needed to understand KTS/RTS-disintegrin evolution. Also, since not all KTS/RTS-disintegrin coding mRNAs are translated and released in viper venom glands (*Macrovipera mauritanica, Echis ocellatus* and *Cerastes vipera*) (Table 5), additional expression studies are required to elucidate KTS/RTS-disintegrin functionality.

Chapter II. RTS-disintegrin-coding gene (*RPTLN*) distribution across Reptilia taxa

Short RTS/KTS disintegrins selectively hit the collagen I and IV binding $\alpha_1\beta_1$ integrin (Calvete et al. 2007b; Brown et al. 2009; Walsh and Marcinkiewicz 2011), and form a distinct clade of recently emerged short disintegrins in viper venoms within genera *Macrovipera* (Marcinkiewicz et al. 2003), and *Daboia* (Kisiel et al. 2004; Olfa et al. 2005). RTS/KTS-disintegrins are released in the snake venom by short-coding mRNAs synthesis in Viperinae venoms (Marcinkiewicz et al. 2003; Kisiel et al. 2004; Olfa et al. 2005; Bazaa et al. 2007). Strikingly, a non-protein-translated mRNA sequence encoding the full-length RTS-disintegrin jerdostatin was originally amplified from *Protobothrops jerdonii* venom gland cDNA library [AY262730] (Sanz et al. 2005) (Figure 52). Subsequently, identical mRNAs have been cloned from a number of Crotalinae and Viperinae venom gland cDNA libraries (Sanz et al. 2006; Bazaa et al. 2007). Equal mRNA encoding jerdostatin was sequenced in the *Daboia russelii* venom gland cDNA library and it was reported in NCBI (National Center of Biotechnology Information) database and registered as FF277034.1 (Table 9).

Table 9. Distribution of mRNA sequence encoding the full-length RTS-disintegrin jerdostatin across Viperidae venom gland.

Species containing jerdostatin coding RNA	GenBank accession code	Bibliography
Protobothrops jerdonii	AY262730	(Sanz et al. 2005)
Cerestes vipera	AM114012.1	(Monleón et al. 2003; Sanz et al. 2006)
Macrovipera mauritanica	AM261813.1	(Bazaa et al. 2007)
Echis ocellatus	AM286798.1	(Bazaa et al. 2007)
Daboia russelii	FF277034.1	by Sai-Ngam, A.

Submitted sequences by Sai-Ngam, A. from "Snake bite and venom research unit, Chula Medical Research Center Faculty of Medicine, Chulalongkorn University, Thailand" are unpublished.

1 Russellistatin, RTS-disintegrin, is released exclusively in *Daboia russelii* venom

Previous venomic studies did not identify any RTS-disintegrin in *C. vipera, M. mauritanica, D. russelii* or *E. ocellatus* venom proteome (Bazaa et al. 2005; Wagstaff et al. 2009; Makran et al. 2012). Moreover, *Daboia russelii* venom proteome was analyzed by (Risch et al. 2009), though, only proteins above 10kDa were identified. Therefore, we could not discard that low molecular mass disintegrin are expressed in *D. russelii* venom proteome.

To test the hypothesis that an RTS-disintegrin with 100% sequence identity to jerdostatin may be expressed in the venom proteome of *D. russelii*, low molecular mass proteins of its venom were analyzed. First, *D. russelii* venom proteins were separated by reverse-phase HPLC (Figure 48).

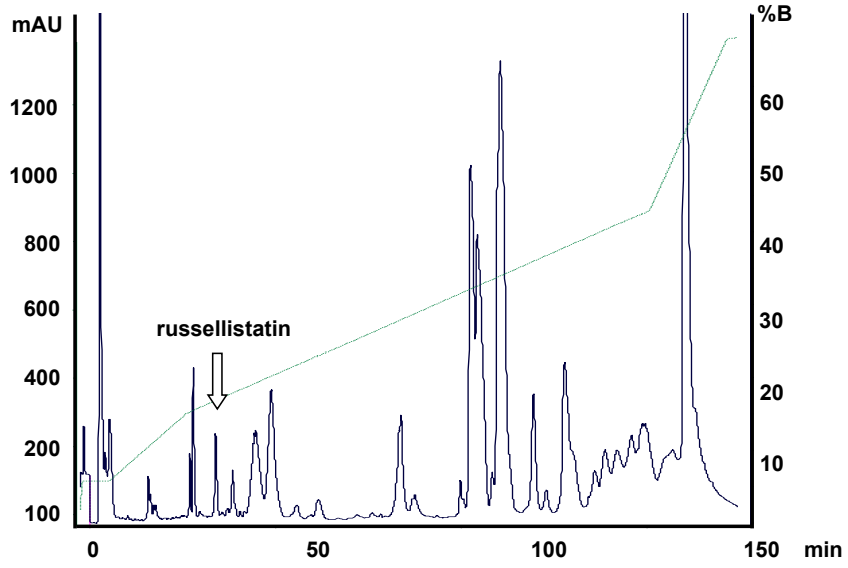


Figure 48. Chromatogram of the *Daboia russelii* venom protein separation by reverse-phase HPLC. Russellistatin peak is pointed in the chromatogram.

Subsequently, possible disintegrin peaks were characterized by electrospray-ionization mass spectrometry. One peak exhibits the expected isotope-averaged molecular mass for a short disintegrin, 4408.5 Da (Figure 49).

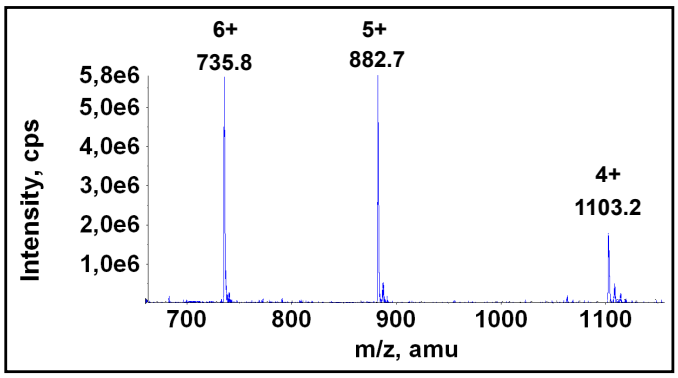


Figure 49. Electrospray-ionization mass spectrometry, of reverse-phase HPLC-isolated likely disintegrin peak, which exhibit the expected isotope-averaged molecular mass for a short disintegrin (4408.5 Da).

To identify the possible RTS-disintegrin sequence, this protein peak (4408.5 Da) was in-gel digested by trypsin. The resulting tryptic fragments were sequenced by MS/MS (Figure 50). The amino acid sequence of this peptide, which we termed russellistatin, is identical to jerdostatin sequence (Sanz et al. 2005) (Figure 52). Noteworthy, *Daboia russelii* is the unique snake, which releases an RTS-disintegrin (russellistatin) in its venom, despite other snakes contain the mRNA coding jerdostatin in their venom glands (Sanz et al. 2006; Bazaa et al. 2007) (Table 9). However, RTS-disintegrins seems to be only released in Viperinae venoms, similar to KTS-disintegrins (Table 9). Supporting the hypothesis that KTS/RTS-disintegrin expression is restricted to Eurasian vipers.

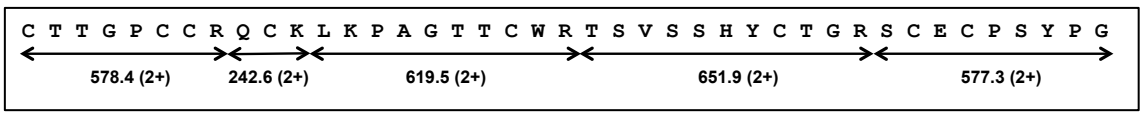


Figure 50. Russellistatin sequence product determined by MS/MS spectrometry sequencing of the fragments generated by tryptic digestion of the likely disintegrin peak.

2 Mature jerdostatin genomic DNA in Reptilia

In addition to the presence of jerdostatin-like sequence in different venom gland cDNAs (Table 9), previous analyses in our laboratory suggest that genomic DNA (gDNA) fragment encoding the mature jerdostatin-like domain, 132bp, have been PCR-amplified from gDNA of a number of taxa across Serpentes, including Viperidae and Elapidae, and lizards (Lacertidae and Iguanidae) (Figure 51) (Sanz-Soler et al., 2012). However, mature jerdostatin-like sequences (Figure 52) were not amplified in Aves (*G. gallus*), Amphibia (*B. orientalis*), and Mammalia (*M. musculus*) (Figure 51), suggesting that this gene may exhibit a reptile-restricted distribution. Therefore, the term *RPTLN* (RePTiLiN) has been proposed in our laboratory to design the gene containing the sequence coding for the full-length jerdostatin (see in Figure 52).

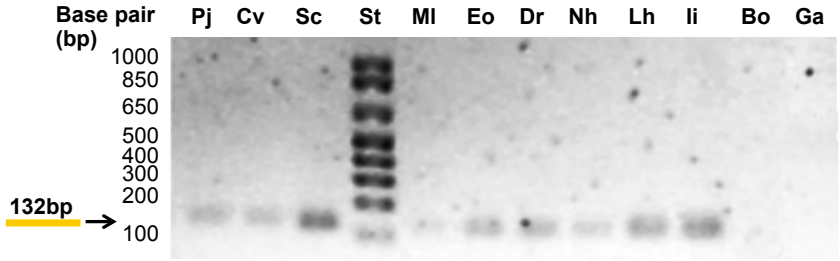


Figure 51. Mature jerdostatin-like DNA sequences across Reptilia. PCR-amplification of jerdostatin-like sequences from genomic DNA of *Protobothrops jerdonii* (Pj), *Crotalus viridis* (Cvi), *Sistrurus catenatus catenatus* (Scc), *Macrovipera lebetina* (Ml), *Echis ocellatus* (Eo), *Daboia russelli* (Dr), *Naja haje* (Nh), *Lacerta hispanica* (Lh), and *Iguana iguana* (Ii). Bo, *Bombina orientalis*; Ga, *Gallus gallus*. Adapted from Sanz-Soler 2012.

Later in this chapter, we provide further evidence for the broad distribution of the *RPTLN* genes across Reptilia, and we report its uneven transcriptional profile in adult lizard and colubrid organs. Moreover, a role for *RPTLN* in the evolution of SVMPs is hypothesized.

3 Intronless *RPTLN* genes represent a broad and reptile-restricted multigene family

In the same direction, full-length *RPTLN* gene was studied. It is notable the simplicity of the *RPTLN* genetic structure, far for being part of a PII-SVMP or a complex genetic structure, no introns were amplified in any of the analyzed species (Figures 52 and 53).

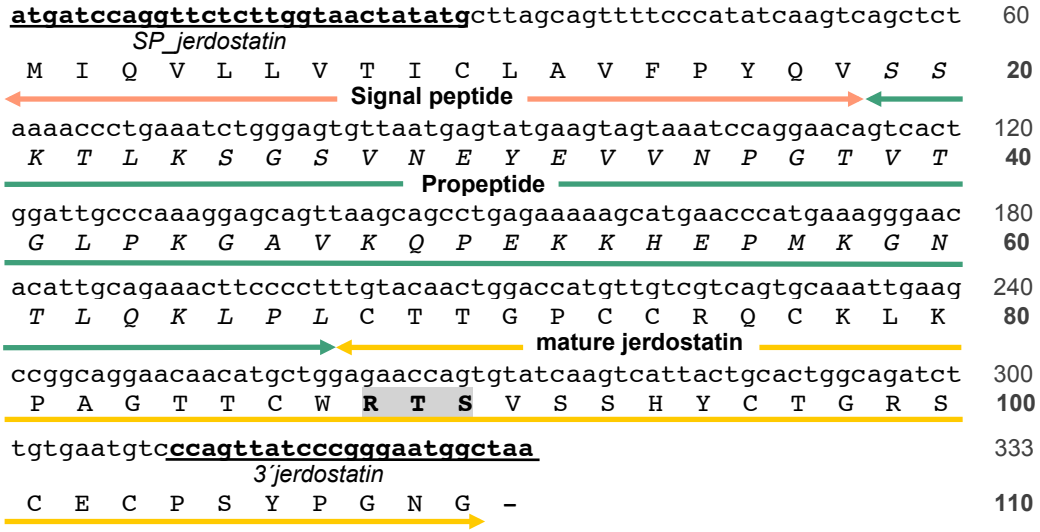


Figure 52. Nucleotide sequence of the genomic DNA (*RPTLN*-1 gene, AY262730) encoding full-length jerdostatin. The limits of the signal peptide, propeptide, and disintegrin domain are indicated. The active RTS tripeptide sequence is highlighted in bold in a grey background. The sequences of the forward (SP_jerdostatin) and reverse (3'jerdostatin) primers used for PCR amplifications are in lower case, highlighted in boldface, and labeled.

Genomic DNA encoding full-length jerdostatin (*RPTLN*-1) (Figure 52) and full-length jerdostatin-like sequences (*RPTLN*-n) were amplified from a number of Anapsida (Testudines), Diapsida (Serpentes, Sauria), and Archosauria (Crocodylia) reptiles (Figure 53) (Tables 10 and 11), but attempts to amplify this gene in amphibians (*Bombina orientalis*), birds (*Gallus gallus, Aletoris rufa, Anas platyrhynchos*), and mammals (*Mus musculus* and *Homo sapiens*) were all unsuccessful. These striking results suggest that *RPTLN* genes exhibit a broad, reptile-specific distribution (Tables 10 and 11) (Figure 53).

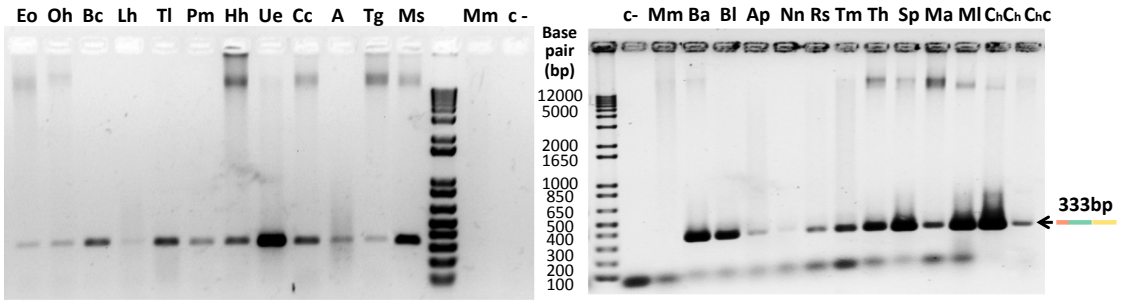


Figure 53. PCR-amplification of *RPTLN* genes from gDNA of *Echis ocellatus (Eo), Ophiophagus hannah (Oh), Boa constrictor (Bc), Lacerta hispanica (Lh), Timon lepidus (Tl), Podarcis muralis (Pm), Heloderma horridum (Hh), Uroplatus ebenaui (Ue), Chamaeleo calyptratus (Cc), Alligator mississippiensis (A), Testudo graeca (Tg), Mauremys sinensis (Ms), Bothrops asper (Ba), Bothriechis lateralis (Bl), Atropoides picadoi (Ap), Naja naja haje (Nn), Rhinechis scalaris (Rs), Tarentola mauritanica (Tm), Testudo Hermanni (Th), Stigmochelys pardalis (Sp), Mauremys annamensis (Ma), Mauremys leprosa (Ml), Chelonoidis chilensis (ChCh), Chelonoidis carbonaria (Chc), and Mus musculus (Mm).* (c-), negative control.

Interestingly, 21 different *RPTLN-like* gene copies were found in some of the studied reptile genomes (Tables 10 and 11). Moreover, in some species, more than two different *RPTLN* gene copies were sequenced, supporting the idea of *RPTLN* gene duplication and multigene family (Table 10).

Table 10. Distribution of *RPTLN* gen copies across reptiles. Full-length sequences are displayed in Figure 54. Nucleotide changes in *RPTLN* genes respect to *RPTLN*-1 [jerdostatin, AY262730] are listed in Table 11. See accession codes in Table 1S (p.171).

Reptile class	Family	Species	*RPTLN* gene copy
Snake	Crotalinae	*P. jerdonii*	*RPTLN-1*
		B. asper	*RPTLN*-1, *RPTLN*-15, *RPTLN*-16
		B. lateralis	*RPTLN*-1, *RPTLN*-7, *RPTLN*-17
	Viperinae	*E. ocellatus*	*RPTLN*-1, *RPTLN*-13
		A. picadoi	*RPTLN*-1, *RPTLN*-10, *RPTLN*-15
	Elapidae	*O. hannah*	*RPTLN*-1, *RPTLN*-6, *RPTLN*-17
		N. haje haje	*RPTLN*-1, *RPTLN*-5
	Colubridae	*R. scalaris*	*RPTLN*-1, *RPTLN*-5, *RPTLN*-18-21
	Boideae	*B. constrictor*	*RPTLN*-1, *RPTLN*-17
Lizard	Lacertidae	*L. hispanica*	*RPTLN*-1, *RPTLN*-9, *RPTLN*-17
		T. lepidus	*RPTLN*-1
		P. muralis	*RPTLN*-1, *RPTLN*-9, *RPTLN*-17-19
		P. hispanica	*RPTLN*-1, *RPTLN-8, RPTLN-11, RPTLN-13, RPTLN-16, RPTLN-21*
	Helodermatidae	*H. horridum*	*RPTLN*-1
	Gekkonidae	*U. ebenaui*	*RPTLN*-1, *RPTLN*-9
		T. mauritanica	*RPTLN*-1, *RPTLN*-11
	Chamaeleonidae	*C. calyptratus*	*RPTLN*-1, *RPTLN*-17
Crocodrile	Crocodylidae	*A. mississipiensis*	*RPTLN*-1, *RPTLN-6, RPTLN-11, RPTLN-12, RPTLN-14*
Tortoise	Testudinidae	*T. greca*	*RPTLN*-1, *RPTLN-3*
		T. hermanii	*RPTLN*-1
		S. pardalis	*RPTLN*-1, *RPTLN*-16, *RPTLN*-18
	Geomydidae	*M. annamensis*	*RPTLN*-1, *RPTLN*-15
		M. sintesis	*RPTLN*-1, *RPTLN*-4, *RPTLN*-7, *RPTLN-18*
		M. leprosa	*RPTLN*-1, *RPTLN*-2, *RPTLN*-5
		C. carbonaria	*RPTLN*-1, *RPTLN*-2
		C. chilensis	*RPTLN*-1, *RPTLN*-16

Table 11. Nucleotide changes in *RPTLN*-n genes respect to *RPTLN*-1 [jerdostatin, AY262730]

RPTLN gene copy	Nucleotide substitutions respect *RPTLN*-1
RPTLN-2	41T>C
RPTLN-3	50A>G
RPTLN-4	50A>G;173A>G;179A>G
RPTLN-5	del65C
RPTLN-6	84T>C
RPTLN-7	87T>C
RPTLN-8	128C>T
RPTLN-9	128C>T; 259T>A
RPTLN-10	145C>T
RPTLN-11	145C>T; 185T>C
RPTLN-12	145C>T; 185T>C; 220T>C
RPTLN-13	155A>G
RPTLN-14	173A>G
RPTLN-15	173A>G; 179A>G
RPTLN-16	173A>G; 179A>G; 255A>G
RPTLN-17	220T>C
RPTLN-18	249A>G
RPTLN-19	249A>G; 260G>A
RPTLN-20	253A>G
RPTLN-21	253A>G; 279T>C

Also remarkable is the structural conservation of these genes (Figure 54) in taxa that had a common ancestor ≥ 250 million years ago (Hedges 1999; Hedges and Vidal 2009; Pyron et al. 2013) (Table 10) (Figure 56). In particular, the nucleotide stretch 1-54 shares >96% identity with nucleotide sequences encoding the signal peptide of snake venom metalloproteinase (SVMP) and short-coding RGD-disintegrin precursors from a large number of Viperinae and Crotalinae snake species; *RPTLN* nucleotides 55-201 show 86-93% identity with pro-peptide-encoding nucleotide sequences for *M. lebetina* [AY835996, X97894] and *D. russelii* [GQ420354] PIII- and PII-SVMPs; and the nucleotide sequence 202-333 only matches homologous sequences from *M. lebetina* [AM114015, AM261813], *C. vipera* [AM114012], and *P. jerdonii* [AY262730] encoding RTS- and KTS-disintegrin domains, which exhibit a high degree of identity (94-100%) in any pairwise comparison.

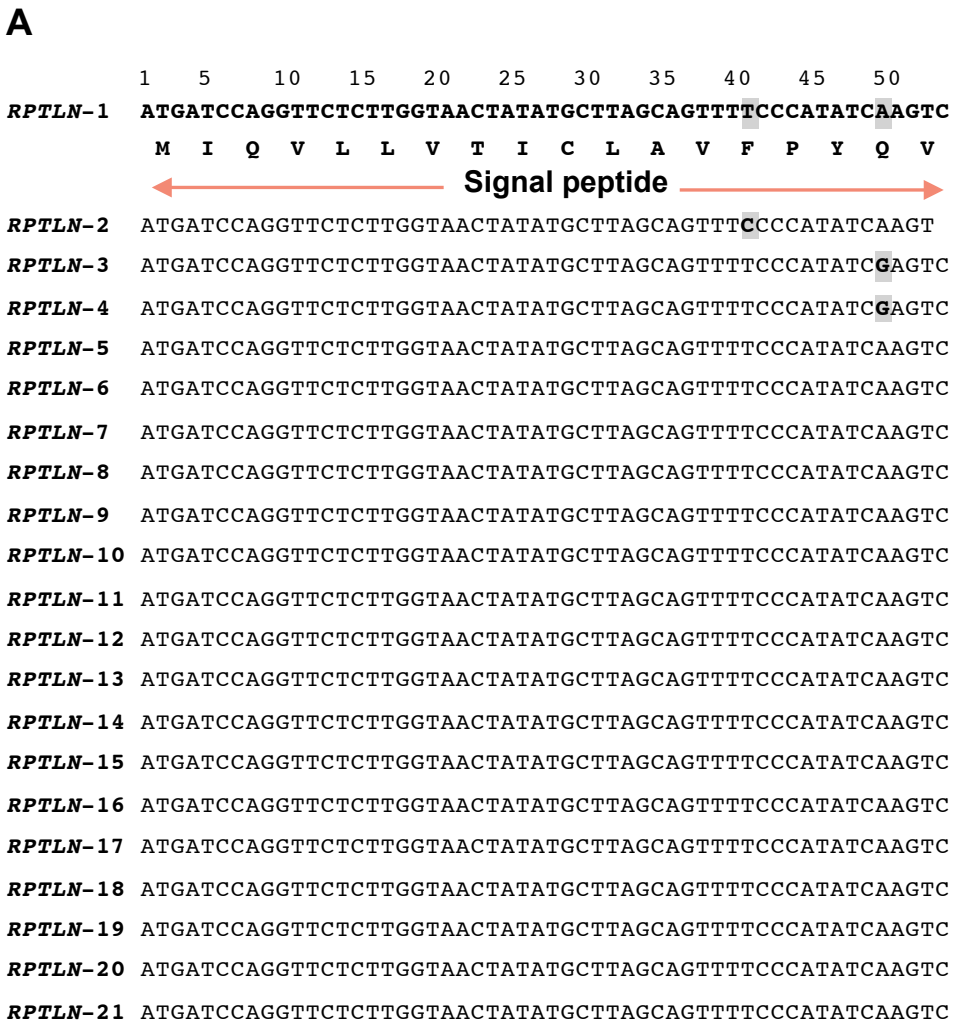


Figure 54. Multiple sequence alignments *RPTLN*-1 and the twenty unique *RPTLN* genes amplified from gDNA or mRNA of the different reptile species listed in Table 10. For convenience, the alignment has been divided into three blocks of sequences, each of which codes for a domain of a hypothetical protein. Nucleotide changes between *RPTLN* sequences are highlighted in boldface and in grey background. The distribution of these sequences across the phylogeny of reptiles is shown in Figure 56.

B

```
          55   60   65   70   75   80   85   90   95  100  105  110  115  120  125  130  135  140  145  150  155  160  165  170  175  180  185  190  195  200
RPTLN-1   AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
          S  S  K  T  L  K  S  G  S  V  N  E  Y  E  V  V  N  P  G  T  V  T  G  L  P  K  G  A  V  K  Q  P  E  K  K  H  E  P  M  K  G  N  T  L  Q  K  L  P  L
                                                                  propeptide
RPTLN-2   AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-3   AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-4   AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAGAGGGAGCACATTGCAGAAACTTCCCCTT
RPTLN-5   AGCTCTAAAA-CCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-6   AGCTCTAAAACCCTGAAATCTGGGAGTGTCAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-7   AGCTCTAAAACCCTGAAATCTGGGAGTGTTAACGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-8   AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCTCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-9   AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCTCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-10  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGTAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-11  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGTAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATCGCAGAAACTTCCCCTT
RPTLN-12  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGTAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATCGCAGAAACTTCCCCTT
RPTLN-13  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAGAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-14  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAGAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-15  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAGAGGGAGCACATTGCAGAAACTTCCCCTT
RPTLN-16  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAGAGGGAGCACATTGCAGAAACTTCCCCTT
RPTLN-17  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-18  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-19  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-20  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
RPTLN-21  AGCTCTAAAACCCTGAAATCTGGGAGTGTTAATGAGTATGAAGTAGTAAATCCAGGAACAGTCACTGGATTGCCCAAAGGAGCAGTTAAGCAGCCTGAGAAAAAGCATGAACCCATGAAAGGGAACACATTGCAGAAACTTCCCCTT
```

C

```
                205  210  215  220  225  230  235  240  245  250  255  260  265  270  275  280  285  290  295  300  305  310  315  320  325  330
RPTLN-1    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
                C  T  T  G  P  C  C  R  Q  C  K  L  K  P  A  G  T  T  C  W  R  T  S  V  S  S  H  Y  C  T  G  R  S  C  E  C  P  S  Y  P  G  N  G  -
                                                                                 disintegrin domain
RPTLN-2    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-3    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-4    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-5    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-6    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-7    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-8    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-9    TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCAGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-10   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-11   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-12   TGTACAACTGGACCATGTCGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-13   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-14   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-15   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-16   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACGTGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-17   TGTACAACTGGACCATGTCGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-18   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGGACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-19   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGGACAACATGCTAGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-20   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAGCATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
RPTLN-21   TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAGCATGCTGGAGAACCAGTGTATCAAGCCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGCTAA
```

The unusual high conservation of *RPTLN* genes across Reptilia suggests a relevant function in reptile biology for this ancient gene family. Whether *RPTLN* genes are i) translated into (body) protein(s), ii) acts as regulatory RNA molecules, or iii) serve other unknown function(s), remains elusive. Of relevance to this point, among the 31 nucleotide changes identified in the 22 *RPTLN* genes listed in Table 11; 6 involve the third base of *RPTLN* codons, whereas 8 and 17 affect first and second codon positions, respectively (Figure 54). In protein-coding DNA sequences, the second-codon position is the most functionally constrained, whereas, due to the degenerate nature of the genetic code, the third-codon position is the least functionally constrained in terms of nucleotide changes (Bofkin and Goldman, 2007). It is thus tempting to hypothesize that the biological role of transcribed *RPTLN* RNA may be strongly dependent on their folded structure. Supporting this hypothesis, the RNAfold WebServer (Gruber *et al.*, 2008) predicted for the full-length *RPTLN*-transcribed 333 bp RNA a stable (-90.59 kcal/mol minimum free energy) secondary structure ensemble (Figure 55).

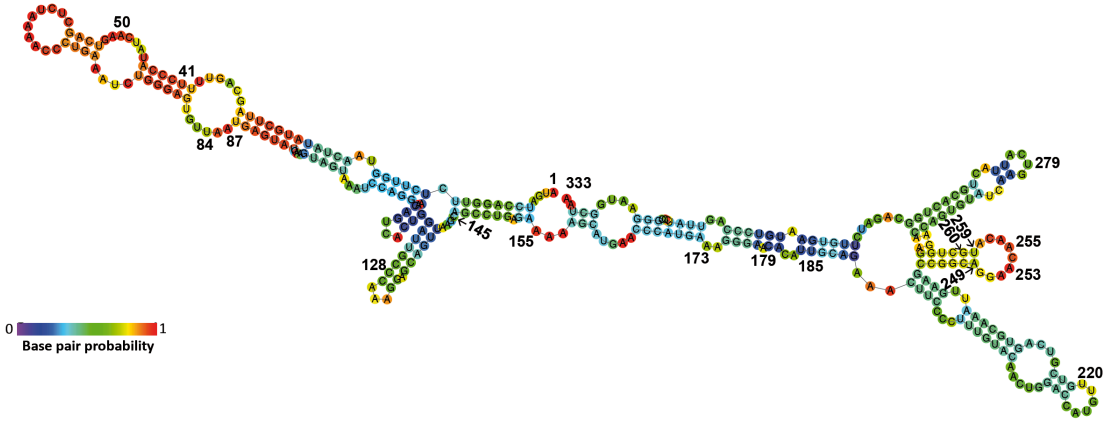


Figure 55. Scheme of the predicted secondary structure of *RPTLN RNA*. Nucleotide positions mutated in any of the *RPTLN*-2-21 genes listed in Table 1 are indicated.

RPTLN genes were found at nodes predating the separation of Toxicofera and Lacertidae (Figure 56), and thus before venom arose in squamate evolution, approximately 170 Mya during the Jurassic period (Fry et al. 2006; 2012). Despite their

broad distribution across the phylogeny of Reptilia (Figure 56), *RPTLN* genes have been found translated into KTS/RTS disintegrins only in the venoms of *M. l. obtusa* (obtustatin (KTS) [P83469; 1MPZ], (Marcinkiewicz et al. 2003; Moreno-Murciano et al. 2003a; Sanz et al. 2008)), *M. lebetina, M. mauritanica* (lebestatin (KTS) [CAJ34939], (Olfa et al. 2005; Makran et al. 2012)), *D. palestinae* (viperistatin (KTS) [P0C6E2], (Kisiel et al. 2004)), and *D. russelii* (russellistatin (RTS), as we had shown previously). Evolutionary relationships reconstruction inferred from mitochondrial DNA sequences dated the emergence of Eurasian viper (genera *Eristicophis*, *Pseudocerastes*, *Vipera*, *Macrovipera*, and *Daboia*) in the early Miocene (23-16 million years ago, Mya) (Lenk 2001), coinciding with the geographical separation of the landmasses Europe, Middle East, and North Africa by the Mediterranean and Parathethys seas (Rögl and Steininger 1983). This evidence indicates that *RPTLN* genes comprise an ancient multigenic family, and that their restricted expression and neofunctionalization in the venom gland of *Macrovipera* and *Daboia* species represent recent events (Figure 56).

A hallmark of the *RPTLN* genes transcribed into protein (KTS- and RTS-disintegrin)-coding mRNAs with respect to those that are not translated is the accumulation of mutations in the C-terminal half of the disintegrin domain (Sanz-Soler et al. 2012) (Figure 38, p.78), which constitutes a conformational functional epitope encompassing the $\alpha_1\beta_1$ integrin-inhibitory loop and the C-terminal tail of KTS/RTS disintegrins (Kisiel et al. 2004). However, the mechanism underlying this non-protein-coding to protein-coding transition remains elusive. It is tempting to speculate that accumulation of nucleotide changes in certain *RPTLN* gene copies may have resulted in destabilization of the transcribed non-coding RNA (ncRNAs) into a translatable, or in the formation of pseudogenes. In this context, it is worth mentioning that ncRNAs transcribed from pseudogenes may play regulatory roles regulating the expression of their parental or non-parental genes (Mighell et al. 2000; Balakirev and Ayala 2003; Trinklein et al. 2007; Muro et al. 2011; Pink et al. 2011).

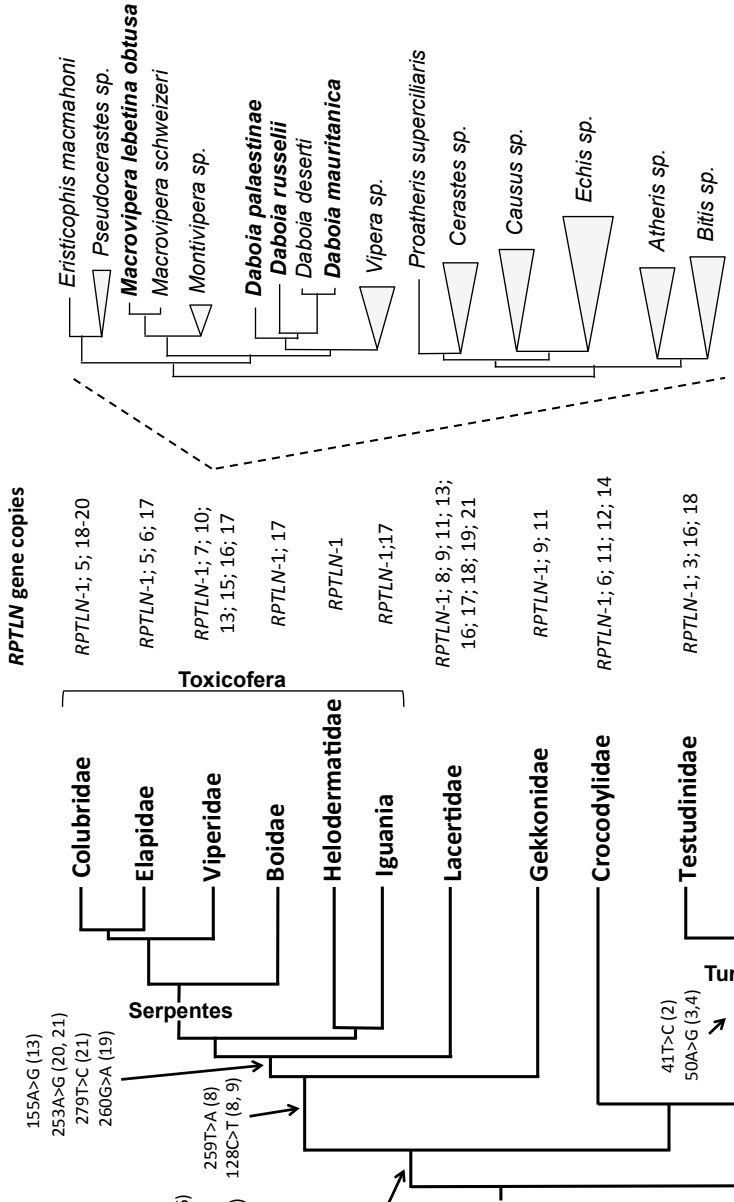


Figure 56. Distribution of the *RPTLN* genes listed in Table 1 in the phylogeny of reptiles. Tree positions where mutations associated with the 20 different *RPTLN* genes arose are indicated. The identity of the particular *RPTLN*-n (n = 2-21) isogenes containing the mutation(s) are specified in parentheses. Viperinae branch is expanded, and those species from which venom RTS/KTS disintegrins have been isolated, are highlighted in bold.

4 A hypothesized role for *RPTLN* in the evolution of SVMPs

The nucleotide sequence encoding the signal peptide (SP) of *RPTLN* genes is highly conserved in SVMP precursor genes of Viperinae, Crotalinae, Elapidae, and Colubridae snake species (Figure 57). On the contrary, this region is not conserved in ADAM (A Disintegrin And Metalloprotease) genes (Figure 57). The closest non-venom ancestor of SVMPs was likely an ADAM28 precursor gene (Casewell 2012) that was recruited into the snake venom gland proteome (Moura-da-Silva et al. 1996) after the divergence of squamate reptiles, lizards and snakes (Fry et al. 2006; 2012) in the Jurassic, ~170-150 million years before present (MYBP) (Hedges and Vidal 2009). The high conservation in extant SVMPs of the SP sequence coded for by *RPTLN* genes strongly suggests that this region may have played a key role in the recruitment and restricted expression of SVMP genes in the venom gland of Caenophidian snakes. In this respect, the exon-intron organization of pre-pro *E. ocellatus* EOC00089-like PIII-SVMP and *A. carolinensis* ADAM28 genes is conserved, and their 17-residue signal peptides are entirely coded for by exon 1 (Sanz *et al*., 2012). Most introns of *A. carolinensis* ADAM 28 contain inserted retroelements capable of invading new genomic sites (Alföldi et al. 2011), particularly short interspersed retrotransposable elements (Sauria SINE) in introns 1, 3-8, 10, 13, 14 and 16, and LINEs (long interspersed elements) in introns 1 and 10. The family of Sauria SINEs are widely distributed among genomes of lizards, snakes, and tuataras (Piskurek et al. 2006; 2009). Sauria SINEs arose more than 200 million years ago, and the members of this family comprise a 5' tRNA-related region, a tRNA-unrelated region, and a 3' tail region identical with Bov-B LINEs (Piskurek et al. 2006). Their retrotransposition depends on reverse transcriptase and endonuclease activities encoded by partner LINEs, and it has been proposed (Piskurek et al. 2006) that Sauria SINEs utilize the enzymatic machinery of Bov-B LINEs for their retrotransposition.

Species	Accession code	nucleotide sequence	Translated AA sequence	% nt identity
		1 5 10 15 20 25 30 35 40 45 50		
Protobothrops jerdonii RPTLN-1	AY262730	ATGATCCAGGTTCTCTTGGTAACTATATGCTTAGCAGTTTTCCCATATCAAG	MIQVLLVTICLAVFPYQ	100
Atheris squamigera SVMP	HF543864	ATGATCCAGGTTCTCTTGGTAATTATATGCTTAGCAGTTTTCCCATATCAAG	MIQVLLVTICLAVFPYQ	98
Viridovipera stejnegeri SVMP	DQ335449	ATGATCCAGGTTCTCTTGGTAATTATATGCTTAGCAGTTTTCCCATATCAAG	MIQVLLVTICLAVFPYQ	98
Gloydius intermedius SVMP	KM435346	ATGATCCAGGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCTTATCAAG	MIQVLLVIICLAVFPYQ	96
Echis pyramidum leakeyi SVMP	GU012290	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCATATCAAG	MIQVLLVTICLAVFPYQ	96
Echis coloratus SVMP	GU012229	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCATATCAAG	MIQVLLVTICLAVFPYQ	96
Macrovipera lebetina SVMP	DQ288157	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCATATCAAG	MIQVLLVTICLAVFPFQ	96
Crotalus adamanteus SVMP	HQ414112	ATGATCCAGGTTCTCTTGGTAACTATATCCTTAGCAGTTTTTCCTTATCAAG	MIQVLLVTISLAVFPYQ	94
Agkistrodon p. leucostoma SVMP	GQ451443	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCTTATCAAG	MIQVLLVTICLAVFPYQ	94
Bothrops insularis SVMP	AY736107	ATGATCCAGGTTCTCTTGGTAACTATATGCTTAGCAGCTTTTCCTTATCAAG	MIQVLLVTICLAAFPYQ	94
Ovophis okinavensis SVMP	AB851968	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTCTTTCCTTATCAAG	MIQVLLVTICLAVFPYQ	94
Gloydius saxatilis SVMP	AY204244	ATGATCCAGGTTCTCTTGGTAACTATATGCTTAGTAATCTTTCCATATCAAG	MIQVLLVTICLVIFPYQ	94
Ophiophagus hannah SVMP	EF065674	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGTGGTTTTTCCATATCAAG	MIQVLLVTICLVVFPYQ	92
Thamnophis sirtalis SVMP	XM_014065955	ATGATCCAAGCTCTCTTGGTAACTATATGCTTAGTGGTTTTTCCATATCAAG	MIQALLVTICLVVFPYQ	90
Anolis carolinensis ADAM28	XM_008119851	ATGATCAAAGCACTCCTTCTGGCTTTGTGCTTCGTGCTTTTCCAGCATCAAG	MIKALLLALCFVLFQHQ	65
Anolis carolinensis ADAM23	XM_008114972	ATGAAGCCGCCTGGAAGGAGAAGTCCTCTGCGGCAGACCTCCTCCTCTTCTC	MKPPGRRSPLRQTSSSS	ns
Anolis carolinensis ADAM9	XM_003226844	ATGGCTGCTTGGAGCTGCTGCGCCTGGCTTTGGCTCCTGTGCGCCCTGAGCG	MAAWSCCAWLWLLCALS	ns
Homo sapiens ADAM28	BC136478	ATGTTGCAAGGTCTCCTGCCAGTCAGTCTCCTCCTCTCTGTTGCAGTAAGTG	MLQGLLPVSLLLSVAVS	38
Homo sapiens ADAM7	NM_003817	ATGCTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAG	MLPGCIFLMILLIPQVK	ns
Gallus gallus ADAM28	XM_001233495	ATGAATAATGTCATTCTCATTTTTGTTGTTTTGCTATGCTTTCTCCATCAAG	MNNVILIFVVLLCFLHQ	44

Figure 57. **Comparison of the nucleotide and encoded signal peptide sequences of *RPTLN*-1, SVMPs, and ADAM molecules. SVMP and ADAM residues differing from the nucleotide or translated amino acid sequence of *RPTLN*-1 are highlighted in bold and in gray background.**

5 Transcription and translation of *RPTLN* genes in *P. muralis*, *P. hispanica* (Lacertidae), and *R. scalaris* (Colubridae) organs

Full-length *RPTLN* transcripts were amplified by PCR in different organs and tissues of *P. muralis*, *P. hispanica*, and *R. scalaris* (Table 12).

Table 12. *RPTLN* gene copies in organs of Colubridae and Lacertidae taxa. Full-length sequences are displayed in Figure 54. Nucleotide changes in *RPTLN* genes respect to *RPTLN*-1 [jerdostatin, AY262730] are listed in Table 11. See accession codes in Table 2S (p.172).

Species	Organ	*RPTLN* gene copy
R. scalaris	Lung	*RPTLN*-1
	Heart	273A>T (*RPTLN*-22)
	Muscle	*RPTLN*-21
	Skin	*RPTLN*-5
P. muralis	Bladder	*RPTLN*-1, *RPTLN*-18
	Liver	*RPTLN*-1
	Lung	*RPTLN*-1, *RPTLN*-17
	Kidney	*RPTLN*-1, *RPTLN*-18
	Muscle	*RPTLN*-1
	Skin	*RPTLN*-18, *RPTLN*-19
	Stomach	*RPTLN*-18
	Heart	*RPTLN*-9, *RPTLN*-18
P. hispanica	Liver	*RPTLN*-1
	Lung	*RPTLN*-11
	Skin	*RPTLN*-1, *RPTLN*-15
	Stomach	*RPTLN*-8, *RPTLN*-13
	Heart	*RPTLN*-1
	Brain	*RPTLN*-21

These results provide the view of a ubiquous distribution of the *RPTLN-like* RNA in several tissues of the Colubridae and Lacertidae (Table 12). Then, *RPTLN-like* RNA levels were compared in different *Podarcis muralis* tissues. Semiquantitative PCR amplification suggested differential transcription levels at different organs of *P. muralis* (Figure 58).

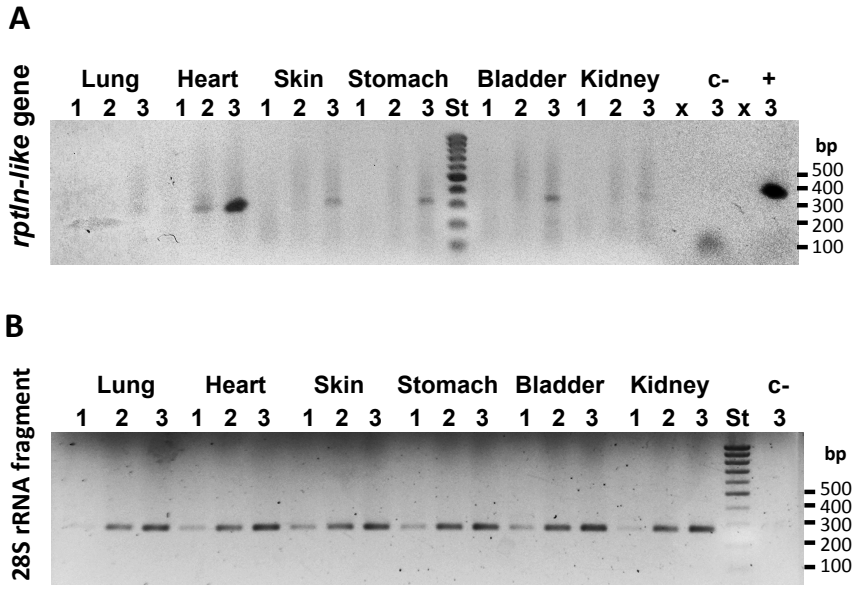


Figure 58. *RPTLN-like* RNA expression in *Podarcis muralis*. (A) Semiquantitative PCR-amplification of *RPTLN-like* from coding DNA (cDNA) of several organ of *Podarcis muralis* (lung, heart, skin, bladder and kidney), *Mus musculus* (Mm) and negative control without DNA (-) PCR-amplification program with (1) 25 reaction cycles, (2) 30 reaction cycles and (3) 35 reaction cycles (B) Semiquantitative PCR-amplification of a housekeeping gene, 28S ribosomal RNA fragment, using as a template cDNA of the different *Podarcis muralis* organs.

However, attempts to quantify the levels of *RPTLN* RNA by quantitative PCR amplification failed due to the low amount of transcripts. Low expression levels may arise from alternative promoter usage of ncRNAs compared to protein-coding RNAs (The Fantom Consortium *et al.*, 2005).

On the other hand, we did not find evidence for *RPTLN* translation in any organ investigated by Western blot analysis of 40-60 µg of total proteins extracted from 200 µg organ homogenates and size-fractionated by ultrafiltration (≥ 10 kDa, 10-3 kDa, and ≤ 3 kDa fractions) (Figure 59). Recombinant r-jerdostatin was used to estimate the immunodetection limit, which was 50 ng (Figure 59, lines J). For comparison, the short KTS-disintegrin obtustatin [P83469] (Sanz et al. 2008), and its homologs, the RTS-disintegrin russellistatin (Figure 48, p.92) (Sanz-Soler et al. 2012) and the KTS-disintegrin lebestatin [Q3BK14] (Makran et al. 2012), comprise 2.8%, 2%, and 7.8% of the total venom proteins of *Macrovipera lebetina obtusa*, *Daboia russelii*, and

Macrovipera (*Daboia*) *mauritanica*, respectively. Furthermore, expression yields for functionally active recombinant jerdostatin (wild-type and mutants) in *E. coli* were about 0.5-2 mg/L of cell culture (Sanz et al. 2005; Sanz-Soler et al. 2012). These data clearly show that *RPTLN* and *RPTLN*-like DNA sequences can be transcribed and translated into functional proteins in different cellular environments.

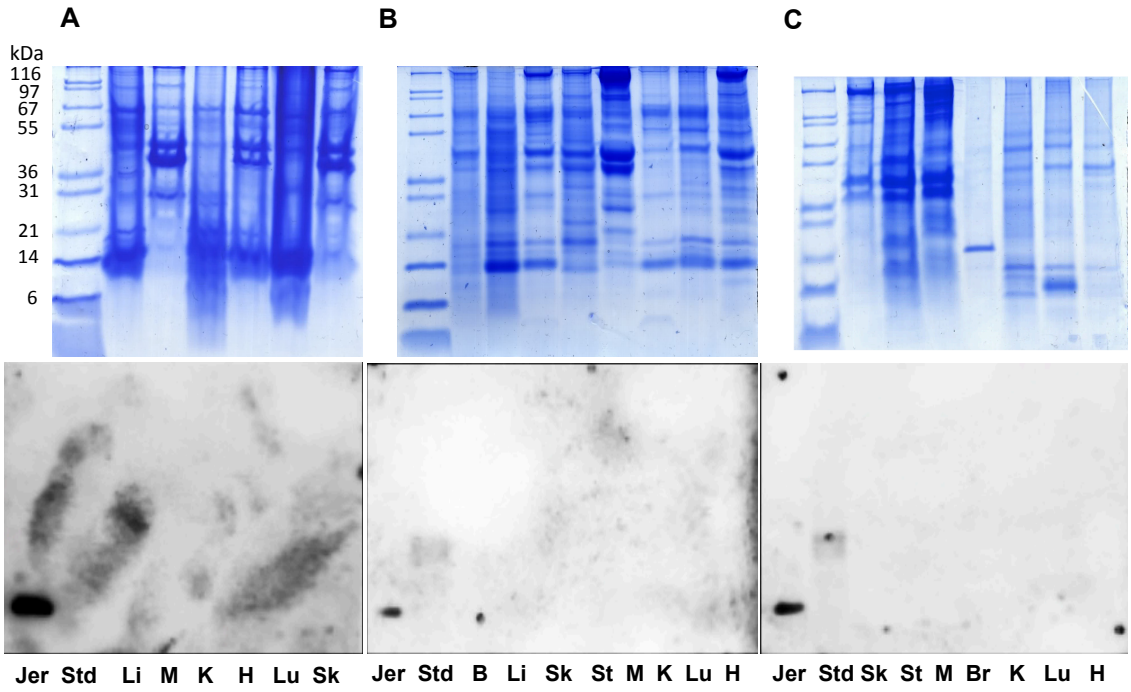


Figure 59. Coomassie blue-stained SDS-PAGE (upper panels) and Western blot analyses (lower panels) probed with anti-jerdostatin PEP160 polyclonal antibodies of 40-60 µg of total proteins extracted from different organs of *R. scalaris* (A), *P. muralis* (B), and *P. hispanica* (C). Positive control, lane Jer, ~50 ng of purified recombinant jerdostatin; Immunoreactivity is only observed against the positive control. Std, molecular mass standard Mark12™ (Invitrogen); Li, liver; M, skeletal muscle; K, kidney; H, heart; Lu, lung; Sk, skin; St, stomach; Br, brain.

Although the possibility that *RPTLN* genes are translated into very low protein concentration can not be ruled out, all available data support the view that *RPTLN* gene copies encode a long (>200 nt) ncRNAs (lncRNAs). Eukaryote genomes include tens of thousands of long noncoding RNAs with little or no protein-coding capacity (Wilusz et al. 2009; Wilusz 2015). Only a limited number of lncRNAs have been functionally characterized. However, paradigms for how lncRNAs exert regulatory functions are beginning to emerge (reviewed by Wilusz et al. 2009; Wilusz 2015). In particular,

lncRNAs are both regulated by unique post-transcriptional control mechanisms, and control various aspects of post-transcriptional processing of mRNAs. These functions often involve the formation of ribonucleoprotein, RNA-RNA, and DNA-RNA complexes. These features of lncRNAs provide a meaning to the conservation of the third-base of *RPTLN* codons, and support our view of the possible biological function of this reptile-specific gene.

6 Concluding remarks, reflexions and perspectives

This work elaborates upon a previous work in chapter I and (Sanz-Soler *et al.*, 2012) in which we show results providing strong support for an independent evolutionary history of the RTS/KTS and the RGD clades of short disintegrins. This assumption is also supported by our findings described in this chapter, in which we report full-length *RPTLN* gene sequences amplified from species at nodes predating the separation of Toxicofera and Lacertidae, and thus preceding the emergence of venom in the evolution of squamate reptiles, ~170 Mya during the Jurassic period (Okuda et al. 2002; Fry 2005; Hedges and Vidal 2009; Reeder et al. 2015) (Figure 60).

The remarkable structural conservation of *RPTLN* genes across Reptilia, their low transcriptional level, and the lack of evidence for *RPTLN* translation in any reptile organ investigated (Figure 60), suggest a yet elusive role for transcribed *RPTLN* RNA as a long non-protein-coding RNA (Figure 61).

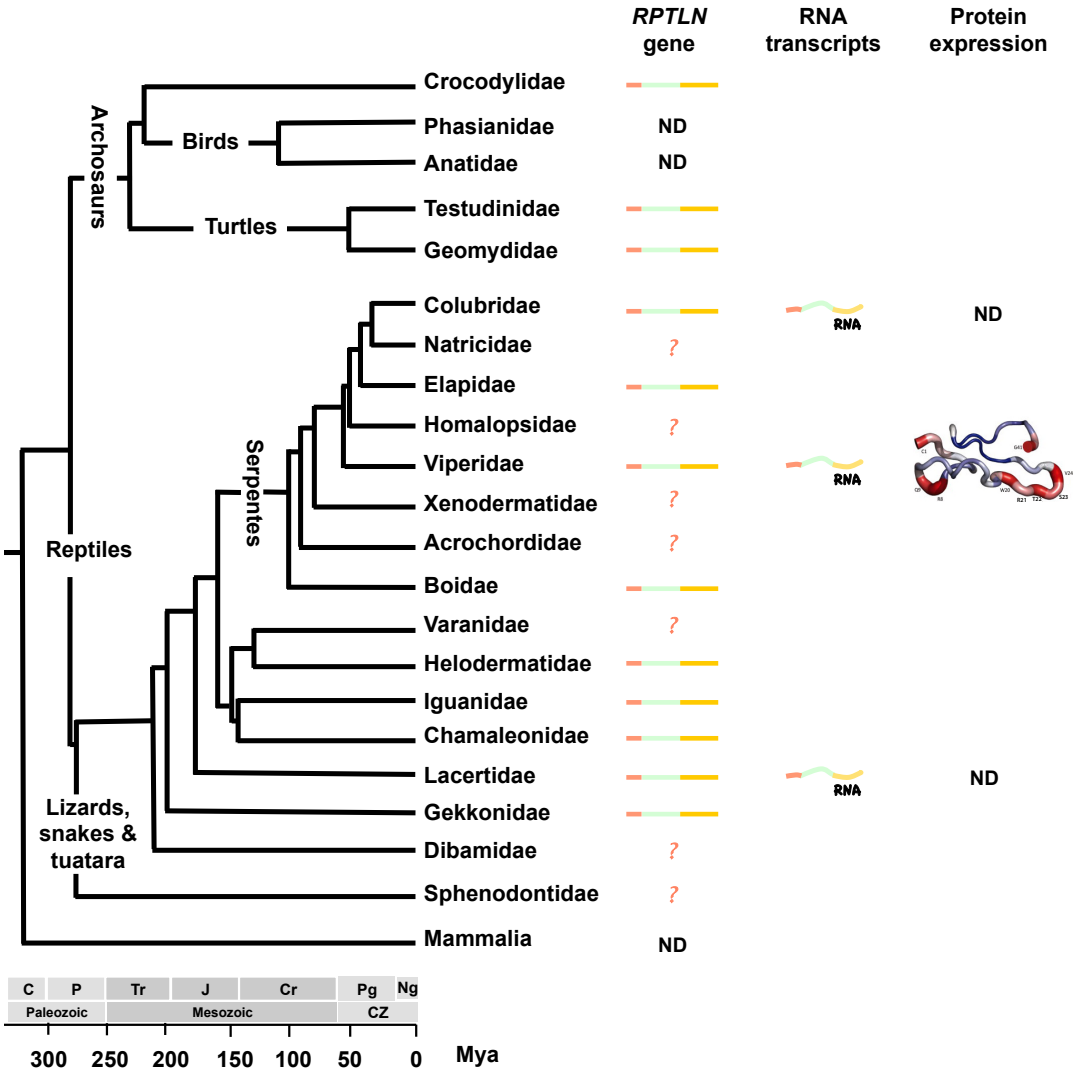


Figure 60. Reptile cladogenesis, including *RPLTN* gene detection (*RPTLN* gene cartoon, signal peptide (coral), propeptide (light green), disintegrin domain (yellow)) and non-detection (ND) in different amniotes species. Non-analyzed families, in *RPTLN* gene studio, are designed by a question mark. RNA cartoons are displayed next to the families, in which *RPTLN* RNA sequences have been detected. Protein expression (in venom gland of Eurasian vipers) is displayed by the jerdostatin structure (Shimokawa et al. 1998; Sanz et al. 2006; Calvete et al. 2009b; Calvete 2010; Carey et al. 2012) and non-detected protein expression by ND. The rest of families were not analyzed, neither RNA transcription nor protein expression. Mammalia represents the out-group. Mya, Millions years ago.

We hypothesize that the high conservation of the SP sequence of *RPTLN* and extant SVMP genes may suggest a functional role for this region in the ancestral recruitment of SVMP gene expression in the venom gland of Caenophidian snakes (Figure 61). The origin of SVMPs has been inferred to have occurred after the split of the Pareatidae from the remaining Caenophidians, ~50-60 Mya, during the Paleogene period of the Cenozoic Era (Calvete et al. 2009b; Calvete 2010; Casewell et al. 2012; Gauthier et al. 2012; Carbajo et al. 2015). The evolutionary path that led to the family of the RGD/XXD-disintegrins from PII-SVMPs has been dissected in some detail at the molecular and structural levels (Nei and Rooney 2005; Juárez et al. 2008; Townsend et al. 2011; Pyron et al. 2013). Neofunctionalization of the *RPTLN* gene to express RTS/KTS disintegrins (Figure 61) represents an independent alternative route from the evolution of PII-SVMP-derived disintegrins, which occurred more recently in venoms of Eurasian vipers within genera *Macrovipera* and *Daboia* (Table 9), in the early Miocene (~23-16 million years ago, Mya) (Juárez et al. 2008; Jones et al. 2013). The finding of an RTS-disintegrin (hitherto termed russellistatin) in the venom of *D. russelii* (Figure 48, 49 and 50, p.92 and 93) also supports this hypothesis.

Figure 61 outlines a cartoon of processes in which we hypothesize that *RPTLN* may have been involved. Understanding the physiological function and evolutionary history of this enigmatic highly conserved (and thus presumably relevant) gene across the phylogeny of reptiles clearly requires further detailed molecular studies. In particular, comparative analysis of the upstream regions of non-protein coding and protein coding *RPTLN* genes may identify nucleotide motifs that contribute to regulate their distinct molecular fates. On the other hand, the hypothesis that *RPTLN* may have played a key role in the recruitment and restricted expression of SVMP genes in the venom gland of Caenophidian snakes predicts that venom gland *RPTLN* and SVMP genes may share tissue-specific regulatory elements. Future genomic studies should support or refuted our hypothesis.

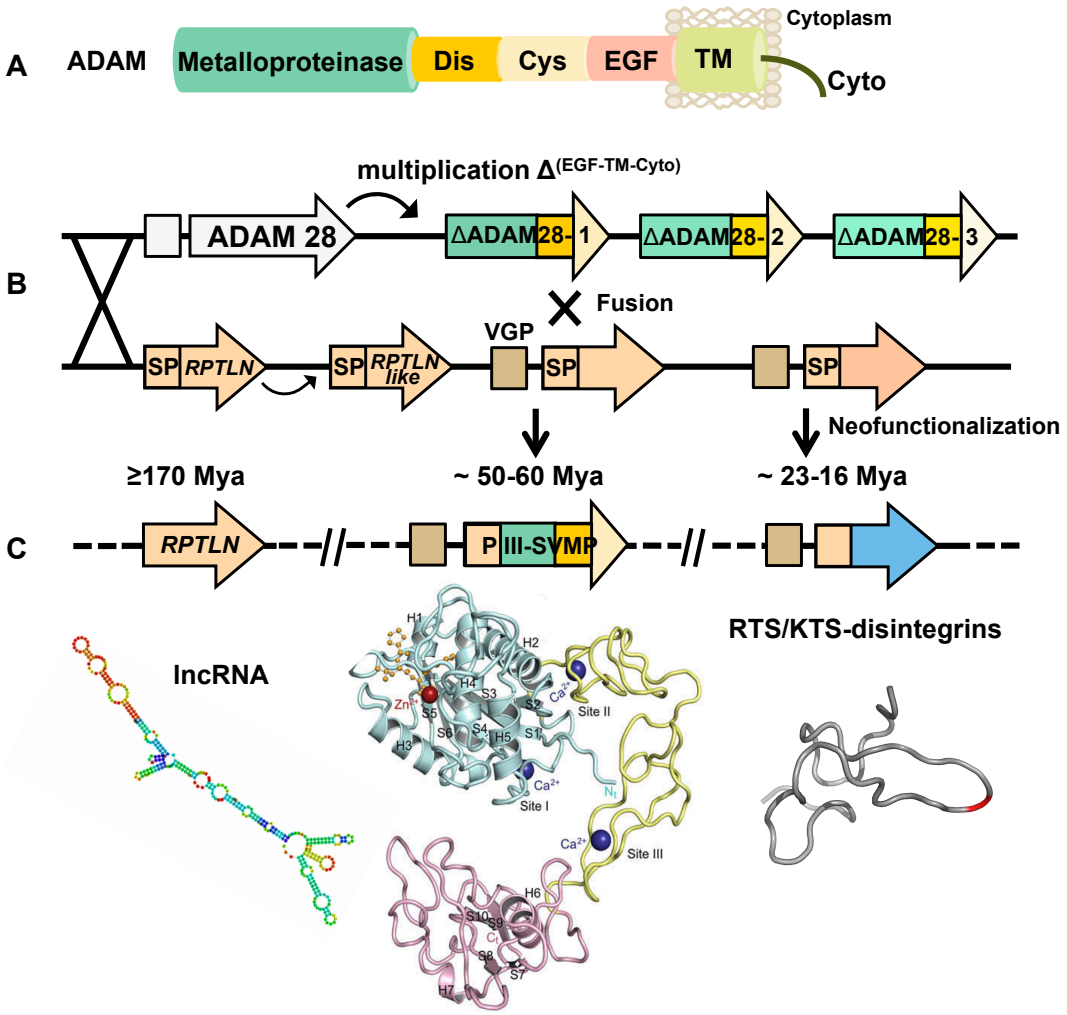


Figure 61. Cartoon of processes in which the *RPTLN* gene is hypothesized to have been involved during its long evolutionary history. A, A Disintegrin And Metalloproteinase (ADAM) different structural domains are displayed in the cartoon. B, Fusion of the ADAM28 extracellular domains-coding gene region $^{\Delta EGF/TM/Cyto}$ ADAM28 and an *RPTLN* gene under a venom gland-specific promoter (VGP) generated a SVMP gene bearing the *RPTLN* signal peptide (SP) sequence and exhibiting venom gland restricted translation (C). C, Timing of the functional processes in which *RPTLN* is hypothesized to have been involved, i.e. a yet elusive role for transcribed *RPTLN* as a long non-protein-coding RNA since ≥ 170 Mya; the recruitment and venom gland-restricted expression of PIII-SVMPs, 50-60 Mya; and more recently (23-16 Mya), its own neofunctionalization into $\alpha_1\beta_1$-inhibitory short RTS/KTS disintegrins in venoms of certain Eurasian viper species within genera *Macrovipera* and *Daboia* (Figure 56). Dis, disintegrin-like domain; Cys, cysteine-rich domain; EGF, epithelial growth factor-like domain; TM, transmembrane domain; cyto, cytoplasmic domain.

Chapter III Supporting the idea of a new venom specific function for disintegrins in viperid snakes envenomation

Snakes use an array of predatory strategies and behaviors for overcoming potentially dangerous prey. Non-venomous snakes utilize constriction and/or jaw holding to subdue prey, whereas venomous snakes immobilize and kill prey through the use of venom. Although venomous snakes utilize the same chemical means (venom) for dispatching prey, the mode of venom delivery (prey envenomation) significantly differs between families of venomous snake. For example, elapid snakes (cobras and coral snakes) commonly strike-and-hold prey, whereas a strike-and-release envenomation is observed in viperid (vipers and pit vipers) snakes. For vipers, the strategy of releasing prey allows the snake to avoid any injury that may arise due to retaliation from a potentially dangerous prey item. However, it adds an extra task of relocating the envenomated prey that has wandered from the attack location. This undertaking can be challenging due to confounding chemical cues of both non-envenomated conspecific and heterospecific prey sources. By using rapid tongue flicking (strike-induced chemosensory searching) to detect, and the vomeronasal organ to analyze volatile and non-volatile chemical cues (Schwenk 1995) (Figure 4, p.11), several snake behavioral studies have addressed how vipers relocate envenomated prey, indicating that these snakes use vomeronasal chemoreception to discriminate between the chemical cues of envenomated and non-envenomated prey sources (Chiszar et al. 1999; Greenbaum et al. 2003; Greenbaum 2004; Chiszar et al. 2008; Saviola et al. 2010). Recent studies in the laboratory of Dr. Stephen Mackessy (University of Northern Colorado), (Saviola et al. 2013), suggest that the venom component responsible for successful recovery of envenomated prey is contained in a specific fraction of proteins in the *Crotalus atrox* venom, in which the medium size disintegrins, crotatroxin-1 and 2, are included.

These results suggested that disintegrins in the venom of *C. atrox* could act as the relocator element during predatory episodes. However, it may be possible that some additional molecules might be contained in this defined venom fraction that were not detected by mass spectrometry. Therefore, to further assess that the disintegrins in *C. atrox* venom are the relocator compound, and to rule out the possibility of other molecules from *C. atrox* venom contributing to prey relocation, we generated and behaviorally tested a recombinant form of crotatroxin. This allowed us to avoid the presence of other elements from *Crotalus atrox* venom.

1 Crotatroxin nucleotide sequence determination.

The stability of mRNA in the snake venom is biologically fascinating (Currier et al. 2012). Therefore, the laboratory of Dr. Mackessy have been developed a new methodology to isolate total RNA from fresh snake venom (Modahl et al., in press) (Saviola et al. 2015). To confirm a new function for disintegrins in Viperidae venoms, crotatroxin-coding DNA was determined following Modahl et al., (*in press*) protocol. RNA was extracted from fresh *Crotalus atrox* venom and subsequently complementary DNA (cDNA) was synthesized and used as a template for the PCR amplification of the crotatroxin coding DNA fragment, using a forward degenerated disintegrin oligonucleotide. The DNA insert was further cloned and sequenced, resulting in the sequence presented in Figure 62.

```
ggagaagaatgtgactgtggctctcctgcaaatccgtgctgcgatgctgcaacc  54
 G  E  E  C  D  C  G  S  P  A  N  P  C  C  D  A  A  T   18
tgtaaactgagaccaggggcacagtgtgcagatggactatgttgtgaccagtgc 108
 C  K  L  R  P  G  A  Q  C  A  D  G  L  C  C  D  Q  C   36
agatttattaaaaaaggaacagtatgccggccagcaaggggtgattggaatgac 162
 R  F  I  K  K  G  T  V  C  R  P  A  R  G  D  W  N  D   54
gatacctgcactggccaatctgctgactgtcccagaaatggcctctatggctaa 216
 D  T  C  T  G  Q  S  A  D  C  P  R  N  G  L  Y  G  -   71
```

Figure 62. Isolated crotatroxin nucleotide sequence and the translated amino acid sequence. RGD motif is labeled in red and highlighted, Cysteine residues are in bold.

2 Recombinant crotatroxin: cloning, expression, purification and mass spectrometry identification

According to the crotatroxin DNA sequences, appropriate specific primers were designed to generate the crotatroxin expression vector, including the cleavage site for the tobacco etch virus (TEV) protease in the N-terminus. Similar to the recombinant ocellatusin construction, the amplified DNA insert was cloned in the expression vector pET32a(+) (Novagen, Madison, WI). pET32a(+)/TEV-crotatroxin positive clones were selected and sequenced to determine that the crotatroxin nucleotide sequence was inserted in the correct reading frame.

```
                1         10         20         30         40         50         60         70
r-crotatroxin   GGEECDCGSPANPCCDAATCKLRPGAQCADGLCCDQCRFIKKGTVCRPARGDWNDDTCTGQSADCPRNGLYG   72
crotatroxin1    AGEECDCGSPANPCCDAATCKLRPGAQCADGLCCDQCRFIKKGTVCRPARGDWNDDTCTGQSADCPRNGLYG   72
crotatroxin2     GEECDCGSPANPCCDAATCKLRPGAQCADGLCCDQCRFIKKGTVCRPARGDWNDDTCTGQSADCPRNGLYG   71
```

Figure 63. Crotatroxin 1 and 2, released in the venom of *C. atrox*, and recombinant crotatroxin sequences alignment. RGD motif is labeled in red and underlined, Cysteine are in bold.

E. coli BL21 bacteria transformed with pET32a(+)/TEV-crotatroxin plasmid overexpressed soluble crotatroxin-thioredoxin-His_6 fusion protein by addition of a final concentration of 1mM IPTG (Figure 25, p.54). r-crotatroxin (Figure 63) was purified from the cell lysate following several purification steps as previously described for the recombinant ocellatusin disintegrins (Figure 64 A, B and C) (Figure 26, p.55). Purification yield of r-crotatroxin was approximately 1mg/L cell culture.

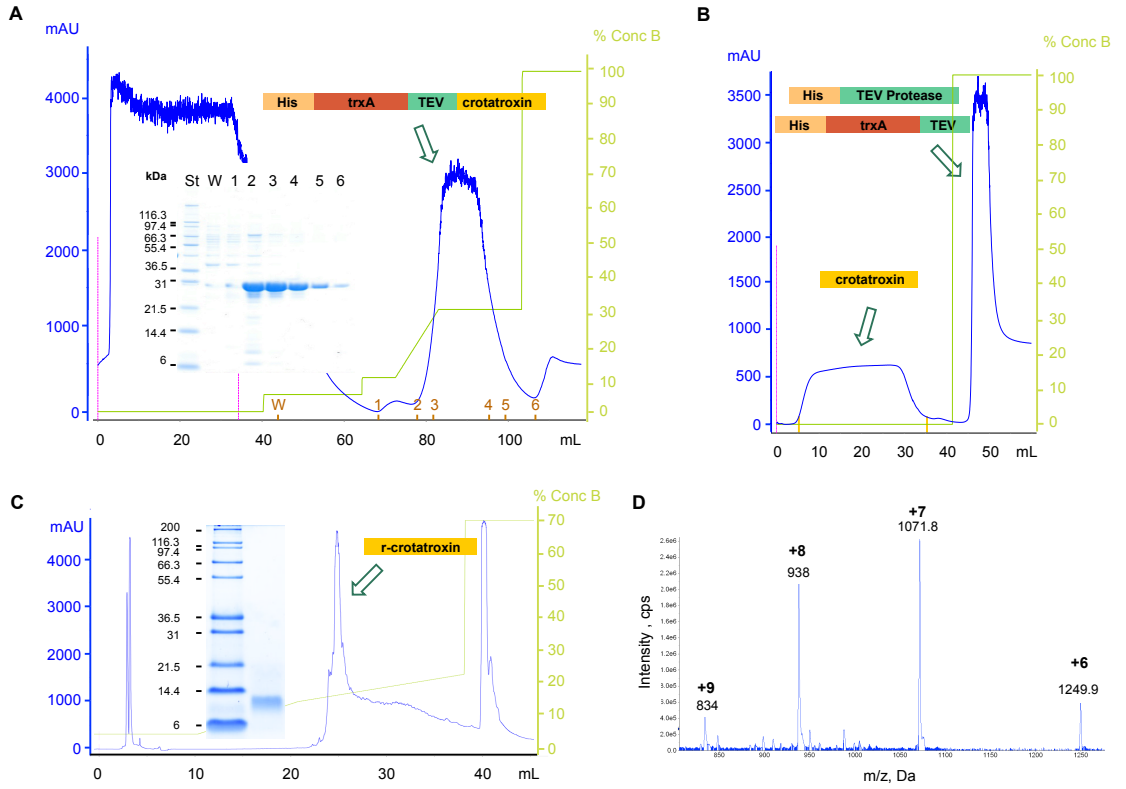


Figure 64. Purification steps of recombinant crotatroxin. A. Soluble fraction of lysated *E.coli* BL21 cells expressing the His_6-thioredoxin-TEV-crotatroxin fusion protein was loaded in the first step of His-Trap affinity chromatography. To identify the fraction containing our construct, eluted fractions were loaded in 12% SDS-PAGE. The arrow points to the fusion protein fraction. B. Second step of His-Trap affinity chromatography after the digestion of the construct (eluted 2-5) by r-TEV-protease. The retained and non-retained protein identities are specified. The flow through fraction, containing the r-crotatroxin was concentrated. C. The concentrate fraction was purified by high performance reverse-phase high performance liquid chromatography (RP-HPLC). The purified r-crotatroxin peak is denoted and analyzed by Tris–Tricine–(10%)SDS-PAGE. Lane 1, r-crotatroxin purified by RP-HPLC from the flow-through of the HisTrap affinity column. Lanes S, molecular weight markers (Mark12™, Invitrogen), whose apparent molecular mass is indicated at the left side of the gels. D. Electrospray-ionization mass spectrometry of RP-HPLC-purified r-crotatroxin. ESI-MS is the experimental mass (7495.9±0.3 Da) and calculated molecular mass (7495.29 Da) corresponds to fully oxidized (4 disulfide bonds) monoisotopic species.

The purity of the isolated r-crotatroxin was assed by 10% Tris-tricine-SDS-PAGE and electrospray-ionization mass spectrometry (Figures 64 C and D). In addition, electrospray-ionization mass spectrometry proved that the experimental molecular mass

(7495.9±0.3 Da) (Figures 64 D) accurately matching the calculated (7495.3 Da) mass for the recombinant crotatroxin with fully oxidized cysteine residues confirming the correct primary sequence and the cysteine pairing.

3 Recombinant crotatroxin as *Crotalus atrox* prey relocator molecule

A suitable bioassay of vomeronasal chemoreception was previously developed for evaluating preference towards envenomated (E) *vs.* non-envenomated (NE) mice carcasses, with snakes showing high rates of tongue flicking directed toward E carcasses, a behavior known as strike-induced chemosensory searching (SICS) (Chiszar et al. 1992; 1999; Greenbaum et al. 2003; Chiszar et al. 2008; Saviola et al. 2013).

To determine that only crotatroxin disintegrins are the molecules, which allow Western diamondback rattlesnakes (*C. atrox*) to discriminate among envenomated (E) and non-envenomated (NE) prey, we offered E and NE mouse carcasses injected with either r-crotatroxin or a saline control, respectively (Figure 29, p.63). When the carcasses were artificially "envenomated" with r-crotatroxin, the mean number of tongue flicks was significantly greater for the E mouse. Total number of tongue flicks was converted to percentage of tongue flicks, to account for natural variation (reflected by s.e.m values) in absolute tongue flicks rate between individual snakes (Table 13). The mean percentage of the tongue flicks toward the envenomated carcass (70%) was significantly higher for the E mouse (t=3.36, df=11, *P-value*=0.003). These results agree with the previous study performed using the fraction III, containing crotatroxin-1 and crotatroxin-2 disintegrins (Saviola et al. 2013) (Table 13), confirming our hypothesis. *C. atrox* discriminates among envenomated (with r-crotatroxin) and non-envenomated mice, having preference for the mouse injected with r-crotatroxin. These results suggest that the crotatroxin disintegrins are the relocator elements in the venom of *Crotalus atrox* rattlesnakes.

Table 13. Rattlesnakes discriminate between non-envenomated and envenomated (with r-crotatroxin) mice

Sample		Tongue flicks mean		*t*	*P-value*	df
	Values	NE mouse	E mouse			
r-crotatroxin	nº	38.4 (9.8)	117.7 (27.4)			
	%	30.0	70.0 (6)	3.36**	0.003	11
Peak III*	nº	25.3 (5.1)	53.6 (7.7)			
***C-atrox* venom**	%	32	68 (3.2)	5.78**	<0.01	10

Mean percent (%) tongue flicks directed at envenomated (E) and non-envenomated (NE) mice, when envenomated mice were injected with r-crotatroxin. Standard Error of the Mean (s.e.m) is indicated in brackets. Single-sample t-test was conducted on mean percentages where the mean percent of tongue flicks directed towards E mice were compared to 50%, the value expected under the null hypothesis; df =n (trials)-1. Because the two means are not independent, the same t value but with the opposite sign would be obtained for each mean. * The same data from the peak III from C. atrox venom Saviola at al. (2013) are indicated. For r-crotatroxin raw data, see Table 3S (p.172). ** P < 0.01.

4 *Crotalus atrox* specific discrimination for a prey relocator molecule

Considering these results, we decided to examine if *C. atrox* could recognize prey injected with a disintegrin that is not naturally found in the venom of this species. The venom of *C. atrox* contains only medium-sized monomeric disintegrins (Calvete et al. 2009a), such as the crotatroxins. Nonetheless, besides medium-sized disintegrins, several vipers contain dimeric as well as long and/or short monomeric disintegrins (Juárez et al. 2008). Disintegrins are classified based on their length, number of disulfide bonds and according to their integrin-binding motif. Depending on the amino acids present in the binding-motif, disintegrins are able to block different integrin receptors (Calvete 2010). At present, it is unknown how the crotatroxins create an olfactory "mark" that snakes are able to recognize, however, Saviola et al. (2013) hypothesized an integrin-mediated release of chemical cues from prey stimulate the vomeronasal system of snakes. Bearing in mind this hypothesis, and considering that crotatroxin includes an RGD integrin-binding motif in the apex of an 11 amino acid loop, identical to the short disintegrin ocellatusin, as well as both disintegrins inhibiting ADP-induced platelet aggregation with similar potency (an IC_{50} of; 17.5 $\times 10^{-8}$M and 16.8 $\times 10^{-8}$ ±0.2 M, respectively (Smith et al. 2002; Galán et al. 2008); we decided to test *C. atrox* response towards recombinant ocellatusin (see chapter I). Envenomated mice were injected with

ocellatusin, and NE with saline solution as previously stated, and both were offered to the *Crotalus atrox* to determine if snakes can differentiate between ocellatusin-E and NE mice. Similar to the previous study with r-crotatroxin, total number of tongue flicks were converted to percentages to control natural variation in absolute tongue flicks rate between snakes. The mean percentage of tongue flicks towards E (35.8%) and NE (64.2%) mice suggest that *C. atrox* do not show preference for E mice injected with ocellatusin. Likewise, there was not significant difference between the percentages of tongue flicks directed towards the E and NE carcasses (P-value > 0.025) (Table 14).

Table 14. Rattlesnakes discriminate between non-envenomated and envenomated (injected with r-disintegrin) mice

Sample	Tongue flicks mean			*t*	*P-value*	df
	Values	NE mouse	E mouse			
r-crotatroxin	nº	38.4 (9.8)	117.7 (27.4)			
	%	30.0	70.0 (6)	3.355**	0.003	11
r-ocellatusin	nº	78 (12.1)	47 (15.9)			
	%	64.2	35.8 (5.5)	-2.11	0.036	7

Mean percent (%) tongue flicks directed at envenomated (E) and non-envenomated (NE) mice, when envenomated mice were injected with r-crotatroxin and r-ocellatusin. Standard Error of the Mean (s.e.m) is indicated in brackets. Single-sample t-test was conducted on mean percentages where mean percent to E mice were compared with 50%, the value expected under the null hypothesis; df =n (trials)-1. Because the two means are not independent, the same t value but with the opposite sign would be obtained for each mean. For r-ocellatusin raw data, see Table 3S (p.172) and 4S (p.173). ** P < 0.025.

These results indicate that *C. atrox* does not recognize mice injected with the short RGD disintegrin, ocellatusin. Its activity does not appear to be enough to release volatile cues in the prey, which *C. atrox* could recognize to discriminate between the E and NE prey items.

Nevertheless, our recombinant crotatroxin behavior analyses validate Saviola et al. (2013) suggestions. The medium-sized crotatroxin disintegrins allow *C. atrox* to distinguish between envenomated and non-envenomated prey sources, presumably by varying the chemical odor of prey integument. Furthermore, our results hint that the differences in N-terminal of r-crotatroxin, crotatroxin-1 and 2 (Figure 63) do not affect for the *C. atrox* recognition.

Ocellatusin exhibits similar platelet aggregation activity when compared to crotatroxin by selectively binding $\alpha_{IIb}\beta_3$ integrin, which is expressed on platelet surfaces (Calvete et al. 1994). However, disintegrins expressing the RGD motif blocks several integrin receptors beyond $\alpha_{IIb}\beta_3$ (Figure 12, p24). The short disintegrin ocellatusin has also high specificity for $\alpha_5\beta_1$ integrin (Smith et al. 2002). However, the medium-sized crotatroxins show higher affinity for integrin $\alpha_V\beta_1$ than for integrins $\alpha_5\beta_1$ and $\alpha_V\beta_3$, contrary to other *Crotalus* medium-sized RGDW-disintegrins, which preferentially block integrins $\alpha_5\beta_1$ and $\alpha_V\beta_3$ (Galán et al. 2008).

The fact that *C. atrox* did not recognize volatile cues released from E mice injected with ocellatusin, might be due to: i) RGD-disintegrins exhibiting specific binding and affinities to integrin receptors, which may not be recognized by short disintegrins, ii) a necessary integration of the different crotatroxin integrin-binding, which might trigger the release of volatile molecules from the prey, making possible the relocation of the prey, iii) Short disintegrins are not present in Crotalinae venom, as they appear to have emerged in Viperinae after the split of Crotalinae taxa (Juárez et al. 2008). Therefore, snakes belonging to the genus *Crotalus* may not recognize the chemical cues (if any) released by the specific binding of short disintegrins, or iv) contrary to the previous assumption, crotatroxin might create an olfactory "mark" independent of an integrin-mediated function.

Darwin (Darwin 1859) suggested that predator diversification may be largely based on selection and that dangerous prey items are a major forcing acting on predatory behaviors. Natural selection has influenced snakes to respond to stimuli that are most likely to lead to successful relocation of envenomated prey (Kardong et al. 1997; Benard 2004; Chiszar et al. 2008; Cooper 2008; Juárez et al. 2008; Mackessy 2008; Fox and Serrano 2008b). This evolutionary pressure might contribute to the high level of variation and rapid evolution of disintegrins (Juárez et al. 2008), being part of a predator-prey a arms race that allows the predator to adapt to the variety of different prey items. Nonetheless, additional Viperidae behavioral studies should be conducted to further confirm this presumption.

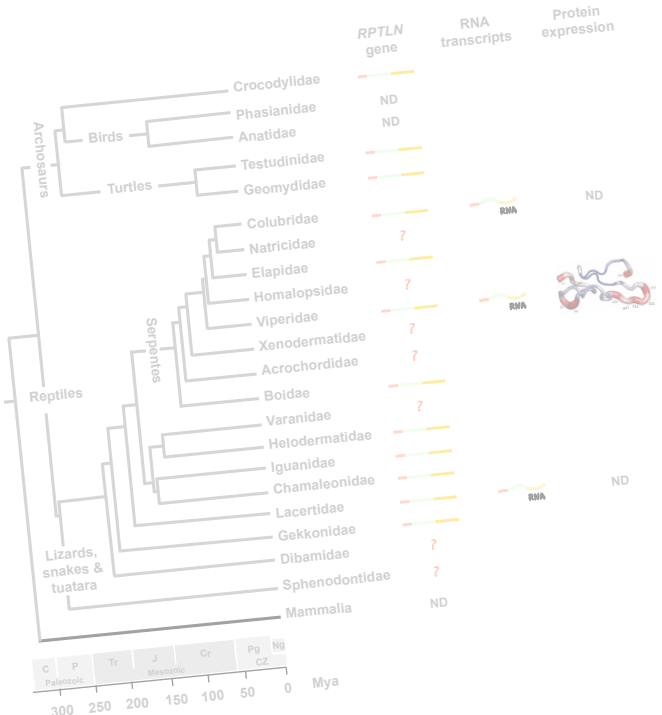
5 Concluding remarks, reflexions and perspectives

Our results agree with those previously reported by Saviola et al. (2013); suggesting that *Crotalus atrox* disintegrins have evolved into multifunctional proteins, which evoke vomeronasally-salient cues, enabling the snake to relocate envenomated prey after the strike. In addition, prey recognition for *C. atrox* appears to be dependent on the disintegrin injected to the E mice, suggesting a major selective advantage for the evolution of free disintegrins among viperid venoms (apparently exclusively), which would be provided by their role in prey relocation. Therefore, in addition to immobilizing, killing and predigesting prey, a "new" biological role of viper venoms appears to be prey relocation, as this behavior is observed by many rattlesnake species (Chiszar et al. 1999; Greenbaum et al. 2003; Greenbaum 2004; Chiszar et al. 2008; Saviola et al. 2010; 2013).

Additionally, Greenbaum et al. (2003) suggested that chemoreception is sensitive to subtle differences in venom biochemistry and may reflect adaptation to improve efficiency of finding envenomated prey (Greenbaum et al. 2003). This is supported by our negative response of *C. atrox* to volatile cues in E mice injected with r-ocellatusin, since *Echis ocellatus,* which express ocellatusin in its venom, and *C. atrox* shared a common ancestor approximately 22-35 millions of years ago (Lenk 2001). Moreover, some *Echis* species exhibit a sit-and-wait foraging strategy (Tsairi and Bouskila 2004) and strike-and-release their prey (Cundall 2009) similar to *Crotalus genus* (Lavin-Murcio et al. 1993; Cooper 1994; Clark 2006; Saviola et al. 2010).

Hence, future behavioral studies of vipers (other pitvipers, *Echis sp*., etc) that express disintegrins in their venoms, and exhibit a strike-and-release foraging strategy, should be done to address if other disintegrins are able to act as "relocator molecule". Moreover, new studies will help elucidate the molecular mechanism(s) by which disintegrins interact with prey tissues and facilitate the relocation of envenomated prey by these vipers.

Conclusions



Conclusions

1. The conservation of the *RPTLN* gene across reptiles, coupled with the structure-function studies of the *Frankenstein* mutants, provides evidence that the ancestral gene that codes for RTS/KTS short disintegrins was recruited in Viperidae venom gland, independently of the canonical short RGD-disintegrin evolutionary pathway.
2. We postulate that *RPTLN* genes comprise an ancient multigene family, and that their restricted expression and neofunctionalization in the venom gland of *Macrovipera* and *Daboia* species represent recent evolutionary events.
3. The detection of *RPTLN* transcripts in different organs of *Podarcis sp.* and *Rhinechis scalaris*, in absence (or low levels) of RTS-disintegrins, combined with the conservation of the *RPTLN* gene in reptiles, suggest a possible function of the *RPTLN* gene as non-coding RNA in reptilian body organs (non-venom gland tissues).
4. The high conservation of the Signal Peptide sequence coded for by *RPTLN* genes in extant SVMPs strongly suggests that this region may have played a key role in the recruitment and restricted expression of SVMP genes in the venom gland of Caenophidian snakes.
5. The functionality of the disintegrin ocellatusin depends on the amino acids G25 and D26 in the RGD active motif, and its inhibitory potency depends on the R24 residue. However, the amino acids in the N-terminal side of the RGD motif, as well as the E47-P50 residues in the carboxyl terminal tail, are not essential *per se* for ocellatusin functionality.
6. Our studies have shed light on the possible function of Viperidae disintegrins, as chemotactic element, to relocate the prey.
7. The *"new"* role of disintegrins as a relocator element of prey following strike-and-release appears to be species-specific, indicating that this particular disintegrin functionality might also contribute to the disintegrin evolution.

Resumen

1 Introducción

El grupo Reptilia, consta de aves y reptiles no aviares. Los reptiles son uno de los grupos de organismos vivos más notables, desde el punto de vista ecológico y evolutivo, que han colonizado la mayor parte del planeta. La mayor diversidad de los reptiles no aviares (96,4%) se concentra en Squamata (lagartos y serpientes). Por otra parte, tortugas y cocodrilos representan el 3,4% y el 0,2%, respectivamente (reptile-database.org) (Uetz and Hošek 2015). En la figura 1 (p.7) se muestra una cladogénesis generada a partir de datos mostrados en (Hedges and Vidal 2009; Pereiraa and Bakera 2009; Shaffer 2009; Shedlock and Edwards 2009; Vidal and Hedges 2009; Vidal et al. 2009), sin embargo, todavía existe controversia en cuanto a cómo y cuando aparecieron algunos taxones.

Las serpientes pertenecen al suborden Serpentes, que representan 3567 especies distribuidas en todos los continentes excepto la Antartida (Uetz and Hošek 2015). Recientemente se han generado filogenias de serpientes, a partir de la integración de datos morfológicos y moleculares (Vidal et al. 2007; Castoe et al. 2009; Pyron et al. 2011; 2013; Reeder et al. 2015). A pesar de estos estudios, la posición relativa de algunas serpientes aún es objeto de debate. No obstante, Viperidae, Elapidae y Colubridae están contenidas en la superfamilia Colubroidea, que incluye a todas las serpientes venenosas conocidas, y a la mayoría de serpientes actuales (>2/3 *sp.*) (Vidal et al. 2009; Pyron et al. 2011; Pyron and Burbrink 2012) (Figura 4, p. 13).

La gran diversidad de serpientes, junto con su plasticidad genética y fenotípica, convierte a las serpientes y sus venenos en objeto de estudio en muchas disciplinas, desde evolución, ecología y comportamiento, a medicina. Concretamente el estudio de los venenos de serpientes es de gran interés en varios aspectos. El conocimiento generado a partir del estudio del veneno, o de alguna de sus proteínas, podría aportar la base para el desarrollo de un nuevo medicamento, o ayudar a la generación de mejores antivenenos. Además, tales estudios podrían revelar nuevos mecanismos moleculares en mamíferos, mediante el uso de toxinas en estudios de receptores celulares como por ejemplo las integrinas.

Los venenos de serpientes contienen una compleja variedad de componentes farmacológicamente activos, con y sin actividad enzimática. De todos ellos, en este trabajo nos hemos centrado en el estudio de una familia de proteínas que no poseen actividad enzimática, las disintegrinas. Esta familia ha servido como modelo para generar varios fármacos que están en el mercado para tratar el síndrome coronario agudo, mediante la inhibición de agregación plaquetaria (por ejemplo, Aggrastat® (Tirofiban) e Integrilin® (Eptifibatidae)) (Granada and Kleiman 2004).

Únicamente los venenos de Viperinae y Crotalinae contienen disintegrinas, una familia de proteínas (41-100 aminoácidos) sintetizadas a partir de RNAs mensajeros cortos (Okuda et al. 2002) o liberadas en el veneno por procesamiento proteolítico de metalloproteasas de tipo PII (PII-SVMP) (Kini and Evans 1992; Fox and Serrano 2005).

La familia de las disintegrinas incluye antagonistas potentes y específicos de los receptores de integrina β_1 y β_3, que han evolucionado en la parte apical de un lazo móvil de 11 residuos. La mayoría de las disintegrinas de cadena simple (largas, medias y cortas) expresan la secuencia RGD que representa el motivo de inhibición básico de la unión de las integrinas a sus ligandos y bloquea su función adhesiva.

La evolución de las disintegrinas RGD, desde las disintegrinas largas, a las más recientes en la evolución, las disintegrinas cortas, implica una reducción de tamaño debida a la pérdida de cisteínas y procesamiento del extremo N-terminal. En todas ellas su actividad inhibidora depende de la apropiada formación de puentes disulfuro entre sus cisteínas, que determina la conformación del lazo móvil que alberga el motivo activo. Además de las disintegrinas canónicas (RGD/XXD), existen otro tipo de disintegrinas, que contienen los motivos KTS/RTS y se encuentran en venenos de víboras Euroasiáticas. Estas disintegrinas bloquean selectivamente la unión de la integrina $\alpha_1\beta_1$ a su ligando (colágeno I y IV) *in vitro* y bloquean la angiogénesis in *vivo* (Marcinkiewicz et al. 2003; Olfa et al. 2005; Brown et al. 2008).

Estudios estructurales mediante RMN de las disintegrinas (K/R)TS (obtustatina y jerdostatina) han revelado que el bucle de unión a la integrina y el extremo C-terminal forman un epítopo funcional que muestra movimientos concertados (Moreno-Murciano et al. 2003b; Carbajo et al. 2011). La base estructural de la selectividad y

especificidad de disintegrinas (K/R)TS por la integrina $\alpha_1\beta_1$, subyace en la forma y tamaño del lazo de unión a integrina de 9 residuos, junto con su composición, flexibilidad y orientación lateral del tripéptido (K/R)TS (Monleón et al. 2003; Calvete et al. 2005; Sanz et al. 2005; Calvete et al. 2007b; Carbajo et al. 2011).

Las disintegrinas (K/R)TS son sintetizadas a partir de ARNs mensajeros cortos que codifican para un péptido señal, propeptido corto y el dominio disintegrina. La disintegrina RTS jerdostatina codificada por el gen *RPTLN* conserva esta estructura génica. Aunque la secuencia nucleotídica que codifica la disintegrina jerdostatina se conserva con identidad del 100% en la glándula del veneno de varias víboras euroasiáticas, *Prothoboptrops jerdonii*, *Cerastes vipera* y *Echis ocellatus* (Sanz et al. 2005; 2006; Bazaa et al. 2007), esta disintegrina RTS no se había encontrado en ningún veneno estudiado hasta la fecha. Varios estudios han discutido la difícil cuestión de explicar cuál podría ser el papel de las disintegrinas RTS con actividad antiangiogénica en el contexto predador-presa (revisado por Walsh and Marcinkiewicz 2011; Calvete 2013b). Recientemente, en el trabajo de Saviola et al. 2013, se sugirió una nueva función para las disintegrinas en el veneno. Puesto que la fracción del veneno que incluye las disintegrinas parece contener el elemento quimiotáctico que utiliza *Crotalus atrox*, para la relocalización de su presa envenenada y liberada de sus fauces.

2 Objetivos

En esta tesis hemos explorado la relación evolutiva de dos clases de disintegrinas cortas, RGD y RTS/KTS, además de estudiar la distribución del gen *RPTLN* y su posible función en reptiles. Para ello se abordaron los siguientes objetivos parciales:

1. Establecer la historia natural de la familia RTS/KTS-disintegrinas, analizando la relación estructura-función de las disintegrinas RGD y RTS/KTS mediante la generación de mutantes quimera.

2. Definir los amino ácidos del lazo activo claves para la función de la disintegrina corta RGD ocellatusina.

3. Caracterizar el gen *RPTLN* completo, que codifica para una disintegrina RTS. Y explorar su distribución en reptiles.

4. Determinar la presencia o ausencia de transcritos del gen *RPTLN* y la posible traducción del dominio disintegrina en otros órganos distintos de la glándula del veneno.

5. Explorar la posible función de las disintegrina crotatroxina, del veneno de *Crotalus atrox*, como elemento relocalizador de la presa, así como la posibilidad de que la disintegrina ocellatusina tenga una función similar.

3 Metodología

Amplificación del gen *RPTLN* mediante reacción en cadena de la polimerasa (PCR)

La reacción en cadena de la polimerasa, conocida como PCR (*polymerase chain reaction*), es una técnica desarrollada en 1983 por Kary Mullis (Bartlett and Stirling 2003) que revolucionó los estudios de biología molecular. Para la identificación del gen *RPTLN* en distintos genomas se llevó a cabo la extracción de ADN genómico a partir de las muestras de sangre y tejido disponibles (Tabla 1, p.41) (Longmire et al. 1997). Posteriormente, se amplificó por PCR el fragmento completo del gen *RPTLN* a partir del ADN genómico de diversos reptiles. Los fragmentos de ADN candidatos se cortaron del gel de agarosa y se purificaron, con el fin de clonarlos y secuenciarlos. De este modo se identificaron los genes *RPTLN*-like contenidos en genomas de una amplia variedad de reptiles.

Identificación de transcritos del gen *RPTLN*

La identificación de transcritos en determinados órganos de culebra (*Rhinechis scalaris*) y lagarto (*Podarcis muralis* y *P. hispanicus*) se llevó a cabo del mismo modo que la determinación del gen *RPTLN* en el ADN genómico de distintos reptiles. Sin embargo, en este caso se usó el ADN complementario (cDNA) como molde en la amplificación por PCR. Para la síntesis de cDNA se llevó a cabo la extracción de ARN de los órganos y a continuación se realizó retrotranscripción del ARN a ADN complementario (cDNA).

Además, para poder observar diferencias de expresión de gen *RPTLN* entre los distintos órganos, se realizó la amplificación por PCR semicuantitativa y PCR cuantitativa.

Análisis de secuencias y deposito en la base de datos NCBI

Las secuencias *RPTLN*-like obtenidas se identificaron utilizando BLASTn, y se depositaron en la base de datos NCBI (Centro Nacional de Información Biotecnológica). Su alineamiento múltiple se realizó utilizando el programa MEGA (Molecular Evolutionary Genetic analysis) (Kumar et al. 2001); y para la predicción de estructura

secundaria de RNA se uso RNAfold WebServer (Mathews et al. 2004; Gruber et al. 2008; Lorenz et al. 2011).

Extracción de proteína y detección de jerdostatin mediante Western Blot

El Western Blot es una técnica analítica ampliamente extendida que se usa para la detección de proteínas específicas en una muestra determinada. Concretamente en este trabajo se pretendía determinar la traducción de los transcritos del gen *RPTLN* detectados en órganos de *Podarcis sp.* y *Rhinechis scalaris.* Para ello se extrajo proteína total de los órganos mediante la homogenización de cada tejido en un tampón de lisis, incluyendo los inhibidores de proteasas necesarios. La detección de jerdostatina en nuestras muestras se realizó mediante Western Blot utilizando el anticuerpo primario anti-PEP160 que reconoce el péptido [CKPSYPGNG] en el extremo C-terminal de la jerdostatina.

Expresión y purificación de disintegrinas recombinantes

Existe una amplia variedad de métodos de generación, expresión y purificación de proteínas. En nuestro caso, diseñamos y generamos plásmidos de expresión en bacterias, para la posterior expresión y purificación de disintegrinas recombinantes de forma soluble y activa.

En el caso de crotatroxin, se determinó la secuencia nucleotídica *de novo.* Una vez conocida la secuencia de nuestras disintegrinas silvestres, se clonaron en un plásmido de expresión específico de bacterias (pET32a(+)), incluyendo la secuencia de reconocimiento de la proteasa TEV con el fin de poder cortar posteriormente la etiqueta de las disintegrinas recombinantes (Figura 23, p.50).

La obtención de los plásmidos de expresión de las disintegrinas mutantes, que incluyen cambios puntuales en su secuencia de ADN, se realizó mediante mutagénesis dirigida.

Los plásmidos de expresión generados se transformaron en la cepa de expresión BL21 de *Escherichia coli,* para la sobreexpresión de cada una de las disintegrinas recombinantes (Figura 25, p.54). Posteriormente, las disintegrinas recombinantes se

purificaron a partir de la fracción soluble del lisado celular mediante varios pasos de purificación por afinidad, digestión por la proteasa TEV, y un último paso de purificación por cromatografía de fase reversa. La pureza de la muestra se testó mediante electroforesis en gel, 10%Tris-tricine SDS-PAGE, y espectrometría de masas (ESI-MS) (Figura 26, p.55).

Purificación de disintegrinas a partir del veneno

Para aislar disintegrinas a partir del veneno, las proteínas solubles del veneno se separaron mediante cromatografía de fase reversa.

Pureza e identificación de las disintegrinas mediante espectrometría de masas

Con el fin de identificar las disintegrinas purificadas a partir de material recombinante o del veneno se midieron las masas de las disintegrinas por espectrometría de masas mediante una fuente de ionización por nanoelectronebulización (Figura 27, p.57). Esto nos permitió confirmar la identidad de todas las disintegrinas, pureza y su correcta purificación. En cada caso, la masa monoisotópica teórica, se comparó con la masa calculada, considerando condiciones nativas, con la formación de 4 puentes disulfuro en el caso de las disintegrinas cortas y 6 en las de tamaño medio.

Purificación de la integrina $\alpha_1\beta_1$ humana e inhibición de la unión de la integrina $\alpha_1\beta_1$ al fragmento CB3 del colágeno IV mediada por las disintegrinas

Las disintegrinas KTS/RTS inhiben la unión de la integrina $\alpha_1\beta_1$ a su ligando natural, colágeno IV/I. Por lo que para testar la capacidad de inhibición de unión de la integrina $\alpha_1\beta_1$ de nuestros mutantes quimera (RTS-ocellatusin) denominados *Frankenstein,* se purificó la integrina $\alpha_1\beta_1$ y se midió la inhibición de unión a su ligando mediante un ensayo ELISA.

El ectodominio de la integrina $\alpha_1\beta_1$ humana se expresó en células *Schneider* de *Drosophila sp.*, y se purificó siguiendo el protocolo descrito previamente por (Eble et al. 2006), posteriormente se testó su funcionalidad mediante una curva de titulación.

La capacidad de inhibición de las disintegrinas se midió mediante un ensayo ELISA en el que el fragmento CB3 del colágeno IV fue inmovilizado en una placa de 96 pocillos y se incubó con una mezcla de la integrina $\alpha_1\beta_1$ a concentración fija y concentraciones crecientes de la disintegrina objeto de estudio. Posteriormente, se cuantificó la unión de la integrina $\alpha_1\beta_1$ por colorimetría y se transformó en porcentaje de no-inhibición. Las disintegrinas lebestatina y ocellatusina se usaron como control positivo y negativo de inhibición de unión de la integrina $\alpha_1\beta_1$, respectivamente.

Preparación de las plaquetas en suspensión y estudio de la inhibición de la agregación plaquetaria (inducida por colágeno I), mediada por disintegrinas

Las disintegrinas que contienen el motivo de unión RGD (Arg-Gly-ASp) son capaces de inhibir la agregación plaquetaria mediante la unión a la integrina $\alpha_{IIb}\beta_3$ y con ello bloquean la unión del fibrinógeno a dicha integrina (Calvete et al. 1994). Este método se ha utilizado históricamente para la testar la capacidad de inhibición plaquetaria de diferentes péptidos.

Para ello se aislaron las plaquetas humanas a partir de muestras de sangre fresca de voluntarios sanos, que no hubieran recibido ningún tratamiento médico que pudiese afectar a la respuesta de las plaquetas. Para la preparación de plaquetas lavadas en suspensión se siguió el protocolo previamente descrito en (Antunes et al. 2010).

Para monitorizar la inhibición de la agregación plaquetaria por las disintegrinas, se midió la transmisión de luz y se transformó en porcentaje de agregación plaquetaria (Figura 28, p.60). El valor de agregación máxima (MA) se utilizó para comparar diferentes potencias inhibidoras. A partir de este valor, se calculó la concentración necesaria para reducir al 50% de la agregación plaquetaria inducida por Col I, con respecto al control (valor IC_{50}).

Experimentos de comportamiento de la serpiente de cascabel, *Crotalus atrox*, frente a las disintegrinas recombinantes crotatroxina y ocellatusina

Con el fin de determinar si la disintegrina crotatroxina es el elemento relocalizador de la presa y si ocellatusina pudiera tener la misma función, se realizaron análisis de comportamiento de las serpientes de cascabel, *Crotalus atrox*, permitiendo que

inyectaran su veneno a un ratón sacrificado previamente. De este modo, se indujo la búsqueda quimiosensora de la presa (Chiszar et al 1992) y se retiró el ratón de la caja donde posteriormente se realizó el ensayo. Se metió en la caja un útil (Figura 29, p.62), que consistía en una base en la que se situaron dos bolsas (de rejilla), separadas entre ellas 4 cm; en una se introdujo un ratón inyectado con la disintegrina recombinante (ratón "envenenado", E); en otra, el ratón control (ratón "no envenenado", NE) (Figura 29, p.62). Se grabaron los ensayos y los videos fueron analizados por un observador que desconocía las condiciones del ensayo. En cada uno de los videos, se contaron el número de "flicks" (proyección/retracción) de la lengua de la serpiente, dirigidos a menos de 1cm del ratón E o NE.

Por último, se llevó a cabo el análisis estadístico del número de "proyección/retracción" de la lengua de *Crotalus atrox* dirigidos hacia el ratón E o NE. Este número se convirtió en porcentaje con el fin de evitar la variabilidad natural de este dato. Estos datos fueron analizados empleando test t de Student para muestras emparejadas.

4 Resultados y discusión:

Capitulo I.

1. Ni la inserción del motivo RTS en diferentes posiciones del lazo funcional de ocellatusina, ni la sustitución del lazo RGD de ocellatusina completo [^{20}**C**KMARGDNMHDY**C**^{32}] por el lazo de la disintegrina RTS-jerdostatina [^{20}**C**WRTS--VSSHY**C**^{32}], son suficientes para conferir la funcionalidad de una disintegrina (R/K)TS, bloqueando la unión de la integrina $\alpha_1\beta_1$ al fragmento CB3 del colágeno IV.

2. El acortamiento del extremo carboxilo terminal, o la sustitución de la región C-terminal de ocellatusina [^{39}**C**PRNPYKG^{46}] por la de jerdostatina [^{39}**C**P*SY*P*GN*G^{46}], en mutantes que contienen el lazo funcional de jerdostatina [^{20}**C**WRTS--VSSHY**C**^{32}], no fueron suficiente para generar un sitio de inhibición de la unión de integrina $\alpha_1\beta_1$ a CB3 (Colágeno IV). Tampoco lo fueron los mismos mutantes en los que se incluyeron cambios en la región alrededor del lazo funcional [^{19}I**C**WRTSVSSHY**C***T*^{33}] y [^{19}*T***C**WRTSVSSHY**C**N^{33}].

3. Al igual que en otras disintegrinas cortas RGD estudiadas previamente, los aminoácidos G25 y D26 del tripéptido activo RGD de ocellatusina, son clave para su actividad bloqueante de la unión de la integrina $\alpha_{IIb}\beta_3$ a fibrinógeno.

4. La sustitución de aminoácidos en el lado N-terminal del lazo activo de ocellatusina, concretamente, M22R y A23T, no son críticos *per se* para el mantenimiento de la función inhibidora de esta disintegrina.

5. La supresión de los aminoácidos E47-P50 de ocellatusina, no afectan a su función y potencia de inhibición de la agregación plaquetaria inducida por colágeno I.

6. A pesar de representar motivos funcionales y que podrían haber surgido evolutivamente con el cambio de un sólo nucleótido en el codón de la arginina (R), los tripéptidos SGD, TGD y GGD no se han encontrado en disintegrinas naturales Ello puede ser debido a que su potencia inhibidora es de dos ordenes de magnitud menor que la de las disintegrinas naturales (RGD, WGD, KGD, MVD). Estos resultados apoyarían la selección natural de las disintegrinas guiada por evolución positiva.

Capitulo II.

1. Russellistatina es la única disintegrina RTS detectada hasta la fecha en un veneno de serpiente.
2. La secuencia completa del gen *RPLTN* está conservada en reptiles no aviares, con alto grado de identidad. Esta conservación en un taxón cuyo ancestro común existió hace mas de 250 millones de años parece indicar que el producto de este gen debe de estar realizando una función básica y específica de reptiles. Bien i) transcribiéndose a proteína, ii) actuando como ARN regulatorio o iii) cumpliendo otra función todavía desconocida.
3. El número de copias génicas diferentes del gen *RPTLN* encontradas en reptiles, y la presencia de varias de ellas en una sola especie, apoyan la idea de duplicación génica en el gen *RPTLN*.
4. El gen *RPTLN* se transcribe en distintos órganos de especies de las familias Lacertidae y Colubridae, pero no se ha detectado su traducción en los mismos, lo cual señala hacía una función como ARN no codificante.
5. Cuando se comparan los péptidos señal de SVMPs con los de las proteínas ADAM y disintegrinas RTS, se observa una conservación clara del péptido señal de las SVMP existentes, y del codificado por el gen *RPTLN* en las disintegrinas RTS. Por el contrario, las proteínas ADAM contienen péptidos señales distintos.

Capitulo III.

1. *Crotalus atrox* reconoce significativamente el ratón inyectado con la disintegrina recombinante crotatroxina con respecto al ratón control.
2. *Crotalus atrox* no es capaz de diferenciar entre el ratón inyectado con la disintegrina recombinante ocellatusina y el ratón control.

5 Conclusiones:

1. La conservación del gen *RPTLN* en reptiles, junto con los estudios estructura-función de los mutantes *Frankenstei*n, aportan evidencias de que el gen ancestral de las disintegrinas cortas RTS/KTS fue reclutado potencialmente en las glándulas del veneno de la familia Viperidae, independientemente de la vía clásica de neofuncionalización de las disintegrinas que contienen el dominio RGD.

2. Postulamos que el gen *RPTLN* constituye una familia multigénica y su expresión y neofuncionalización está restringida a la glándula del veneno de las especies *Macrovipera* y *Daboia*, lo cual representa un evento reciente en la evolución.

3. La presencia de mARN del gen *RPTLN* en diferentes órganos de *Podarcis sp.* y *Rhinechis scalaris*, y la ausencia (o baja expresión) de distintegrina RTS en los mismos órganos, además de la conservación del gen *RPTLN*, sugiere una posible función del gen *RPTLN* como ARN no codificante (ncARN) en órganos de reptiles (no en tejidos de producción de veneno).

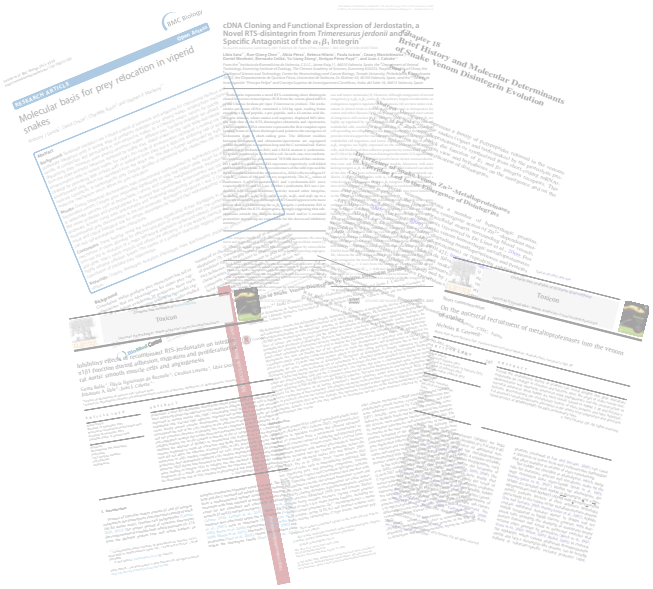
4. La conservación de la secuencia del péptido señal en las SVMPs existentes, también conservado el gen *RPTLN*, sugiere que esta región pudo haber jugado un papel en el reclutamiento de las SVMP y en su expresión restringida en la glándula del veneno de serpientes Caenophidian.

5. La funcionalidad de ocellatusina depende de los aminoácidos G25 y D26, y su potencia inhibidora del residuo R24. Ni los aminoácidos situados en el lado N-terminal del motivo RGD, en el lazo activo, ni los residuos E47-P50, en la cola C-terminal, son críticos *per se* para la funcionalidad de ocellatusina.

6. Nuestro datos han arrojado luz en cuanto a una nueva posible función de las disintegrinas, en Viperidae, como elemento quimiotáctico para relocalizar la presa.

7. El papel de las disintegrinas como elemento relocalizador de la presa parece ser específico de especie, indicando que esta función también podría contribuir a la evolución de las disintegrinas.

Bibliography



Aird SD. Ophidian envenomation strategies and the role of purines. Toxicon. 2002 Apr;40(4):335–93.

Aird SD, Aggarwal S, Villar-Briones A, Tin MM-Y, Terada K, Mikheyev AS. Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly. BMC Genomics. 2015;16:647.

Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature. 2011 Sep 19;477(7366):587–91.

Angulo Y, Castro A, Lomonte B, Rucavado A, Fernández J, Calvete JJ, et al. Isolation and characterization of four medium-size disintegrins from the venoms of Central American viperid snakes of the genera *Atropoides, Bothrops, Cerrophidion* and *Crotalus*. Biochimie. 2014 Dec 1;107(Pt B):376–84.

Antunes TC, Yamashita KM, Barbaro KC, Saiki M, Rodríguez S. Comparative analysis of newborn and adult *Bothrops jararaca* snake venoms. Toxicon. 2010 Dec 1;56(8):1443–58.

Arlinghaus FT, Eble JA. C-type lectin-like proteins from snake venoms. Toxicon. 2012 Sep 15;60(4):512–9.

Balakirev ES, Ayala FJ. Pseudogenes: Are They "Junk" or Functional DNA? Annu Rev Genet. Annual Reviews. 2003 Dec;37(1):123–51.

Bartlett JMS, Stirling D. A Short History of the Polymerase Chain Reaction. New Jersey: Humana Press; 2003 Aug 1;226:3–6.

Bazaa A, Juárez P, Marrakchi N, Bel Lasfer Z, Ayeb El M, Harrison RA, et al. Loss of introns along the evolutionary diversification pathway of snake venom disintegrins evidenced by sequence analysis of genomic DNA from *Macrovipera lebetina transmediterranea* and *Echis ocellatus*. J Mol Evol. 2007 Feb;64(2):261–71.

Bazaa A, Marrakchi N, Ayeb El M, Sanz L, Calvete JJ. Snake venomics: Comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes, Cerastes vipera* and *Macrovipera lebetina*. Proteomics. 2005 Nov;5(16):4223–35.

Benard MF. Predator-induced phenotypic plasticity in organisms with complex life histories. Annual Review of Ecology. 2004;35:651–73.

Benton MJ. Phylogeny of the major tetrapod groups: morphological data and divergence dates. J Mol Evol. 1990 May;30(5):409–24.

Bolás G, de Rezende FF, Lorente C, Sanz L, Eble JA, Calvete JJ. Inhibitory effects of recombinant RTS-jerdostatin on integrin $\alpha_1\beta_1$ function during adhesion, migration and proliferation of rat aortic smooth muscle cells and angiogenesis. Toxicon. 2014 Mar;79:45–54.

Brown MC, Eble JA, Calvete JJ, Marcinkiewicz C. Structural requirements of KTS-disintegrins for inhibition of $\alpha_1\beta_1$ integrin. Biochem J. 2009 Jan 1;417(1):95–101.

Brown MC, Staniszewska I, Del Valle L, Tuszynski GP, Marcinkiewicz C. Angiostatic activity of obtustatin as $\alpha_1\beta_1$ integrin inhibitor in experimental melanoma growth. Int J Cancer. 2008 Nov 1;123(9):2195–203.

Burghardt GM. Chemical perception in reptiles. Advances in chemoreception. In: Johnson Jr JW, Moulton DG, Turk A (Eds.) Communication by Chemical Signals. New York: Appleston- Century-Crofts, 241-308.

Calvete JJ. Structure-function correlations of snake venom disintegrins. Current Pharmaceutical design. 2005;11(7):829–35.

Calvete JJ. Brief history and molecular determinants of snake venom disintegrin evolution. In: Kini RM, Clemetson KJ, Markland FS, McLane MA, Morita T, (Eds.), Toxins and hemostasis: From Bench to Bedside. Amsterdam: Springer, 2010. pp. 285–300.

Calvete JJ. Snake venomics: from the inventory of toxins to biology. Toxicon. 2013a Dec 1;75:44–62.

Calvete JJ. The continuing saga of snake venom disintegrins. Toxicon. 2013b Feb;62:40–9.

Calvete JJ. Consideraciones sobre la interpretación de espectros de masas. In: Corrales F and Calvete JJ (Eds). Manual de proteómica. Sociedad Española de Proteómica; 2014. pp. 173–86.

Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG. Exploring the venom proteome of the western diamondback rattlesnake, *Crotalus atrox*, via snake venomics and combinatorial peptide ligand library approaches. J Proteome Res. 2009a Jun 5;8(6):3055–67.

Calvete JJ, Fox JW, Agelan A, Niewiarowski S, Marcinkiewicz C. The presence of the WGD motif in CC8 heterodimeric disintegrin increases its inhibitory effect on $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$, and $\alpha_5\beta_1$ integrins. Biochemistry. 2002 Feb;41(6):2014–21.

Calvete JJ, Juárez P, Sanz L. Snake venomics. Strategy and applications. J Mass Spectrom. 2007a;42(11):1405–14.

Calvete JJ, Juárez P, Sanz L. Snake venomics and disintegrins. Portrait and evolution of a family of snake venom integrin antagonists. In: Mackessy SP, (Ed.), Handbook of Venoms and Toxins of Reptiles. Boca Ratón: CRC Press, Taylor & Francis, 2009b. pp. 337–537 (Chapter 17).

Calvete JJ, Marcinkiewicz C, Monleón D, Esteve V, Celda B, Juárez P, et al. Snake venom disintegrins: evolution of structure and function. Toxicon. 2005 Jun 15;45(8):1063–74.

Calvete JJ, Marcinkiewicz C, Sanz L. KTS and RTS-disintegrins: anti-angiogenic viper venom peptides specifically targeting the $\alpha_1\beta_1$ integrin. Current Pharmaceutical design. 2007b;13(28):2853–9.

Calvete JJ, McLane MA, Stewart GJ, Niewiarowski S. Characterization of the cross-linking site of disintegrins albolabrin, bitistatin, echistatin, and eristostatin on isolated human platelet integrin GPIIb/IIIa. Biochem Biophys Res Commun. 1994 Jul;202(1):135–40.

Calvete JJ, Moreno-Murciano MP, Theakston RDVG, Kisiel DG, Marcinkiewicz C. Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulphide bond engineering. Biochem J. 2003 Jun;372(3):725–34.

Cao Y, Sorenson MD, Kumazawa Y, Mindell DP, Hasegawa M. Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. Gene. 2000 Dec;259(1-2):139–48.

Carbajo RJ, Sanz L, Mosulén S, Pérez A, Marcinkiewicz C, Pineda-Lucena A, et al. NMR structure and dynamics of recombinant wild type and mutated jerdostatin, a selective inhibitor of integrin $\alpha_1\beta_1$. Proteins. 2011 Jun;79(8):2530–42.

Carbajo RJ, Sanz L, Pérez A, Calvete JJ. NMR structure of bitistatin – a missing piece in the evolutionary pathway of snake venom disintegrins. FEBS J. 2015 Jan;282(2):341–60.

Carey CM, Bueno R, Gutierrez DA, Petro C, Lucena SE, Sanchez EE, et al. Recombinant rubistatin (r-Rub), an MVD disintegrin, inhibits cell migration and proliferation, and is a strong apoptotic inducer of the human melanoma cell line SK-Mel-28. Toxicon. 2012 Feb;59(2):241–8.

Carroll RL. Problems of the ancestry of turtles. In: Brinkman DB, holroyd PA and Gardner JD (Eds). Morphology and Evolution of Turtles. Dordrecht: Springer, 2013. pp. 19–36.

Casewell NR. On the ancestral recruitment of metalloproteinases into the venom of snakes. Toxicon. 2012 Sep 15;60(4):449–54.

Casewell NR, Huttley GA, Wuster W. Dynamic evolution of venom proteins in squamate reptiles. Nat Commun. 2012;3:1066.

Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wuster W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. Mol Biol Evol. 2011 Aug;28(9):2637–49.

Casewell NR, Wagstaff SC, Wuster W. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. PNAS. 2014;111(25):9205–10.

Casewell NR, Wuster W, Vonk FJ, Harrison RA, Fry BG. Complex cocktails: the evolutionary novelty of venoms. Trends in Ecology & Evolution. 2013 Apr;28(4):219–29.

Castoe TA, Daza JM, Smith EN, sasa M, Kuch U, Campbell JA, et al. Comparative phylogeography of pitvipers suggests a consensus of ancient Middle American highland biogeography. Journal of Biogeography. 2009 Nov;36:88–103.

Chen Y, Suri AK, Kominos D, Sanyal G, Naylor AM, Pitzenberger SM, et al. Three-dimensional structure of echistatin and dynamics of the active site. J Biomol NMR. 1994 May;4(3):307–24.

Chen Y-C, Cheng C-H, Shiu J-H, Chang Y-T, Chang Y-S, Huang C-H, et al. Expression in *Pichia pastoris* and characterization of echistatin, an RGD-containing short disintegrin. Toxicon. 2012 Dec;60(8):1342–8.

Chiszar D, Lee RKK, Radcliffe CW, Smith HM. Searching behaviors by rattlesnakes following predatory strikes. In: J.A. Campbell and E.D. Brodie Jr., (Eds.) Biology of the Pitvipers. Tyler,Texas: Selva, 1992. pp. 369–82.

Chiszar D, Taylor SV, Radcliffe CW, Smith HM, O'Connell B. Effects of Chemical and Visual Stimuli upon Chemosensory Searching by Garter Snakes and Rattlesnakes. Journal of Herpetology. 1981 Oct 31;15(4):415.

Chiszar D, Walters A, Smith HM. Rattlesnake preference for envenomated prey: species specificity. Journal of Herpetology. 2008;42(4):764–7.

Chiszar D, Walters A, Urbaniak J, Smith HM, Mackessy SP. Discrimination between Envenomated and Nonenvenomated Prey by Western Diamondback Rattlesnakes (*Crotalus atrox*): Chemosensory Consequences of Venom. American Society of Ichthyologists and Herpetologists (ASIH). 1999 Aug;1999(3):640–8.

Chouhan BS, Käpylä J, Denessiouk K, Denesyuk A, Heino J, Johnson MS. Early Chordate Origin of the Vertebrate Integrin αI Domains. PLoS ONE. 2014 Nov 19;9(11):e112064.

Clark RW. Timber rattlesnakes (*Crotalus horridus*) use chemical cues to select ambush sites. J Chem Ecol. 2004;30(3):607–17.

Clark RW. Post-strike behavior of timber rattlesnakes (*Crotalus horridus*) during natural predation events. Ethology. 2006 Nov;112(11):1089–94.

Cooper WE. Chemical discrimination by tongue-flicking in lizards: A review with hypotheses on its origin and its ecological and phylogenetic relationships. J Chem. 1994 Feb;20(2):439–87.

Cooper WE. Tandem evolution of diet and chemosensory responses in snakes. Amphibia-Reptilia. 2008;29:393–8.

Cossins D. From Toxins to Therapeutics. The Scientist. 2013. Available from: http://www.the-scientist.com/?articles.view/articleNo/34745/title/From-Toxins-to-Therapeutics/

Cowles RB, Phelan RL. Olfaction in Rattlesnakes. American Society of Ichthyologists and Herpetologists (ASIH). 1958 Jun 18;1958(2):77–83.

Cundall D. Viper Fangs: Functional Limitations of Extreme Teeth. Physiological and Biochemical Zoology. 2009 Jan;82(1):63–79.

Currier RB, Calvete JJ, Sanz L, Harrison RA, Rowley PD, Wagstaff SC. Unusual stability of messenger RNA in snake venom reveals gene expression dynamics of venom replenishment. PLoS ONE. 2012;7(8):e41888.

Cushman DW, Ondetti MA. Design of angiotensin converting enzyme inhibitors. Nat Med. 1999 Oct;5(10):1110–3.

Darwin C. On the origin of species by means of natural selection. London: John Murray. 1859.

Desfilis E, Font E, Belekhova M, Kenigfest N. Afferent and efferent projections of the dorsal anterior thalamic nuclei in the lizard *Podarcis hispanica* (Sauria, Lacertidae). Brain Res Bull. 2002 Feb;57(3-4):447–50.

Diochot S, Baron A, Salinas M, Douguet D, Scarzello S, Dabert-Gay A-S, et al. Black mamba venom peptides target acid-sensing ion channels to abolish pain. Nature. 2012 Oct;490(7421):552–5.

Ducancel F, Durban J, Verdenaud M. Transcriptomics and venomics: implications for medicinal chemistry. Future Medicinal Chemistry. 2014 Oct;6(15):1629–43.

Eble JA, Kassner A, Niland S, Mörgelin M, Grifka J, Grässel S. Collagen XVI harbors an integrin $\alpha_1\beta_1$ recognition site in its C-terminal domains. J Biol Chem. 2006 Sep;281(35):25745–56.

Escalante T, Rucavado A, Fox JW, Gutiérrez JM. Key events in microvascular damage induced by snake venom hemorrhagic metalloproteinases. J Proteomics. 2011 Aug;74(9):1781–94.

Fahmi L, Makran B, Pla D, Sanz L, Oukkache N, Lkhider M, et al. Venomics and antivenomics profiles of North African. J Proteomics. 2012 Apr;75(8):2442–53.

Ferreira SH. A bradykinin-potentiating factor (BPF) present in the venom of *Bothrops jararaca*. British Journal of Pharmacology. 1965;24:163–9.

Fox JW, Serrano SMT. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon. 2005 Jun;45(8):969–85.

Fox JW, Serrano SMT. Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. Proteomics. 2008a Feb;8(4):909–20.

Fox JW, Serrano SMT. Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. FEBS Journal. 2008b Jun;275(12):3016–30.

Fry BG. From genome to "venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Research. 2005 Feb 14;15(3):403–20.

Fry BG, Casewell NR, Wuster W, Vidal N, Young B, Jackson TNW. The structural and functional diversification of the Toxicofera reptile venom system. Toxicon. 2012 Sep 15;60(4):434–48.

Fry BG, Wuster W. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Mol Biol Evol. 2004 May;21(5):870–83.

Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annu Rev Genomics Hum Genet. 2009a;10:483–511.

Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SFR, et al. Evolution of an Arsenal: Structural and Functional Diversification of the Venom System in the Advanced Snakes (Caenophidia). Molecular & Cellular Proteomics. 2008 Oct 14;7(2):215–46.

Fry BG, Undheim EAB, Ali SA, Jackson TNW, Debono J, Scheib H, et al. Squeezers and leaf-cutters: differential diversification and degeneration of the venom system in toxicoferan reptiles. Molecular & Cellular Proteomics. 2013 Jul;12(7):1881–99.

Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SFR, et al. Early evolution of the venom system in lizards and snakes. Nature. 2006 Feb 2;439(7076):584–8.

Fry BG, Vidal N, van der Weerd L, Kochva E, Renjifo C. Evolution and diversification of the Toxicofera reptile venom system. J Proteomics. 2009b Mar;72(2):127–36.

Fry BG, W ster W, Kini RM, Brusic V, Khan A, Venkataraman D, et al. Molecular Evolution and Phylogeny of Elapid Snake Venom Three-Finger Toxins. J Mol Evol. 2003 Jul 1;57(1):110–29.

Furry K, Swain T, Chiszar D. Strike-Induced Chemosensory Searching and Trail following by Prairie Rattlesnakes (*Crotalus viridis*) Preying upon Deer Mice (*Peromyscus maniculatus*): Chemical Discrimination among Individual Mice. Herpetologica. 1991;47(1):69–78.

Galán JA, Sanchez EE, Rodríguez-Acosta A, Soto JG, Bashir S, McLane MA, et al. Inhibition of lung tumor colonization and cell migration with the disintegrin crotatroxin 2 isolated from the venom of *Crotalus atrox*. Toxicon. 2008 Jun;51(7):1186–96.

Gan ZR, Gould RJ, Jacobs JW, Friedman PA, Polokoff MA. Echistatin. A potent platelet aggregation inhibitor from the venom of the viper, *Echis carinatus*. J Biol Chem. 1988 Dec;263(36):19827–32.

Garmy-Susini B, Varner JA. Roles of Integrins in Tumor Angiogenesis and Lymphangiogenesis. Lymphatic Research and Biology. 2008 Dec;6(3-4):155–63.

Gauthier JA, Kearney M, Maisano Anderson J, Rieppel O, Behlke ADB. Assembling the Squamate Tree of Life: Perspectives from the Phenotype and the Fossil Record. Bulletin of the Peabody Museum of Natural History. 2012 May;53(1):3–308.

Geim AK, Dubonos SV, Grigorieva IV, Novoselov KS, Zhukov AA, Shapoval SY. Microfabricated adhesive mimicking gecko foot-hair. Nat Mater. 2003 Jun 1;2(7):461–3.

Ghazaryan NA, Ghulikyan LA, Kishmiryan AV, Kirakosyan GR, Nazaryan OH, Ghevondyan TH, et al. Anti-tumor effect investigation of obtustatin and crude *Macrovipera lebetina obtusa* venom in S-180 sarcoma bearing mice. European Journal of Pharmacology. 2015 Jul;764:340–5.

Gillingham JC, Clark DL. Snake tongue-flicking: transfer mechanics to Jacobson's organ. Canadian Journal of Zoology. 1981;59:1651–7.

Gonçalves-Machado L, Pla D, Sanz L, Jorge RJB, Leitão-De-Araújo M, Alves MLM, et al. Combined venomics, venom gland transcriptomics, bioactivities, and antivenomics of two *Bothrops jararaca* populations from geographic isolated regions within the Brazilian Atlantic rainforest. J Proteomics. 2015 Apr;135(C):73–89.

Granada JF, Kleiman NS. Therapeutic use of intravenous eptifibatide in patients undergoing percutaneous coronary intervention: acute coronary syndromes and elective stenting. Am J Cardiovasc Drugs. 2004;4(1):31–41.

Greenbaum E. The influence of prey-scent stimuli on predatory behavior of the North American copperhead *Agkistrodon contortrix* (Serpentes: Viperidae). Behavioral Ecology. 2004;15(2):345–50.

Greenbaum E, Galeva N, Jorgensen M. Venom variation and chemoreception of the viperid *Agkistrodon contortrix*: evidence for adaptation? J Chem Ecol. 2003 Aug;29(8):1741–55.

Greene HW. The ecological and behavioral context for pitviper evolution. In: J.A. Campbell and E.D. Brodie Jr., (Eds.) Biology of the Pitvipers. Tyler,Texas: Selva, 1992, 7–118.

Greene HW, Fogden M, Fogden P. Snakes: the evolution of mystery in nature. Vol. 351. Berkeley: University of California Press; 1997. pp. 76–7.

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA Websuite. Nucleic Acids Research. 2008 May;36(Web Server):70–4.

Grzimek B. Reptiles (Vol. 7). In: Grzimek's animal life encyclopedia. 2nd Ed. Hutchins M and Schlager, N (Eds.). Farmington Hills: Thomson Gale; 2003.

Guindon S, Gascuel O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Systematic Biology. 2003 Oct 1;52(5):696–704.

Gutiérrez JM, Burnouf T, Harrison RA, Calvete JJ, Brown N, Jensen SD, et al. A Call for Incorporating Social Research in the Global Struggle against Snakebite. de Silva HJ, editor. PLoS Negl Trop Dis. 2015 Sep 17;9(9):e0003960.

Gutiérrez JM, Burnouf T, Harrison RA, Calvete JJ, Kuch U, Warrell DA, et al. A multicomponent strategy to improve the availability of antivenom for treating snakebite envenoming. Bull World Health Organ. 2014 Mar 4;92(7):526–32.

Gutiérrez JM, Lomonte B, León G, Alape-Girón A, Flores-Díaz M, Sanz L, et al. Snake venomics and antivenomics: Proteomic tools in the design and control of antivenoms for the treatment of snakebite envenoming. J Proteomics. 2009 Mar 6;72(2):165–82.

Gutiérrez JM, Solano G, Pla D, Herrera M, Segura Á, Villalta M, et al. Assessing the preclinical efficacy of antivenoms: from the lethality neutralization assay to antivenomics. Toxicon. 2013 Jul;69:168–79.

Gutiérrez JM, Theakston RDVG, Warrell DA. Confronting the Neglected Problem of Snake Bite Envenoming: The Need for a Global Partnership. Plos Med. 2006;3(6):e150.

Gutiérrez JM, Williams D, Fan HW, Warrell DA. Snakebite envenoming from a global perspective: Towards an integrated approach. Toxicon. 2010 Dec 15;56(7):1223–35.

Habib AG, Kuznik A, Hamza M, Abdullahi MI, Chedi BA, Chippaux J-P, et al. Snakebite is Under Appreciated: Appraisal of Burden from West Africa. PLoS Negl Trop Dis. 2015 Sep;9(9):e0004088.

Halpern M. Nasal chemical senses in reptiles: structure and function. Hormones, Brain and Behavior. In: Gans C, Crews D, (Eds.). Biology of the Reptilia; 1992. pp. 424–75 (Vol. 18).

Harburger DS, Calderwood DA. Integrin signalling at a glance. J Cell Sci. 2009 Jan 15;122(Pt 2):159–63.

Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. Genome Biol Evol. 2014a Aug;6(8):2088–95.

Hargreaves AD, Swain MT, Logan DW, Mulley JF. Testing the Toxicofera: Comparative transcriptomics casts doubt on the single, early evolution of the reptile venom system. Toxicon. 2014b Dec 15;92(C):140–56.

Harrison RA, Hargreaves A, Wagstaff SC, Faragher B, Lalloo DG. Snake Envenoming: A Disease of Poverty. White J, editor. PLoS Neglected Tropical Diseases 2009 Dec 22;3(12):e569.

Harvey AL. Twenty years of dendrotoxins. Toxicon. 2001 Jan;39(1):15–26.

Harvey AL. Toxins 'R'Us: more pharmacological tools from nature's superstore. Trends in pharmacological sciences. 2002;25(5):201–3.

Harvey AL. Toxins and drug discovery. Toxicon. 2014 Dec 15;92:193–200.

Harvey AL, Bradley KN, Cochran SA, Rowan EG, Pratt JA, Quillfeldt JA, et al. What can toxins tell us for drug discovery? Toxicon. 1998 Nov;36(11):1635–40.

Hayes WK, Kaiser II, Duvall D. The mass of venom expended by prairie rattlesnakes when feeding on rodent prey. In: J.A. Campbell and E.D. Brodie Jr., (Eds.) Biology of the Pitvipers. Tyler,Texas: Selva, 1992. pp. 383–8.

Hedges SB. A Molecular Phylogeny of Reptiles. Science. 1999 Feb 12;283(5404):998–1001.

Hedges SB, Vidal N. Lizards, snakes, and amphisbaenians (Squamata). In: Hedges SB and Kumar S (Eds). The timetree of life. 2009;:383–9.

Hedges SB. Amniote phylogeny and the position of turtles. BMC Biol. 2012;10(1):64.

Humphries JD, Byron A, Humphries MJ. Integrin ligands at a glance. J Cell Sci. 2006 Oct 1;119(Pt 19):3901–3.

Hynes RO. Integrins: bidirectional, allosteric signaling machines. Cell. 2002 Sep 20;110(6):673–87.

Iwabe N, Hara Y, Kumazawa Y, Shibamoto K, Saito Y, Miyata T, et al. Sister group relationship of turtles to the bird-crocodilian clade revealed by nuclear DNA-coded proteins. Mol Biol Evol. 2005 Apr;22(4):810–3.

Jackson TNW, Casewell NR, Fry BG. Response to "Replies to Fry et al. (Toxicon 2012, 60/4, 434–448). Part A. Analyses of squamate reptile oral glands and their products: A call for caution in formal assignment of terminology designating biological function." Toxicon. 2012 Nov 17;:1–14.

Janes DE, Organ CL, Fujita MK, Shedlock AM, Edwards SV. Genome Evolution in Reptilia, the Sister Group of Mammals. Annu Rev Genomics Hum Genet. 2010 Sep;11(1):239–64.

Jin H, Varner J. Integrins: roles in cancer development and as treatment targets. Br J Cancer. 2004 Feb 9;90(3):561–5.

Johnson MS, Lu N, Denessiouk K, Heino J, Gullberg D. Integrins during evolution: evolutionary trees and model organisms. Biochim Biophys Acta. 2009 Apr;1788(4):779–89.

Jones ME, Anderson CL, Hipsley CA, Müller J, Evans SE, Schoch RR. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). BMC Evol Biol. 2013;13:208.

Juárez P, Bolás G, de Rezende FF, Calvete JJ, Eble JA. Recombinant expression in human cells of active integrin $\alpha_1\beta_1$-blocking RTS-disintegrin jerdostatin. Toxicon. 2010 Nov;56(6):1052–8.

Juárez P, Comas I, González-Candelas F, Calvete JJ. Evolution of snake venom disintegrins by positive Darwinian selection. Mol Biol Evol. 2008 Nov;25(11):2391–407.

Juárez P, Sanz L, Calvete JJ. Snake venomics: Characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. Proteomics. 2004 Feb;4(2):327–38.

Juárez P, Wagstaff SC, Oliver J, Sanz L, Harrison RA, Calvete JJ. Molecular cloning of disintegrin-like transcript BA-5A from a *Bitis arietans* venom gland cDNA library: a putative intermediate in the evolution of the long-chain disintegrin bitistatin. J Mol Evol. 2006a Jul;63(1):142–52.

Juárez P, Wagstaff SC, Sanz L, Harrison RA, Calvete JJ. Molecular Cloning of *Echis ocellatus* Disintegrins Reveals Non-Venom-Secreted Proteins and a Pathway for the Evolution of Ocellatusin. J Mol Evol. 2006b Jul 7;63(2):183–93.

Junqueira-de-Azevedo ILM, Bastos CMV, Ho PL, Luna MS, Yamanouye N, Casewell NR. Venom-related transcripts from *Bothrops jararaca* tissues provide novel molecular insights into the production and evolution of snake venom. Mol Biol Evol. 2014 Dec;32(3):754–66.

Kapoor VK. Natural toxins and their therapeutic potential. Indian J Exp Biol. 2010 Mar;48(3):228–37.

Kardong KV. Predatory strike behavior of the rattlesnake, *Crotalus viridis oreganus*. Journal of Comparative Psychology.1986;100(3):304.

Kardong KV. Replies to Fry et al. (Toxicon 2012, 60/4, 434-448) Part B. Properties and biological roles of squamate oral products: The "venomous lifestyle" and preadaptation. Toxicon. 2012 Oct;60(5):964–6.

Kardong KV, Kiene TL, Bels V. Evolution of trophic systems in squamates. Netherlands Journal of Zoology. 1997;47(4):411–27.

Kasturiratne A, Wickremasinghe AR, de Silva N, Gunawardena NK, Pathmeswaran A, Premaratna R, et al. The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. Plos Med. 2008 Nov 4;5(11):e218.

Kelly CMR, Barker NP, Villet MH. Phylogenetics of advanced snakes (Caenophidia) based on four mitochondrial genes. Systematic Biology. 2003 Aug;52(4):439–59.

Kern A, Eble JA, Golbik R, Kühn K. Interaction of type IV collagen with the isolated integrins $\alpha_1\beta_1$ and $\alpha_2\beta_1$. Eur J Biochem. 1993 Jul;215(1):151–9.

Khoshnoodi J, Pedchenko V, Hudson BG. Mammalian collagen IV. Microsc Res Tech. 2008 May;71(5):357–70.

King GF. Venoms as a platform for human drugs: translating toxins into therapeutics. Expert Opinion on Biological Therapy. 2011 Sep 23;11(11):1469–84.

Kini RM. Anticoagulant proteins from snake venoms: structure, function and mechanism. Biochem J. 2006 Aug 1;397(3):377–87.

Kini RM, Evans HJ. Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon. 1992 Mar;30(3):265–93.

Kisiel DG, Calvete JJ, Katzhendler J, Fertala A, Lazarovici P, Marcinkiewicz C. Structural determinants of the selectivity of KTS-disintegrins for the $\alpha_1\beta_1$ integrin. FEBS Lett. 2004 Nov 19;577(3):478–82.

Koh CY, Kini RM. From snake venom toxins to therapeutics--cardiovascular examples. Toxicon. 2012 Mar 15;59(4):497–506.

Kordis D, Gubensek F. Adaptive evolution of animal toxin multigene families. Gene. 2000 Dec 30;261(1):43–52.

Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics. 2001 Dec;17(12):1244–5.

Lavin-Murcio P, Robinson BG, Kardong KV. Cues Involved in Relocation of Struck Prey by Rattlesnakes, *Crotalus viridis oreganus*. Herpetologica. 1993 Dec 1;49(4):463–9.

Lawson R, Slowinski JB, Crother BI, Burbrink FT. Phylogeny of the Colubroidea (Serpentes): New evidence from mitochondrial and nuclear genes. Molecular Phylogenetics and Evolution. 2005 Nov;37(2):581–601.

Le Blanc JCY, Hager JW, Ilisiu AMP, Hunter C, Zhong F, Chu I. Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications. Proteomics. 2003 Jun;3(6):859–69.

Lenk P. Evolutionary Relationships among the True Vipers (Reptilia: Viperidae) Inferred from Mitochondrial DNA Sequences. Molecular Phylogenetics and Evolution. 2001 Apr;19(1):94–104.

Lewis RJ, Garcia ML. Therapeutic potential of venom peptides. Nat Rev Drug Discov. 2003 Oct;2(10):790–802.

Lin F-Y, Zhu J, Eng ET, Hudson NE, Springer TA. β-Subunit Binding Is Sufficient for Ligands to Open the Integrin $\alpha_{IIb}\beta_3$ Headpiece. J Biol Chem. 2016 Feb 26;291(9):4537–46.

Lomonte B, Fernández J, Sanz L, Angulo Y, Sasa M, Gutiérrez JM, et al. Venomous snakes of Costa Rica: biological and medical implications of their venom proteomic profiles analyzed through the strategy of snake venomics. J Proteomics. 2014 Jun 13;105:323–39.

Longmire JL, Maltbie M, Baker RJ. Use of "lysis buffer" in DNA isolation and its implication for museum collections. The museum of Texas tech university; 1997. pp. 1–4.

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:26.

Lu X, Lu D, Scully MF, Kakkar VV. Snake venom metalloproteinase containing a disintegrin-like domain, its structure-activity relationships at interacting with integrins. Curr Med Chem Cardiovasc Hematol Agents. 2005 Jul;3(3):249–60.

Mackessy SP. Venom composition in rattlesnakes: trends and biological significance. In: Hayes WK, Beaman KR, Cardwell MD, P BS, (Eds.). The biology of rattlesnakes; Loma Linda, CA: Loma Linda University Press, 2008. pp. 495–510.

Mackessy SP. Evolutionary trends in venom composition in the Western Rattlesnakes (Crotalus viridis sensu lato): Toxicity vs. tenderizers. Toxicon. 2010 Jul;55(8):1463–74.

Makran B, Fahmi L, Pla D, Sanz L, Oukkache N, Lkhider M, et al. Snake venomics of *Macrovipera mauritanica* from Morocco, and assessment of the para-specific immunoreactivity of an experimental monospecific and a commercial antivenoms. J Proteomics. 2012 Apr 18;75(8):2431–41.

Marcinkiewicz C, Lobb RR, Marcinkiewicz MM, Daniel JL, Smith JB, Dangelmaier C, et al. Isolation and characterization of EMS16, a C-Lectin type protein from *Echis multisquamatus* venom, a potent and selective inhibitor of the $\alpha_2\beta_1$ Integrin. Biochemistry. 2000 Aug;39(32):9859–67.

Marcinkiewicz C, Vijay-Kumar S, McLane MA, Niewiarowski S. Significance of RGD loop and C-terminal domain of echistatin for recognition of $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$ integrins and expression of ligand-induced binding site. Blood. 1997 Aug 15;90(4):1565–75.

Marcinkiewicz C, Weinreb PH, Calvete JJ, Kisiel DG, Mousa SA, Tuszynski GP, et al. Obtustatin: A potent selective inhibitor of $\alpha_1\beta_1$ integrin *in vitro* and angiogenesis *in vivo*. Cancer Research. 2003 May 1;:2020–3.

Markland FS. Snake venom fibrinogenolytic and fibrinolytic enzymes: an updated inventory. Registry of exogenous hemostatic factors of the scientific and standardization committee of the international society on thrombosis and haemostasis. Thromb Haemost. 1998 Mar;79(3):668–74.

Markland FSJ, Swenson S. Snake venom metalloproteinases. Toxicon. 2013 Feb;62:3–18.

Marsh NA. Diagnostic uses of snake venom. Haemostasis. 2001 May;31(3-6):211–7.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. PNAS. 2004 May;101(19):7287–92.

McLane MA, Marcinkiewicz C, Vijay-Kumar S, Wierzbicka-Patynowski I, Niewiarowski S. Viper venom disintegrins and related molecules. Proc Soc Exp Biol Med. 1998 Nov;219(2):109–19.

McLane MA, Vijay-Kumar S, Marcinkiewicz C, Calvete JJ, Niewiarowski S. Importance of the structure of the RGD-containing loop in the disintegrins echistatin and eristostatin for recognition of $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$ integrins. FEBS Lett. 1996 Aug;391(1-2):139–43.

Menez AE. Perspectives in Molecular Toxinology. Chichester, UK: John Wiley & Sons, LTD; 2002.

Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. FEBS Lett. 2000 Feb 25;468(2-3):109–14.

Modahl, C. M., Mackessy, S.P. (2016) Full-length venom protein cDNA sequences from venom-derived mRNA: exploring compositional variation and adaptive multigene evolution. PLoS Neglected Tropical Diseases (*in press*)

Momic T, Arlinghaus FT, Arien-Zakay H, Katzhendler J, Eble JA, Marcinkiewicz C, et al. Pharmacological aspects of *Vipera xantina palestinae* venom. Toxins. 2011 Nov 1;3(11):1420–32.

Monleón D, Esteve V, Kovacs H, Calvete JJ, Celda B. Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. Biochem J. 2005 Apr 1;387(1):57–66.

Monleón D, Moreno-Murciano MP, Kovacs H, Marcinkiewicz C, Calvete JJ, Celda B. Concerted motions of the integrin-binding loop and the C-terminal tail of the non-RGD disintegrin obtustatin. J Biol Chem. 2003 Nov 14;278(46):45570–6.

Moreno-Murciano MP, Monleón D, Calvete JJ, Celda B, Marcinkiewicz C. Amino acid sequence and homology modeling of obtustatin, a novel non-RGD-containing short disintegrin isolated from the venom of *Vipera lebetina obtusa*. Protein Sci. 2003a Feb;12(2):366–71.

Moreno-Murciano MP, Monleón D, Marcinkiewicz C, Calvete JJ, Celda B. NMR solution structure of the non-RGD disintegrin obtustatin. J Mol Biol. 2003b May;329(1):135–45.

Moura-da-Silva AM, Theakston RDVG, Cramptonl JM. Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: Gene duplication and divergence of a common ancestor rather than convergent evolution. J Mol Evol. 1996 Sep;43(3):263–9.

Muro EM, Mah N, Andrade-Navarro MA. Functional evidence of post-transcriptional regulation by pseudogenes. Biochimie. 2011 Nov;93(11):1916–21.

Navdaev A, Lochnit G, Eble JA. The rhodocetin αβ subunit targets GPIb and inhibits von Willebrand factor induced platelet activation. Toxicon. 2011;57(7-8):1041–8.

Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. PNAS. 1997 Jul 22;94(15):7799–806.

Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 2005;39:121–52.

Ogawa T, Chijiwa T, Oda-Ueda N, Ohno M. Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom. Toxicon. 2005 Jan;45(1):1–14.

Okuda D, Koike H, Morita T. A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. Biochemistry. 2002 Dec 3;41(48):14248–54.

Olfa K-Z, José L, Salma D, Amine B, Najet SA, Vidal N, et al. Lebestatin, a disintegrin from *Macrovipera* venom, inhibits integrin-mediated cell adhesion, migration and angiogenesis. Laboratory Investigation. 2005 Dec;85(12):1507–16.

Olivera BM, Teichert RW. Diversity of the neurotoxic *Conus* peptides: a model for concerted pharmacological discovery. Mol Interv. 2007 Oct;7(5):251–60.

Ovadia M. Isolation and characterization of a hemorrhagic factor from the venom of the snake *Atractaspis engaddensis* (Atractaspididae). Toxicon. 1987;25(6):621–30.

Peichoto ME, Tavares FL, Rodríguez S, Mackessy SP. Venom proteomes of South and North American opisthoglyphous (Colubridae and Dipsadidae) snake species: a preliminary approach to understanding their biological roles. Comp Biochem Physiol Part D Genomics Proteomics. 2012 Dec;7(4):361–9.

Pereiraa SL, Bakera AJ. Waterfowl and gamefowl (Galloanserae). In: Hedges SB and Kumar S (Eds). The timetree of life. 2009;:415–8.

Pincheira-Donoso D, Bauer AM, Meiri S, PPeter U. Global Taxonomic Diversity of Living Reptiles. PLoS ONE. 2013;8(3):e59741.

Pink RC, Wicks K, Caley DP, Kathleen EP, Jacobs L, Carter DRF. Pseudogenes: Pseudo-functional or key regulators in health and disease? RNA. Cold Spring Harbor Laboratory Press; 2011 May 1;17(5):792–8.

Piskurek O, Austin CC, Okada N. Sauria SINEs: Novel short interspersed retroposable elements that are widespread in reptile genomes. J Mol Evol. 2006 Apr 11;62(5):630–44.

Piskurek O, Nishihara H, Okada N. The evolution of two partner LINE/SINE families and a full-length chromodomain-containing *Ty3/Gypsy* LTR element in the first reptilian genome of *Anolis carolinensis*. Gene. 2009 Jul 15;441(1–2):111–8.

Pyron RA, Burbrink FT. Extinction, ecological opportunity, and the origins of global snake diversity. Evolution. 2012 Jan;66(1):163–78.

Pyron RA, Burbrink FT, Colli GR, de Oca ANM, Vitt LJ, Kuczynski CA, et al. The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees.. Molecular Phylogenetics and Evolution. 2011 Feb 1;58(2):329–42.

Pyron RA, Burbrink FT, Wiens JJ. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol Biol. 2013 Apr 29;13(1):1.

Reeder TW, Townsend TM, Mulcahy DG, Noonan BP, Wood PL Jr, Sites JW Jr, et al. Integrated Analyses Resolve Conflicts over Squamate Reptile Phylogeny and Reveal Unexpected Placements for Fossil. PLoS ONE. 2015 Mar 24;10(3):e0118199.

Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM, et al. Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. Molecular Phylogenetics and Evolution. 2003 Nov;29(2):289–97.

Reyes-Velasco J, Card DC, Andrew AL, Shaney KJ, Adams RH, Schield DR, et al. Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. Mol Biol Evol. 2014 Dec 19;32(1):173–83.

Rieppel O. Turtles as hopeful monsters. Bioessays. 2001 Nov;23(11):987–91.

Risch M, Georgieva D, Bergen von M, Jehmlich N, Genov N, Arni RK, et al. Snake venomics of the Siamese Russell's viper (Daboia russelli siamensis). J Proteomics. 2009 Mar 6;72(2):256–69.

Rögl F, Steininger FF. Vom Zerfall der Tethys zu Mediterran und Paratethys. Die neogene Paläogeographie und Palinspastik des zirkum-mediterranen Raumes. Wien: Annalen des Naturhistorischen Museums in Wien; 1983. pp. 135–63 (Vol. 85/A).

Sanchez EE, Rodríguez-Acosta A, Palomar R, Lucena SE, Bashir S, Soto JG, et al. Colombistatin: a disintegrin isolated from the venom of the South American snake (*Bothrops colombiensis*) that effectively inhibits platelet aggregation and SK-Mel-28 cell adhesion. Arch Toxicol. 2009 Mar;83(3):271–9.

Sanz L, Ayvazyan N, Calvete JJ. Snake venomics of the Armenian mountain vipers *Macrovipera lebetina obtusa* and *Vipera raddei*. J Proteomics. 2008 Jul;71(2):198–209.

Sanz L, Bazaa A, Marrakchi N, Pérez A, Chenik M, Bel Lasfer Z, et al. Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina transmediterranea* venom gland cDNA libraries: insight into the evolution of the snake venom integrin-inhibition system. Biochem J. 2006 Apr 15;395(2):385–92.

Sanz L, Chen R-Q, Pérez A, Hilario R, Juárez P, Marcinkiewicz C, et al. cDNA cloning and functional expression of jerdostatin, a novel RTS-disintegrin from *Trimeresurus jerdonii* and a specific antagonist of the $\alpha_1\beta_1$ integrin. J Biol Chem. 2005 Dec 9;280(49):40714–22.

Sanz-Soler R, Lorente C, Company B, Sanz L, Juárez P, Pérez A, et al. Recombinant expression of mutants of the *Frankenstein* disintegrin, RTS-ocellatusin. Evidence for

the independent origin of RGD and KTS/RTS disintegrins. Toxicon. 2012 Sep 15;60(4):665–75.

Saviola AJ, Chiszar D, Bealor MT. Response of western diamondback rattlesnakes (*Crotalus atrox*) to chemical cues of mice (*Mus musculus*) of different genders and reproductive status. The Psychological Record. 2010;60:217–26.

Saviola AJ, Chiszar D, Busch C, Mackessy SP. Molecular basis for prey relocation in viperid snakes. BMC Biol. 2013;11:20.

Saviola AJ, Lamoreaux WE, Opferman R. Chemosensory response of the threatened eastern indigo snake (*Drymarchon couperi*) to chemical and visual stimuli. Herpetological Conservation and Biology. 2011 Dec;6(3):449–54.

Saviola AJ, McKenzie VJ, Chiszar D. Chemosensory responses to chemical and visual stimuli in five species of colubrid snakes. Acta Herpetologica. 2012;7(1):91–103.

Saviola AJ, Modahl CM, Mackessy SP. Disintegrins of *Crotalus simus tzabcan* venom: Isolation, characterization and evaluation of the cytotoxic and anti-adhesion activities of tzabcanin, a new RGD disintegrin. Biochimie. 2015 Sep;116:92–102.

Scarborough J. Pharmacy and Drug Lore in Antiquity: Greece, Rome, Byzantium. Social History of Medicine. VT: Ashgate Variorum, 2010, pp. XXVIII-345

Scarborough RM, Rose JW, Hsu MA, Phillips DR. Barbourin. A GPIIb-IIIa-specific integrin antagonist from the venom of *Sistrurus m. barbouri*. Journal of Biological chemistry. 1991;266(15):9359–62.

Scarborough RM, Rose JW, Naughton MA, Phillips DR, Nannizzi L, Arfsten A, et al. Characterization of the integrin specificities of disintegrins isolated from American pit viper venoms. Journal of Biological chemistry. 1993 Jan 25;268(15):1068–1065.

Scheen AJ. Pharmacokinetics and Clinical Use of Incretin-Based Therapies in Patients with Chronic Kidney Disease and Type 2 Diabetes. Clin Pharmacokinet. 2014 Oct 21;54(1):1–21.

Schoch RR, Sues H-D. A Middle Triassic stem-turtle and the evolution of the turtle body plan. Nature. 2015 Jun 24;523(7562):584–7.

Schwenk K. Of tongues and noses: chemoreception in lizards and snakes. Trends in Ecology & Evolution. 1995 Jan;10(1):7–12.

Sebé-Pedrós A, Roger AJ, Lang FB, King N, Ruiz-Trillo I. Ancient origin of the integrin-mediated adhesion and signaling machinery. Proceedings of the National Academy of Sciences. 2010 Jun 1;107(22):10142–7.

Seifert P. Toxicology in classical antiquity. Hippokrates. 1954 Jun 30;25(12):389–93.

Shaffer HB. Turtles (Testudines). In: Hedges SB and Kumar S (Eds). The timetree of life. 2009;:398–401.

Shedlock AM, Botka CW, Zhao S, Shetty J, Zhang T, Liu JS, et al. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. PNAS. 2007 Feb 20;104(8):2767–72.

Shedlock AM, Edwards SV. Amniotes (amniota). Hedges SB, Kumar S, editors. The timetree of life. 2009;:375–9.

Shimokawa K, Jia LG, Shannon JD, Fox JW. Isolation, sequence analysis, and biological activity of atrolysin E/D, the non-RGD disintegrin domain from *Crotalus atrox* venom. Arch Biochem Biophys. 1998 Jun 15;354(2):239–46.

Shultz D. Snakebites deadly as other diseases in West Africa. Science. 2015 Sep 26. http://www.sciencemag.org/news/2015/09/snakebites-deadly-other-diseases-west-africa

Smith JB, Theakston RDG, Coelho ALJ, Barja-Fidalgo C, Calvete JJ, Marcinkiewicz C. Characterization of a monomeric disintegrin, ocellatusin, present in the venom of the Nigerian carpet viper, *Echis ocellatus*. FEBS Lett. 2002 Feb 13;512(1-3):111–5.

Smith KJ, Jaseja M, Lu X, Williams JA, Hyde EI, Trayer IP. Three-dimensional structure of the RGD-containing snake toxin albolabrin in solution, based on 1H NMR spectroscopy and simulated annealing calculations. Int J Pept Protein Res. 1996 Sep;48(3):220–8.

Staniszewska I, Walsh EM, Rothman VL, Gaathon A, Tuszynski GP, Calvete JJ, et al. Effect of VP12 and viperistatin on inhibition of collagen-receptor-dependent melanoma metastasis. Cancer Biol Ther. 2009 Aug;8(15):1507–16.

Takeda S, Takeya H, Iwanaga S. Snake venom metalloproteinases: structure, function and relevance to the mammalian ADAM/ADAMTS family proteins. Biochim Biophys Acta. 2012 Jan;1824(1):164–76.

Tan NH, Saifuddin MN. Isolation and characterization of a hemorrhagin from the venom of *Ophiophagus hannah* (king cobra). Toxicon. 1990;28(4):385–92.

Townsend TM, Mulcahy DG, Noonan BP, Sites JW, Kuczynski CA, Wiens JJ, et al. Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. Molecular Phylogenetics and Evolution. 2011 Nov;61(2):363–80.

Trinklein ND, Karaöz U, Wu J, Halees A, Force Aldred S, Collins PJ, et al. Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome. Genome Research. 2007 Jun;17(6):720–31.

Tropea JE, Cherry S, Waugh DS. Expression and purification of soluble His(6)-tagged TEV protease. Methods Mol Biol. 2009;498:297–307.

Tsairi H, Bouskila A. Ambush site selection of a desert snake (*Echis coloratus*) at and oasis. Herpetologica. 2004 Mar;60(1):13–23.

Tucker GC. Integrins: Molecular targets in cancer therapy. Curr Oncol Rep. 2006 Apr;8(2):96–103.

Uetz P, Hošek J. The Reptile Database. PP U, Hošek J, editors. Uetz P; 2015. Available from: http://www.reptile-database.org

Vidal N, Delmas A-S, Carter DRF, Cruaud C, Couloux A, Hedges SB. The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. Comptes Rendus Biologies. 2007 Feb;330(2):182–7.

Vidal N, Hedges SB. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. C R Biologies. 2005 Oct;328(10-11):1000–8.

Vidal N, Hedges SB. The molecular evolutionary tree of lizards, snakes, and amphisbaenians. C R Biologies. 2009 Mar 9;332(2-3):129–39.

Vidal N, Rage J-C, Coloux A, Hedges SB. Snakes (Serpentes). In: Hedges SB and Kumar S (Eds). The timetree of life. 2009;:390–7.

Vonk FJ, Admiraal JF, Jackson K, Reshef R, de Bakker MAG, Vanderschoot K, et al. Evolutionary origin and development of snake fangs. Nature. 2008 Jul 31;454(7204):630–3.

Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJR, et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proceedings of the National Academy of Sciences. 2013 Dec 17;110(51):20651–6.

Vonk FJ, Jackson K, Doley R, Madaras F, Mirtschin PJ, Vidal N. Snake venom: From fieldwork to the clinic. Bioessays. 2011 Jan 27;33(4):269–79.

Wagstaff SC, Sanz L, Juárez P, Harrison RA, Calvete JJ. Combined snake venomics and venom gland transcriptomic analysis of the ocellated carpet viper, *Echis ocellatus*. J Proteomics. 2009 Jan 30;71(6):609–23.

Walsh EM, Marcinkiewicz C. Non-RGD-containing snake venom disintegrins, functional and structural relations. Toxicon. 2011 Sep 15;58(4):355–62.

Weinstein AS, Keyler DE, White J. Replies to Fry et al. (Toxicon 2012, 60/4, 434-448) Part A. Analyses of squamate reptile oral glands and their products: A call for caution in formal assignment of terminology designating biological function. Toxicon. 2012 Oct 1;60(5):954–63.

Werneburg I, Sánchez-Villagra MR. Timing of organogenesis support basal position of turtles in the amniote tree of life. BMC Evol Biol. 2009;9:82.

Wierzbicka-Patynowski I, Niewiarowski S, Marcinkiewicz C, Calvete JJ, Marcinkiewicz MM, McLane MA. Structural requirements of echistatin for the recognition of $\alpha_V\beta_3$ and $\alpha_V\beta_1$ integrins. J Biol Chem. 1999 Dec 31;274(53):37809–14.

Williams D, Gutiérrez JM, Harrison R, Warrell DA, White J, Winke KD, et al. The Global Snake Bite Initiative: an antidote for snake bite. Vol. 375. The lancet; 2010. pp. 89–91.

Williams DJ, Gutiérrez JM, Calvete JJ, Wüster W, Ratanabanangkoon K, Paiva O, et al. Ending the drought: New strategies for improving the flow of affordable, effective antivenoms in Asia and Africa. J Proteomics. 2011 Aug 24;74(9):1735–67.

Wilusz JE. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. Biochim Biophys Acta. 2015 Jun 23;:1–11.

Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes & Development. 2009 Jul 1;23(13):1494–504.

Xiong JP. Crystal Structure of the Extracellular Segment of Integrin $\alpha_V\beta_3$ in Complex with an Arg-Gly-Asp Ligand. Science. 2002 Mar 7;296(5565):151–5.

Yang L-J, Niu B, Zhang D, Yang T. Substitution of the Echistatin Amino Acid Motif RGDD with KGDW Enhances Inhibition of Platelet Aggregation and Thrombogenesis. International Journal of Peptide Research and Therapeutics. 2015 Jun 11;:1–8.

Young BA. Evaluating hypotheses for the transfer of stimulus particles to Jacobson's organ in snakes. Brain Behav Evol. 1993;41(3-5):203–9.

Zardoya R, Meyer A. The evolutionary position of turtles revised. Naturwissenschaften. 2001 May 1;88(5):193–200.

Annex


Toxicon
Recombinant expression of mutants of the Frankenstein disintegrin, RTS-ocellatusin. Evidence for the independent origin of RGD and KTS/RTS disintegrins
Manuscripts submitted to Integrative and Comparative Biology
Integrative and Comparative Biology
DISTRIBUTION OF RPTLN GENES ACROSS REPTILIA. HYPOTHESIZED ROLE FOR RPTLN IN THE EVOLUTION OF SVMPs

Supplemental material

Table 1S . Accession numbers of *RPTLN* gene copies across reptiles. Full-length sequences are displayed in Figure 54. Nucleotide changes in *RPTLN* genes respect to *RPTLN*-1 [jerdostatin, AY262730] are listed in Table 11.

Species	*RPTLN* gene copy
P. jerdonii	*RPTLN-1*
B. asper	*RPTLN*-1 (KU563546), *RPTLN*-15 (KU563589), *RPTLN*-16 (KU563592)
B. lateralis	*RPTLN*-1 (KU563547), *RPTLN*-7 (KU563579), *RPTLN*-17 (KU563595)
E.ocellatus	*RPTLN*-1 (KU563548), *RPTLN*-13 (KU563587)
A. picadoi	*RPTLN*-1 (KU563549), *RPTLN*-10 (KU563583), *RPTLN*-15 (KU563590)
O. hannah	*RPTLN*-1 (KU563550), *RPTLN*-6 (KU563577), *RPTLN*-17 (KU563596)
N. haje haje	*RPTLN*-1 (KU563551), *RPTLN*-5 (KU563575)
R. scalaris	*RPTLN*-1 (KU563552), *RPTLN*-5 (KU563608), *RPTLN*-18 (KU563600), *RPTLN*-19 (KU563603), *RPTLN*-20 (KU563604), *RPTLN*-21 (KU563617)
B. constrictor	*RPTLN*-1 (KU563553), *RPTLN*-17 (KU563597)
L. hispanica	*RPTLN*-1 (KU563554), *RPTLN*-9 (KU563581), *RPTLN*-17 (KU563598)
T. lepidus	*RPTLN*-1 (KU563555)
P. muralis	*RPTLN*-1 (KU563556), *RPTLN*-9 (KU563610), *RPTLN*-17 (KU563614), *RPTLN*-18 (KU563615),*RPTLN*-19 (KU563616)
P. hispanica	*RPTLN*-1 (KU563557), *RPTLN*-8 (KU563609), *RPTLN*-11(KU563611), *RPTLN*-13 (KU563612), *RPTLN*-16 (KU563613), *RPTLN*-21 (KU563618)
H. horridum	*RPTLN*-1 (KU563558)
U. ebenaui	*RPTLN*-1 (KU563559), *RPTLN*-9 (KU563582)
T. mauritanica	*RPTLN*-1 (KU563560), *RPTLN*-11 (KU563584)
C. calypratus	*RPTLN*-1 (KU563561), *RPTLN*-17 (KU563599)
A. mississipiensis	*RPTLN*-1 (KU563562), *RPTLN*-6 (KU563578), *RPTLN*-11 (KU563585), *RPTLN*-12 (KU563586), *RPTLN*-14 (KU563588)
T. greca	*RPTLN*-1 (KU563563), *RPTLN*-3 (KU563573)
T. hermanni	*RPTLN*-1 (KU563564)
S. pardalis	*RPTLN*-1 (KU563565), *RPTLN*-16 (KU563593), *RPTLN*-18 (KU563601)
M. annamensis	*RPTLN*-1 (KU563566), *RPTLN*-15 (KU563591)
M. sintesis	*RPTLN*-1 (KU563567), *RPTLN*-4 (KU563574), *RPTLN*-7 (KU563580), *RPTLN*-18 (KU563602)
M. leprosa	*RPTLN*-1 (KU563568), *RPTLN*-2 (KU563571), *RPTLN*-5 (KU563576)
C. carbonaria	*RPTLN*-1 (KU563569), *RPTLN*-2 (KU563572)
C. chilensis	*RPTLN*-1 (KU563570), *RPTLN*-16 (KU563594)

Table 2S. Accession numbers of *RPTLN* gene copies amplified from mRNAs. Full-length sequences are displayed in Figure 54. Nucleotide changes in *RPTLN* genes respect to *RPTLN*-1 [jerdostatin, AY262730] are listed in Table 11.

Species	*RPTLN* gene copy
R. scalaris	*RPTLN-1* (KU563605), *RPTLN*-22 (273A>T) (KU563619), *RPTLN*-21 (KU563617), *RPTLN*-5 (KU563608)
P. muralis	*RPTLN*-1 (KU563606), *RPTLN*-9 (KU563610), *RPTLN*-17 (KU563614), *RPTLN*-18 (KU563615), *RPTLN*-19 (KU563616)
P. hispanica	*RPTLN*-1 (KU563607), *RPTLN*-8 (KU563609), *RPTLN*-11 (KU563611), *RPTLN*-13 (KU563612), *RPTLN*-16 (KU563613), *RPTLN*-21 (KU563618)

Table 3S. Raw data collected and raw data percent, for behavioral experiments consisted of paired trials using non-envenomated vs envenomated (r-crotatroxin) mouse. Trials were of 10 minutes duration, and the number of tongue flicks directed toward one or the other mouse was recorded and analyzed by an observer blind to the condition. Standard Error of the Mean (s.e.m) is indicated next to the mean

Trial number	Snake tongue flicks directed to		Snake tongue flicks (%)directed to	
	NE mouse (PBS)	E mouse (r-crotatroxin)	NE mouse (PBS)	E mouse (r-crotatroxin)
1	21	58	26.6	73.4
2	36	148	19.6	80.4
3	34	48	41.5	58.5
4	102	222	31.5	68.5
5	70	221	24.1	75.9
6	30	68	30.6	69.4
7	20	23	46.5	53.5
8	79,3	25	79.3	20.7
9	0	207	0	100.0
10	38	76	33.3	66.7
11	8	25	24.2	75.8
12	6	291	2,0	98.0
Mean	38.4 (9.8)	117.7 (27.8)	30.0	70.0 (6)

Table 4S. Raw data collected and raw data percent, for behavioral experiments consisted of paired trials using non-envenomated vs envenomated (r-ocellatusin) mouse. Trials were of 10 minutes duration, and the number of tongue flicks directed toward one or the other mouse was recorded and analyzed by an observer blind to the condition. Standard Error of the Mean (s.e.m) is indicated next to the mean

Trial number	Snake flicks directed to		Snake flicks (%) directed to	
	NE mouse (PBS)	E mouse (r-ocellatusin)	NE mouse (PBS)	E mouse (r-ocellatusin)
1	134	6	95.7	4.3
2	131	31	80.9	19.1
3	129	125	50.8	49.2
4	44	32	57.9	42.1
5	63	111	36.2	63.8
6	41	13	75.9	24.1
7	35	31	53.0	47.0
8	47	27	63.5	36.5
Average (%)	78 (12.1)	47 (15.9)	64.2	35.8 (5.5)

Publication I

Toxicon 60 (2012) 665–675

Contents lists available at SciVerse ScienceDirect

Toxicon

journal homepage: www.elsevier.com/locate/toxicon

Recombinant expression of mutants of the *Frankenstein* disintegrin, RTS-ocellatusin. Evidence for the independent origin of RGD and KTS/RTS disintegrins

Raquel Sanz-Soler [a,1], Carolina Lorente [a], Beatriz Company [a], Libia Sanz [a], Paula Juárez [a], Alicia Pérez [a], Yun Zhang [b], Yang Jin [b], Runqiang Chen [c], Johannes A. Eble [d], Juan J. Calvete [a,*], Gema Bolás [a,*,1]

[a] *Instituto de Biomedicina de Valencia, C.S.I.C., Jaume Roig 11, 46010 Valencia, Spain*
[b] *Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China*
[c] *La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA*
[d] *Center for Molecular Medicine, Department of Vascular Matrix Biology, Frankfurt-am-Main, Germany*

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form 5 May 2012
Accepted 17 May 2012
Available online 4 June 2012

Keywords:
Recombinant RGD-disintegrin
Ocellatusin
Recombinant RTS/KTS-disintegrin
Jerdostatin
Evolution of disintegrins

ABSTRACT

The requirements to transform a short disintegrin of the RGD clade into an RTS disintegrin, were investigated through the generation of recombinant mutants of ocellatusin in which the RGD tripeptide was substituted for RTS in different positions along the integrin-specificity loop. Any attempt to create an active integrin $\alpha_1\beta_1$ inhibitory motif within the specificity loop of ocellatusin was unsuccessful. Replacing the whole RGD-loop of ocellatusin by the RTS-loop of jerdostatin was neither sufficient for confering $\alpha_1\beta_1$ binding specificity to this ocellatusin-RTS *Frankenstein*[2] mutant. Factors other than the integrin-binding loop sequence *per se* are thus required to transform a disintegrin scaffold from the RGD clade into another scaffold from the RTS/KTS clade. Moreover, our results provide evidences, that the RTS/KTS short disintegrins have potentially been recruited into the venom gland of Eurasian vipers independently from the canonical neofunctionalization pathway of the RGD disintegrins. PCR-amplifications of jerdostatin-like sequences from a number of taxa across reptiles, including snakes (Crotalinae, Viperinae, and Elapidae taxa) and lizards (Lacertidae and Iguanidae) clearly showed that genes coding for RTS/KTS disintegrins existed long before the split of Lacertidae and Iguania, thus predating the recruitment of the SVMP precursors of disintegrins, providing strong support for the view of an independent evolutionary history of the RTS/KTS and the RGD clades of short disintegrins.

1. Introduction

Snakes of the subfamilies Viperinae (vipers) and Crotalinae (pitvipers and rattlesnakes) of Viperidae have developed in their venoms a broad spectrum of integrin receptor antagonists; with the exception of the $\alpha_2\beta_1$ integrin, which is targeted by a number of C-type lectin-like proteins (Ogawa et al., 2005), inhibitory motifs towards β_1 and β_3 integrins have evolved in different members of the

* Corresponding authors. Tel.: +34 96 339 1778; fax: +34 96 369 0800.
E-mail addresses: jcalvete@ibv.csic.es (J.J. Calvete), gbolas@ibv.csic.es (G. Bolás).

[1] These authors contributed equally.

[2] The term *Frankenstein* used here as a synonym for "a chimeric protein made with pieces from different molecules", is a tribute to Mary W. Shelley's "Frankenstein; or, the Modern Prometheus", Lackington, Hughes, Harding, Mavor & Jones, Gradifco, Switzerland, 1818.

0041-0101/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
10.1016/j.toxicon.2012.05.010

disintegrin family (Calvete et al., 2009; Calvete, 2010). Disintegrins, a family of small (40–100 amino acids), cysteine-rich polypeptides broadly distributed in the venoms of vipers and rattlesnakes (reviewed by Calvete et al., 2009; Calvete, 2010), are released into viper venoms by the proteolytic processing of PII snake venom metalloprotease (SVMP) precursors (Kini and Evans, 1992), or synthesized from short-coding mRNAs (Okuda et al., 2002). Disintegrins have been structurally classified according to their length and number of disulfide bonds into long, medium-sized, dimeric and short subfamilies. The evolutionary pathway(s) of disintegrins involved minimization of both the genomic and protein structures, including, respectively, the stepwise loss of introns and pairs of cysteine linkages from long precursors to the most recent short disintegrin scaffold (Calvete et al., 2009; Calvete, 2010). Functionally, disintegrins have evolved by positive Darwinian evolution guided by the adaptation of a conformational epitope (the integrin recognition loop and the C-terminal tail) to the active site of the targeted integrin receptors (Juárez et al., 2008). The RGD tripeptide has been inferred to represent the ancestral integrin recognition motif, which emerged during the Paleogene period of the Cenozoic Era (approximately 54–64 Mya) from a subgroup of PIII-SVMP bearing the RDE<u>C</u>D sequence (Juárez et al., 2008) which originated by recruitment, duplication, and neofunctionalization of a cellular ADAM-7 or 28 ancestor gene (Moura-Da-Silva et al., 1996; Fry, 2005; Fry et al., 2006, 2008; Vidal et al., 2009; Casewell, 2012).

Following the deletion of the (underlined) PIII-lineage-specific Cys residue, conversion of RDE into RGD can be accomplished with a minimum of two mutations (Calvete et al., 2009; Calvete, 2010). Structural studies of a number of short (1RO3) and medium-sized (kistrin, 1N4Y; flavoridin, 1FVL; albolabrin (Smith et al., 1996); salmosin, 1IQ2, 1L3X; rhodostomin, 1Q7J, 1Q7I, 2PJI, 2PJF, 1JYP; trimestatin, 1J2L) and dimeric disintegrins (1Z1X, 1RMR, 1TEJ, 3CO5) have revealed that their RGD/KGD integrin inhibitory tripeptides have evolved at the apex of a mobile loop protruding 14–17 ≈ from the disintegrin protein scaffold and maintained in the active conformation by the appropriate pairing of cysteine residues. Currently known integrin-blocking motifs include RGD, which blocks integrins $\alpha_8\beta_1$, $\alpha_5\beta_1$, $\alpha_v\beta_1$, $\alpha_v\beta_3$, and $\alpha_{IIb}\beta_3$; MLD targets the $\alpha_4\beta_1$, $\alpha_4\beta_7$, $\alpha_3\beta_1$, $\alpha_6\beta_1$, $\alpha_7\beta_1$ and $\alpha_9\beta_1$ integrins; VGD and MGD impair the function of the $\alpha_5\beta_1$ integrin; KGD inhibits the $\alpha_{IIb}\beta_3$ integrin with a high degree of selectivity; WGD has been reported to be a potent inhibitor of the RGD-dependent integrins $\alpha_5\beta_1$, $\alpha_v\beta_3$, and $\alpha_{IIb}\beta_3$; the adhesive function of the latter integrin is also blocked by MVD (Sanz et al., 2006; Calvete et al., 2009; Calvete, 2010).

The crystal structure of the extracellular segment of integrin αvβ3 in complex with an RGD ligand (Xiong et al., 2002) revealed that the peptide fits into a crevice between the αV propeller and the β3 A-domain. The Arg side-chain is held in place by interactions with αV carboxylates, the Gly residue makes several hydrophobic interactions with αV, and the Asp ligand interacts primarily with βA residues. Thus, the conserved aspartate residue might be responsible for the binding of disintegrins to integrin receptors which share a β subunit, while the two other residues of the integrin-binding motif (RG, KG, MG, WG, ML, MV, VG) may dictate the primary integrin-recognition specificity, with residues flanking the active tripeptide finely tuning the potency and integrin receptor selectivity of disintegrins (McLane et al., 1996; Wierzbicka-Patynowski et al., 1999; reviewed by Calvete, 2005; Calvete et al., 2005). High-resolution NMR studies (Monleón et al., 2005) provided a structural ground for the biochemically defined functional synergy between the RGD loop and the C-terminal region of echistatin (Marcinkiewicz et al., 1997).

Short RTS/KTS disintegrins, which selectively target the integrin $\alpha_1\beta_1$ (Calvete et al., 2007), form a distinct clade within the short disintegrin subfamily (Fig. 1). Compared to all known disintegrin structures, in which the RGD motif is located at the apex of an eleven residue hairpin loop, the active RTS/KTS tripeptide is oriented towards a side of a nine residue integrin-binding loop (Moreno-Murciano et al., 2003). Structure-function correlation studies have shown that the selectivity of KTS-disintegrins for the $\alpha_1\beta_1$ integrin resides within a conformational epitope encompassing the integrin-binding loop and the C-terminal tail (Moreno-Murciano et al., 2003; Monleón et al., 2003; Kisiel et al., 2004). The potency of recombinant KTS-disintegrin obtustatin also depended on the residue C-terminally adjacent to the active motif (Brown et al., 2009).

The striking similar functional requirements and structural differences between RGD and RTS/KTS short disintegrins prompted us to investigate a possible transformation route of RGD-ocellatusin into RTS-jerdostatin. The lack of inhibitory activity of the different constructions, even when the whole RGD-loop of ocellatusin was replaced by the RTS-loop of jerdostatin, suggested that RTS/KTS short disintegrins could have been recruited independently of the canonical RGD disintegrins. PCR-amplifications of jerdostatin-like sequences from a number of taxa across reptiles support this hypothesis.

2. Materials and methods

2.1. Materials

Recombinant soluble human integrin α1β1 was produce in transfected *Drosophila* Schneider cells and isolated as previously described (Eble et al., 2006). The Collagen IV fragment CB3 was generated as described (Kern et al., 1993). Antiserum PEP160 against the C-terminal tail of jerdostatin was produced in rabbit by Abintek Biopharma S.L. (Parque Tecnológico de Bizkaia, Derio, Bizkaia, Spain) using a standard immunization protocol and the synthetic peptide CKPSYPGNG conjugated to keyhole limpet hemocyanin (KLH) as immunogen. Anti-CKPSYPGNG antibodies were affinity-purified on a peptide-Sepharose column. Anti-human integrin β1 subunit antibody was raised in rabbit (Eble et al., 1993). KTS-disintegrin lebestatin [Q3BK14] was isolated from the venom of *Macrovipera lebetina* as described (Bazaa et al., 2005).

Blood sample from *Iguana iguana* was kindly provided by Dr. Vet. Med. David García, Hospital Veterinario, Burjassot (Valencia, Spain). Tissues from *Lacerta hispanica* were obtained from specimens caught in Valencia. Tissues from *Naja haje haje* and *Daboia russelli* were kindly provided by

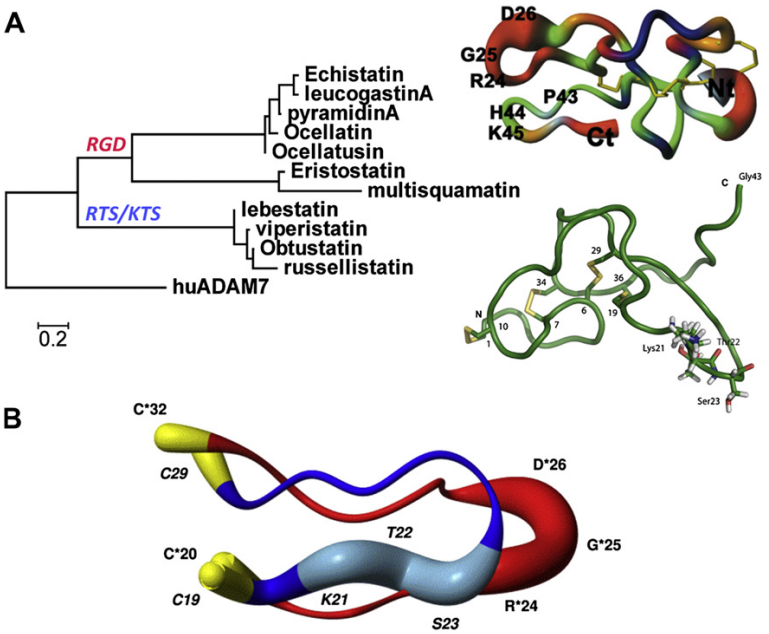


Fig. 1. Inferred phylogeny within the short disintegrin subfamily. Panel A, cladistic relationships between the RGD and the KTS/RTS short disintegrins were inferred through Neighbor-Joining using maximum-likelihood distances using the PHYML program (Guindon and Gascuel, 2003), with the disintegrin-like domain of human ADAM-7 serving as outgroup. This distribution parallels the phylogenetic tree of the species in whose venoms short disintegrins have been characterized. Expression of RGD-disintegrins appear to be restricted to Echis taxa, whereas RTS/KTS-disintegrins have been only reported in venoms from Eurasian vipers, genera Vipera, Macrovipera, and Daboia. Molecular models correspond to the NMR solution structures of the RGD-disintegrin echistatin (1RO3, Monleón et al., 2005) and the RTS-disintegrin jerdostatin (2W9O, Carbajo et al., 2011). Panel B, superposition of the integrin-binding loops of echistatin (red) and obtustatin (blue), highlighting their different size and the different topology of their active motifs, RGD and KTS (Moreno-Murciano et al., 2003). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the private zoological collection owner, César Olmos (Cullera, Valencia). Genomic DNA from *Sistrurus catenatus catenatus* and *Crotalus viridis* were kindly provided by Prof. H.L. Gibbs (Department of Evolution, Ecology and Organismal Biology, Ohio State University, USA). Genomic DNA from *Echis ocellatus* was kindly provided by Dr. Robert A. Harrison (Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, UK). *Bungarus multicinctus* was collected in South China and genomic DNA samples were isolated from the liver with a genomic DNA isolation kit (TaKaRa, Dalian) according to the manufacture's protocol. *Gallus gallus* (domestic chicken) and *Bombina orientalis* (Fire-bellied toad) were purchased at local animal dealers.

2.2. Design and cloning of r-jerdostatin and r-ocellatusin

The full length cDNA jerdostatin sequence was PCR-amplified using pRc_CMV/FlagJerdostatin vector (Juárez et al., 2010) as template. The forward primer, 5′-ATGGTACC**GAGAATCTTTACTTCCAAGGA**GCCCCACTAGCCGAC-3′, contained a KpnI restriction site (underlined) and the cleavage site for the tobacco etch virus (TEV) protease (in bold). The reverse primer, 5′-GCCTCGAGTA**TTA**GCCATTCCCGGGATAAC-3′, included a restriction site for XhoI (underlined), the last six C-terminal residues of jerdostatin and a stop codon (in bold).

Wild-type ocellatusin was PCR-amplified from an *E. ocellatus* venom gland cDNA (Juárez et al., 2006) using the forward primer 5′-GGAGATCTC**GAGAATCTTTACTTCCAAGGA**GACTGTGAATCTGGACC-3′ and the reverse primer 5′-GTAAGCTT**CTA**CGGATCATGTTCGCCTTTG-3′. The primers included, respectively, BglII and HindIII restriction sites (underlined), the TEV cleavage site (forward primer, in bold) and a stop codon (reverse primer, in bold). The PCR-amplification protocol included initial denaturation at 94 °C for 2 min, followed by 35 cycles of denaturation (94 °C for 30 s), annealing (60 °C for 30 s), extension (72 °C for 30 s) and a final extension step for 7 min at 72 °C. The PCR products were purified using the Illustra GFX Gel Band Purification Kit (GE Healthcare, Buckinghamshire, UK), cloned in a pGEM-T vector (Promega, Madison, WI, USA), and then used to transform *Escherichia coli* DH5α strain cells (Novagen, Madison, WI) by electroporation using an Eppendorf 2510 electroporator. Positive clones were selected by growing the transformed cells in Luria-Broth (LB) medium containing 100 μg/ml ampicillin. Positive clones were confirmed by PCR and sequenced on an Applied Biosystems model 377 DNA sequencer.

To construct an expression vector of disintegrin-thioredoxin-His_6 fusion protein the pGEM-T/disintegrin

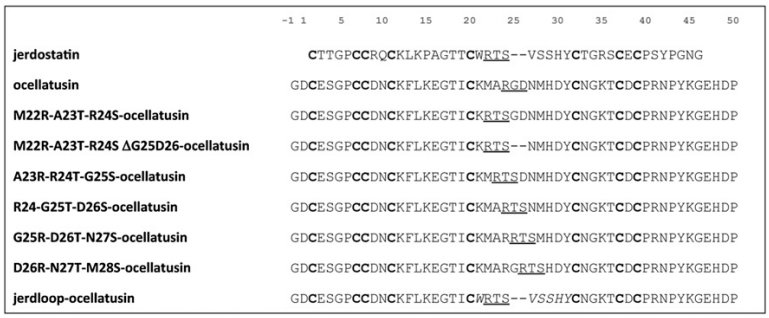


Fig. 2. Amino acid sequences and nomenclature of wild-type jerdostatin, wild-type ocellatusin and the *Frankenstein* RTS-ocellatusin mutants. Cysteine residues are in bold, and the (R/K)TS and RGD motifs are underlined. Residue labeled −1 corresponds to the last residue of the TEV protease cleavage site, ENLYFQG, inserted between the His_6 tag and the disintegrin sequence.

plasmids and the pET32a vector (Novagen, Madison, WI) were respectively digested with KpnI/XhoI and BglII/HindIII restriction enzimes for 24 h at 37 °C. The disintegrin fragments and the linear pET32a vector were purified by agarose gel electrophoresis using the Illustra GFX Gel Band Purification kit (GE Healthcare), and ligated with T4 DNA ligase (Invitrogen) overnight at 4 °C. *E. coli DH5α* strain cells were transfected with this construct, and positive clones were checked by PCR and DNA sequencing.

2.3. Generation of r-RTS-ocellatusin mutants

Hybrid ("*Frankenstein*") ocellatusin–jerdostatin constructions (Fig. 2) were generated by site-directed mutagenesis using recombinant ocellatusin as template (Sanz et al., 2005) and the iProof™ High Fidelity Master Mix kit. To this end, the plasmid pET32a/ocellatusin was used as template for PCR amplification (denaturation at 94 °C for 2 min, followed by 12 cycles of denaturation for 30 s at 94 °C, annealing for 60 s at 55 °C, extension for 12 min at 68 °C, and a final extension for 10 min at 68 °C), using different complementary primers for each mutant. Additionally, to obtain the jerdloop-ocellatusin mutant we generated the intermediate construction pET32/K21W-M22R-A23T-R24S ΔG25D26-ocellatusin that served as template for the final PCR mutagenesis. The forward primers used are listed in Table 1. The PCR products were digested with DpnI and used to transform electrocompetents *E. coli* DH5α cells. Plasmids were isolated with DNA Purification System Withard® Plus SV Minipreps (Promega) and sequenced.

2.4. Recombinant expression and purification of r-jerdostatin, r-ocellatusin and the Frankenstein ocellatusin mutants

E. coli BL21 strain cells (Novagen, Madison, WI) were transfected with the pET32a/disintegrin constructs. The presence of the disintegrin–thioredoxin fusion constructs in positive clones was checked by PCR. Positive *E. coli BL21* clones were grown overnight at 37 °C in LB medium containing 100 μg/ml ampicillin, followed by a 1:50 (v/v) dilution in the same medium until an OD_{600} of around 0.8–1 was reached. Expression of the recombinant fusion proteins was then induced by addition of isopropyl-D-thiogalactosidase (IPTG) to a final concentration of 1 mM, and incubation of the cell suspensions for 4 h at 37 °C. Cells were pelleted by centrifugation (4000*g* for 30 min), resuspended in the same volume of 20 mM sodium phosphate, 150 mM NaCl, pH 7.4, washed three times with this buffer, and resuspended in 50 mL of 20 mM sodium phosphate, 250 mM NaCl, pH 7.4. Cells were lyzed by sonication (15 cycles of 15 s sonication/1 min resting) in an ice bath. The lysates were centrifuged at 10,000 g for 30 min at 4 °C, and the soluble and the insoluble fractions were analyzed by SDS-PAGE using 12% polyacrylamide gels under reducing conditions. Disintegrin-thioredoxin-His_6 fusion proteins were purified from the soluble fraction of the lysate by affinity chromatography using an ÄKTA Basic chromatograph equipped with a 5 ml HisTrap HP column (Amersham Biosciences) equilibrated in 20 mM sodium phosphate, 250 mM NaCl, pH 7.4. Bound protein was eluted with a linear gradient of 50–500 mM imidazole. Eluted fractions were checked by SDS-PAGE and those containing

Table 1
List of forward primer sequences (5′–3′) used to PCR-amplify the *Frankenstein* ocellatusin constructions displayed in Fig. 2. Nucleotides coding for the mutated residues are in boldface.

M22R-A23T-R24S-ocellatusin	AGGAACAATATGCAAG**AGGACAAGC**GGTGATAACATGC
M22R-A23T-R24S ΔG25D26-ocellatusin	CAATATGCAAG**AGGACAAGC**AACATGCATGATTACTGC
A23R-R24T-G25S-ocellatusin	GGAACAATATGCAAGATG**AGAACGAGT**GATAACATGCATG
R24-G25T-D26S-ocellatusin	ATGCAAGATGGCA**AGGACTAGT**AACATGCATGATTACTGC
G25R-D26T-N27S-ocellatusin	TGCAAGATGGCAAGG**CGTACTAGC**ATGCATGATTACTG
D26R-N27T-M28S-ocellatusin	ATGGCAAGGGGT**CGTACCAGC**CATGATTACTG
K21W-M22R-A23T-R24S ΔG25D26-ocellatusin	TGAAGGAAGGAACAATATGC**TGGAGGACA**AGCAACATGCATG
Jerdloop-ocellatusin	TGC**TGGAGGACAAGCGTCAGCAGTCAT**TACTGCAATGG

the fusion protein were pooled, dialyzed against 20 mM sodium phosphate, 250 mM NaCl, pH 7.4, and digested overnight at 4 °C with TEV-His_6 protease (1:20, w/w). The recombinant disintegrins were separated from thioredoxin-His_6 and TEV-His_6 by chromatography on a HisTrap column (as above), the flow-through fractions were concentrated using Amicon filtration membranes with a pore size of 3000 Da (Millipore, MA, USA), and the proteins purified by reverse-phase HPLC on a C18 column (250 × 4.6 mm, 5 μm, Waters, MA, USA) equilibrated with 0.05% TFA in water (solution A) and eluted with a 45 min linear gradient of 5–70% acetonitrile in 0.05% TFA. The purity of the isolated protein was assessed by SDS-PAGE and electrospray-ionization mass spectrometry using a QTrap 2000 instrument (Applied Biosystems) equipped with a nanoelectrospray source (Proxeon, Denmark). Protein concentration was determined using the bicinchoninic acid protein method (BCA™ Protein Assay, Pierce).

2.5. Western blot analysis

Reverse-phase purified jerdostatin was analyzed by Tris–Tricine–SDS-(10%)PAGE under reducing conditions and immobilized onto PVDF membrane (Hybond-P, GE Healthcare) using a semi-dry electrotranfer device. The membranes, blocked with 5% (w/v) non-fat dried milk in PBS (20 mM Na_2HPO_4 pH 7,5, 150 mM NaCl) for 1 h at 25 °C, were incubated with a 1:500 (v/v) dilution of anti-PEP160 polyclonal antibodies in 5% non-fat dried milk/PBS-T (20 mM Na_2HPO_4 pH 7,5, 150 mM NaCl, 0.1% (v/v) Tween-20) for 1 h at 25 °C. The membranes were then washed three times with PBS-T followed by incubation with a 1:10,000 (v/v) dilution of peroxidase-conjugated anti-rabbit IgG (Sigma) in 5% non-fat dried milk in PBS-T. After 3× washing with PBS-T, the membranes were developed using the chemiluminescence ECL Plus kit (GE Healthcare).

2.6. Platelet aggregation assay

Human platelets were isolated from fresh blood from healthy volunteers as described (Navdaev et al., 2011). Platelet aggregation was monitored in a CHRONO-LOG aggregometer with continuous stirring at 37 °C. The reaction mixture contained 400 μl of freshly prepared washed human platelets (cell count adjusted to $\sim 5 \times 10^8$ per ml), 1 μl $CaCl_2$ (1M), $MgCl_2$ (1M), 1 μl of PGE1 20 μM, and 1 μl of either PBS or recombinant ocellatusin serial dilutions from 2.1 mg/l to 2.1 μg/ml (0.37 μM–0.37 nM). The aggregation was initiated by adding 2 μL of 1 mg/mL collagen I 1min after the other components had been mixed at 37 °C. The reaction was allowed to proceed for at least 3 min. The extent of inhibition of aggregation was expressed as the percentage of the rate of aggregation observed in the absence of inhibitor.

2.7. Inhibition of soluble α1β1 integrin binding to CB3 by wild-type and Frankenstein disintegrins

The inhibitory activity of the purified recombinant disintegrins was carried out following the protocol described by Juárez et al. (2010). Briefly, the collagen type IV fragment CB3 was immobilized on a 96-well plate overnight at 4 °C in TBS/Mg^{2+} (20 mM Tris, 150 mM NaCl, pH 7.5 containing 2 mM $MgCl_2$) at a concentration of 5 μg/ml. The plate was washed three times with TBS/Mg^{2+} and non-specific binding sites blocked with 1%(v/v) BSA in TBS/Mg^{2+} at room temperature for 1 h. 3.5 μg/ml of soluble α1β1 integrin, dissolved in TBS/Mg^{2+} were mixed with different concentrations of recombinant disintegrins. The mixtures were added to the plate and incubated for 2 h at room temperature. The plate was then washed twice with 50 mM Hepes (pH 7.5) containing 150 mM NaCl, 2 mM $MgCl_2$ and 1 mM $MnCl_2$, and the bound integrin was fixed with 2.5% (v/v) glutaraldehyde in the same buffer for 10 min at room temperature. For α1β1 integrin detection a primary rabbit anti-β1 antiserum (1:2000) was employed. After 3μ washing with PBS, goat anti-rabbit IgG conjugated with alkaline phosphatase (AP, 1:2000) was added. 4-nitrophenyl phosphate disodium salt hexahydrate (Sigma) was used as AP substrate, colour was developed at room temperature and quantified in an ELISA plate reader at 405 nm.

2.8. PCR-amplification of jerdostatin-like DNA sequences

Genomic DNA was extracted from fresh tissues or blood samples essentially as described (Bazaa et al., 2005). PCR-amplification of jerdostatin-like sequences was done in a final volume of 25 μl containing 50 ng of genomic DNA, 0.5 units of each DNA and Pfu polymerases (Biotools), and 1 μl (0.4 μM final concentration) of forward primer 5′-TGT ACA ACT GGA CCA TGT TGT C (CTTGPCC) and reverse primer 3′-TAG CCA TTC CCG GGA TAA CTG (SYPGNG*). The PCR protocol included denaturation at 94 °C for 10 min, followed by 35 cycles of denaturation for 45 s at 94 °C, annealing for 45 s at 50 °C, extension for 1 min at 72 °C, and a final extension for 10 min at 72 °C. Amplification of jerdostatin from *Prothrobothrops jerdonii* and jerdostatin-like DNA sequence from *B. multicinctus* venom gland cDNA libraries was carried out at Kumming Institute of Zoology. The PCR protocol was carried out in a final volume of 50 μl containing 200 pmol of each primer (forward: 5′-CCAAATCCAG(C/T)CTCCAAAATG-3′; reverse: 5′-TTCCAG/TCTCCATTGTTG(G/T)TTA-3′), using 0.5 μl RT reaction mixture as DNA template. After denaturation at 94 °C for 5 min, rTaq polymerase (TaKaRa Biotech., Dalian) was added, followed by 33 cycles (94 °C for 30 s, 50 °C 30 s, 72 °C 2 min), and ended with 72 °C for 10 min. After agarose gel electrophoresis analysis, the DNA fragment about 300 bp was extracted and inserted into pGEM-T vector for sequencing.

3. Results and discussion

3.1. Design rationale

Short RGD-disintegrins appear to be restricted to African and Asian *Echis* and *Eristicophis* species, and represent the most recent members of the disintegrin family (Juárez et al., 2008; Calvete, 2010), which may have evolved after the radiation of Viperinae during the late Oligocene or the early Miocene, between 22 and 24 Mya

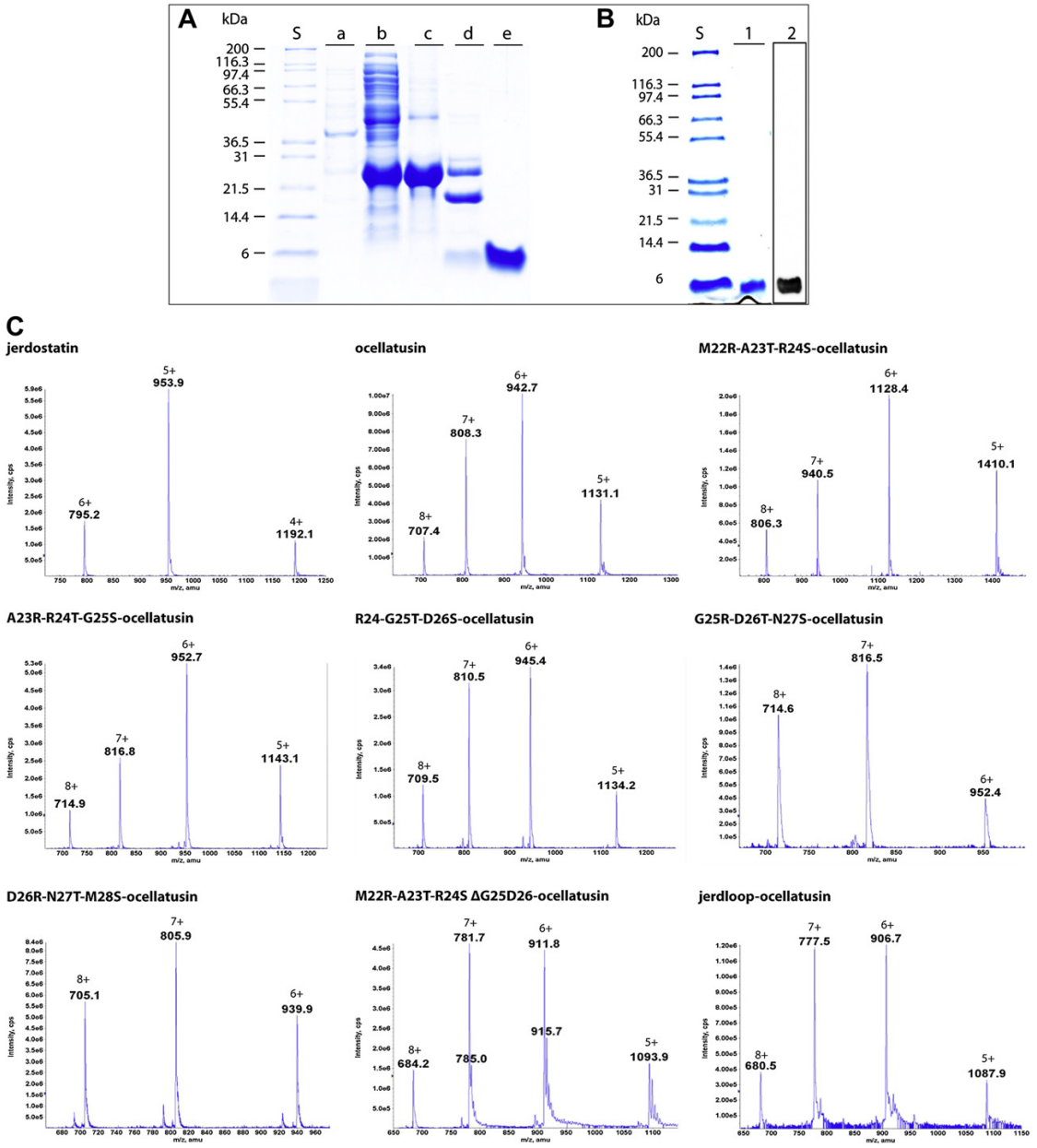


Fig. 3. Expression and purification of recombinant disintegrins. Panel A, analysis by Tris–Tricine–SDS-(10%) PAGE of the overexpression and purification steps of r-M22R-A23T-R24S-ocellatusin. Lanes a and b, insoluble and soluble fractions, respectively, of lysates of *E. coli* BL21 cells expressing the disintegrin-TEV-thioredoxin-His_6 fusion protein. Lane c, HisTrap affinity-purified disintegrin-thioredoxin-His_6 (24 kDa). Lane d, digestion products of r-M22R-A23T-R24S-ocellatusin-TEV-thioredoxin-His_6 fusion protein after incubation with TEV protease. Lane e, r-M22R-A23T-R24S-ocellatusin purified by reverse-phase HPLC from the flow-through of the HisTrap affinity column of the protein mixture shown in lane d. Panel B, Coomassie blue-stained SDS-PAGE (lane **1**), and Western blot analysis using anti-PEP160 antibodies (lane **2**) of reverse-phase HPLC-purified jerdostatin. Lanes **S**, molecular weight markers (Mark12™, Invitrogen), whose apparent molecular mass is indicated at the left side of the gels. Panel C, electrospray-ionization mass spectrometry of reverse-phase HPLC-purified wild-type recombinant jerdostatin and ocellatusin, and the *Frankenstein* disintegrins listed in Fig. 2.

Table 2
Experimental (ESI-MS) and calculated molecular masses of the recombinant disintegrins. Calculated masses correspond to fully oxidized (4 disulfide bonds) monoisotopic species.

Recombinant protein	ESI-MS (Da)	Calculated monoisotopic mass (Da) (4SS bonds)
Jerdostatin	4764.7 ± 0.5	4765.4
Ocellatustin	5650.8 ± 0.5	5650.3
M22R-A23T-R24S-ocellatusin	5636.7 ± 1.5	5636.2
M22R-A23T-R24S ΔG25D26-ocellatusin	5465.0 ± 0.7	5464.1
A23R-R24T-G25S-ocellatustin	5710.9 ± 0.7	5710.3
R24-G25T-D26S-ocellatustin	5666.7 ± 1.3	5666.3
G25R-D26T-N27S-ocellatustin	5708.6 ± 0.2	5708.4
D26R-N27T-M28S-ocellatustin	5634.9 ± 0.3	5634.3
Jerdloop-ocellatusin	5435.1 ± 1.0	5435.0

(Castoe et al., 2009), at a time when eastern North America and Eurasia were widely separated across the Atlantic, whereas northeastern Asia and Alaska remained connected via the Bering land bridge. The proposed mechanism for the emergence of the short RGD-disintegrin ocellatusin from a short-coding dimeric disintegrin precursor involves just two nucleotide mutations (Juárez et al., 2006). Furthermore, the most parsimonious nucleotide substitution model required for the emergence of all known XXD disintegrin's integrin inhibitory sequences from an ancestral RGD motif (Juárez et al., 2008) involves a minimum of three mutations (Calvete, 2010). However, circumstantial evidence suggest that KTS/RTS disintegrins may not follow this canonical evolutionary scenario. Hence, i) no putative dimeric disintegrin precursor has been found in the few species from which KTS/RTS-disintegrins have been isolated or cloned, i.e. in *M. lebetina* (lebestatin, Bazaa et al., 2005), *Vipera palestinae* (viperistatin, Kisiel et al., 2004), *M. lebetina obtusa* (obtustatin, Marcinkiewicz et al., 2003; Sanz et al., 2008), *Prothrobothops jerdonii* (jerdostatin, unpublished observations), *D. russelli* (russellistatin, unpublished observations); ii) whereas the integrin-inhibitory loops of XXD disintegrins are absolutely conserved in residue length (11 amino acids) and harbor the active tripeptide at the tip, the active tripeptides of jerdostatin (RTS, Carbajo et al., 2011) and obtustatin (KTS, Moreno-Murciano et al., 2003; Monleón et al., 2003) are oriented towards the side of nine-residue integrin-binding loops.

To investigate the minimal requirements to transform the RGD-disintegrin ocellatusin into the $\alpha_1\beta_1$-blocking RTS-disintegrin jerdostatin, *Frankenstein* disintegrins were designed in which the RGD tripeptide was substituted for RTS (R24-G25T-D26S-ocellatusin). In addition, mutants A23R-R24T-G25S-ocellatusin, G25R-D26T-N27S-ocellatusin, D26R-N27T-M28S-ocellatusin, and M22R-A23T-R24S-ocellatusin were designed to vary the topology of the engineered RTS motif from the tip of the loop (as in ocellatusin) towards a lateral position (as in jerdostatin). The length of the integrin-binding loop of ocellatusin was preserved in these four *Frankenstein* mutants. Another mutant, M22R-A23T-R24S-ΔG25D26-ocellatusin, was designed to shorten the length of the integrin-binding loop from 11 to 9 residues (as in jerdostatin) while maintaining the lateral topology of the RTS motif. Finally, the full sequence of the inhibitory loop of ocellatusin was replaced by the complete amino acid sequence of jerdostatin's integrin-binding loop in the jerdloop-ocellatusin mutant (Table 1).

3.2. Recombinant expression and activity of the Frankenstein mutants

E. coli BL21 cells transformed with pET32a/TEVDisintegrin plasmids overexpressed soluble thioredoxin-His_6 fusion proteins (Fig. 3A and B). Purification yields of r-jerdostatin, r-ocellatusin (wild-type and mutants) were about 0.5–1 mg/L of cell culture. Electrospray-ionization mass spectrometry (Fig. 3C), proved that the experimental molecular masses accurately matched the calculated masses for the recombinant disintegrins with fully oxidized cysteine residues (Table 2). This confirmed the correct primary sequence.

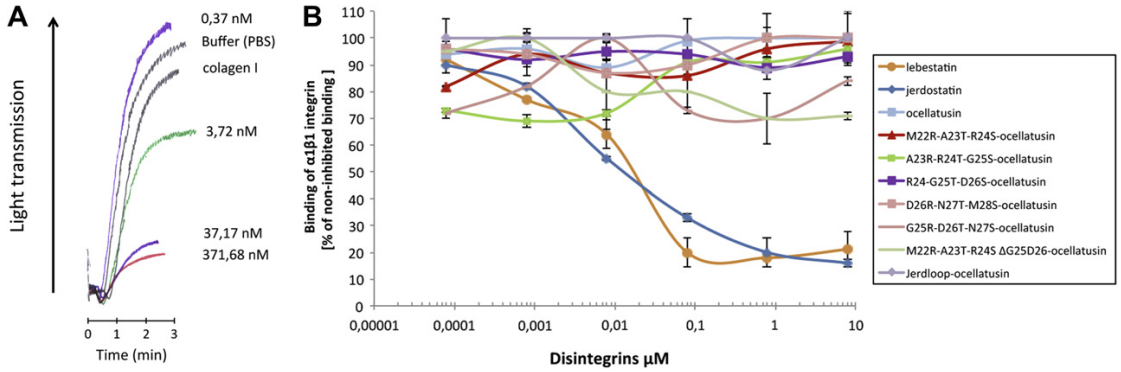


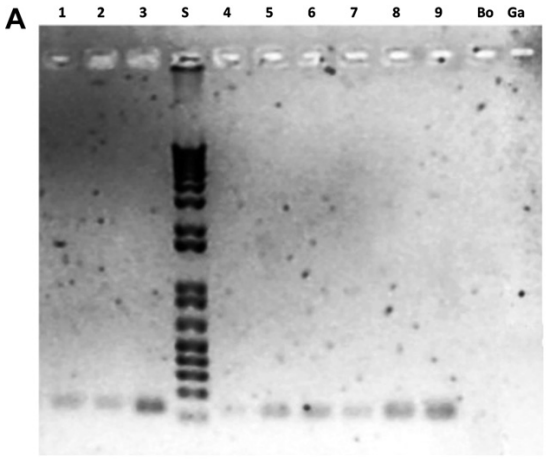
Fig. 4. Inhibitory activity of recombinant disintegrins. Panel A, Concentration-dependent blocking of collagen-induced platelet aggregation by recombinant ocellatusin. Panel B, Inhibition of the binding of integrin α1β1 to CB3 fragment of collagen IV was assessed by incubating soluble integrin α1β1 (3.5 µg/mL) with increasing concentrations of r-jerdostatin, lebestatin, r-ocellatusin and r-RTS-ocellatusin mutants in 96-wells plates coated with 5 µg/mL of CB3. Bound integrin was detected by ELISA. Lebestatin and wild-type r-ocellatusin were used as positive and negative inhibition controls, respectively.

3.3. Inhibition of collagen-induced platelet aggregation by recombinant ocellatusin

Recombinant ocellatusin inhibited the collagen-induced aggregation of human washed platelets in a dose-dependent manner with an IC_{50} of 3.5×10^{-9} M (Fig. 4A). This result, confirming that r-ocellatusin inhibited platelet aggregation with a potency similar to that reported for the natural short RGD-disintegrin echistatin (IC_{50} of 3×10^{-8} M) (Gan et al., 1988), strongly suggests that the recombinant disintegrin may have folded into the same biologically-active conformation than its venom-isolated homologue (1RO3, Monleón et al., 2005).

3.4. Recombinant Frankenstein disintegrins do not exhibit $\alpha 1\beta 1$ integrin inhibitory activity

In agreement with previous studies (Juárez et al., 2010), both recombinant jerdostatin and lebestatin (positive controls, Olfa et al., 2005) blocked in a concentration-dependent and divalent ion-independent manner the high-affinity interaction between the soluble hetero-dimeric ectodomain of $\alpha_1\beta_1$ and the CB3 fragment of collagen IV (Fig. 4B). In contrast, wild-type r-ocellatusin (negative control) did not show any inhibitory activity. Similarly, none of the *Frankenstein* RTS-disintegrins blocked the binding of $\alpha_1\beta_1$ to the CB3 collagen fragment.



B

SNAKES

Crotalinae

```
          1        10        20        30        40        50        60        70        80        90       100       110       120       129
Pj        TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTGTCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
            C  T  T  G  P  C  C  R  Q  C  K  L  K  P  A  G  T  T  C  W  R  T  S  V  S  S  H  Y  C  T  G  R  S  C  E  C  P  S  Y  P  G  N  G
Cvi       TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAGGCCGGCAGGAACAACATGCTGGAAAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTACCAGTTATCCCGGGAATGGC*
            C  T  T  G  P  C  C  R  Q  C  K  L  R  P  A  G  T  T  C  W  K  T  S  V  S  S  H  Y  C  T  G  R  S  C  E  C  T  S  Y  P  G  N  G
Scc       TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTGTCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
```

Viperinae

```
Ml        TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTGTCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
Cv        TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
Eo        TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
Dr        TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAAAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
```

Elapidae

```
Bm        TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGGTCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
Nnh1      TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
Nnh2      TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAAAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
```

LIZARDS

Lacertidae

```
Lh        TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
```

Iguanidae

```
Ii        TGTACAACTGGACCATGTTGTCGTCAGTGCAAATTGAAGCCGGCAGGAACAACATGCTGGAGAACCAGTGTATCAAGTCATTACTGCACTGGCAGATCTTGTGAATGTCCCAGTTATCCCGGGAATGGC*
```

Fig. 5. Jerdostatin-like DNA sequences across Reptilia. Panel A, PCR-amplification of jerdostatin-like sequences from genomic DNA of *Protobothrops jerdonii* (Pj) (1), *Crotalus viridis* (Cvi) (2), *Sistrurus catenatus catenatus* (Scc) (3), *Echis ocellatus* (Eo) (4), *Daboia russelli* (Dr) (6), *Naja haje haje* (Nhh) (5), *Lacerta hispanica* (Lh) (7), and *Iguana iguana* (Ii) (8). Panel B, jerdostatin-like DNA sequences from the DNA amplified in A. Sequences from *Macrovipera lebetina* (Ml) and *Cerastes vipera* (Cv) are from Bazaa et al. (2007) and Sanz et al. (2006). Nucleotides differing from the *P. jerdonii* sequence (jerdostatin) are highlighted in red and boldface. Bo, *Bombina orientalis*; Ga, *Gallus gallus*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The lack of inhibitory activity strongly suggests that neither the insertion of the RTS motif in different positions of the integrin-binding loop of an XXD short-disintegrin scaffold, nor the replacement of the whole RGD-loop of ocellatusin (KMARGDNMHDY) by the RTS-loop of jerdostatin (WRTSVSSHY) are sufficient for confering $\alpha_1\beta_1$ binding specificity. Clearly, factors other than the integrin-binding loop sequence may modulate its active conformation and/or provide additional elements involved in determining the disintegrin' selectivity and specificity for integrin $\alpha_1\beta_1$. In this respect, it is worth to mention that the integrin-binding loop and the C-terminal tail of both obtustatin (Monleón et al., 2003) and echistatin (Monleón et al., 2005) display concerted motions. Replacement of echistatin's C-terminal sequence ^{44}HKGPAT^{49} with that of another RGD-disintegrin, eristostatin (WNG) decreased but did not abolished the inhibitory potential of the mutated echistatin on ADP-induced platelet aggregation (Wierzbicka-Patynowski et al., 1999). As a whole, these data indicate that the C-terminal tail may act in synergy with the integrin-binding loop to modulate the high affinity and selectiveness of disintegrins for their target integrin receptors. Further mutants are underway in our laboratory to address the contribution of the C-terminal region for transforming a disintegrin scaffold from the RGD clade into another from the RTS/KTS clade (Fig. 1). On the other hand, another reading of our results indicates that the (K/R)TS-disintegrins must have diverged from a common precursor to the RGD-disintegrins in an unusually accelerated pace, or that they have been recruited independently of the canonical XXD disintegrins.

3.5. Evidence for a non-canonical evolution of RTS/KTS disintegrins

The origin of SVMPs has been inferred to have occurred after the split of the Pareatidae from the remaining Caenophidians, approximately 54–64 Mya during the Paleogene period of the Cenozoic Era (Fry et al., 2006, 2008; Vidal et al., 2009; Pyron and Burnbrink, 2012; Casewell, 2012). The presence of PIII-SVMPs in the venoms of Viperidae, Elapidae, Colubridae and Atractaspididae supports the view that an early recruitment event of an ancestral ADAM gene predated the radiation of the advanced snakes (Fry et al., 2009). On the other hand, metalloproteinases of class PII (which contain a disintegrin domain at the carboxyl terminus of the catalytic domain) occur only in viperids and may thus represent a derivation from ancestral PIII-SVMP genes subsequently to the split of Viperidae, an event which has been dated around the Cretaceous-Tertiary boundary, approximately 60 Mya (Wüster et al., 2008). The emergence and diversification of the disintegrin family should be limited to this time frame and PII-SVMP-derived disintegrins may be regarded as Viperidae-specific toxins.

Using jerdostatin-specific primers, highly conserved jerdostatin-like DNA sequences were amplified from both snakes, Crotalinae, Viperinae, and Elapidae taxa, and lizards, Lacertidae and Iguanidae (Fig. 5). These results, clearly showing that intronless genes coding for RTS/KTS disintegrins existed long before the split of Lacertidae and Iguania, thus predating the recruitment of the SVMP precursors of disintegrins (Fry et al., 2006), strongly

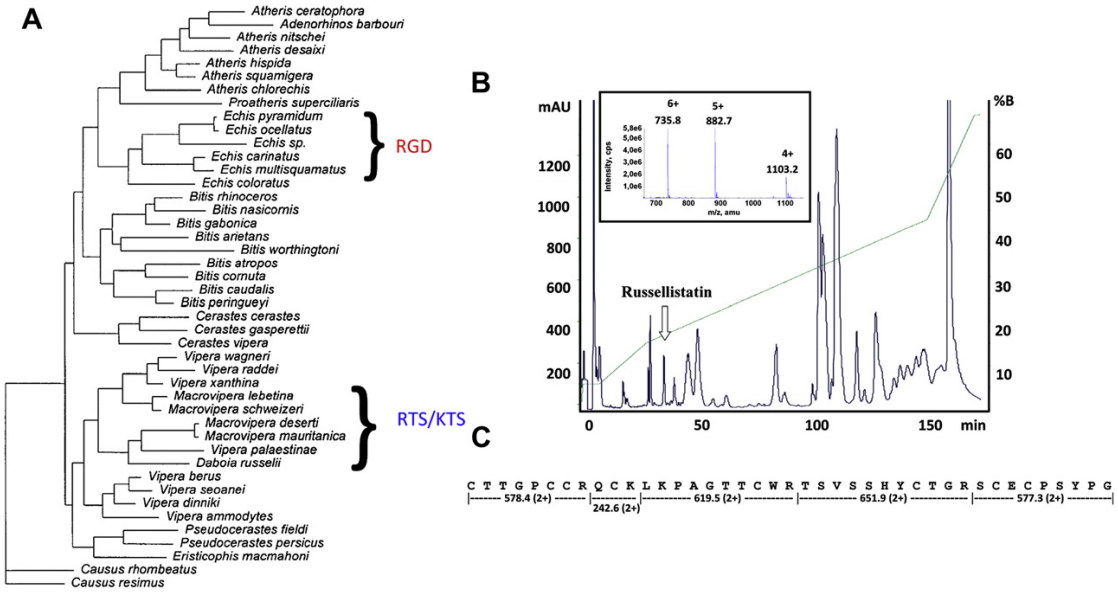


Fig. 6. Proteomic characterization of the RTS-disintegrin russellistatin. Panel A, Phylogram of the true vipers constructed using a maximum-likelihood approach with Causus serving as outgroup (Lenk et al., 2001; Pyron et al., 2011). Clades expressing RGD and RTS/KTS short disintegrins are indicated. To test the hypothesis that RTS/KTS disintegrins evolved in the clade of Eurasian vipers, the venom proteins of *D. russelli* were separated by reverse-phase HPLC (B) and the isolated proteins were characterized by electrospray-ionization mass spectrometry (insert). A peak exhibiting the expected isotope-averaged molecular mass for a short disintegrin (4408.5 Da) was reduced, carbamidomethylated, digested with trypsin, and the resulting tryptic fragments sequenced by MS/MS (C). The amino acid sequence of this protein, termed russellistatin, is identical to that of jerdostatin from *P. jerdonii* (Sanz et al., 2005).

support the view of an independent evolutionary history of the RTS/KTS and the RGD clades of short disintegrins (Fig. 1).

4. Reflexions and perspectives

Ongoing results from our laboratories indicate that jerdostatin-like genes are widely distributed across the orders Crocodilia, Testudines (turtles), and Squamata (lizards and snakes) of Reptilia. However, the jerdostatin-like sequence was not amplified in *G. gallus* (Aves) and *B. orientalis* (Amphibia) (Fig. 5A), suggesting that this gene, which we propose to term *RPTLN*, may exhibit a reptile-specific distribution. The strong conservation of *RPTLN* across reptiles (Fig. 5B) points to a relevant function in reptilian biology. However, besides for the few viperid species which express RTS/KTS disintegrins into their venoms, the question of whether the *RPTLN* gene is translated into (a body) protein (or acts as regulatory? RNA) remains unexplored. Another intriguing reflexion pertains the current evidence that, in spite of the expanded existence of the RPTLN gene, RTS/KTS disintegrins appear to be uniquely translated into the venom proteomes of the relatively recent clade of Eurasian vipers, including the genera Vipera, Macrovipera, and Daboia (Lenk et al., 2001) (Fig. 6A). KTS-disintegrins have been reported in *V. palestinae* (Kisiel et al., 2004), *M. lebetina transmediterranea* (Bazaa et al., 2005; Olfa et al., 2005), *M.l. obtusa* (Marcinkiewicz et al., 2003), and *Macrovipera mauritanica* (Makran et al., 2012). The finding of an RTS-disintegrin (hitherto termed russellistatin) in the venom of *D. russelli* (Fig. 6B and C) supports this hypothesis. However, the evolutionary pressure to express integrin $\alpha_1\beta_1$-specific inhibitors in these taxa appears to be difficult to rationalize in the context of a predator-prey arms race. The hypothesis that the molecular machinery operating on the neo-functionalization of disintegrin scaffolds evolved only in Viperinae deserves future detailed investigations.

Acknowledgements

This work has been financed by grants BFU2010-17373 (from the Ministerio de Economía y Competitividad, Madrid, Spain), PROMETEO/2010/005 (from the Generalitat Valenciana, Valencia, Spain), and grant SFB815 project A6 from Deutsche Forschungsgemeinschaft (to J.A.E.). This work is part of the PhD theses of GB (recipient of a JAE-Predoctoral fellowship) and RS-S (FPI predoctoral fellows).

Conflict of interest statement

The authors declare that there are no conflicts of interest regarding the preparation of this manuscript.

References

Bazaa, A., Marrakchi, N., El Ayeb, M., Sanz, L., Calvete, J.J., 2005. Snake venomics: comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes, Cerastes vipera* and *Macrovipera lebetina*. Proteomics 5, 4223–4235.

Bazaa, A., Juárez, P., Marrakchi, N., Bel Lasfer, Z., El Ayeb, M., Harrison, R.A., Calvete, J.J., Sanz, L., 2007. Loss of introns along the evolutionary diversification pathway of snake venom disintegrins evidenced by sequence analysis of genomic DNA from *Macrovipera lebetina transmediterranea* and *Echis ocellatus*. J. Mol. Evol. 64, 261–271.

Brown, M.C., Eble, J.A., Calvete, J.J., Marcinkiewicz, C., 2009. Structural requirements of KTS-disintegrins for inhibition of $\alpha_1\beta_1$ integrin. Biochem. J. 417, 95–101.

Calvete, J.J., 2005. Structure-function correlations of snake venom disintegrins. Curr. Pharm. Des 11, 829–835.

Calvete, J.J., Marcinkiewicz, C., Monleón, D., Esteve, V., Celda, B., Juárez, P., Sanz, L., 2005. Snake venom disintegrins: evolution of structure and function. Toxicon 45, 1063–1074.

Calvete, J.J., Marcinkiewicz, C., Sanz, L., 2007. KTS and RTS-disintegrins: anti-angiogenic viper venom peptides specifically targeting the $\alpha_1\beta_1$ integrin. Curr. Pharm. Des. 13, 2853–2859.

Calvete, J.J., Juárez, P., Sanz, L., 2009. Snake venomics and disintegrins. Portrait and evolution of a family of snake venom integrin antagonists. In: Mackessy, S.P. (Ed.), Handbook of Venoms and Toxins of Reptiles. CRC Press, Taylor & Francis, Boca Ratón, pp. 337–357 (Chapter 17).

Calvete, J.J., 2010. Brief history and molecular determinants of snake venom disintegrin evolution. In: Kini, R.M., Markland, F., McLane, M. A., Morita, T. (Eds.), Toxins and Hemostasis: From Bench to Bedside. Springer, Amsterdam, pp. 285–300 (Chapter 18).

Carbajo, R.J., Sanz, L., Mosulén, S., Pérez, A., Marcinkiewicz, C., Pineda-Lucena, A., Calvete, J.J., 2011. NMR structure and dynamics of recombinant wild type and mutated jerdostatin, a selective inhibitor of integrin $\alpha_1\beta_1$. Proteins Struct. Funct. Genet. 79, 2530–2542.

Casewell, N.R. , 2012. On the ancestral recruitment of metalloproteinases into the venom of snakes. Toxicon 60, 449–454.

Castoe, T.A., Daza, J.M., Smith, E.N., Sasa, M., Kuch, U., Campbell, J.A., Chippindale, P.T., Parkinson, C.L., 2009. Comparative phylogeography of pitvipers suggests a consensus of ancient Middle American highland biogeography. J. Biogeogr. 36, 88–103.

Eble, J.A., Golbik, R., Mann, K., Kühn, K., 1993. The $\alpha_1\beta_1$ integrin recognition site of the basement membrane collagen molecule $[\alpha1(IV)]_2$ $\alpha2(IV)$. EMBO J. 12, 4795–4802.

Eble, J.A., Kassner, A., Niland, S., Mörgelin, M., Grifka, J., Grässel, S., 2006. Collagen XVI harbors an integrin $\alpha_1\beta_1$ recognition site in its C-terminal domains. J. Biol. Chem. 281, 25745–25756.

Fry, B.G., 2005. From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res. 15, 403–420.

Fry, B.G., Vidal, N., Norman, J.A., Vonk, F.J., Scheib, H., Ramjan, S.F., Kuruppu, S., Fung, K., Hedges, S.B., Richardson, M.K., Hodgson, W.C., Ignjatovic, V., Summerhayes, R., Kochva, E., 2006. Early evolution of the venom system in lizards and snakes. Nature 439, 584–588.

Fry, B.G., Scheib, H., van der Weerd, L., Young, B., McNaughtan, J., Ramjan, S.F., Vidal, N., Poelmann, R.E., Norman, J.A., 2008. Evolution of an arsenal: structural and functional diversification of the venom system in the advanced snakes (Caenophidia). Mol. Cell. Proteomics 7, 215–246.

Fry, B.G., Vidal, N., van derWeerd, L., Kochva, E., Renjifo, C., 2009. Evolution and diversification of the Toxicofera reptile venom system. J. Proteomics 72, 127–136.

Gan, Z.-R., Gould, R.J., Jacobs, J.W., Friedman, P.A., Polokoff, M.A., 1988. Echistatin. A potent platelet aggregation inhibitor from the venom of the viper, Echis carinatus. J. Biol. Chem. 263, 19827–19832.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Juárez, P., Wagstaff, S.C., Sanz, L., Harrison, R.A., Calvete, J.J., 2006. Molecular cloning of *Echis ocellatus* disintegrins reveals non-venom-secreted proteins and a pathway for the evolution of ocellatusin. J. Mol. Evol. 63, 183–193.

Juárez, P., Comas, I., González-Candelas, F., Calvete, J.J., 2008. Evolution of snake venom disintegrins by positive Darwinian selection. Mol. Biol. Evol. 25, 2391–2407.

Juárez, P., Bolás, G., de Rezende, F.F., Calvete, J.J., Eble, J.A., 2010. Recombinant expression in human cells of active integrin $\alpha_1\beta_1$-blocking RTS-disintegrin jerdostatin. Toxicon 56, 1052–1058.

Kern, A., Eble, J.A., Golbik, R., Kühn, K., 1993. Interaction of type IV collagen with the isolated integrins $\alpha_1\beta_1$ and $\alpha_2\beta_1$. Eur. J. Biochem. 215, 151–159.

Kini, R., Evans, H.J., 1992. Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon 30, 265–293.

Kisiel, D.G., Calvete, J.J., Katzhendler, J., Fertala, A., Lazarovici, P., Marcinkiewicz, C., 2004. Structural determinants of the selectivity of KTS-disintegrins for the $\alpha_1\beta_1$ integrin. FEBS Lett. 577, 478–482.

Lenk, P., Kalyabina, S., Wink, M., Joger, U., 2001. Evolutionary relationships among the true vipers (Reptilia: Viperidae) inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol. 19, 94–104.

Makran, B., Fahmi, L., Pla, D., Sanz, L., Oukkache, N., Lkhider, M., Ghalim, N., Calvete, J.J., 2012. Snake venomics of *Macrovipera mauritanica* from Morocco, and assessment of the para-specific immunoreactivity of an experimental monospecific and a commercial antivenoms. J. Proteomics.

Marcinkiewicz, C., Vijay-Kumar, S., McLane, M.A., Niewiarowski, S., 1997. Significance of RGD loop and C-terminal domain of echistatin for recognition of $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$ integrins and expression of ligand-induced binding site. Blood 90, 1565–1575.

Marcinkiewicz, C., Weinreb, P.H., Calvete, J.J., Kisiel, D.G., Mousa, S.A., Tuszynski, G.P., Lobb, R.R., 2003. Obtustatin: a potent selective inhibitor of $\alpha_1\beta_1$ integrin in *vitro* and angiogenesis i*n vivo*. Cancer Res. 63, 2020–2023.

McLane, M.A., Vijay-Kumar, S., Marcinkiewicz, C., Calvete, J.J., Niewiarowski, S., 1996. Importance of the structure of the RGD-containing loop in the disintegrins echistatin and eristostatin for recognition of $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$ integrins. FEBS Lett. 391, 139–143.

Monleón, D., Moreno-Murciano, M.P., Kovacs, H., Marcinkiewicz, C., Calvete, J.J., Celda, B., 2003. Concerted motions of the integrin-binding loop and the C-terminal tail of the non-RGD disintegrin obtustatin. J. Biol. Chem. 278, 45570–45576.

Monleón, D., Esteve, V., Kovacs, H., Calvete, J.J., Celda, B., 2005. Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. Biochem. J. 387, 57–66.

Moreno-Murciano, M.P., Monleón, D., Marcinkiewicz, C., Calvete, J.J., Celda, B., 2003. NMR solution structure of the non-RGD disintegrin obtustatin. J. Mol. Biol. 329, 135–145.

Moura-Da-Silva, A.M., Theakston, R.D.G., Crampton, J.M., 1996. Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: gene duplication and divergence of a common ancestor rather than convergent evolution. J. Mol. Evol. 43, 263–269.

Navdaev, A., Lochnit, G., Eble, J.A., 2011. The rhodocetin αβ subunit targets GPIb and inhibits von Willebrand factor induced platelet activation. Toxicon 57, 1041–1048.

Ogawa, T., Chijiwa, T., Oda-Ueda, N., Ohno, M., 2005. Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom. Toxicon 45, 1–14.

Okuda, D., Koike, H., Morita, T., 2002. A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. Biochemistry 41, 14248–14254.

Olfa, K.-Z., José, L., Salma, D., Amine, B., Najet, S.A., Nicolas, A., Maxime, L., Raoudha, Z., Kamel, M., Jacques, M., Jean-Marc, S., Mohamed, E.A., Naziha, M., 2005. Lebestatin, a disintegrin from Macrovipera venom, inhibits integrin-mediated cell adhesion, migration and angiogenesis. Lab. Invest. 85, 1507–1516.

Pyron, R.A., Burbrink, F.T., Colli, G.R., Montes de Oca, A.N., Vitt, L.J., Kuczynski, C.A., Wiens, J.J., 2011. The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. Mol. Phylogenet. Evol. 58, 329–342.

Pyron, R.A., Burnbrink, F.T., 2012. Extinction, ecological opportunity, and the origins of global snake diversity. Evolution 66, 163–178.

Sanz, L., Chen, R.-Q., Pérez, A., Hilario, R., Juárez, P., Marcinkiewicz, C., Monleón, D., Celda, B., Xiong, Y.-L., Pérez-Payá, E., Calvete, J.J., 2005. cDNA cloning and functional expression of jerdostatin, a novel RTS-disintegrin from *Trimeresurus jerdonii* and a specific antagonist of the $\alpha_1\beta_1$ integrin. J. Biol. Chem. 280, 40714–40722.

Sanz, L., Bazaa, A., Marrakchi, N., Pérez, A., Chenik, M., Bel Lasfer, Z., El Ayeb, M., Calvete, J.J., 2006. Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina transmediterranea* venom gland cDNA libraries. Insight into the evolution of the snake venom's integrin inhibition system. Biochem. J. 395, 385–392.

Sanz, L., Ayvazyan, N., Calvete, J.J., 2008. Snake venomics of the Armenian mountain vipers *Macrovipera lebetina obtusa* and *Vipera raddei*. J. Proteomics 71, 198–209.

Smith, K.J., Jaseja, M., Lu, X., Williams, J.A., Hyde, E.I., Trayer, I.P., 1996. Three-dimensional structure of the RGD-containing snake toxin albolabrin in solution, based on ^{1}H NMR spectroscopy and simulated annealing calculations. Int. J. Pept. Protein Res. 48, 220–228.

Vidal, N., Rage, J.-C., Couloux, A., Hedges, S.B., 2009. Snakes (Serpentes). In: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford Univ. Press, pp. 390–397.

Wierzbicka-Patynowski, I., Niewiarowski, S., Marcinkiewicz, C., Calvete, J. J., Marcinkiewicz, M.M., McLane, M.A., 1999. Structural requirements of echistatin for the recognition of $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins. J. Biol. Chem. 274, 37809–37814.

Wüster, W., Peppin, L., Pook, C.E., Walker, D.E., 2008. A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). Mol. Phylogenet. Evol. 49, 445–459.

Xiong, J.P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L., Arnaout, M.A., 2002. Crystal structure of the extracellular segment of integrin alphaVbeta3 in complex with an Arg–Gly–Asp ligand. Science 296, 151–155.

Publication II

Accepted 13th April 2016 (in press)

in Integrative and Comparative Biology

(OXFORD University press)

DISTRIBUTION OF *RPTLN* GENES ACROSS REPTILIA. HYPOTHESIZED ROLE FOR *RPTLN* IN THE EVOLUTION OF SVMPs

Raquel SANZ-SOLER, Libia SANZ, Juan J. CALVETE*

Laboratorio de Venómica Estructural y Funcional, Instituto de Biomedicina de Valencia, CSIC, Jaime Roig 11, 46010 Valencia (Spain)

Running title: Distribution of *RPTLN* genes across reptiles

Keywords: RTS-disintegrin; jerdostatin; reptile-specific gene; *RPTLN* gene.

* Author to whom correspondence should be addressed: Juan J. Calvete, Laboratorio de Venómica Estructural y Funcional, Instituto de Biomedicina de Valencia, CSIC, Jaime Roig 11, 46010 Valencia (Spain). E-mail: jcalvete@ibv.csic.es; Tel.: +34 96 339 1778; Fax: +34 96 369 0800.

SYNOPSIS

We report the cloning, full-length sequencing, and broad distribution of reptile-specific *RPTLN* genes across a number of Anapsida (Testudines), Diapsida (Serpentes, Sauria), and Archosauria (Crocodylia) taxa. The remarkable structural conservation of *RPTLN* genes in species that had a common ancestor more than 250 million years ago, their low transcriptional level, and the lack of evidence for *RPTLN* translation in any reptile organ investigated, suggest for this ancient gene family a yet elusive function as long non-coding RNAs. The high conservation in extant snake venom metalloproteinases (SVMPs) of the signal peptide sequence coded for by *RPTLN* genes strongly suggests that this region may have played a key role in the recruitment and restricted expression of SVMP genes in the venom gland of Caenophidian snakes, some 60-50 Mya. More recently, 23-16 Mya, the neofunctionalization of an *RPTLN* copy in the venom gland of snakes of the genera *Macrovipera* and *Daboia* marked the beginning of the evolutionary history of a new family of disintegrins, the RTS/KTS short-disintegrins., inhibitors of $\alpha_1\beta_1$ integrin collagen-binding. This evolutionary scenario predicts that venom gland *RPTLN* and SVMP genes may share tissue-specific regulatory elements. Future genomic studies should support or refuted this hypothesis.

INTRODUCTION

A broad spectrum of β_1 and β_3 integrin receptor antagonists has evolved in venoms of Viperidae and Elapidae snake species. These include proteins of different scaffolds and receptor selectivity: C-type lectin-like molecules, such as EMS16 (Marcinkiewicz et al. 2000), rhodocetin (Eble and Tuckwell 2003), and VP12 (Staniszewska et al. 2009; Momic et al. 2011), selectively inhibit the adhesive functions of the collagen-binding $\alpha_2\beta_1$ integrin (Arlinghaus and Eble 2012); mambin (McDowell et al. 1992; Sutcliffe et al. 1994), dendroaspin (Williams et al. 1993), angustatin and H-toxin TA_2 (Oyama and Takahashi 2015), are short-chain three-finger toxin (3FTx) homologues isolated from venoms of *Dendroaspis* species that potently and specifically target the $\alpha_{IIb}\beta_3$ integrin and inhibit platelet aggregation; and disintegrins (Gould et al. 1990).

Disintegrins are a broad group of small, cysteine-rich polypeptides (Calvete et al. 2009; Calvete 2013) synthesized from short-coding mRNAs (Okuda et al. 2002) or released into the venom of Viperinae (vipers) and Crotalinae (pitvipers) snakes by the proteolytic processing of PII-SVMP precursors (Kini and Evans 1992). According to their polypeptide length and number and pattern of disulfide bonds, the disintegrin family comprise four subfamilies, long-chain (~84-residue cross-linked by 7 intramolecular disulfide linkages), medium-sized (~70 amino acids and 6 intramolecular cystine bonds), homo- and heterodimers of subunits of about 67 residues with 10 cysteines involved in the formation of 4 intra-chain disulfides and 2 inter-chain cystine linkages, and short disintegrins composed of 41-51 residues crosslinked by 4 disulfide bonds (Juárez et al. 2008). This structural diversification, from the ancestral long disintegrins to the more recently evolved short disintegrins, evolved subsequently to the emergence of Viperidae as a distinct taxonomical group of advanced snakes ~37 million years ago, in the Eocene epoch of the Cenozoic era, and involved reduction of the size of the disintegrin fold, including the stepwise loss of pairs of cysteine linkages and processing of the N-terminal region (Carbajo et al. 2015). Concomitant to the structural divergence of the disintegrin scaffold, a restricted panel of β_1 and β_3 integrin inhibitory motifs has emerged via positive Darwinian evolution (Calvete 2010; Carbajo et al. 2015). Most single-chain disintegrins express, at the apex of a mobile 11-residue loop protruding 14-17 Å from the protein core, the basal RGD sequence, which represents the motif that blocks the adhesive function of integrins $\alpha_8\beta_1$, $\alpha_5\beta_1$, $\alpha_v\beta_1$, $\alpha_v\beta_3$, and $\alpha_{IIb}\beta_3$. A few medium-sized disintegrins bear a KGD sequence that inhibits the $\alpha_{IIb}\beta_3$ integrin with a high degree of selectivity. Dimeric disintegrins exhibit the largest variability in their integrin recognition motifs, including, in addition to RGD and KGD, MLD, which targets the $\alpha_4\beta_1$, $\alpha_4\beta_7$, $\alpha_3\beta_1$, $\alpha_6\beta_1$, $\alpha_7\beta_1$ and $\alpha_9\beta_1$ integrins; VGD and MGD, which impair the function of the $\alpha_5\beta_1$ integrin; and WGD, a potent inhibitor of the RGD-dependent integrins $\alpha_5\beta_1$, $\alpha_v\beta_3$, and $\alpha_{IIb}\beta_3$ (Calvete 2009; 2010).

Short RTS/KTS disintegrins selectively hit the collagen I and IV binding $\alpha_1\beta_1$ integrin (Calvete et al. 2007; Brown et al. 2009; Walsh and Marcinkiewicz 2011), and form a distinct clade of recently emerged short disintegrins in viper venoms within genera *Macrovipera* (Marcinkiewicz et al. 2003), and *Daboia* (Kisiel et al. 2004; Olfa et al. 2005; Sanz-Soler et al. 2012). Strikingly, a non-protein-translated mRNA sequence encoding the full-length RTS-disintegrin jerdostatin was originally amplified from a *Protobothrops jerdonii* venom gland cDNA library [AY262730] (Sanz et al. 2005). Subsequently, identical messages have been cloned from a number of taxa across Serpentes, including Crotalinae, Viperinae, and Elapidae (Sanz et al. 2006; Bazaa et al. 2007). Genomic DNA fragment encoding jerdostatin-like sequences have been amplified in lizards (Lacertidae and Iguanidae) (Sanz-Soler et al. 2012). However, jerdostatin-like sequences could not be amplified in Aves (*G. gallus*), Amphibia (*B. orientalis*), and Mammalia (*M. musculus, H. sapiens*) suggesting that this intronless gene, for which the term *RPTLN* (RePTiLiN) has been proposed (Sanz-Soler et al. 2012), may exhibit a reptile-restricted

distribution. Here, we provide further evidence for the broad distribution of the *RPTLN* genes across Reptilia, and report its uneven transcriptional profile in adult lizard and colubrid organs. A role for *RPTLN* in the evolution of SVMPs is hypothesized.

MATERIALS AND METHODS

Scheme 1 summarizes and introduces the various methodological approaches (described in detail below) employed in this study to address specific questions about the evolution and possible function(s) of the *RPTLN* genes.

Materials

Blood samples from king cobra (*Ophiophagus hannah*), boa constrictor (*Boa constrictor*), ocellated lizard *(Timon lepidus lepidus*), beaded lizard (*Heloderma horridum exasperatum*), spearpoint leaf-tail gecko (*Uroplatus ebenaui*), veiled chameleon (*Chamaeleo calypratus*), Greek tortoise (*Testudo graeca graeca*), Hermann's tortoise (*Testudo hermanni*), leopard tortoise (*Stigmochelys pardalis*), Annam leaf turtle (*Mauremis annamensis*), Chinese stripe-necked turtle (*Mauremis sintesis*), red-footed tortoise (*Chelonoidis carbonaria*), Chaco tortoise (*Chelonoidis chilensis*), and liver tissue from American alligator (*Alligator missisipiensis*), were kindly provided by José María López (Sociedad Herpetológica Valenciana). Tail tissue and/or organs were sampled from Spanish lizard (*Lacerta hispanica*), Iberian wall lizard (*Podarcis hispanica*), ladder snake (*Rhinechis scalaris*), and Moorish gecko (*Tarentola mauritanica*), which were collected in the metropolitan area of Valencia. Tissues from Egyptian cobra (*Naja haje haje*) were kindly donated by César Olmos (Entomozoo Cullera, Valencia). Genomic DNA from ocellated carpet viper (*Echis ocellatus*) was kindly provided by Dr. Robert A. Harrison (Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, UK). Genomic DNA from the terciopelo pitviper (*Bothrops asper*), side-striped palm-pitviper (*Bothriechis lateralis*), and Picado's jumping pitviper (*Atropoides picadoi*) were generous donated by Dr. Mahmood Sasa (Instituto Clodomiro Picado, University of Costa Rica). Prof. Enrique Font (Laboratorio de Etología, Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia) provided common wall lizard (*Podarcis muralis*) individuals. Genomic DNA from domestic mouse (*Mus musculus*) was obtained from laboratory animals. Anonymous laboratories provided mouse-ear cress (*Arabidopsis thaliana*) and human (*Homo sapiens sapiens*) genomic DNA samples. DNA from *Bombina orientalis* was obtained from a specimen purchased in a local pet store. Blood samples from chicken (*Gallus gallus*), red-legged partridge (*Aletoris rufa*), and domestic duck (*Anas platyrhynchos*) were kindly provided by local veterinarian Dr. Carlos Nuñez (Valencia).

Recombinant jerdostatin and anti-rjerdostatin antibodies

Recombinant jerdostatin (r-jerdostatin) was produced in *E. coli* BL21 strain cells (Novagen, Madison, WI) as previously described in detail (Sanz-Soler et al. 2012). Antiserum PEP160 against the C-terminal tail of jerdostatin (^{35}CKPSYPGNG^{43}) was generated in rabbit by Abintek Biopharma, S.L. (Parque Tecnológico de Bizkaia, Derio, Bizkaia, Spain) using a standard immunization protocol and the synthetic peptide CKPSYPGNG conjugated to keyhole limpet hemocyanin (KLH) as immunogen. Anti-CKPSYPGNG antibodies were affinity-purified on a peptide-Sepharose column.

Isolation of genomic DNA

Blood cells and tissues were incubated overnight at 55˚C in lysis buffer (100 mM Tris, 25 mM EDTA, 100 mM NaCl, 0.5% SDS) and 0.2 µg/µL proteinase K (Sigma-Aldrich®). Genomic DNA (gDNA) was isolated using phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) extraction and precipitated by adding 1/10 vol of 3M sodium acetate (pH 5.2) and 2 vol of 100% ethanol. gDNA was resuspended in TE buffer (10 mM Tris, pH 8, 1 mM EDTA) (modified from Longmire et al. 1997).

Tissue preparation, RNA isolation, and reverse transcription

Organs from *Rhinechis scalaris* (lung, heart, skeletal muscle, skin), *Podarcis muralis* (bladder, liver, lung, kidney, skeletal muscle, skin, stomach, heart), and *Podarcis hispanica* (liver, lung, skin, stomach, heart, brain) were dissected, minced manually, and stored in RNA*later®* (Sigma-Aldrich). Total RNA extraction was performed using the TRIzol method following the manufacturer's (Life Technologies, NY, USA) recommended protocol. Total RNA was treated with RNAase-free DNAaseI following the manufacturer's (Thermo Scientific) protocol. DNAseI-treated total RNA integrity was assessed by electrophoresis in a 2% agarose gel. One µg of the RNA was reverse-transcribed to first strand cDNA, using oltigo$(dT)_{18}$ and the RevertAid H Minus First Strand coding DNA (cDNA) Synthesis Kit (Thermo Scientific). The cDNA was stored at -80˚ C until used. Reverse transcriptase minus (RT-) negative control was performed to verify the absence of gDNA in the RNA sample. The RT- control contained the same reaction mixture used for reverse transcription except for the RevertAid reverse transcriptase, which was substituted for 1 µl of RNAse-free water. Non-template control (NTC) was also included to discard reagent contaminations. The NTC reaction contained the same reagent than the reverse transcription reaction, excepting the RNA template.

PCR-amplification of DNA and RNA sequences

RPTLN sequences were PCR-amplified in a final volume of 25 µL containing 0.02 units of iProof™ High-Fidelity DNA Polymerase (Bio-Rad), 1.5 mM $MgCl_2$, 0.2 µM of each forward (SP_jerdostatin) 5´-ATGATCCAGGTTCTCTTGGTAACTATATG-3´ [MIQVLLVTI] and reverse (3´jerdostatin) 5´-TAGCCATTCCCGGGATAACTGG-3´ [PSYPGNG] primers (Fig. 1), and 100 ng of gDNA, or 1 µL of cDNA, RT- or NTC, as template. PCR protocol included denaturation at 98˚C for 2 min, followed by 35 cycles of denaturation (10 s at 98˚C), annealing (20 s at 58˚C), extension (40 s at 72˚C), and a final extension step for 5 min at 72˚C. One µL of Mili-Q® water, without template, was used as negative control in every PCR-amplification. The PCR reaction mixture was run in 1% agarose gel, and the candidate 333 bp DNA band was excised, purified using the GeneClean®Turbo Kit (MP BioMedicals, LLC) or Illustra GFX Gel Band Purification Kit (GE Healthcare, Buckinghamshire, UK), and cloned into a pGEM-T (Promega, Madison, WI, USA), or a pJET1.2/blunt vector (Thermo Scientific). *E. coli* DH5α cells (Novagen, Madison, WI, USA) were transformed by electroporation using an Eppendorf 2510 electroporator following the manufacturer´s instructions. Positive clones, selected by growing the transformed cells in Luria-Broth (LB) medium containing 100 µg/ml ampicillin, were confirmed by PCR amplification using the vector-specific primers, and the PCR-amplified positive clones were sequenced (using an Applied Biosystems model 377 DNA sequencer).

PCR-amplification of the housekeeping gene 28S ribosomal RNA (rRNA)

Amplification of a partial sequence of the 28S rRNA gene was used as an internal control in every *RPTLN* gene expression study. The forward (5´-GTAACGCAGGTGTCCTAAGG-3´) and reverse (5´- CGCTTGGTGAATTCTGCTTC-3´) rRNA28S primers were designed based on the partial

sequence of *Anolis carolinensis* 28S ribosomal RNA gene [AY859623]. The homologous 275 bp sequences of *Podarcis muralis* [KU556683], *Podarcis hispanica* and *Rhinechis scalaris* were amplified using the FirePol® DNA Polymerase (Solis BioDyne) protocol. PCR-amplification was performed using an initial denaturation step at 94˚C for 5 min, followed by 35 cycles of denaturation (20 s at 94˚C), annealing (20 s at 58˚C), extension (30s at 72˚C), and a final extension step for 5 min at 72˚C.

PCR-amplification of intron 7 of fibrinogen β-chain

PCR-amplification of a 296 bp fragment of intron 7 of fibrinogen β-chain was performed as a double check to confirm the absence of amplicons arising from contaminating gDNA. *Podarcis muralis* and *Podarcis hispanicus* cDNA were used as templates in the FirePol® (Solis BioDyne) DNA Polymerase protocol, with Intron7 FGB_Podarcis 5´-GGATCATGCTGTCAGGCTGG-3´ and Intron7 FGB_Podarcis 5´-CAGTGGTACCTTGGGTTAAGAAC-3´ as forward and reverse primers, respectively. These primers were designed from the sequence of intron 7 of the *Podarcis muralis* haplotype B80 β-fibrinogen (FGB) gene [EU269550]. PCR-amplification was performed using an initial denaturation step (94˚C for 5 min), followed by 35 cycles of denaturation (20 s at 94˚C), annealing (20 s at 60˚C), extension (30 s at 72˚C), and a final extension step for 5 min at 72˚C. *Rhinechis scalaris* cDNA was also subject to PCR-amplification of a 1519 bp sequence of intron 7 of fibrinogen β-chain [KU556682], using forward (7IFGB 5´-AGAGACAATGATGGATGGTAAG-3´) and reverse (7IFGB 5´-CCTTTTGGGATCTGGGTGTA-3´) primers, designed based on the β-fibrinogen intron 7 sequence of *Trimeresurus sp*. species [AF517209]. FirePol® DNA Polymerase (Solis BioDyne) protocol consisted of an initial denaturation cycle at 94˚C for 5 min, followed by 35 cycles of denaturation (20 s at 94˚C), annealing (20 s at 56˚C), extension (90 s at 72˚C), and a final extension step for 5 min at 72˚C.

Semiquantitative PCR and Real-time Quantitative RT-PCR

Semiquantitative PCR was performed by electrophoresis in 1% agarose gel following the conventional PCR-amplifications of the *RPTLN* gene and the 28S rRNA fragment for 25, 30, and 35 cycles. Real-time quantitative RT-PCR was performed in duplicate using Light Cycler FastStart DNA Master SYBR green I (Roche) in a Light Cycler 480 (Roche), following the manufacture's protocol using SP_jerdostatin and 3´jerdostatin primers for *RPTLN*-like amplification and forward and reverse rRNA28S primers for the housekeeping fragment. 1 µg of total RNA from Podarcis muralis heart (in a final volume of 20 µl) was transcribed into cDNA, and 3µL (0.15 µg of RNA) of the reaction mixture were used as template. pMD18-T/*RPTLN* and pJET1.2/28SrRNA plasmids were used as positive controls and internal standard, respectively.

Sequence analysis

Sequence similarity searches were done using BLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple sequence alignment was performed using MEGA (Molecular Evolutionary Genetic Analysis; http://www.megasoftware.net).

Prediction of RNA secondary structure was performed using the RNAfold WebServer (http://rna.tbi.univie.ac.at) at the Institute for Theoretical Chemistry, University of Vienna (Mathews et al. 2004; Gruber et al. 2008; Lorenz et al. 2011).

Protein extraction and Western blotting

For protein extraction, 5 mm-thick portions of minced tissues or organs where homogenized in lysis buffer (10 mM Tris, pH 7.5, 1 mM EDTA, 1 mM $MgCl_2$, 10% glicerol, 5 mM β-mercaptoetanol, containing a tablet of EDTA-free protease inhibitor (Roche) per 50mL of lysis buffer and 0.4 mM Pefabloc SC (AEBSF) (Roche)), using an Ultra-Turrax® (Ika® Werke) homogenizator. SDS was added to a final concentration of 2% (w/v), the samples were vortexed for ~15 seconds, incubated for 10 min on ice. This process was repeated twice, the samples were centrifuged at 14,000xg for 20 min at 4˚C, and the supernatants were transferred to clean tubes. Aliquots of 40-60 μg of total proteins extracted from 200 μg of organ/tissue homogenates of *R. scalaris* (liver, skeletal muscle, kidney, heart, lung, skin), *P. muralis* (bladder, liver, skeletal muscle, stomach, skeletal muscle, kidney, lung, heart), and *P. hispanicus* (skin, stomach, skeletal muscle, brain, lung, heart) and 50 ng of r-jerdostatin were analyzed in a 10% Tris-tricine SDS-PAGE gel under reducing conditions. Replicate gels were i) stained with Coomasiee Blue R250 and ii) electrotrasferred to PVDF membrane (Hybond-P, GE Healthcare) using a semi-dry electrotransfer device. PVDF membranes were blocked in 5% (w/v) non-fat dried milk in PBS (20 mM Na_2HPO_4, 150 mM NaCl, pH 7.5) overnight at 4˚C, and incubated for 1 h at 25˚C with a 1:500 (v/v) dilution of anti-PEP160 polyclonal antibodies in 5% non-fat dried milk/PBS-Tween (20 mM Na_2HPO_4 pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20). The membranes were then washed three times with PBS-Tween followed by incubation (1 h at 25˚C) with a 1:5000 (v/v) dilution of peroxidase-conjugated anti-rabbit IgG (Sigma) in PBS/5% non-fat dried milk. After 3 times washing with PBS-Tween, the membranes were developed using the chemiluminescence ECL Prime kit (GE Healthcare).

Another set of total protein extracts were fractionated into ≥ 10 kDa, 10-3 kDa, and ≤ 3 kDa fractions using Microcon® YM-10K (Amicon Bioseparations, Millipore) and Amicon Ultracel®-3K (Merck Millipore Ltd.) centrifugal filters. These fractions were run in a 10%Tris-tricine–SDS-PAGE gel, electroblotted onto PVDF membrane, and the blots developed as described above.

Accession codes

All nucleotide sequences gathered in this work have been deposited with the NCBI database under accession codes KU556682 (partial 1519 bp sequence of intron 7 of fibrinogen β-chain of *P. scalaris*), KU556683 (partial 275 bp sequence of 28S ribosomal RNA from *Podarcis muralis*) and KU563546-KU563619 (*RPTLN* 2-21 sequences from different organisms listed in Table 1 and 3.

RESULTS AND DISCUSSION

Intronless RPTLN genes represent a broad and reptile-restricted multigene family

Genomic DNA encoding full-length jerdostatin (*RPTLN*-1) (Fig.1) and full-length jerdostatin-like sequences (*RPTLN*-n) were amplified from a number of Anapsida (Testudines), Diapsida (Serpentes, Sauria), and Archosauria (Crocodylia) reptiles (Fig. 2), but attempts to amplify this gene in amphibians (*Bombina orientalis*), aves (*Gallus gallus*), and mammals (*Mus musculus* and *Homo sapiens*) were all unsuccessful. These striking results suggest that *RPTLN* genes exhibit a broad, reptile-specific distribution. Also remarkable is the structural conservation of these genes (Fig. 3) in taxa that had a common ancestor ≥ 250 million years ago (Hedges and Poling 1999; Vidal and Hedges 1999; Pyron et al. 2013) (Fig. 4, Table 1). In particular, the nucleotide stretch 1-54 shares >96% identity with nucleotide sequences encoding the signal peptide of snake venom metalloproteinase (SVMP) and short-

coding RGD-disintegrin precursors from a large number of Viperinae and Crotalinae snake species; *RPTLN* nucleotides 55-201 show 86-93% identity with pro-peptide-encoding nucleotide sequences for *Macrovipera lebetina* [AY835996, X97894] and *Daboia r. russellii* [GQ420354] PIII- and PII-SVMPs; and the nucleotide sequence 202-333 only matches homologous sequences from *M. lebetina* [AM114015, AM261813], *C. vipera* [AM114012], and *P. jerdonii* [AY262730] encoding RTS- and KTS-disintegrin domains, which exhibit a high degree of identity (94-100%) in any pairwise comparison. The unusual high conservation of *RPTLN* genes across Reptilia suggests a relevant function in reptile biology for this ancient gene family. Whether *RPTLN* genes are i) translated into (body) protein(s), ii) acts as regulatory RNA molecules, or iii) serve other unknown function(s), remains elusive. Of relevance to this point, among the 31 nucleotide changes identified in the 22 *RPTLN* genes listed in Table 2, 6 involve the third base of *RPTLN* codons, whereas 8 and 17 affect first and second codon positions, respectively (Fig.3). In protein-coding DNA sequences, the second-codon position is the most functionally constrained, whereas, due to the degenerate nature of the genetic code, the third-codon position is the least functionally constrained in terms of nucleotide changes (Bofkin and Goldman 2007). It is thus tempting to hypothesize that the biological role of transcribed *RPTLN* RNA may be strongly dependent on their folded structure. Supporting this hypothesis, the RNAfold WebServer (Gruber et al. 2008) predicted for the full-length *RPTLN*-transcribed 333 bp RNA a stable (-90.59 kcal/mol minimum free energy) secondary structure ensemble (Fig. 5).

RPTLN genes were found at nodes predating the separation of Toxicofera and Lacertidae (Fig. 4). Toxicofera, from the Greek "those who bear toxins", is a clade of scaled reptiles that includes the snakes, Anguimorpha (monitor lizards, gila monster, and alligator lizards) and Iguania (iguanas, agamas, and chameleons), and thus *RPTLN* gene(s) existed long before venom arose in squamate evolution, approximately 170 Mya during the Jurassic period (Fry et al. 2006; 2012). Despite their broad distribution across the phylogeny of Reptilia (Fig. 4), *RPTLN* genes have been found translated into KTS/RTS disintegrins only in the venoms of *M. l. obtusa* (obtustatin (KTS) [P83469; 1MPZ], Marcinkiewicz et al. 2003; Moreno-Murciano et al. 2003; Monleón et al. 2003; Sanz et al. 2008), *M. mauritanica* (lebestatin (KTS) [CAJ34939], Olfa et al. 2005; Sanz et al. 2006; Makran et al. 2012), *D. palestinae* (viperistatin (KTS) [P0C6E2], Kisiel et al. 2004), and *D. russellii* (russellistatin (RTS), Sanz-Soler et al. 2012). Evolutionary relationships reconstruction inferred from mitochondrial DNA sequences dated the emergence of Eurasian viper (genera *Eristicophis*, *Pseudocerastes*, *Vipera*, *Macrovipera*, and *Daboia*) in the early Miocene (23-16 million years ago, Mya) (Lenk et al. 2001), coinciding with the geographical separation of the landmasses Europe, Middle East, and North Africa by the Mediterranean and Parathethys seas (Rögl and Steininger 1983). This evidence indicates that *RPTLN* genes comprise an ancient multigenic family, and that their restricted expression and neofunctionalization in the venom gland of *Macrovipera* and *Daboia* species represent recent events (Fig. 4).

A hallmark of the *RPTLN* genes transcribed into protein (KTS- and RTS-disintegrin)-coding mRNAs with respect to those that are not translated is the accumulation of mutations in the C-terminal half of the disintegrin domain (Sanz-Soler et al. 2012), which constitutes a conformational functional epitope encompassing the $\alpha_1\beta_1$ integrin-inhibitory loop and the C-terminal tail of KTS/RTS disintegrins (Kisiel et al. 2004). However, the mechanism underlying this non-protein-coding to protein-coding transition remains elusive. It is tempting to speculate that accumulation of nucleotide changes in certain *RPTLN* gene copies may have resulted in destabilization of the transcribed non-coding RNA (ncRNAs) into a translatable, or in the formation of pseudogenes. In this context, it is worth mentioning that ncRNAs transcribed from pseudogenes may play regulatory roles regulating the expression of their parental or non-parental genes

A hypothesized role for RPTLN in the evolution of SVMPs

The nucleotide sequence encoding the signal peptide (SP) of *RPTLN* genes is highly conserved in SVMP precursor genes of Viperinae, Crotalinae, Elapidae, and Colubridae snake species (Fig. 6). On the contrary, this region is not conserved in ADAM (A Disintegrin And Metalloprotease) genes (Fig. 6). The closest non-venom ancestor of SVMPs was likely an ADAM28 precursor gene (Casewell 2012) that was recruited into the snake venom gland proteome (Moura da Silva et al. 1996) after the divergence of squamate reptiles, lizards and snakes (Fry et al. 2006; 2012) in the Jurassic, ~170-150 million years before present (MYBP) (Hedges and Vidal 2009). The high conservation in extant SVMPs of the SP sequence coded for by *RPTLN* genes strongly suggests that this region may have played a key role in the recruitment and restricted expression of SVMP genes in the venom gland of Caenophidian snakes. In this respect, the exon-intron organization of pre-pro *E. ocellatus* EOC00089-like PIII-SVMP and *A. carolinensis* ADAM28 genes is conserved, and their 17-residue signal peptides are entirely coded for by exon 1 (Sanz et al. 2012). Most introns of *A. carolinensis* ADAM 28 contain inserted retroelements capable of invading new genomic sites (Alföldi et al. 2011), particularly short interspersed retrotransposable elements (Sauria SINE) in introns 1, 3-8, 10, 13, 14 and 16, and LINEs (long interspersed elements) in introns 1 and 10. The family of Sauria SINEs are widely distributed among genomes of lizards, snakes, and tuataras (Piskurek et al. 2006; 2009). Sauria SINEs arose more than 200 million years ago, and the members of this family comprise a 5' tRNA-related region, a tRNA-unrelated region, and a 3' tail region identical with Bov-B LINEs (Piskurek et al. 2006). Their retrotransposition depends on reverse transcriptase and endonuclease activities encoded by partner LINEs, and it has been proposed (Piskurek et al. 2006) that Sauria SINEs utilize the enzymatic machinery of Bov-B LINEs for their retrotransposition.

Transcription and translation of RPTLN genes in P. muralis, P. hispanica (Lacertidae), and R. scalaris (Colubridae) organs

Full-length *RPTLN* transcripts were amplified by PCR in different organs and tissues of *P. muralis*, *P. hispanica*, and *R. scalaris* (Table 3). Semiquantitative PCR amplification suggested differential transcription levels at different organs of *P. muralis*. However, attempts to quantify the levels of *RPTLN* RNA by quantitative PCR amplification failed due to the low amount of transcripts. Low expression levels may arise from alternative promoter usage of ncRNAs compared to protein-coding RNAs (The Fantom Consortium et al. 2005). On the other hand, we did not find evidence for *RPTLN* translation in any organ investigated by Western blot analysis of 40-60 μg of total proteins extracted from 200 μg organ homogenates (Fig. 7) or size-fractionated by ultrafiltration (≥ 10 kDa, 10-3 kDa, and ≤ 3 kDa fractions). Recombinant jerdostatin (Sanz et al. 2005) was used to estimate the immunodetection limit, which was ≤ 50 ng (Fig. 7, lanes Jer). For comparison, the short KTS-disintegrin obtustatin [P83469] (Sanz et al. 2008), and its homologs, the RTS-disintegrin russellistatin (Sanz-Soler et al. 2012) and the KTS-disintegrin lebestatin [Q3BK14] (Makran et al. 2012), comprise 2.8%, 2%, and 7.8% of the total venom proteins of *Macrovipera lebetina obtusa*, *Daboia russelii*, and *Macrovipera* (*Daboia*) *mauritanica*, respectively. Furthermore, expression yields for functionally active recombinant jerdostatin (wild-type and mutants) in *E. coli* were about 0.5-2 mg/L of cell culture (Sanz et al. 2005; Sanz-Soler et al. 2012). These data clearly show that *RPTLN* and *RPTLN*-like DNA sequences can be transcribed and translated into functional proteins in different cellular environments. Although the possibility that *RPTLN* genes are translated into very low protein concentration can not be ruled out, all available data support the view that *RPTLN* gene copies encode a long (>200 nt) ncRNAs (lncRNAs). Eukaryote genomes include tens of thousands of long noncoding RNAs with little or no protein-coding capacity (Wilusz et al. 2009; Wilusz 2015). Only a limited number of lncRNAs have

been functionally characterized. However, paradigms for how lncRNAs exert regulatory functions are beginning to emerge (reviewed by Wilusz et al. 2009; Wilusz 2015). In particular, lncRNAs are both regulated by unique post-transcriptional control mechanisms, and control various aspects of post-transcriptional processing of mRNAs. These functions often involve the formation of ribonucleoprotein, RNA-RNA, and DNA-RNA complexes. These features of lncRNAs provide a meaning to the conservation of the third-base of *RPTLN* codons, and support our view of the possible biological function of this reptile-specific gene.

CONCLUDING REMARKS AND PERSPECTIVES

This work elaborates upon a previous work (Sanz-Soler et al. 2012) in which we showed that *RPTLN* genes coding for RTS/KTS disintegrins existed long before the split of Lacertidae and Iguania, thus predating the recruitment of the SVMP precursors of disintegrins, and providing strong support for an independent evolutionary history of the RTS/KTS and the RGD clades of short disintegrins. Now, we report full-length *RPTLN* gene sequences amplified from species at nodes predating the separation of Toxicofera and Lacertidae, and thus preceding the emergence of venom in the evolution of squamate reptiles, ~170 Mya during the Jurassic period (Fry et al. 2006; 2012). The remarkable structural conservation of *RPTLN* genes across Reptilia, their low transcriptional level, and the lack of evidence for *RPTLN* translation in any reptile organ investigated, suggest a yet elusive role for transcribed *RPTLN* RNA as a long non-protein-coding RNA (Fig. 5). We hypothesize that the high conservation of the SP sequence of *RPTLN* and extant SVMP genes may suggest a functional role for this region in the ancestral recruitment of SVMP gene expression in the venom gland of Caenophidian snakes (Fig. 8). The origin of SVMPs has been inferred to have occurred after the split of the Pareatidae from the remaining Caenophidians, ~50-60 Mya, during the Paleogene period of the Cenozoic Era (Fry et al. 2006; 2008; 2009; Vidal et al. 2009; Pyron and Burnbrink 2012; Casewell 2012). The evolutionary path that led to the family of the RGD/XXD-disintegrins from PII-SVMPs has been dissected in some detail at the molecular and structural levels (Juárez et al. 2008; Calvete 2010; Carbajo et al. 2015). Neofunctionalization of the *RPTLN* gene to express RTS/KTS disintegrins (Fig.6) represents an independent alternative route from the evolution of PII-SVMP-derived disintegrins, which occurred more recently in venoms of Eurasian vipers within genera *Macrovipera* and *Daboia* (Sanz-Soler et al. 2012), in the early Miocene (~23-16 million years ago, Mya) (Lenk et al. 2001). Figure 8 outlines a cartoon of processes in which we hypothesize that *RPTLN* may have been involved. Understanding the physiological function and evolutionary history of this enigmatic highly conserved (and thus presumably relevant) gene across the phylogeny of reptiles clearly requires further detailed molecular studies. In particular, comparative analysis of the upstream regions of non-protein coding and protein coding *RPTLN* genes may identify nucleotide motifs that contribute to regulate their distinct molecular fates. On the other hand, the hypothesis that *RPTLN* may have played a key role in the recruitment and restricted expression of SVMP genes in the venom gland of Caenophidian snakes predicts that venom gland *RPTLN* and SVMP genes may share tissue-specific regulatory elements. Future genomic studies should support or refuted our hypothesis.

ACKNOWLEDGMENTS

This study was supported by grant BFU2013-42833-P from the Ministerio de Economía y Competitividad, Madrid (Spain). This work is part of the PhD thesis of RS-S, who gratefully acknowledges the granting of a predoctoral fellowship from the FPI (Formación de Personal Investigador) program from the Ministerio de Economía y Competitividad, Madrid (Spain).

BIBLIOGRAPHY

Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD, Ray DA. et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477:587-91.

Arlinghaus FT, Eble JA. 2012. C-type lectin-like proteins from snake venoms. Toxicon 60:512-9.

Balakirev ES, Ayala FJ. 2003. Pseudogenes: are they "junk" or functional DNA? Annu Rev Genet. 37:123-51.

Bazaa A, Juárez P, Marrakchi, N, Bel Lasfer Z, El Ayeb M, Harrison RA, Calvete JJ, Sanz L. 2007. Loss of introns along the evolutionary diversification pathway of snake venom disintegrins evidenced by sequence analysis of genomic DNA from *Macrovipera lebetina transmediterranea* and *Echis ocellatus*. J Mol Evol. 64:261-71.

Bofkin L, Goldman N. 2007. Variation in evolutionary processes at different codon positions. Mol Biol Evol. 24:513-21.

Brown MC, Eble JA, Calvete JJ, Marcinkiewicz, C. 2009. Structural requirements of KTS-disintegrins for inhibition of $\alpha_1\beta_1$ integrin. Biochem J. 417:95-101.

Calvete JJ, Marcinkiewicz C, Sanz L. 2007. KTS and RTS-disintegrins: anti-angiogenic viper venom peptides specifically targeting the $\alpha_1\beta_1$ integrin. Curr Pharm Des. 13:2853-9.

Calvete JJ. 2010. Brief history and molecular determinants of snake venom disintegrin evolution. In *Toxins and Hemostasis: from Bench to Bedside* (Kini RM, Markland F, McLane MA, Morita T. eds.) pp. 285-300, Springer Verlag, Amsterdam.

Calvete JJ. 2013. The continuing saga of snake venom disintegrins. Toxicon 62:40-9.

Carbajo RJ, Sanz L, Pérez A, Calvete JJ. 2015. NMR structure of bitistatin – a missing piece in the evolutionary pathway of snake venom disintegrins. FEBS J. 282:341-60.

Casewell NR. 2012. On the ancestral recruitment of metalloproteinases into the venom of snakes. Toxicon 60:449-54.

Eble JA, Tuckwell DS. 2003. The $\alpha_2\beta1$ integrin inhibitor rhodocetin binds to the A-domain of the integrin α_2 subunit proximal to the collagen-binding site. Biochem J. 376:77-85.

Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, et al. 2006. Early evolution of the venom system in lizards and snakes. Nature 439:584-8.

Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SF, Vidal N, Poelmann RE, Norman JA. 2008. Evolution of an arsenal: structural and functional diversification of the venom system in the advanced snakes (Caenophidia). Mol Cell Proteomics 7:215-46.

Fry BG, Vidal N, van derWeerd L, Kochva E, Renjifo C. 2009. Evolution and diversification of the Toxicofera reptile venom system. J Proteomics 72:127-36.

Fry BG, Casewell NR, Wüster W, Vidal N, Young B, Jackson TN. 2012. The structural and functional diversification of the Toxicofera reptile venom system. Toxicon 60:434-48.

Gould RJ, Polokoff MA, Friedman PA, Huang TF, Holt JC, Cook JJ. Niewiarowski S. 1990. Disintegrins: a family of integrin inhibitory proteins from viper venoms. Proc Soc Exp Biol Med. 195:168-71.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. 2008. The Vienna RNA websuite. Nucleic Acids Res. 36(Web Server issue):W70-4.

Hedges SB, Poling LL. 1999. A molecular phylogeny of reptiles. Science 283:998-1001.

Hedges SB, Vidal N. 2009. Lizards snakes and amphisbaenians (Squamata). In *The Timetree of Life* (Hedges SB, Kumar S, eds.), pp. 383-9, University Press Oxford.

Juárez P, Comas I, González-Candelas F, Calvete JJ. 2008. Evolution of snake venom disintegrins by positive Darwinian selection. Mol Biol Evol. 25:2391-407.

Kini R, Evans HJ. 1992. Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon 30:265-93.

Kisiel DG, Calvete JJ, Katzhendler J, Fertala A, Lazarovici P, Marcinkiewicz C. 2004. Structural determinants of the selectivity of KTS-disintegrins for the $\alpha_1\beta_1$ integrin. FEBS Lett. 577:478-82.

Lenk, P, Kalyabina S, Wink M, Joger U. 2001. Evolutionary relationships among the true vipers (Reptilia: Viperidae) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol. 19:94-104.

Longmire JL, Maltbie M, Baker RJ. 1997. Use of "lysis buffer" in DNA isolation and its implication for museum collections. Occasional Papers of the Museum of Texas Technological University 163:1-3.

Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms for Molecular Biology 6:26.

Makran B, Fahmi L, Pla D, Sanz L. Oukkache N, Lkhider M, Ghalim N, Calvete JJ. 2012. Snake venomics of *Macrovipera mauritanica* from Morocco and assessment of the para-specific immunoreactivity of an experimental monospecific and a commercial antivenoms. J Proteomics 75:2431-41.

Marcinkiewicz C, Lobb RR, Marcinkiewicz MM, Daniel JL, Smith JB, Dangelmaier C. Weinreb PH, Beacham DA, Niewiarowski S. 2000. Isolation and characterization of EMS16 a C-lectin type protein from *Echis multisquamatus* venom a potent and selective inhibitor of the $\alpha_2\beta_1$ integrin. Biochemistry 39:9859-67.

Marcinkiewicz C, Weinreb PH, Calvete JJ, Kisiel DG, Mousa SA, Tuszynski GP, Lobb RR. 2003. Obtustatin: a potent selective inhibitor of $\alpha_1\beta_1$ integrin *in vitro* and angiogenesis *in vivo*. Cancer Res 63:2020-3.

Mathews, DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci USA 101:7287-92.

McDowell RS, Dennis MS, Louie A, Shuster M, Mulkerrin MG, Lazarus RA. 1992. Mambin a potent glycoprotein IIb-IIIa antagonist and platelet aggregation inhibitor structurally related to the short neurotoxins. Biochemistry 31:4766-72.

Mighell AJ, Smith NR, Robinson PA, Markham AF. 2000. Vertebrate pseudogenes. FEBS Lett. 468:109-14.

Momic T, Arlinghaus FT, Arien-Zakay H, Katzhendler J, Eble JA, Marcinkiewicz C, Lazarovici P. 2011. Pharmacological aspects of *Vipera xantina palestinae* venom. Toxins 3:1420-32.

Monleón D, Moreno-Murciano MP, Kovacs H, Marcinkiewicz C, Calvete JJ, Celda B. 2003. Concerted motions of the integrin-binding loop and the C-terminal tail of the non-RGD disintegrin obtustatin. J Biol Chem. 278:45570-6.

Moreno-Murciano MP, Monleón D, Marcinkiewicz C, Calvete JJ, Celda B. 2003. NMR solution structure of the non-RGD disintegrin obtustatin. J Mol Biol. 329:135-45.

Moura-da-Silva, AM, Theakston RD, Crampton JM. 1996. Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: gene duplication and divergence of a common ancestor rather than convergent evolution. J Mol Evol. 43:263-9.

Muniz JR, Ambrosio AL, Selistre-de-Araujo HS, Cominetti MR, Moura-da-Silva AM, Oliva G, Garratt RC, Souza DH. 2008. The three-dimensional structure of bothropasin, the main hemorrhagic factor from *Bothrops jararaca* venom: insights for a new classification of snake venom metalloprotease subgroups. Toxicon 52:807-16.

Muro EM, Mah N. Andrade-Navarro MA. 2011. Functional evidence of post-transcriptional regulation by pseudogenes. Biochimie 93:1916-21.

Okuda D, Koike H, Morita T. 2002. A new gene structure of the disintegrin family: a subunit of

dimeric disintegrin has a short coding region. Biochemistry 41:14248-54.

Olfa K-Z, José L, Salma D, Amine B, Najet SA, Nicolas A, Maxime L, Raoudha Z, Kamel M, Jacques M, et al. 2005. Lebestatin a disintegrin from *Macrovipera* venom inhibits integrin-mediated cell adhesion migration and angiogenesis. Lab Invest. 85:1507-16.

Oyama E, Takahashi H. 2015. Purification and characterization of two platelet-aggregation inhibitors named angustatin and H-toxin TA_2 from the venom of *Dendroaspis angusticeps*. Toxicon 93:61-7.

Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Carter DR. 2011. Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 17:792-8.

Piskurek O, Austin CC, Okada N. 2006. Sauria SINEs: Novel short interspersed retroposable elements that are widespread in reptile genomes. J Mol Evol. 62:630-44.

Piskurek O, Nishihara H, Okada N. 2009. The evolution of two partner LINE/SINE families and a full-length chromodomain-containing Ty3/Gypsy LTR element in the first reptilian genome of *Anolis carolinensis*. Gene 441:111-8.

Pyron RA, Burnbrink FT. 2012. Extinction ecological opportunity and the origins of global snake diversity. Evolution 66:163-78.

Pyron RA, Burbrink FT, Wiens JJ. 2013. A phylogeny and revised classification of Squamata including 4161 species of lizards and snakes. BMC Evol Biol. 13:93.

Rögl F, Steininger FF. 1983. Vom Zerfall der Thetys zu Mediterran und Paratethys. Ann Naturhist Mus Wien. 85/A:135-63.

Sanz L, Chen R-Q, Pérez A, Hilario R, Juárez P, Marcinkiewicz C, Monleón D, Celda B, Xiong Y-L, Pérez-Payá E. et al. 2005. cDNA cloning and functional expression of jerdostatin a novel RTS-disintegrin from *Trimeresurus jerdonii* and a specific antagonist of the $\alpha_1\beta_1$ integrin. J Biol Chem. 280:40714-22.

Sanz L, Bazaa A, Marrakchi N, Pérez A, Chenik M, Bel Lasfer Z, El Ayeb M, Calvete JJ. 2006. Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina transmediterranea* venom gland cDNA libraries: insight into the evolution of the snake venom integrin-inhibition system. Biochem J. 395:385-92.

Sanz L, Aybazyan N, Calvete JJ. 2008. Snake venomics of the Armenian mountain vipers *Macrovipera lebetina obtusa* and *Vipera raddei*. J Proteomics 71:198-209.

Sanz L, Harrison RA, Calvete JJ. 2012. First draft of the genomic organization of a PIII-SVMP gene. Toxicon 60:455-69.

Sanz-Soler R, Lorente C, Company B, Sanz L, Juárez P, Pérez A, Zhang Y, Jin Y, Chen R, Eble JA, Calvete JJ, Bolás G. 2012. Recombinant expression of mutants of the *Frankenstein* disintegrin RTS-ocellatusin. Evidence for the independent origin of RGD and KTS/RTS disintegrins. Toxicon 60:665-75.

Staniszewska I, Walsh EM, Rothman VL, Gaathon A, Tuszynski GP, Calvete J.J, Lazarovici P, Marcinkiewicz C. 2009. Effect of VP12 and viperistatin on inhibition of collagen-receptor-dependent melanoma metastasis. Cancer Biol Ther 8:1507-16.

Sutcliffe MJ, Jaseja M, Hyde EI, Lu X, Williams JA. 1994. Three-dimensional structure of the RGD-containing neurotoxin homologue dendroaspin. Nat Struct Biol. 1:802-7.

The FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group 2005. The Transcriptional Landscape of the Mammalian Genome Science 309:1559-63.

Trinklein ND, Karaöz U, Wu J, Halees A, Force Aldred S, Collins PJ, Zheng D, Zhang ZD, Gerstein MB, Snyder M, et al. 2007. Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome. Genome Res. 17:720-31.

Vidal N, Hedges SB. 2009. The molecular evolutionary tree of lizards snakes and amphisbaenians. C R Biol. 332:129-39.

Vidal N, Rage J-C, Couloux A, Hedges SB. 2009. Snakes (Serpentes). In: The Timetree of Life (Hedges

SB and Kumar S, eds.), Oxford Univ Press, pp. 390-7.

Walsh, EM, Marcinkiewicz C. 2011. Non-RGD-containing snake venom disintegrins functional and structural relations. Toxicon 58:355-62.

Williams JA, Lu X, Rahman, S, Keating C, Kakkar V. 1993. Dendroaspin: a potent integrin receptor inhibitor from the venoms of *Dendroaspis viridis* and *D. jamesonii*. Biochem Soc Trans. 21:73S.

Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 23:1494-1504.

Wilusz JE. 2015. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. Biochim Biophys Acta doi: 10.1016/j.bbagrm.2015.06.003.

LEGENDS TO FIGURES

Scheme 1. Summary of the various methodological approaches (described in detail in the Materials and Methods section) employed in this study to address specific questions about the evolution and possible structure-function correlations of the reptile-specific *RPTLN* genes.

Figure 1. Nucleotide sequence of the genomic DNA (*RPTLN*-1 gene, AY262730) encoding full-length jerdostatin. The limits of the signal peptide, prodomain, and disintegrin domain are indicated. The active RTS tripeptide sequence is highlighted in bold in a grey background. The sequences of the forward (SP_jerdostatin) and reverse (3'jerdostatin) primers used for PCR amplifications are in lower case, highlighted in boldface, and labeled.

Figure 2. PCR-amplification of *RPTLN* genes from gDNA of *Echis ocellatus* (Eo), *Ophiophagus hannah* (Oh), *Boa constrictor* (Bc), *Lacerta hispanica* (Lh), *Timon lepidus* (Tl), *Podarcis muralis* (Pm), *Heloderma horridum exasperatum* (Hh), *Uroplatus ebenaui* (Ue), *Chamaeleo calyptratus* (Cc), *Alligator missisipiensis* (A), *Testudo graeca* (Tg), *Mauremys sinensis* (Ms), *Bothrops asper* (Ba), *Bothriechis lateralis* (Bl), *Atropoides picadoi* (Ap), *Naja naja haje* (Nn), *Rhinechis scalaris* (Rs), *Tarentola mauritanica* (Tm), *Testudo Hermanni* (Th), *Stigmochelys pardalis* (Sp), *Mauremys annamensis* (Ma), *Mauremys leprosa* (Ml), *Chelonoidis chilensis* (ChCh), *Chelonoidis carbonaria* (Chc), and *Mus musculus* (Mm). (-), negative control.

Figure 3. Multiple sequence alignments of *RPTLN*-1 and the twenty unique *RPTLN* genes amplified from gDNA or mRNA of the different reptile species listed in Table 1. For convenience, the alignment has been divided into three blocks of sequences, each of which codes for a domain of a hypothetical protein. Nucleotide changes between *RPTLN* sequences are highlighted in boldface and in grey background. The distribution of these sequences across the phylogeny of reptiles is shown in Figure 4.

Figure 4. Distribution of the *RPTLN* genes listed in Table 1 in the (simplified) phylogeny of reptiles. Tree positions where mutations associated with the 20 different *RPTLN* genes arose are indicated. The identity of the particular *RPTLN*-n (n = 2-21) isogenes containing the mutation(s) are specified in parentheses. Viperinae branch is expanded, and those species from which venom RTS/KTS disintegrins have been isolated, are highlighted in bold.

Figure 5. Scheme of the predicted secondary structure of full-length *RPTLN* RNA molecule. Nucleotide positions mutated in any of the *RPTLN*-2-21 genes listed in Table 2 are indicated.

Figure 6. Comparison of the nucleotide and encoded signal peptide sequences of *RPTLN*-1, SVMPs, and ADAM molecules. SVMP and ADAM residues differing from the nucleotide or translated amino acid sequence of *RPTLN*-1 are highlighted in bold and in gray background.

Figure 7. Coomassie blue-stained SDS-PAGE (upper panels) and Western blot analysis (lower panels) probed with anti-jerdostatin PEP160 polyclonal antibodies of 40-60 μg of total proteins extracted from different organs of *R. scalaris* (**A**), *P. muralis* (**B**), and *P. hispanica* (**C**). Positive control, lane Jer, ~50 ng of purified recombinant jerdostatin; Immunoreactivity is only observed against the positive control. Std, molecular mass standard Mark12™ (Invitrogen); Li, liver; M, skeletal muscle; K, kidney; H, heart; Lu, lung; Sk, skin; St, stomach; Br, brain.

Figure 8. Cartoon of processes in which the *RPTLN* gene is hypothesized to have been involved during its long evolutionary history. **A**, Fusion of the ADAM28 extracellular domains-coding gene region $_{\Delta}^{EGF/TM/Cyto}$ADAM28 and an *RPTLN* gene under a venom gland-specific promoter (VGP) generated a SVMP gene bearing the *RPTLN* signal sequence and exhibiting venom gland restricted translation (**B**). The timing of the functional processes in which *RPTLN* is hypothesized to have been involved is indicated in million years ago (Mya), i.e. a yet elusive role for transcribed *RPTLN* as a long non-protein-coding RNA since ≥ 170 Mya; the recruitment and venom gland-restricted expression of PIII-SVMPs, 50-60 Mya; and more recently (23-16 Mya), its own neofunctionalization into $\alpha_1\beta_1$-inhibitory short RTS/KTS disintegrins in venoms of certain Eurasian viper species within genera *Macrovipera* and *Daboia* (Fig. 4). Disi, disintegrin-like domain; Cys, cysteine-rich domain; EGF, epithelial growth factor-like domain; TM, transmembrane domain; cyto, cytoplasmic domain; SP, signal peptide. The structural model of PIII-SVMP has been adapted from the crystallographic structure of bothropasin (Muniz et al. 2008).

Table 1

Distribution of *RPTLN* gene copies across reptiles. Full-length sequences are displayed in Fig.3. Nucleotide changes in *RPTLN* genes respect to *RPTLN*-1 [jerdostatin, AY262730] are listed in Table 2.

Reptile class	**Family**	**Species**	***RPTLN* gene copy**
Snake	Crotalinae	*P. jerdonii*	*RPTLN-1*
		B. asper	*RPTLN*-1 (KU563546), *RPTLN*-15 (KU563589), *RPTLN*-16 (KU563592)
		B. lateralis	*RPTLN*-1 (KU563547), *RPTLN*-7 (KU563579), *RPTLN*-17 (KU563595)
	Viperinae	*E. ocellatus*	*RPTLN*-1 (KU563548), *RPTLN*-13 (KU563587)
		A. picadoi	*RPTLN*-1 (KU563549), *RPTLN*-10 (KU563583), *RPTLN*-15 (KU563590)
	Elapidae	*O. hannah*	*RPTLN*-1 (KU563550), *RPTLN*-6 (KU563577), *RPTLN*-17 (KU563596)
		N. haje haje	*RPTLN*-1 (KU563551), *RPTLN*-5 (KU563575)
	Colubridae	*R. scalaris*	*RPTLN*-1 (KU563552), *RPTLN*-5 (KU563608), *RPTLN*-18 (KU563600), *RPTLN*-19 (KU563603), *RPTLN*-20 (KU563604), *RPTLN*-21 (KU563617)
	Boideae	*B. constrictor*	*RPTLN*-1 (KU563553), *RPTLN*-17 (KU563597)
Lizard	Lacertidae	*L. hispanica*	*RPTLN*-1 (KU563554), *RPTLN*-9 (KU563581), *RPTLN*-17 (KU563598)
		T. lepidus	*RPTLN*-1 (KU563555)
		P. muralis	*RPTLN*-1 (KU563556), *RPTLN*-9 (KU563610), *RPTLN*-17 (KU563614), *RPTLN*-18 (KU563615),*RPTLN*-19 (KU563616)
		P. hispanica	*RPTLN*-1 (KU563557), *RPTLN*-8 (KU563609), *RPTLN*-11(KU563611), *RPTLN*-13 (KU563612), *RPTLN*-16 (KU563613), *RPTLN*-21 (KU563618)
	Holodermatidae	*H. horridum*	*RPTLN*-1 (KU563558)
	Gekkonidae	*U. ebenaui*	*RPTLN*-1 (KU563559), *RPTLN*-9 (KU563582)
		T. mauritanica	*RPTLN*-1 (KU563560), *RPTLN*-11 (KU563584)
	Chamaeleonidae	*C. calypratus*	*RPTLN*-1 (KU563561), *RPTLN*-17 (KU563599)
Crocodrile	Crocodylidae	*A. mississipiensis*	*RPTLN*-1 (KU563562), *RPTLN*-6 (KU563578), *RPTLN*-11 (KU563585), *RPTLN*-12 (KU563586), *RPTLN*-14 (KU563588)
Tortoise	Testudinidae	*T. greca*	*RPTLN*-1 (KU563563), *RPTLN*-3 (KU563573)
		T. hermanni	*RPTLN*-1 (KU563564)
		S. pardalis	*RPTLN*-1 (KU563565), *RPTLN*-16 (KU563593), *RPTLN*-18 (KU563601)
	Geomydidae	*M. annamensis*	*RPTLN*-1 (KU563566), *RPTLN*-15 (KU563591)
		M. sintesis	*RPTLN*-1 (KU563567), *RPTLN*-4 (KU563574), *RPTLN*-7 (KU563580), *RPTLN*-18 (KU563602)
		M. leprosa	*RPTLN*-1 (KU563568), *RPTLN*-2 (KU563571), *RPTLN*-5 (KU563576)
		C. carbonaria	*RPTLN*-1 (KU563569), *RPTLN*-2 (KU563572)
		C. chilensis	*RPTLN*-1 (KU563570), *RPTLN*-16 (KU563594)

Table 2
Nucleotide changes in *RPTLN*-n genes respect to *RPTLN*-1 [encoding jerdostatin, AY262730]

RPTLN-2	41T>C
RPTLN-3	50A>G
RPTLN-4	50A>G;173A>G;179A>G
RPTLN-5	D65C
RPTLN-6	84T>C
RPTLN-7	87T>C
RPTLN-8	128C>T
RPTLN-9	128C>T; 259T>A
RPTLN-10	145C>T
RPTLN-11	145C>T; 185T>C
RPTLN-12	145C>T; 185T>C; 220T>C
RPTLN-13	155A>G
RPTLN-14	173A>G
RPTLN-15	173A>G; 179A>G
RPTLN-16	173A>G; 179A>G; 255A>G
RPTLN-17	220T>C
RPTLN-18	249A>G
RPTLN-19	249A>G; 260G>A
RPTLN-20	253A>G
RPTLN-21	253A>G; 279T>C

Table 3

RPTLN gene copies in organs of Colubridae and Lacertidae taxa

Species	Organ	*RPTLN* gene copy
R. scalaris	Lung	*RPTLN*-1 (KU563605)
	Heart	*273A>T (RPTLN*-22*)* (KU563619)
	Muscle	*RPTLN*-21 (KU563617)
	Skin	*RPTLN*-5 (KU563608)
P. muralis	Bladder	*RPTLN*-1 (KU563606), *RPTLN*-18 (KU563615)
	Liver	*RPTLN*-1 (KU563606)
	Lung	*RPTLN*-1 (KU563606), *RPTLN*-17 (KU563614)
	Kidney	*RPTLN*-1 (KU563606), *RPTLN*-18 (KU563615)
	Muscle	*RPTLN*-1 (KU563606)
	Skin	*RPTLN*-18 (KU563615), *RPTLN*-19 (KU563616)
	Stomach	*RPTLN*-18 (KU563615)
	Heart	*RPTLN*-9 (KU563610), *RPTLN*-18 (KU563615)
P. hispanica	Liver	*RPTLN*-1 (KU563607)
	Lung	*RPTLN*-11 (KU563611)
	Skin	*RPTLN*-1 (KU563607), *RPTLN*-16 (KU563613)
	Stomach	*RPTLN*-8 (KU563609), *RPTLN*-13 (KU563612)
	Heart	*RPTLN*-1 (KU563607)
	Brain	*RPTLN*-21 (KU563618)

Scheme 1

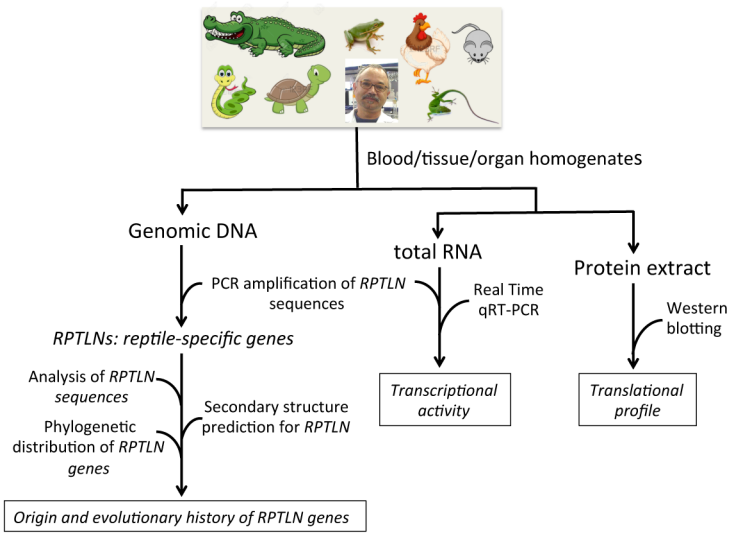


Figure 1

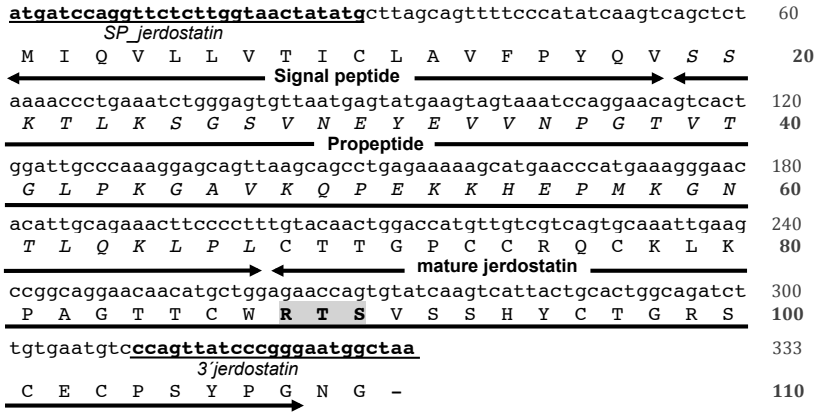


Figure 2

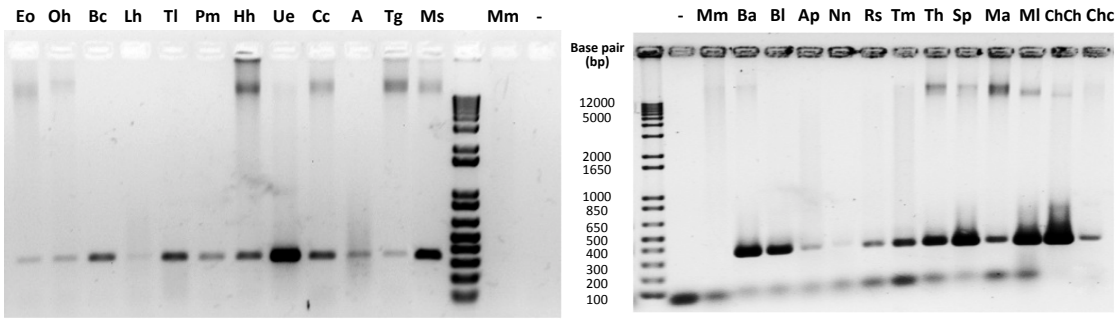


Figure 3

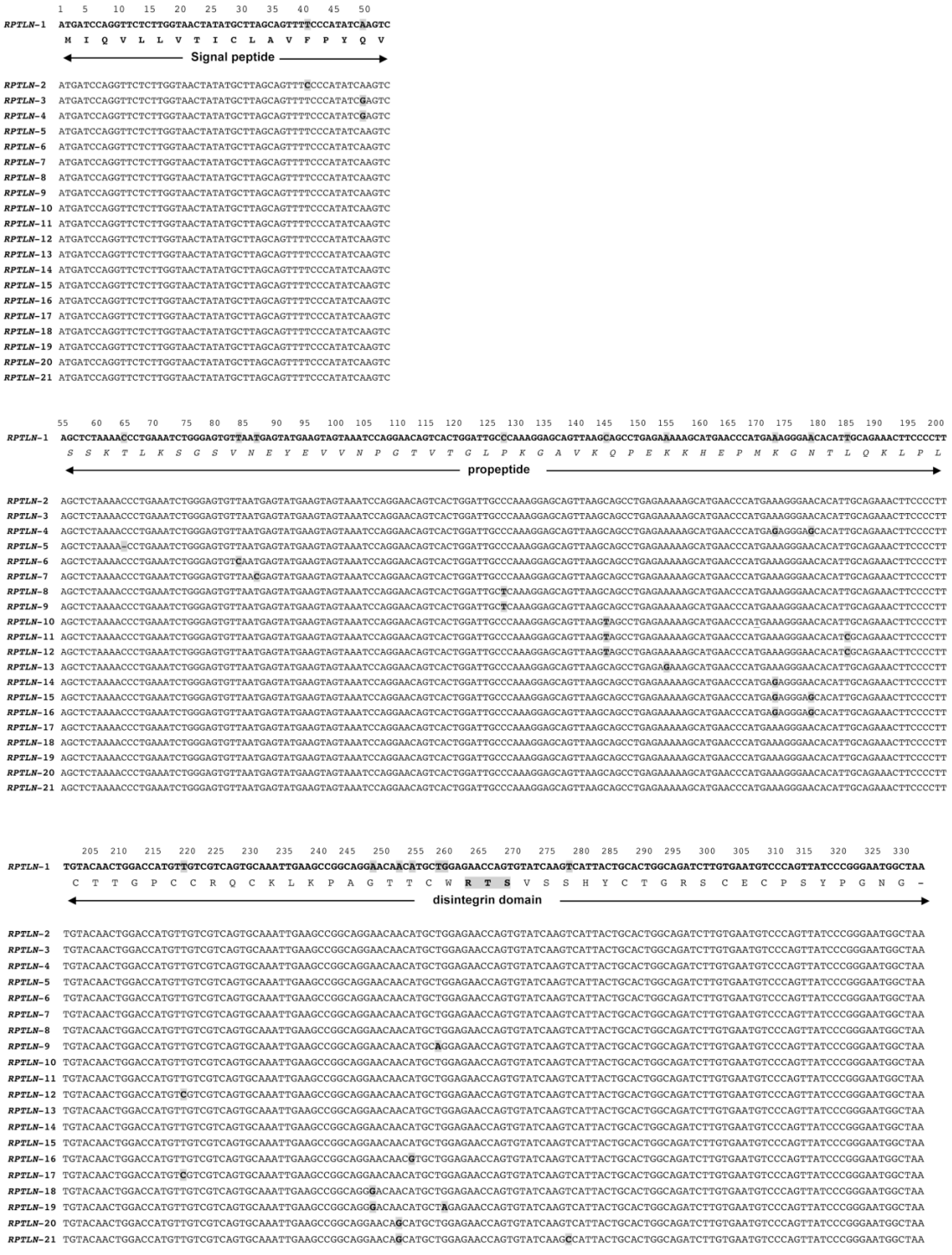

Signal peptide
propeptide
disintegrin domain

Figure 4

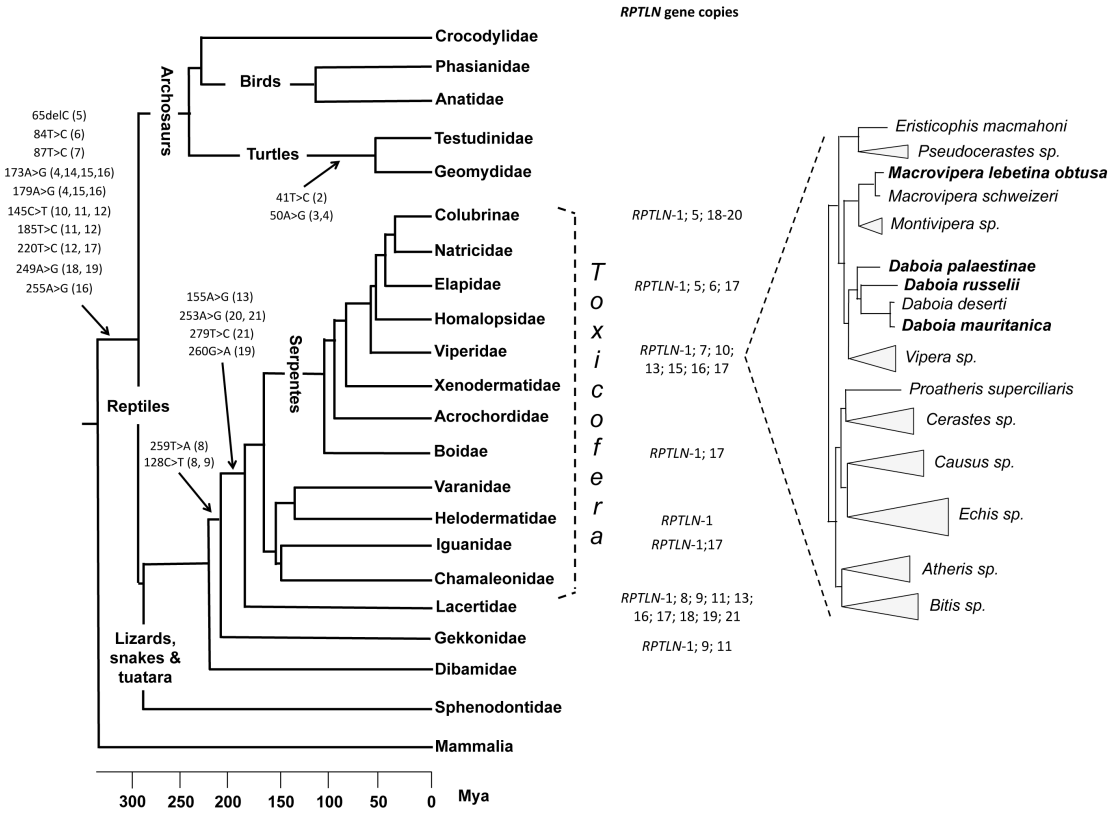


Figure 5

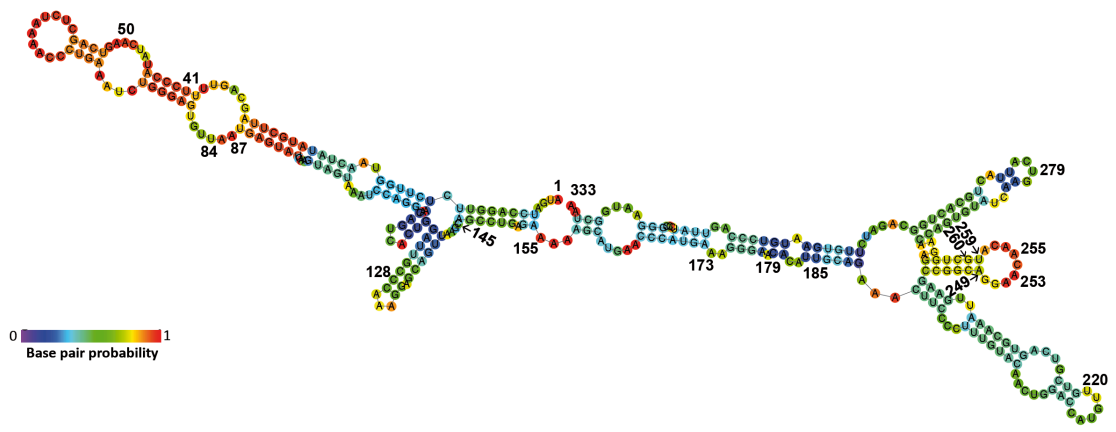


Figure 6

Species	Accession code	nucleotide sequence	Translated AA sequence	% nt identity
		1 5 10 15 20 25 30 35 40 45 50		
Protobothrops jerdonii RPTLN-1	AY262730	ATGATCCAGGTTCTCTTGGTAACTATATGCTTAGCAGTTTTCCCATATCAAG	MIQVLLVTICLAVFPYQ	100
Atheris squamigera SVMP	HF543864	ATGATCCAGGTTCTCTTGGTAATTATATGCTTAGCAGTTTTCCCATATCAAG	MIQVLLVTICLAVFPYQ	98
Viridovipera stejnegeri SVMP	DQ335449	ATGATCCAGGTTCTCTTGGTAATTATATGCTTAGCAGTTTTCCCATATCAAG	MIQVLLVTICLAVFPYQ	98
Gloydius intermedius SVMP	KM435346	ATGATCCAGGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCTTATCAAG	MIQVLLVTICLAVFPYQ	96
Echis pyramidum leakeyi SVMP	GU012290	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCATATCAAG	MIQVLLVTICLAVFPYQ	96
Echis coloratus SVMP	GU012229	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCATATCAAG	MIQVLLVTICLAVFPYQ	96
Macrovipera lebetina SVMP	DQ288157	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCATATCAAG	MIQVLLVTICLAVFPFQ	96
Crotalus adamanteus SVMP	HQ414112	ATGATCCAGGTTCTCTTGGTAACTATATCCTTAGCAGTTTTTCCTTATCAAG	MIQVLLVTISLAVFPYQ	94
Agkistrodon p. leucostoma SVMP	GQ451443	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTTTTTCCTTATCAAG	MIQVLLVTICLAVFPYQ	94
Bothrops insularis SVMP	AY736107	ATGATCCAGGTTCTCTTGGTAACTATATGCTTAGCAGCTTTTCCTTATCAAG	MIQVLLVTICLAAFPYQ	94
Ovophis okinavensis SVMP	AB851968	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGCAGTCTTTCCTTATCAAG	MIQVLLVTICLAVFPYQ	94
Gloydius saxatilis SVMP	AY204244	ATGATCCAGGTTCTCTTGGTAACTATATGCTTAGTAATCTTTCCATATCAAG	MIQVLLVTICLVIFPYQ	94
Ophiophagus hannah SVMP	EF065674	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGTGGTTTTTCCATATCAAG	MIQVLLVTICLVVFPYQ	92
Thamnophis sirtalis SVMP	XM_014065955	ATGATCCAAGCTCTCTTGGTAACTATATGCTTAGTGGTTTTTCCATATCAAG	MIQALLVTICLVVFPYQ	90
Anolis carolinensis ADAM28	XM_008119851	ATGATCAAAGCACTCCTTCTGGCTTGTGCTTCGTGCTTTTCCAGCATCAAG	MIKALLLALCFVLFQHQ	65
Anolis carolinensis ADAM23	XM_008114972	ATGAAGCCGCCTGGAAGCAGAAGTCCTCTGCGGCAGACCTCCTCCTCTTCTC	MKPPGRRSPLRQTSSSS	ns
Anolis carolinensis ADAM9	XM_003226844	ATGGCTGCTTGGAGCTGCTGCGCCTGGCTTTGGCTCCTGTGCGCCCTGAGCG	MAAWSCCAWLWLLCALS	ns
Homo sapiens ADAM28	BC136478	ATGTTGCAAGGTCTCCTGCCAGTCAGTCTCCTCCTCTGTTGCAGTAAGTG	MLQGLLPVSLLLSVAVS	38
Homo sapiens ADAM7	NM_003817	ATGCTTCCCGGGTGTATATTCTTGATGATTTTACTCATTCCTCAGGTTAAAG	MLPGCIFLMILLIPQVK	ns
Gallus gallus ADAM28	XM_001233495	ATGAATAATGTCATTCTCATTTTTGTTGTTTTGCTATGCTTTCTCCATCAAG	MNNVILIFVVLLCFLHQ	44

Figure 7

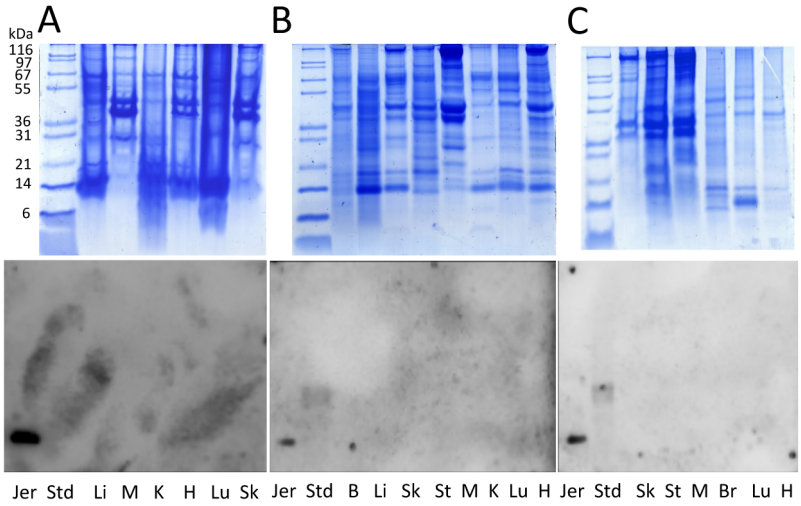


Figure 8